I. R. IRAN

ISSN: 1728-1431
e-ISSN: 1735-9244

International Journal of Engineering

Journal Homepage: www.ije.ir

MERC

TRANSACTIONS A: Basics

Volume 34, Number 07, July 2021

Materials and Energy Research Center

In the Name of God

INTERNATIONAL JOURNAL OF ENGINEERING
Transactions A: Basics

International **J**ournal of **E**ngineering *Transactions A: Basics* (ISSN 1728-1431) (EISSN 1735-9244)
International **J**ournal of **E**ngineering *Transactions B: Applications* (ISSN 1728-144X) (EISSN 1735-9244)
International **J**ournal of **E**ngineering *Transactions C: Aspects* (ISSN 2423-7167) (EISSN 1735-9244)
Web Sites: www.ije.ir & www.ijeir.info E-mails: ije.editor8@gmail.com, Tel: (+9821) 88771578, Fax: (+9821) 88773352
Materials and Energy Research Center (MERC)

ISSN: 1728-1431, e-ISSN: 1735-9244

International Journal of Engineering, Volume 34, Number 7, July 2021

CONTENTS:

ISSN: 1728-1431, e-ISSN: 1735-9244

International Journal of Engineering, Volume 34, Number 7, July 2021

ISSN: 1728-1431, e-ISSN: 1735-9244

International Journal of Engineering, Volume 34, Number 7, July 2021

International Journal of Engineering

Journal Homepage: www.ije.ir

Nitrogen Starvation Effect Versus its Excess on the Performance of *Arthrospira maxima* in Zarrouk's Medium

N. Mirhosseini[a], R. Davarnejad*[a], A. Hallajisani[b], O. Tavakoli[c], E. Cano-Europa[d]

[a] *Department of Chemical Engineering, Faculty of Engineering, Arak University, Arak, Iran*
[b] *Biofuel Laboratory, Caspian Faculty of Engineering, College of Engineering, University of Tehran, Tehran, Iran*
[c] *School of Chemical Engineering, College of Engineering, University of Tehran, Tehran, Iran*
[d] *Laboratorio de Metabolismo, Departamento de Fisiología, Escuela Nacional de Ciencias Biológicas, Instituto Politecnico Nacional, Código Postal Ciudad de México, Mexico*

PAPER INFO

Paper history:
Received 10 November 2019
Received in revised form 23 December 2019
Accepted 04 January 2020

Keywords:
Nitrogen Starvation
General Factorial Design
Performance
Arthrospira Maxima
Zarrouk

ABSTRACT

Providing a new cheap medium in terms of biomass and pigment production from *Arthrospira maxima* is an important issue. In order to determine the effect of nitrogen starvation versus its excess, urea was used as a cheap nitrogen source in two different modes (alternative and additive) at four different concentrations (0, 1.25, 2.5, 5 gL^{-1}). It was indicated that an alternative method is better than additive method due to accelerate the ammonia synthesis (NH_3) and pH changes in the period of the growth that is directly related to biomass and pigments production. Moreover, the intensive production of biomass concentration, PC, APC, C_a, and $C_{(X+C)}$ content were 1.403, 0.074, 0.093,10.72, and 3.17 mgL^{-1} with nitrogen starvation medium. Urea considered being one of the inhibition sources of biomass growth due to the formation of NH_3. Analyzing final results by general factorial design concluded that omitting nitrogen sources had the potential possibility for growing ***A.*** *maxima* in order to reduce the production costs in large-scale cultivation,which yields better performance than cost-effective Zarrouk's medium.

doi: *10.5829/ije.2021.34.07a.01*

NOMENCLATURE

A	Absorbance at different wavelength	PC	Phycocyanin
APC	Allophycocyanin	OD_t	Optical density at time t
C_a	Chlorophyll-a	OD_0	Optical density at time t_0
C_b	Chlorophyll-b	R	Correlation coefficient
$C_{(x+c)}$	Total carotenoids	DT	Doubling time (d)
C_t	Biomass concentration at time t (gL^{-1})	U	Urea
C_0	Biomass concentration at time t_0 (gL^{-1})	V	Total volume of cultivation (L)
mM	Molar concentration	X_m	Maximum dry weight (gL^{-1})
M	Addition method of urea	X_i	Initial cell dry weight (gL^{-1})
N_t	Total quantity of added nitrogen (g)	$Y_{X/N}$	Nitrogen-cell conversion (gg^{-1})
K	Mean daily division rate (division d^{-1})	**Greek Symbols**	
P	Productivity ($gL^{-1}d^{-1}$)	μ_{max}	Maximum specific growth rate (d^{-1})

1. INTRODUCTION

Algae are as biological sources with a broad variety of applications [1]. Interest in the applications of natural colors has increased due to their non-toxic, non-allergic, and antimicrobial effects [2]. The natural pigments are not only colorful but also have physiological activity compared with the synthetic colorants [3]. The ability of natural pigments extraction from microalga has been receiving more attention recently [4, 5]. Currently,

*Corresponding Author Institutional Email: R-Davarnejad@araku.ac.ir (R. Davarnejad)

Please cite this article as: N. Mirhosseini, R. Davarnejad, A. Hallajisani, O. Tavakoli, E. Cano-Europa, Nitrogen Starvation Effect Versus its Excess on the Performance of *Arthrospira maxima* in Zarrouk's Medium, International Journal of Engineering, Transactions A: Basics Vol. 34, No. 7, (2021) 1557-1568

Spirulina sp. is one of the most widely available microalga sources that contain a higher value of natural pigments like chlorophyll, carotenoids, and C-phycocyanin [6]. ***A.*** *maxima* as a photosynthetic, filamentous, spiral-shaped, multicellular, and blue-green *Cyanobacterium* was consumed as a food in some counties such as Mexico since 400 years ago [7]. The biomass production rate of *Spirulina* is mainly dependent on nutrient availability, temperature, and light [8]. The effect of various environmental conditions on biomass and pigment production of ***A.*** *maxima* were reported in the literature [9]. Furthermore, different mediums were successfully used for the microalgae growth like Zarrouk [10], modified Zarrouk [11], CFTRI, JPJM [12], Bangladesh medium [13], and Tofu wastewater [14]. Nutrient limitationis one of the major promising strategies to change and control the microalgae cell cycle and the biochemical production of lipid and other pigments [15]. The essential nutrients are an organic or inorganic carbon source as well as nitrogen, phosphorus, and iron in the course of alga growth [16]. Nitrogen is one of the remarkable factors which has an impact on both growth and pigment production of microalgae [17]. Sodium nitrate is a common nitrogen source used in the spirulina cultivation. Moreover, the feasibility of replacement of basic nitrogen sources of medium with low-cost alternatives such as urea, ammonium sulfate, and ammonium chloride in *S. platensis* cultivation was already considered [18–20].

Few studies were conducted on the cultivation of ***A.*** *maxima* based on the urea (nitrogen starvation) effect according to high production of pigments and low production cost under Mexico city climate [21]. Urea was used as an additive and alternative nitrogen source for investigating the effect of supplemented nitrogen source compared to the nitrogen starvation (as urea) [22–24]. In addition, the incorporation of bicarbonate and nitrate is an important cost-effectiveness technique in large-scale cultivation of *Arthrospira* [25].

In this research, the responses of ***A***. *maxima* are studied in terms of nitrogen starvation and urea concentrations on biomass growth, productivity, specific growth rate, and pigment production (including phycocyanin, allophycocyanin, carotenoids, and chlorophylls pigments). Finally, a low-cost effective medium is developed based on inexpensive raw material to be used on a future large-scale production.

2. MATERIALS AND METHODS

2. 1. Microorganism and Culture Condition *Arthrospira maxima* CIB 79 was provided from the National Polytechnic University (National Polytechnic University, Mexico City, Mexico), which was grown in an axenic batch culture of Zarrouk's medium. Figure 1 shows a microscopic images of axenic unialga. The stock and test cultures flask carried out at laboratory temperature (28-32 °C) under a white fluorescent light with a photoperiod of 24 h (80µmol$m^{-2}s^{-1}$) and pH control free (8.30-10.30). During the process of growth, the cultures continuously aerated with supporting an air pump [AC-9602 (RESUN, Mexico)] to accelerate the growth process. Urea was consumed at various concentrations (0, 1.25, 2.5, and 5 gL^{-1}) by the medium. The cultivation environment was prepared either with (additive) or without (alternative) basic nitrogen source of Zarrouk's medium. Then, urea was added to the cultivation at various concentrations. Concentration of the used urea was based on the amount of the standard nitrogen source of Zarrouk's medium (2.5 gL^{-1}) for comparing the basic Zarrouk's medium with nitrogen starvation (0 gL^{-1}). In fact, concentration of 1.25 and 5 gL^{-1} was chosen as a mean and double concentration of basic nitrogen source to study the effect of extra nitrogen source versus nitrogen starvation. Zarrouk medium with following chemicals and composition $NaHCO_3$ 16.8 gL^{-1}, $NaNO_3$ 2.5 gL^{-1}, K_2HPO_4 0.5 gL^{-1}, K_2SO_4 1.0 gL^{-1}, NaCl 1.0 gL^{-1}, $MgSO_4.7H_2O$ 0.2 gL^{-1}, EDTA-$Na_2.2H_2O$ 0.08 gL^{-1}, $CaCl_2.2H_2O$ 0.04 gL^{-1}, and $FeSO_4.2H_2O$ 0.01 gL^{-1}, micronutrient elements solution (H_3BO_3 2.86 gL^{-1}, $MnCl_2.4H_2O$ 1.81 gL^{-1}, $ZnSO_4.7H_2O$ 0.222 gL^{-1}, MoO_3 0.01 gL^{-1}, $CoCl_2.6H_2O$ 0.01 gL^{-1}, $CuSO_4.5H_2O$ 0.079 gL^{-1}) 1.0 mLL^{-1} was used as a standard medium of *Spirulina* cultivation. All the chemicals were purchased from Merck and Sigma-Aldrich Company.Treatments were cultivated in a 125 mL flask which were inoculated with 40 ml of ***A***. *maxima* inoculumin an initial concentration of ≈0.4 OD. The biomass growth and pigment production were recorded once during the maximum 8 days of cultivation. The experiment was performed during 2019-2020 in June.

2. 2. Phycobiliproteins and Photosynthetic Analysis For pigment extraction, the cell walls were ruptured using five freezing and thawing cycles [26]. Thus, the samples were ferozen at -20 °C for 1 h and then thawed at room temperature for 45 min. The samples were centrifuged (with velocity of 14/14 R refrigerated centrifuge, China) at 10000 rpm for 10 min at 4 °C where the pellet was transferred to vials. Then, the supernatant was collected for detecting absorbance of

Figure 1. Microscopic images of axenic unialga

PC and APC at 620 and 652 nm, respectively. The pellet was picked after centrifugation for photosynthetic pigment measurements like total chlorophylls and carotenoids. Then, it was oven-dried at almost 25 °C for 3 h. Moreover, the pellet was dissolved in 600 µL of an acetone and chloroform mixture (70:30 v/v). Chlorophylls and carotenoids were also quantified at 470, 645, and 662 nm. The pigment extraction was carried out under the dim light condition to avoid its degradation. The pigments concentration of phycobiliproteins (PC and APC), chlorophyll-a (C_a), chlorophyll-b (C_b), and total carotenoids [$C_{(X+C)}$] contents were then calculated using equations (1) to (5) by Bennett and Bogorad [27], Parson and Strickland [28], and Wellburn [29]:

$$\mathrm{PC}\ (\mathrm{gL}^{-1}) = (\mathrm{A}_{620} - 0.474\ \mathrm{A}_{652})/5.34 \quad (1)$$

$$\mathrm{APC}\ (\mathrm{gL}^{-1}) = (\mathrm{A}_{652} - 0.208\ \mathrm{A}_{620})/5.09 \quad (2)$$

$$\mathrm{C_a}\ (\mathrm{mgL}^{-1}) = 11.24 \times \mathrm{A}_{662} - 2.04 \times \mathrm{A}_{645} \quad (3)$$

$$\mathrm{C_b}(\mathrm{mgL}^{-1}) = 20.13 \times \mathrm{A}_{645} - 4.19 \times \mathrm{A}_{662} \quad (4)$$

$$\mathrm{C}_{(\mathrm{x+C})}\ (\mathrm{mgL}^{-1}) = \big((1000 \times \mathrm{A}_{470}) - (1.09\mathrm{C_a} - 63.14\mathrm{C_b})\big)/214 \quad (5)$$

A corresponds to the measuring absorbance in the desired wavelengths (470, 620, 645, 652 and 662 nm), PC, APC, C_a, C_b and $C_{(X+C)}$, are phycocyanin, allophycocyaninchlorophyll-a, chlorophyll-b, and Carotenoid, respectively.

2. 3. Analytical and Statistical Analysis The biomass concentration (cell dry weight) was determined by measuring optical density at 674 nm during the sampling time. Optical density [by three replications (OD)] was recorded by a spectrophotometer (Thermo Scientific, Multiskan Go, England). This was used for the pigments estimation (PC and APC), as well. LEGEND MICRO 17 spectrophotometer (Thermo Scientific model, Germany) was applied for spectrophotometric measurement of chlorophylls and carotenoids. Moreover, the maximum specific growth rate (μ_{max}), doubling time (DT), and productivity (P) were determined from the maximum biomass concentration at the end of each run when the algal cultures were in exponential phase by applying the following equations [30]:

$$\mu_{max}(\mathrm{d}^{-1}) = \frac{\ln(C_t) - \ln(\mathrm{C_0})}{t - t_0}, \mathrm{DT\ (d)} = \frac{0.693}{\mu} \quad (6)$$

$$P(\mathrm{gL}^{-1}\mathrm{d}^{-1}) = \frac{C_t - \mathrm{C_0}}{t - t_0} \quad (7)$$

where, C_t and C_0 are the biomass concentrations at time of t and t_0.

The mean daily division rate (K, division d^{-1}) was estimated using the following equation [31]:

$$K(\text{division d}^{-1}) = \frac{3.3}{st} \times \log OD_t - \log OD_0 \quad (8)$$

where, t is cultivation days, OD_t is the optical density at time of t, and OD_0 is the optical density at time of t_0 [11].

In the present study, a general full factorial design was used to obtain all possible ways of levels for two chosen independent variables as shown in Table 1. Moreover, an experimental design was carried out to examine the correlations. Endpoint values of PC, APC, C_a, and $C_{(X+C)}$ were analyzed by two-way ANOVAs. The values in all figures represent the mean and error ± SEM of three replicates.

Multiple comparisons, Levene and Tukey tests of ANOVA, was used to determine whether the mean of a dependent variable is the same in unrelated groups using MiniTab19. On the other hand, the biomass growth data was employed to evaluate the normality test using Anderson-Darling (AD), Ryan-Joiner (RJ), and Kolmogorov–Smirnov (KS) model before the analysis.

3. RESULTS AND DISCUSSION

3. 1. Estimation of *A.maxima* Biomass Concentrtion The dried biomass concentration of ***A.*** *maxima* was evaluated by measuring the optical density at 674 nm using a calibration curve. The measured biomass concentration linearly related to the optical density that was described by the equation of $Y(gL^{-1})=0.58X-0.0201$ (with $R^2=0.997$). where, Y and X are biomass concentration and optical density, respectively.

3. 2. Biomass and Pigment Production The results were examined in two groups with urea as an additive source and with it as an alternative source. Therefore, a standard cultivation medium was performed to analyze the best results. Addition of urea in basic nitrogen source of Zarrouk's medium was made according to the method proposed by Chouhan and coworkers to obtain the influence of the additional amount of nitrogen source on biomass growth and pigment production compared to Zarrouk's medium and nitrogen starvation [32]. The biomass growth curve in the case of additive nitrogen source was graphed according to Figure 2 (a and b). According to these figures,

TABLE 1. 2^4 general factorial design

Factors	Levels	Values
U[1]	4	0, 1.25, 2.5, 5
M[2]	2	0, 1

[1]: concentration of urea, [2]: addition method

significant changes were not observed in the biomass growth in the medium with urea at various concentrations. Similar increment in biomass growth at various concentrations of urea was found for the first four days of cultivation, as well. Obviously, a large difference in biomass growth and cellular growing between medium with nitrogen stravations and urea was shown during six days of cultivation. However, the exponential phase of biomass growth started at the optical density of 0.43 for medium with urea but, the cell number of standard Zarrouk's medium increased. Furthermore, the medium with urea as a nitrogen source reached exponential phase of the growth within four days of cultivation. Biomass concentration decreased with increasing the concentration of urea after the fourth day. In fact, urea did not yield the highest biomass dry weight when an additive nitrogen source was added. The best growth in biomass composition occurred on the medium with limited amount of urea.

In the second group, urea was substituted with basic nitrogen source of Zarrouk's medium according to a method proposed by Madkour and his coworkers. [33]. The influence of nitrogen starvation (0 gL^{-1}), replacing high concentration of basic sodium nitrate of Zarrouk's medium with the mean amount of urea (1.25 gL^{-1}), was investigated. It was also compared with urea at the same amount of sodium nitrate of Zarrouk's medium (2.5 gL^{-1}) and with pure urea as a cheap nitrogen source instead of sodium nitrate (5 gL^{-1}) on the biomass growth and pigments production of microalga. Figure 3 displays the effect of alternative nitrogen source on optical density and biomass growth. According to Figure 3 (a and b),

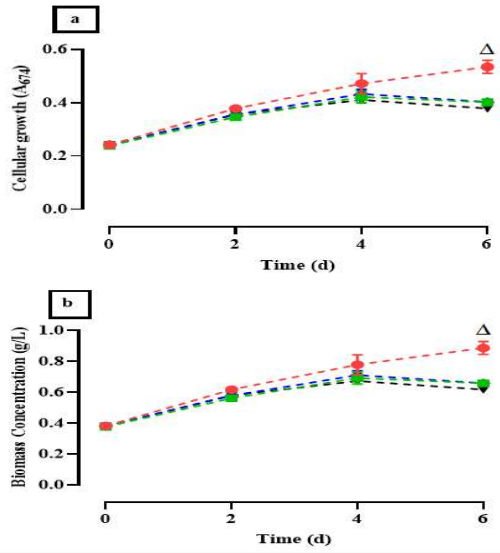


Figure 2. Effect of additive nitrogen source on cellular growth (a), biomass concentration (b) (●: 0 gL^{-1}, ■: 1.25 gL^{-1},♦: 2.5 gL^{-1},▼: 5 gL^{-1}). Δ symbol displays the best value of medium with standard Zarrouk's medium

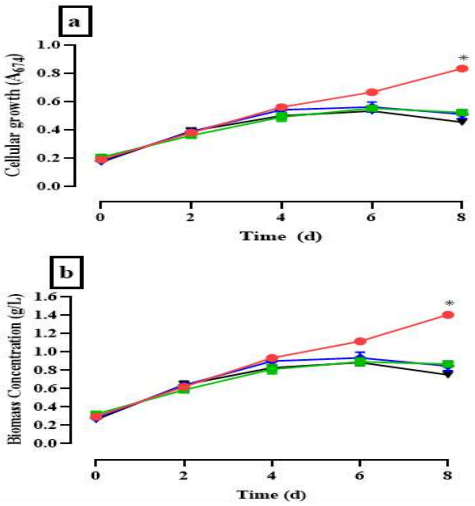


Figure 3. Effect of alternative nitrogen source on cellular growth (a), biomass concentration (b) (●: 0 gL^{-1}, ■: 1.25 gL^{-1},♦: 2.5 gL^{-1},▼: 5 gL^{-1}). * symbol displays the best value of medium with standard Zarrouk's medium

biomass and cellular growth reached the exponential phase with the urea concentration increment which was maximal at the lowest concentration of urea (0 gL^{-1}). In this group, the experiment was stopped due to reaching the exponential phase in microalgae growth in eight day. Furthermore, the biomass concentration was about 1.115 gL^{-1}at 0 gL^{-1} concentration of urea in this group while it was almost 1.26 times higher than that of the observed with 0 gL^{-1} concentration of urea in the additive group. In the current study, it was observed that addition of urea in the medium caused a reduction in biomass growth. Thus, the biomass composition grew when the medium was without nitrogen source.

Then, an original seed culture divided into two or more equal new cells according to Table 2 [34]. According to this table, the mean daily diversion rate was higher than that of a sample containing 0 gL^{-1} urea. Furthermore, the highest value was found for Zarrouk's medium without nitrogen source (sodium nitrate and urea).

Figure 4 shows analysis of the growth parameters in the various experiments. The effect of additive and alternative nitrogen sources during six days of cultivation was shown in this figure. According to this figure, a further increase of urea concentration from 0 to 5 gL^{-1} will decrease μ_{max} and P when urea is added as an additive nitrogen source. Furthermore, μ_{max} was higher than that of it in the standard Zarrouk's medium when urea was added as an additive nitrogen source. Moreover, it was more than that of it in a medium with 1.25 gL^{-1} urea, more than that of it in a medium containing 2.5 gL^{-1} urea and more than that of it in a medium containing 5 gL^{-1} urea, respectively. atmospheric nitrogen amount supplied by

TABLE 2. Mean daily division for ***A***. *maxima* at different concentrations of urea

Time (D)	Urea C (gL^{-1})	Biomass[a]	K_{alt}[a]	Biomass[b]	K_{add}[b]
2	0	0.619	0.026	0.625	0.072
	1.25	0.641	0	0.56	0
	2.5	0.635	0.065	0.569	0
	5	0.641	0.093	0.58	0
4	0	0.933	0.514	0.752	0.328
	1.25	0.827	0.431	0.699	0.328
	2.5	0.899	0.523	0.749	0.334
	5	0.827	0.516	0.694	0.307
6	0	1.115	0.624	0.737	0.389
	1.25	0.884	0.545	0.644	0.4
	2.5	0.935	0.606	0.634	0.389
	5	0.884	0.615	0.6	0.383
8	0	1.404	0.688	-	-
	1.25	0.75	0.57	-	-
	2.5	0.846	0.623	-	-
	5	0.75	0.624	-	-

Urea C: Urea Concentation, [a]: urea was added as an alternative source, [b]: urea was added as an additive source. Concentration of 0 gL^{-1} in additive group pointed at Zarrouk's medium and concentration of 0 gL^{-1} in alternative group pointed at nitrogen starvation medium.

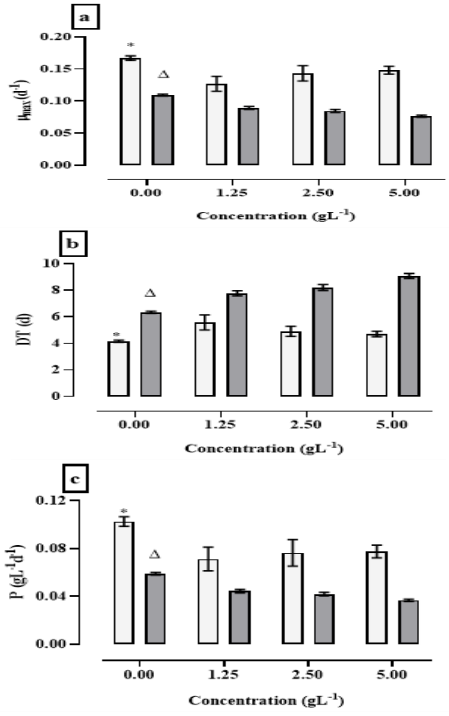


Figure 4. Changes in growth parameters when urea was added as an additional (dark gray bar) and alternative (light gray bar) sources during six days of cultivation. (a) Maximum specific growth rate (μ_{max}), (b) doubling time (DT), and (c) productivity (P). Δ displays the standard Zarrouk's medium. ⁎ displays the best value of medium with nitrogen starvation. The horizontal axis (X-axis) shows numeric values of urea concentration.

the aeration source. However, a dramatical enhancement was observed for μ_{max}(0.195 d^{-1}) and P (0.139$gL^{-1}d^{-1}$) after 8 days of cultivation in medium with nitrogen starvation but, DT decreased to 3.54 day. Furthermore, the results revealed that second-best result for biomass growth and pigment production was obtained from the treatment with 2.5 gL^{-1} of urea concentration when urea was added as an alternative nitrogen source. μ_{max}, DT, and P of this treatment (2.5 gL^{-1}) was 0.143d^{-1}, 4.90 day, and 0.076$gL^{-1}d^{-1}$, respectively. The best result for μ_{max}, DT, and P were found from the group with urea as an alternative source in all concentrations and also their amounts were highest in the medium with nitrogen starvation compared to the medium either with urea or with the mixture of sodium nitrate and urea. The μ_{max} and P were assessed to be approximately double in the medium with nitrogen starvation (0.195 d^{-1}, 0.139 $gL^{-1}d^{-1}$) compared to the medium with urea starvation (standard Zarrouk's medium, 0.109 d^{-1}, 0.060 $gL^{-1}d^{-1}$).

In fact, nitrogen did not play as an essential role in maintaining high μ_{max} (0.167 d^{-1}) and P (0.103 $gL^{-1}d^{-1}$) in the lowest DT (4.16 d) among the nutrients required for ***A***. *maxima* growth and the

As shown in Figure 5 (a and b), the amount of PC and APC increased with nitrogen starvation compared to the medium urea as an additive and alternative sources. PC (0.0734 mgL^{-1}), APC (0.093 mgL^{-1}), C_a (10.717 mgL^{-1}), and $C_{(X+C)}$ (3.171 mgL^{-1}) contents in the medium with nitrogen starvation were almost higher than that of it in the control group with a basic amount of nitrogen. According to Figure 5 (c and d), the amounts of $C_{(X+C)}$ and C_a increased when the medium was without a nitrogen source. Furthermore, the amounts of PC and APC were decreased by increasing urea concentration (and biomass production). Moreover, the PC and APC contents almost were at the same value for 1.25 and 2.5 gL^{-1} urea concentration when this was added as an additive nitrogen source. The medium with 2.5 gL^{-1} urea as an alternative source slightly increased biomass growth and pigment content.

3. 4. Statistical Analysis and Experimental Design

In the present study, an experimental design is carried out to examine the correlations between the variables that affect biomass and pigment production. Table 3 shows normal distribution of biomass growth, and large difference in mean value. It may be due to sampling variance which may decrease by the urea concentration increment. The data follow a normal distribution if p-value is greater than 0.05 in Anderson-Darling test.

Significance level was set at $p > 0.15$ and $p > 0.1$ for KS and RJ tests, respectively. The experimental results of degree of freedom and adjusted sum of square for variables and their interactions are presented in Table 4. Figure 6 shows the Pareto and residual normal probability plots of the variables and their interactions.

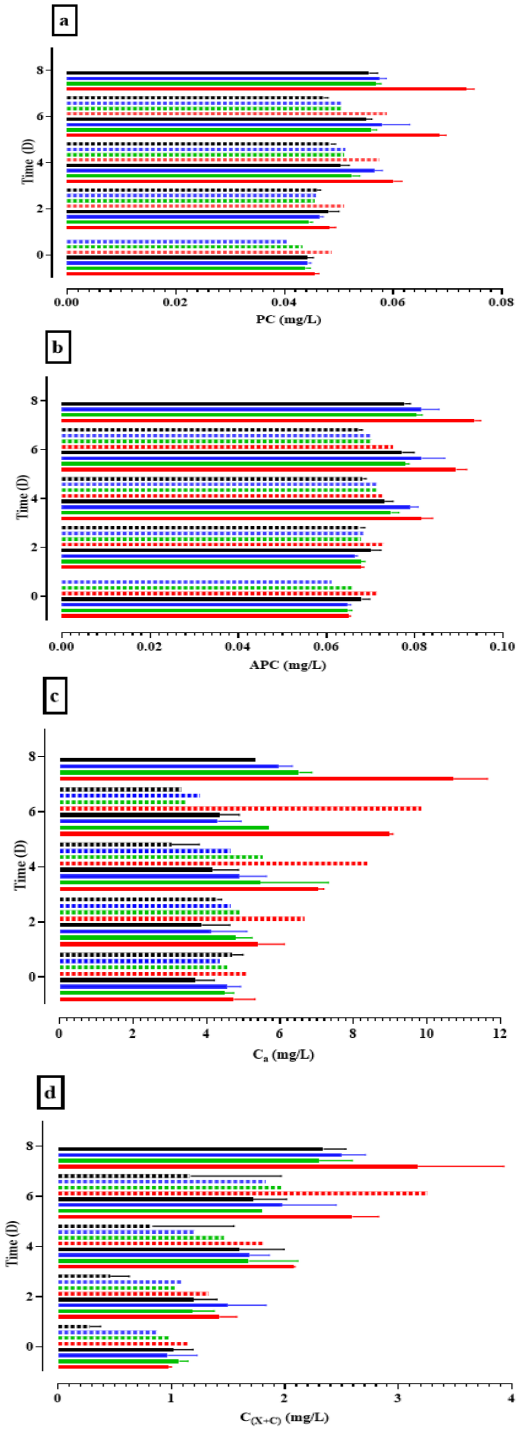


Figure 5. Effect of urea concentration on PC (a), APC (b), chlorophyll-a (c) and $C_{(X+C)}$ (d) contents with time. The dashed line represents urea as an additional source and solid line represents urea as an additive source. (red line: 0 gL^{-1}, green line: 1.25 gL^{-1},blue line: 2.5 gL^{-1}, black line: 5 gL^{-1})

According to Figure 6, a variable which is greater than baseline (1.746) will affect the target.

TABLE 3. Means, StDev and p-values in the KS, RJ and AD tests

Urea con (gL^{-1})	Add mode	Mean	StDev	KS	RJ	AD	AD p-value
0 (Zarrouk's medium)		0.659	0.160	0.191	0.997	0.220	0.733
1.25	Add	0.576	0.148	0.211	0.978	0.218	0.614
2.5		0.583	0.151	0.236	0.979	0.219	0.612
5		0.562	0.129	0.304	0.945	0.323	0.309
0 (Nitrogen starvation medium)		0.873	0.428	0.146	0.997	0.143	0.922
1.25	Alt	0.694	0.241	0.284	0.929	0.405	0.206
2.5		0.712	0.206	0.281	0.945	0.378	0.248
5		0.672	0.246	0.249	0.924	0.403	0.209

TABLE 4. Degree of freedom and adjusted sum of square in regression analysis

Sources	DF	ADJ SS
U	3	0.564
M	1	0.585
U&M interaction	3	0.256
Error	16	0.017
Total	23	1.424

U: Urea concentration, M: Addition method, DF: Degree of freedom

Furthermore, U and M are most important than the interaction between them.The residual values were marked in normal probability plot of each response and the fittest red-line crossing these points. The p-value data show individual effects of variables and their interactions. p-values with a low probability ($p < 0.1$) indicate a very high significance for the corresponding coefficients. The effects of addition methods for urea (M) and almost urea concentration (U) statistically to be significant in the applied range. On the other hand, the carotenoid value and the combination effect of UM interactions did not have a statistical significance in the applied range. The interaction of variables for the concentration of zero showed a lower statistical significance for the responses whereas p-value amount increased by increasing the urea concentration.

As shown in Table 5, urea addition as an alternative source was the best method. The positive terms of coefficient factor have a synergistic effect on the response while the negative terms have an antagonistic effect on it. This means that yield increases by increasing the variables. As shown in Table 6, urea at the maximum concentration has the highest reducing effect on producing C_a while nitrogen starvation medium has the

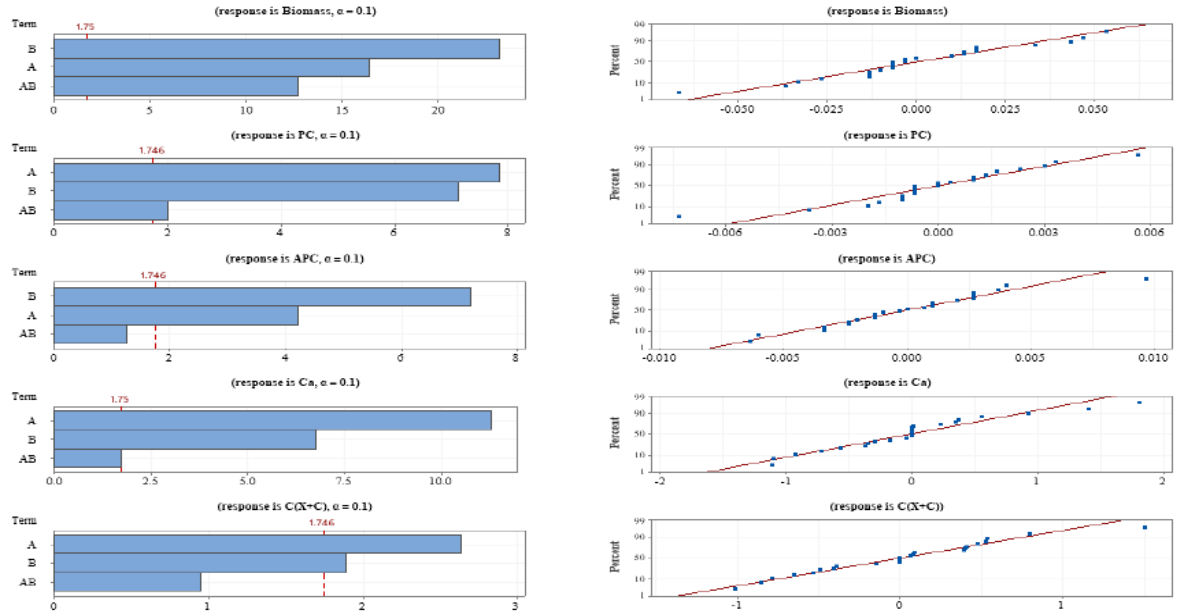


Figure 6. Pareto chart for the standardized effects and residual normal probability plot of each response (factor A, B, AB indicate U, M and interaction of U and M, respectively)

largest coefficient effect (due to having a positive sign) in biomass and pigment production. According to Table 7, R^2 value for biomass (98.77%) indicates the best fitting for the proposed model.

The accuracy of total carotenoids significantly decreased. This may be due to low amount of its concentration in *Spirulina*. The mathematical models for the responses as function of two variables were investigated as Equations (9) to (13):

Biomass amount= 0.80958 + 0.2604 $U_{0.00}$ - 0.0596 $U_{1.25}$- 0.0646 $U_{2.50}$-0.1363 U_5+ 0.15625 M_0 -0.15625 M_1+ 0.1771 $U\&M_{0.00,0}$ - 0.1771 $U\&M_{0.00,1}$ - 0.0496 $U\&M_{1.25,0}$+ 0.0496 $U\&M_{1.25,1}$-0.0446 $U\&M_{2.50,0}$+ 0.0446 $U\&M_{2.50,1}$- 0.0829 $U\&M_{5.00,0}$ + 0.0829 $U\&M_{5.00,1s}$ (9)

PC amount= 0.056125 + 0.01004 U0.00 - 0.00229 U1.25- 0.00262 U2.50- 0.00513 U5.00+ 0.004375 M0- 0.004375 M1+ 0.00312 U&M0.00,0- 0.00312 U&M0.00,1- 0.00121 U&M1.25,0+ 0.00121 U&M1.25,10.00154 U&M2.50,0+ 0.00154 U&M2.50,1- 0.00037 U&M5.00,0+ 0.00037 U&M5.00,1 (10)

APC amount= 0.076458 + 0.00788 U0.00 - 0.00229 U1.25- 0.00229 U2.50- 0.00329 U5.00+ 0.006042 M0- 0.006042 M1+ 0.00296 U&M0.00,0- 0.00296 U&M0.00,1- 0.00121 U&M1.25,0+ 0.00121 U&M1.25,1- 0.00188 U&M2.50,0 + 0.00188 U&M2.50,1 + 0.00012 U&M5.00,0 - 0.00012 U&M_.00,1 (11)

TABLE 5. p-value of each factor and interaction between U&M

Response	P- VALUE						
	U			M	U&M		
	0	1.25	2.5		0	1.25	2.5
Biomass	0	0	0	0	0	0.001	0.002
PC	0	0.047	0.025	0	0.01	0.274	0.167
APC	0	0.135	0.135	0	0.059	0.419	0.216
Ca	0	0	0	0	0.027	0.204	0.771
$C_{(X+C)}$	0.002	0.400	0.400	0.077	0.219	0.630	0.916

TABLE 6. Analysis of variance of biomass and pigments production

Sources	U			M	U&M			Constant
	0	1.25	2.5	0	0	1.25	2.5	
Biomass	0.2604	-0.0596	-.0.646	0.15625	0.1771	-0.0496	-0.0446	0.80958
PC	0.01004	-0.00229	-0.00262	0.00437	0.00312	-0.00121	-0.00154	0.05612
APC	0.00788	-0.00229	-0.00229	0.00604	0.00296	-0.00121	-0.00188	0.07645
Ca	4.030	-1.285	-1.370	1.152	-0.721	0.391	-0.088	6.255
$C_{(X+C)}$	0.888	-0.214	-0.196	0.270	-0.317	-0.122	0.027	2.326

TABLE 7. R^2 values obtained from the 2^4 factorial experiment matrix

	Biomass	PC	APC	Ca	$C_{(X+C)}$
R^2 (%)	98.77	90.46	84.32	93.73	56.21

$$C_a \text{ amount} = 6.255 + 4.030\ U_{0.00} - 1.285\ U_{1.25} - 1.370\ U_{2.50} - 1.375\ U_{5.00} + 1.152\ M_0 - 1.152\ M_1 - 0.721\ UM_{0.00,0} + 0.721\ UM_{0.00,1} + 0.391\ UM_{1.25,0} - 0.391\ UM_{1.25,1} - 0.088\ UM_{2.50,0} + 0.088\ UM_{2.50,1} + 0.417\ UM_{5.00,0} - 0.417\ UM_{5.00,1} \quad (12)$$

$$C_{(X+C)} \text{ amount} = 2.326 + 0.888\ U_{0.00} - 0.214\ U_{1.25} - 0.196\ U_{2.50} - 0.478\ U_{5.00} + 0.270\ M_0 - 0.270\ M_1 - 0.317\ U\&M_{0.00,0} + 0.317\ U\&M_{0.00,1} - 0.122\ U\&M_{1.25,0} + 0.122\ U\&M_{1.25,1} + 0.027\ U\&M_{2.50,0} - 0.027\ U\&M_{2.50,} + 0.412\ U\&M_{5.00,0} - 0.412\ U\&M_{5.00,1} \quad (13)$$

Figure 7 shows the mean effect plot for pigmnet and biomass content versus addition mode and urea concentration. The effect of biomass variation with changing urea concentration in Zarrouk's medium depicted in Figure 7 (a and b). Figure 7(b) shows Levene's test of biomass growth versus urea concentration. Levene's test is used to verify the normality of variances. The Levene's test statistically showed a significant differences between biomass growth changes and urea concentration (P = 0). According to Figure 7(c and d), PC and APC were in maximum at the lowest concentration of urea. The pigments content was higher than that of it when urea was added as an alternative source. Furthermore, a low value of F-value (F= 0.91) implies that there is no significant difference between the mean of carotenoids and urea concentrations. However, a higher ratio of F-value (F=131.09) indicates that the variation among groups means (alternative and additive group) are greatly different from each other compared to the variation of the individual observations in each group. The mean value of C_a reaches a stationary phase by increasing the concentration of urea in alternative and additive nitrogen source from 1.25 to 5 gL^{-1}. This indicates that nitrogen starvation has the highest mean value for both groups of addition mode while it is started to be declined, there after.

This study aims to evaluate the impact of all variables and their interactions on the biomass and pigments production. Moreover, the optimal conditions were also obtained and illustrated in Table 8. For the optimal conditions validation, an experiment was carried out at the same conditions during 8 days of cultivation.

According to the microscopic analysis, the cells breakdown was intensified by increasing the concentration of urea and incubation time. In addition, the cells breakdown was extremely intensified when urea as an additive source was added. The optical density was raised during the first 4 days, due to existing more N_2 in the culture medium when sodium nitrate and urea were

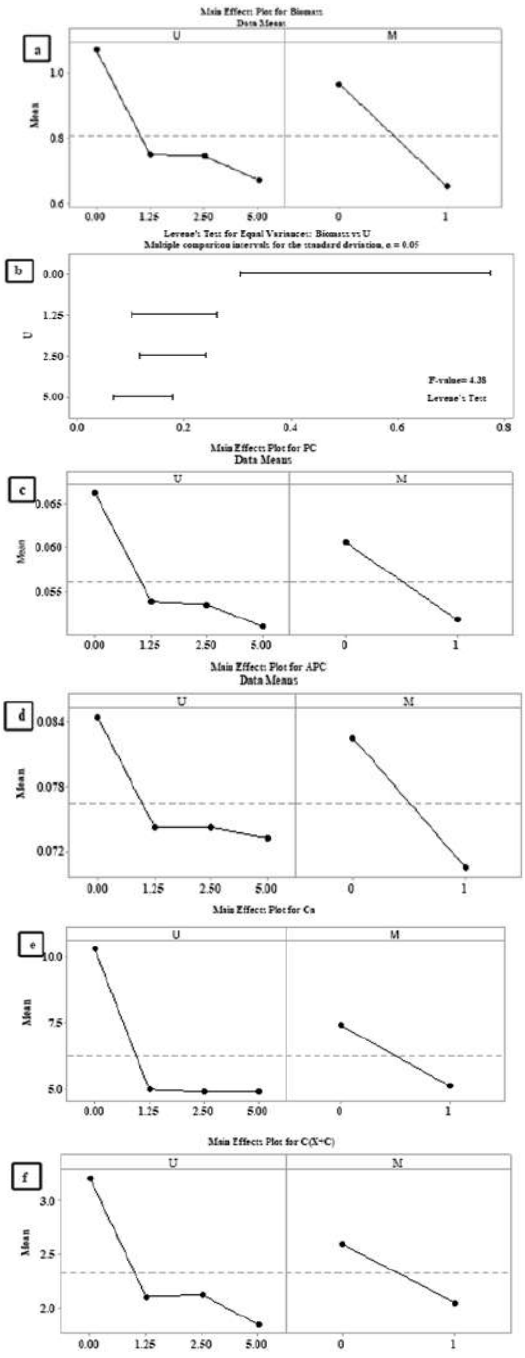


Figure 7. Mean effect plot for biomass growth (a&b), PC (c), APC (d), C_a (e) and $C_{(X+C)}$ (f) content versus addition mode and urea concentration. (0, 1.25, 2.5 and 5 (gL^{-1}) are urea concentration (U) and addition method of urea (M) was at 0 and 1 (gL^{-1}) as alternative and additive source of Zarrouk's medium, respectively

TABLE 8. Optimal conditions and their responses

Cond	U	M	Biomass fit	PC fit	APC fit	$C_{(X+C)}$ fit	C_a fit
mo	0	0	1.403	0.074	0.094	3.167	10.717
Ex	0	0	1.436	0.078	0.103	3.214	10.689

Cond: Condition; Optimal data based on (mo) the model, (Ex) the experiment (validation)

simultaneously added. In contrast, cultures supplemented with urea showed slow growth rate compared to the traditional Zarrouk mediumand the medium with an alternative nitrogen source, respectively. Since, urea is as an excellent nitrogen source and can successfully be metabolized by algae, *Spirulina* efficiently utilizes ammonia nitrate [18, 19]. The inhibition effect of urea can slightly be marked due to enzymatic hydrolysis of it by the urease enzyme [35]. Furthermore, the concentration of sodium nitrate in Zarrouk medium (2.5 gl^{-1}) reduces without losing the productivity as an important cost-saving factor in large-scale mass production of *S. platensis* [36]. Urea concentration increment in the medium (due to binding two amide groups) causes the cell division rate reduction. Cost et al. [24] used urea as a nitrogen source for growing *S. platensis*. They found that the production rate is in minimum when nitrate source of medium is replaced by urea. It also was found that *A. maxima* culture dies in the tanks with a very strong ammonia odor. In fact, nitrogen in culture is naturally reduced by the biological process of nitrification-denitrification [24]. According to Abou-Eleala et al. [37], an aqueous solution of ammonia consists of two forms (NH_4^+ and NH_3). Their formation depends on culture pH and temperature. NH_4^+ ions were produced when pH value was below 8.75 while NH_3 molecules were produced when pH value was above 9.75 [21]. In addition, the presence of NH_3 increased the toxicity of biomass culture [38]. In our experimental research, it was concluded that a high concentration of urea in culture medium may cause death of *A .maxima* by increasing incubtion time. Soletto et al. [39] used urea as an alternative nitrogen source which was more suitable than the classic nitrate-based culture medium for *S. platensis* growth. However, the use of urea in fed-batch Schlosser culture improved the biomass growth and increased the feeding rate of urea from 0.56 mM to 1.7 mM to prevent from the accumulation of ammonia in the medium during the cultivation process [39]. Moreover, urea contains two nitrogen atoms in a molecule with more than 45% of nitrogen while standard Zarrouk's medium was containing 16% of the nitrogen in sodium nitrate [40]. Abd El-Baky [41] found that nitrogen concentration reduction in *S. platensis* culture decreases phycocyanin, chlorophyll and protein contents due to breaking down the chloroplasts. The concentration of chlorophyll increased when nitrogen was used in the *S. platensis* culture medium, as well [41]. According to the current research, chlorophyll-a content decreased with increasing theurea concentration. Although, nitrogen sources did not play a significant role in chlorophyll-a and carotenoids production but, their contents stayed higher with nitrogen starvation during the cultivation process. However, the other researchers demonstrated that the concentration of sodium nitrate in Zarrouk's medium was easily replaced by sodium nitrate (0.3 gL^{-1}) for the *S. platensis* growth. Furthermore, they modified Zarrouk's mediumas urea subsituation [42]. In several researches, nitrogen starvation increased the lipid fraction of some microalgaes such as *S. platesis* [43–46]. Rodrigues et al. [35] concluded cyanobacteria initially used the nitrogen source of medium for the cells growth. Then, its excess used for the protein production [19]. This is in good agreement with the results obtained from the current research on the *A. maxima* growth with nitrogen starvation in the medium. de Castro et al. [47] demonstrated that the concentration of sodium nitrate in the Zarrouk's medium will be reduced by increasing the biomass growth while a higher biomass production depends on the bicarbonate sodiumamount. Moreover, a research showed that the highest lipids content and biomass production of *Spirulina platensis* without nitrogen source during 31 days of cultivation may be obtained while according to the present research, nitrogen had no significant effect on the biomass production enhancement. Its reason may be due to the applied range of this reaserch [48].

4. CONCLUSION

In this research, biomass and pigment production of A. maxima were measured to investigated some correlations between the effect of addition methods and urea concentrations. For this purpose, the experimental design was carried out by the general factorial method and the results were analyzed under urea concentrations in the range of 0-5 gL-1, urea addition methods as an additive, and as an alternative. The results showed that urea concentration was the most effective factor in the biomass and pigments production. Therefore, the biomass and pigments production was disrupted and significantly decreased for urea concentrations over 1.25 gL-1 although urea introduced into the culture medium had no positive effect under the conditions tested in this study. Moreover, urea (from 1.25 to 5 gL-1) as an additive nitrogen source could limit the intensive growth of microalgae. However, culture cells will breakdown by the urea concentration increment in presence of NH3 but, this may poison the microalgas. Furthermore, sodium nitrate is one of the most cost-effective sources in ***A***. *maxima* cultivation was used. The effects of binary interactions of parameters on the biomass and pigments

production were investigated by two-way ANOVA, as well. p-value showed that some independent variables such as U&M considerably affected all responses expect the carotenoid pigment which the other factors could control its production. It was concluded that biomass and pigments were highly produced through the alternative method. This may be due to accelerating NH_3 molecules and changing pH during the growth process. The maximum biomass growth (1.403 gL^{-1}), the highest specific growth rate (0.167 d^{-1}) with minimum doubling time (4.16 d) and maximum biomass productivity (0.103 $gL^{-1}d^{-1}$) with nitrogen starvation source were investigated. It is concluded that atmospheric nitrogen can grow microalga up to convert into the pigments. Besides, the sodium nitrate omission will decrease the production costs of large-scale cultivation.

5. REFERENCES

1. Hadiyanto, H., Soetrisnanto, D., Christwardhana, M., "Phytoremediation of Palm Oil Mill Effluent Using Pistia Stratiotes Plant and Algae Spirulina sp for Biomass Production", ***International Journal of Engineering, Transactions C: Aspects***, Vol. 27, No. 12, (2014), 1809-1814. DOI: 10.5829/idosi.ije.2014.27.12c.02
2. Sen, T., Barrow, C. J., and Deshmukh, S. K., "Microbial pigments in the food industry-challenges and the way forward", ***Frontiers in Nutrition***, Vol. 6, No. 7, (2019), 1-14. doi: 10.3389/fnut.2019.00007
3. Oplatowska-Stachowiak, M., and Elliott, C.T., "Food colors: Existing and emerging food safety concerns", ***Critical Reviews in Food Science and Nutrition***, Vol. 57, No. 3, (2017), 524-548. doi: 10.1080/10408398.2014.889652.
4. Mary Leema, J.T., Kirubagaran, R., Vinithkumar, N.V., Dheenan, P. S., and Karthikayulu, S., "High value pigment production from Arthrospira (Spirulina) platensis cultured in seawater", ***Bioresource Technology***, Vol. 101, No. 23, (2010), 9221-9227. doi:10.1016/j.biortech.2010.06.120
5. Griffiths, M. J., Garcin, C., van Hille, R. P., and Harrison, S. T. L., "Interference by pigment in the estimation of microalgal biomass concentration by optical density", ***Journal of Microbiological Methods***, Vol. 85, No. 2, (2011), 119-123. DOI: 10.1016/j.mimet.2011.02.005
6. Park, W. S., Kim, H. J., Li, M., Lim, D. H., Kim, J., Kwak, S. S., Kang, C. M., Ferruzzi, M. G., and Ahn, M. J., "Two classes of pigments, carotenoids and c-phycocyanin, in spirulina powder and their antioxidant activities", ***Molecules***, Vol. 23, No. 2065, (2018), 1-11. doi: 10.3390/molecules23082065.
7. Q. Ashton Acton., "*Cyclohexenes—Advances in Research and Application*": *2013 Edition: ScholarlyBrief. Atlanta, Georgia*, (2013), 228.
8. Carvalho, A.P., and Malcata, F.X., "Kinetic Modeling of the Autotrophic Growth of Pavlova lutheri: Study of the Combined Influence of Light and Temperature". ***Biotechnology Progress***, Vol. 19, No. 4, (2003), 1128-1135. doi: 10.1021/bp034083+.
9. De Oliveira, M.A.C.L., Monteiro, M.P.C., Robbs, P.G., and Leite, S.G.F., "Growth and chemical composition of Spirulina maxima and Spirulina platensis biomass at different temperatures", ***Aquaculture International***, Vol. 7, No. 4, 261-275. DOI:10.1023/A:1009233230706
10. Raoof, B., Kaushik, B. D. D., and Prasanna, R., "Formulation of a low-cost medium for mass production of Spirulina", ***Biomass and Bioenergy***, Vol. 30, No. 6, (2006), 537-542. DOI: 10.1016/j.biombioe.2005.09.006
11. Rajasekaran, C., Ajeesh, C. P. M., Balaji, S., Shalini, M., Siva, R., Das, R., Fulzele, D. P., and Kalaivani, T., "Effect of Modified Zarrouk's Medium on Growth of Different Spirulina Strains", ***Agriculture Technology and Biological Sciences Walailak Journal Science and Technology***, Vol. 13, No. 1, (2016), 67-75. DOI:10.14456/VOL13ISS2PP
12. Salunke, K.J., Magar, S.A., Joshi, R.R., and Wadikar, M.S, "Comparative study on growth of Spirulina platensis on different culture media", ***Bioscience Discovery***, Vol. 7, No. 1, (2016), 90-92. DOI: 10.1016/j.egyr.2019.02.009
13. Khatun, R., Noor, P., Akhter, N., Jahan, M., Hossain, M., and Munshi, J., "Spirulina Culture in Bangladesh XI: Selection of a Culture Medium, Suitable for Culturing a Local Strain of *Spirulina",**Bangladesh Journal of Scientific and Industrial Research***, Vol. 41, No. 3, (1970), 227-234. DOI:10.3329/BJSIR.V41I3.293
14. Syaichurrozi, I., Jayanudin, J., "Effect of Tofu Wastewater Addition on The Growth and Carbohydrate-Protein-Lipid Content of Spirulina platensis", ***International Journal of Engineering, Transactions B: Applications***, Vol. 30, No. 11, (2017), 1631-1638. DOI: 10.5829/ije.2017.30.11b.02
15. Mirhosseini, N., Davarnejad, R., Hallajisani, A., Cano-Europa, E., Tavakoli, O., Franco-Colín, M., Blas-Valdivia. V., "Cultivations of Arthrospira maxima (Spirulina) using ammonium sulfate and sodium nitrate as an alternative nitrogen sources", ***Iranian Journal of Fisheries Sciences***, Vol. 20, No. 2, (2021), 475-489. DOI: 10.22092/ijfs.2021.351071.0
16. Grobbelaar JU., "*Handbook of Microalgal Culture: Biotechnology and Applied Phycology*", *Wiley* (Amos Richmond ed.), (2004), 95-115. DOI: 10.1002/biot.201000433
17. Çelekli, A., and Yavuzatmaca, M., "Predictive modeling of biomass production by Spirulina platensis as function of nitrate and NaCl concentrations", ***Bioresource Technology***, Vol. 100, No. 5, (2009), 1847-1851. doi: 10.1016/j.biortech.2008.09.042.
18. Bezerra, R.P., Matsudo, M.C., Converti, A., Sato, S., and De Carvalho, J.C.M., "Influence of ammonium chloride feeding time and light intensity on the cultivation of Spirulina (Arthrospira) platensis", ***Biotechnology and Bioengineering***, Vol. 100, No. 2, (2008), 297-305. doi: 10.1002/bit.21771.
19. Ferreira, L.S., Rodrigues, M.S., Converti, A., Sato, S., and Carvalho, J.C.M. "A new approach to ammonium sulphate feeding for fed-batch Arthrospira (Spirulina) platensis cultivation in tubular photobioreactor", ***Biotechnology Progress***, Vol. 26, No. 5, (2010), 1271-1277. doi: 10.1002/btpr.457.
20. Avila-Leon, I., Chuei Matsudo, M., Sato, S., and de Carvalho, J. C. M., "Arthrospira platensis biomass with high protein content cultivated in continuous process using urea as nitrogen source", ***Journal of Applied Microbiology***, Vol. 112, No. 6, (2012), 1086-1094. doi: 10.1111/j.1365-2672.2012.05303.x.
21. Affan, M. A., Lee, D. W., Al-Harbi, S. M., Kim, H. J., Abdulwassi, N. I., Heo, S. J., Oh, C., Park, H. S., Ma, C. W., Lee, H. Y. and Kang, D. H., "Variation of Spirulina maxima biomass production in different depths of urea-used culture medium", ***Brazilian Journal of Microbiology***, Vol. 46, No. 4, (2015), 991-1000. DOI: 10.1590/S1517-838246420140188
22. Sukumaran, P., Nulit, R., Halimoon, N., Simoh, S., Omar, H., and Ismail., A., "Formulation of Cost-effective Medium Using Urea as a Nitrogen Source for Arthrospira platensis Cultivation under Real Environment", ***Annual Research and Review in Biology***, Vol. 22, No. 2, (2018), 1-12. DOI: 10.9734/ARRB/2018/38182
23. Danesi, E.D.G., Rangel-Yagui, C.O., Sato, S., and de Carvalho, J.C.M., "Growth and content of spirulina platensis biomass chlorophyll cultivated at different values of light intensity and

temperature using different nitrogen sources", ***Brazilian Journal of Microbiology***, Vol. *42, No.* 1, (2011), 362-373. doi: 10.1590/S1517-83822011000100046

24. Cost, J.A.V., Cozz, K.L., Oliveira, L., and Magagnin, G., "Different nitrogen sources and growth responses of Spirulina platensis in microenvironments", ***World Journal of Microbiology and Biotechnology***, Vol. 17, No. 5, (2001), 439-442. DOI: 10.1023/A:1011925022941
25. Jung, J.Y., Yang, J.W., Kim, K., Hwang, K.T., Jung, S.M.G., and Kwon, J.H., "Cost-efficient cultivation of Spirulina platensis by chemical absorption of CO_2 into medium containing NaOH", ***Korean Journal of Chemical Engineering***, Vol. 32, No. 11, (2015), pp 2285-2289. DOI: 10.1007/s11814-015-0073-3
26. Horváth, H., Kovács, A.W., Riddick, C., and Présing, M., "Extraction methods for phycocyanin determination in freshwater filamentous cyanobacteria and their application in a shallow lake", ***European Journal of Phycology***, Vol. 48, No. 3, (2013), 278-286. DOI: 10.1080/09670262.2013.821525
27. Bennett, A., and Bogobad, L., "Complementary chromatic adaptation in a filamentous blue-green alga", ***Journal of Cell Biology***, Vol. 58, No. 2, (1973), 419-435. doi: 10.1083/jcb.58.2.419.
28. Parsons, T.R., and Strickland, J.D.H., "Discussion of Spectrophotometric Determination of Marine-Plant Pigments, with Revised Equations for Ascertaining Chlorophylls and Carotenoids", ***Journal of Marine Research***, Vol. 21, (1963), 155-163. DOI:10.1016/0011-7471(65)90662-5
29. Wellburn, A.R., "The Spectral Determination of Chlorophylls a and b, as well as Total Carotenoids, Using Various Solvents with Spectrophotometers of Different Resolution", ***Journal of Plant Physiology***, Vol. 144, No. 3, (1994), 307-313. DOI: 10.1016/S0176-1617(11)81192-2
30. Ilknur, A.k., "Effect of an organic fertilizer on growth of blue-green alga Spirulina platensis", ***Aquaculture International***, Vol. 20, (2012), 413-422. DOI: 10.1007/s10499-011-9473-5
31. Vo, T.S., Ngo, D.H., and Kim, S.K., "Nutritional and Pharmaceutical Properties of Microalgal Spirulina". *Handbook of Marine Microalgae:* ***Biotechnology Advances,*** (2015), Chapter 19, 299-308.
32. Chouhan, S., Kaithwas, V., Kachouli, R., and Bhargava, S., "Productivity of the cyanobacterium spirulina platensis in cultures using high bicarbonate and different nitrogen sources", ***American Journal of Plant Physiology***, Vol. 8, No. 1, (2013), 17-31. DOI: 10.3923/ajpp.2013.17.31
33. Madkour, F.F., Kamil, A.E.W., and Nasr, H.S., "Production and nutritive value of Spirulina platensis in reduced cost media", ***Egyptian Journal of Aquatic Research***, Vol. 38, No. 1, (2012), 51-57. DOI: 10.1016/j.ejar.2012.09.003
34. Casem, M.L., "Book Series: Problem Sets in Biological and Biomedical Sciences", *Case Studies in Cell Biology*, (2016), 327-343.
35. Rodrigues, M.S., Ferreira, L.S., Converti, A., Sato, S., and Carvalho, J.C.M., "Fed-batch cultivation of Arthrospira (Spirulina) platensis: Potassium nitrate and ammonium chloride as simultaneous nitrogen sources", ***Bioresource Technology***, Vol. 101, No. 12, (2010), 4491-4498. DOI: 10.1016/j.biortech.2010.01.054
36. Abou-Elela, S. I., El-Sayed, M. M. H., El-Gendy, A. S., Abou-Taleb, E. M., "Comparative Study of Disinfection of Secondary Treated Wastewater Using Chlorine, UV and Ozone", ***Journal of Applied Sciences Research***, Vol. 8, No. 10, (2012), 5190-5197.
37. Converti, A., Scapazzoni, S., Lodi A., and Carvalho J.C.M. "Ammonia and urea removed by Spirulina platensis", ***Journal of Industrial Microbiolosgy and Biotechnology***, Vol. 33, (2006), 8-16. DOI: 10.1007/s10295-005-0025-8.
38. Soletto, D., Binaghi, L., Lodi, A., Carvalho, J.C.M., and Converti, A., "Batch and fed-batch cultivations of Spirulina platensis using ammonium sulphate and urea as nitrogen sources", ***Aquaculture***, Vol. 243, No. 1-4, (2005), 217-224. DOI: 10.1016/j.aquaculture.2004.10.005
39. Colla, L.M., Oliveira Reinehr, C., Reichert, C., Costa, J.A.V.,"Production of biomass and nutraceutical compounds by Spirulina platensis under different temperature and nitrogen regimes", ***Bioresource Technology***,Vol. 98, No. 7, (2007), 1489-1493. DOI: 10.1016/j.biortech.2005.09.030
40. Abd El-Baky, H.H., "Over Production of Phycocyanin Pigment in Blue Green Alga Spirulina sp. and It`s Inhibitory Effect on Growth of Ehrlich Ascites Carcinoma Cells", ***Journal of Medical Sciences***, Vol. 3, No. 4, (2003), 314-324. DOI: 10.3923/jms.2003.314.324
41. Salbitani, G., Carfagna, S., "Ammonium utilization in microalgae: A sustainable method for wastewater treatment", ***Sustainability***, Vol. 13, No. 2, (2021), 1-17. DOI: 10.3390/su13020956
42. Danesi, E.D.G., Rangel-Yagui, C.D.O., De Carvalho, J.C.M., and Sato, S., "An investigation of effect of replacing nitrate by urea in the growth and production of chlorophyll by Spirulina platensis", ***Biomass and Bioenergy***, Vol. 23, No. 4, (2002), 261-269. DOI: 10.1016/S0961-9534(02)00054
43. Soni, R.A., Sudhakar, K., & Rana, R.S., "Comparative study on the growth performance of Spirulina platensis on modifying culture media", ***Energy Reports***, Vol. *5*, (2019), 327-336. DOI: 10.1016/j.egyr.2019.02.009
44. Uslu, L., Işik, O., Koç, K., and Göksan, T., "The effects of nitrogen deficiencies on the lipid and protein contents of Spirulina platensis", ***African Journal of Biotechnology***, Vol. 10, No. 3, (2011), 386-389. DOI: 10.5897/AJB10.1547
45. Illman, A. M., Scragg, A. H., and Shales, S. W., "Increase in Chlorella strains calorific values when grown in low nitrogen medium", ***Enzyme and Microbial Technology***, Vol. *27*, No. 8, (2000), 631-635. DOI: 10.1016/S0141-0229(00)00266-0
46. de Castro, G.F.P.da.S., Rizzo, R.F., Passos, T.S., Santos, B.N.C.d., Dias, D.D.S., Domingues, J.R., Araújo, K.G.da.L., "Biomass production by Arthrospira platensis under different culture conditions", ***Food Science and Technology***, Vol. 35, No. 1, (2015), 18-24. DOI: 10.1590/1678-457X.6421
47. Ak, B., Işık, O., Uslu, L., Azgın, C., "The effect of stress due to nitrogen limitation on lipid content of phaeodactylum tricornutum (Bohlin) cultured outdoor in photobioreactor", ***Turkish Journal of Fisheries and Aquatic Sciences***, Vol. 15, (2015), 647-652. DOI: 10.4194/1303-2712-v15_3_09

Persian Abstract

چکیده

تهیه محیط کشت جدید ارزان قیمت و تولید بیشتر زیست‌توده و رنگدانه از ریزجلبک (***A.maxima***) Arthrospira maxima از اهمیت بسزایی برخوردار است. برای انجام این تحقیق، اوره به عنوان یک منبع نیتروژن ارزان قیمت در دو حالت (جایگزین و افزودنی) , چهار غلظت مختلف (۰، ۱/۲۵، ۲/۵، ۵ گرم بر لیتر) برای تعیین تاثیرفقر منبع نیتروژن در مقابل میزان مازاد منبع نیتروژنی استفاده شد. مشخص شد که روش افزودن اوره به عنوان منبع جایگزین ارجح‌تر از روش افزودنی مازاد است. که این امر ممکن است به دلیل سرعت بخشیدن به سنتز آمونیاک (NH_3) و تغییرات pH در طول رشد باشد که تقریباً تأثیر مستقیمی بر تولید زیست‌توده و رنگدانه دارد. علاوه بر این، بیشترین تولید زیست‌توده، فایکوسیانین، آلوفیکوسیانین، کلروفیل-a و کاروتنوئید ۱/۴۰۳، ۰/۰۷۴ ، ۰/۰۹۳، ۱۰/۷۲ و ۳/۱۷ میلی گرم بر لیتر از محیط کشت با فقر منبع نیتروژنی حاصل گردید. اوره به دلیل تشکیل NH_3 یکی از منابع مهار رشد زیست توده محسوب می‌شود. تجزیه و تحلیل نتایج نهایی به روش فاکتوریل جزئی اثبات کرد که حذف منابع نیتروژن امکان رشد ***A.maxima*** را به منظور کاهش هزینه‌های تولید کشت در مقیاس وسیع دارد که عملکرد بهتری نسبت به محیط اقتصادی مقرون به صرفه زاروک دارد.

IJE TRANSACTIONS A: Basics Vol. 34, No. 7, (July 2021) 1569-1575

International Journal of Engineering

Journal Homepage: www.ije.ir

Synthesis of Mesoporous SiO_2 Xerogel from Geothermal Sludge using Sulfuric Acid as Gelation Agent

H. Widiyandari*[a], P. Pardoyo[b], J. Sartika[b], O. A. Putra[a], A. Purwanto[c], L. Ernawati[d]

[a] *Department of Physics, Faculty of Mathematics and Natural Sciences, Sebelas Maret University, Indonesia*
[b] *Department of Chemistry, Faculty of Science and Mathematics, Diponegoro University, Indonesia*
[c] *Department of Chemical Engineering, Faculty of Engineering, Sebelas Maret University, Indonesia*
[d] *Department of Chemical Engineering, Institut Teknologi Kalimantan, Balikpapan, Indonesia*

PAPER INFO

Paper history:
Received 28 January 2021
Received in revised form 17 April 2021
Accepted 07 May 2021

Keywords:
SiO_2 Mesoporous
Geothermal Sludge
Gelation Agent
Alkaline Extraction
High Specific Surface Area

ABSTRACT

A large amount of sludge is produced by the geothermal brine at the Dieng Geothermal power plant, exceeding 165 tons per month. This sludge is generally not utilized, except for use in landfills. The precipitate (sludge) is primarily composed of silica. The aim of this research is to synthesis mesoporous silica (SiO_2) xerogel from geothermal sludge (GS) and to investigate the effects of pH as an effort to elevate the economic value of sludge through alkaline extraction followed by acidification. SiO_2 xerogel was prepared by extracting the GS to become sodium silicate (Na_2SiO_3) assisted by a base NaOH and precipitated using H_2SO_4 as a gelation agent. The FTIR analysis of the SiO_2 xerogel showed a group of silanol (Si-OH) and siloxane (Si-O-Si). The XRD analysis indicated that SiO_2 xerogel was amorphous. Furthermore, it was observed from nitrogen absorption-desorption using BET (Breneur-Emmet-Teller) method test that decreased pH tends to the specific surface area increase, and the pore size becomes decrease. The largest specific surface area observed at SiO_2 xerogel prepared at pH of 5.5 reached 400.10 m^2/g with a pore size of 4.5 nm. The pore sized for all cases was in the range of 4 ~12 nm, indicating that the SiO_2 xerogels were mesoporous. Pore size of the as-prepared silica affected the thermal stability property of the sample.

***doi**: 10.5829/ije.2021.34.07a.02*

1. INTRODUCTION

Finding new generations of energy sources highly recommended due to an increase in energy demand worldwide. The one alternative solution to overcome potential problems of worldwide energy deficiency is using nepw and renewable energy (NRE). One of the promising NRE technology for developing countries such as Indonesia is geothermal energy. Geothermal energy harnesses the energy generated below the surface of the earth via steam generation. Such energy source is available abundantly and does not depend on the availability of fossils-based fuel [1]. One of the problems faced while producing electricity from geothermal energy is solid and liquid wastes (sludge). [2]. The liquid waste was handled by re-injection into the earth's surface layer. [3]. The solid waste also needs to be handled due to the clogging effect of SiO_2 deposit inside the well tube, causing a reduction of power generation. The production of geothermal sludge (GS) in one of the geothermal power plants in Indonesia could exceed 165 tons per month while its utilization is minimal

SiO_2 has good chemical stability, is not soluble in water, and resistant to high temperatures. In general, SiO_2 can be obtained from inorganic materials and organic materials. The most widely known silica is silica TEOS (Tetraethyl orthosilicate) and TMOS (Tetramethyl orthosilicate), which have the advantage of being able to bind aggregate rock into monolithic material. However, this silica has a weakness, both of which have prices that are relatively expensive, difficult to obtain, and not environmentally friendly [4]. Based on the disadvantages

*Corresponding Author Institutional Email: hendriwidiyandari@staff.uns.ac.id (H. Widiyandari)

Please cite this article as: H. Widiyandari, P. Pardoyo, J. Sartika, O. A. Putra, A. Purwanto, L. Ernawati, Synthesis of Mesoporous Silica Xerogel from Geothermal Sludge using Sulfuric Acid as Gelation Agent, International Journal of Engineering, Transactions A: Basics Vol. 34, No. 7, (2021) 1569-1575

of silica source, TEOS and TMOS, the previous researchers conducted a lot of research to obtain alternative silica by utilizing natural ingredients that were not too useful before.

SiO_2 can be produced in several forms including fused quartz, crystals, irritated silica (or pyrogenic silica, colloidal SiO_2, SiO_2 gel and aerogels [5-7]. SiO_2 aerogel is a substance made from silicon and is the low density solid. It is formed from 99.8% air and rigid foam with a density of 3 mg cm^{-3}, low density, high surface area and low thermal conductivity [8]. With the advancing technology, SiO_2 from the waste could also be utilized as precursors of nanosilica [9,10] or mesoporous SiO_2 xerogel which could be applied as catalysts [11,12], adsorbents [13–15], ultrafiltration [16,17], drug delivery [18,19] and other applications [20-22].

Previous researchers have done a lot of research to synthesis SiO_2 using various plants such as rice husk waste [7,8]. However, there are still few researchers conducted research by utilizing geothermal sludge (GS) as an alternative raw material for SiO_2. SiO_2 from wastes or natural resources could be obtained through facile alkaline extraction and acidic precipitation method [23-25]. This process and its product characteristics are strongly affected by pH level during the synthesis. Muljani et al. [26] have succesfully produced mesoporous SiO_2 gels from geothermal sludge HCl and tartaric acid as gelling agents. In the sol-gel preparation, others hydrolysis catalysts were used such as HNO_3 and H_2SO_4 as gelation agent. It has also been reported that sulfated SiO_2 as solid acid catalyst is one of the modified silica gel products by reacting silica gel with sulfuric acid. The utilization of SO_4^{2-} anion display an increase in acidic properties of silica as well as physical properties [27, 28].

Different from the previous research, herein, we improved a modification of synthetic route to prepare mesoporous SiO_2 xerogel with geothermal sludge (GS) from Dieng Mountain as raw material. The modification is done to maximize the work of SiO_2 gel, especially for catalysts in the acidification process using H_2SO_4. We further investigate the effects of pH during the formation of mesoporous SiO_2 xerogel. This work is done as our contribution and effort to elevate the economic value of geothermal sludge (GS).

2. MATERIALS AND METHODS

2. 1. Material and Synthesis The materials used in this study were Mount Dieng Geothermal Sludge (GS), Indonesia. Solvents and reagent used were analytical grade of ethanol (Merck, 96%, distilled water, sodium hydroxide (NaOH, Merck), and sulfuric acid (H_2SO_4, Merck 96%).

The GS was dried in a oven at 80°C for 60 minutes, ground and screened using 80 and 200 meshes filters. 10 g of the prepared solid waste was extracted in reflux using 1.25 M NaOH (Merck) solution for an hour at 80°C and stirred continuously at 300 rpm. Extraction solution was cooled to room temperature and filtered using Whattman number 42 filter paper. The filtrate was sodium silicate. The sodium silicate solution was dissolved 5 times in demineralized water. 1 N solution of H_2SO_4 were added into the dissolved solution until the pH reached 10. The as-prepared solution was aged for 2 hours and reintroduced with 1N H_2SO_4 until the solution reaches various pH level (9, 8, 7, 6 and 5.5). After the desired pH level achieved, each solution was aged for 18 hours. The formed gel was dissolved using 300 mL of demineralized water and filtered by vacuum. The residue was dried in a oven at 80°C for 12 hours and washed three times using demineralized water. The washed residue was dried furtherly at 100°C for 24 hours. The SiO_2 xerogel was obtained. Figure 1 shows the schematic diagram of the material preparation.

2. 2. Material Characterization The functional groups of starting material and samples were analyzed using Shimadzu IR Prestige 21 FTIR (Fourier Transform Infra-Red). X-ray diffraction pattern of starting material and SiO_2 xerogel sampel were studied using Shimadzu XRD-7000 Maxima X-Ray Diffractometer (XRD) with CuKα (λ= 0.154 nm), scanning speed of 0.02°, 2θ angle range of 10° – 90°, and applied power of 3 kW. The surface area were analyzed using BET (Breneur Emmet Teller) method using ASAP 2020 V4.20E Surface Area Analyzer (SSA). The morphology was characterized by Scanning Electron Microscopy (JEOL, JCM-7000 NeoScope™ Benchtop SEM) with voltage of 10 kV. TG/DSC (Hitachi STA200RV) was used to characterize the thermal stability, adjusted to a temperature range from 25 to 1000 °C in an air atmosphere at a heating rate of 10 °C/min.

3. RESULTS

The formation of SiO_2 xerogel from PLTP (Geothermal Power Plant) Dieng by alkaline extraction are expressed

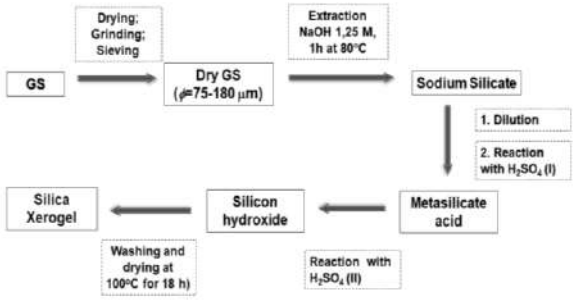


Figure 1. Experimental procedure for the preparation of mesoporous SiO_2 xerogel from geothermal sludge

in Equations (1), (2) and (3) [4, 29].

$$SiO_{2_{(s)}} + 2NaOH_{(aq)} \rightarrow Na_2SiO_{3_{(aq)}} + H_2O_{(l)} \quad (1)$$

During the extraction of silica, the silica reacts and dissolved into this alkaline NaOH solution to form sodium silicate (Na_2SiO_3) solution. To reduce the particle sizes of SiO_2 xerogel, demineralized water is added to the sodium silicate solution to increase its concentration. Thus, when precipitation occurs, small precipitate is formed due to slow nucleation process.

$$Na_2SiO_{3_{(aq)}} + H_2SO_{4_{(aq)}} \rightarrow H_2SiO_{3_{(aq)}} + Na_2SO_{4_{(aq)}} \quad (2)$$

The next process is a metasilicate acid formation by adding aqueous sulfuric acid H_2SO_4 into the sodium silicate solution. With the increasing amount of H^+, a polymerization reaction of silicate acid occurs.

$$H_2SiO_{3_{(aq)}} + H_2O_{(aq)} \rightarrow Si(OH)_{4_{(aq)}} \quad (3)$$

Gel formation occurs due to hydrolysis reaction of polymeric silicate acid with demineralized water while the xerogel were formed due to the drying of gel at 100°C for 24 hours. The porous xerogel is formed. Infra-red adsorption from difference group functions in geothermal sludge and prepared samples are shown in Figure 2. FTIR characterization of each samples are based on the study performed by Brinker and Scherer. The adsorption on wave number 3400 cm^{-1}, 1630 cm^{-1}, 1430 cm^{-1}, 1099 cm^{-1} and 950 cm^{-1} show stretch vibration of –OH from Silicanol (Si-OH) or H_2O, -OH from Si-OH, C-H, Si-O from Siloxane (Si-O-Si), Si-O from Si-OH respectively. Wavenumber of 790 cm^{-1} shows symmetrical Si-O from Si-O-Si stretch vibration, wavenumber of 450 cm^{-1} and 2300 cm^{-1} show bending vibration of Si-O=Si [30–32].

The increasing of adsorption Intensity at wavenumber 1630 cm^{-1} from sample S0 (waste), S1 (pH 9), S2 (pH 7) and S3 (pH 5.5) happened due to the proton content in the solution during the synthesis forming Si-OH. High level of pH causes the reaction tendency to form Si-O-Si (siloxane) proved by the increasing intensity at 1060 cm^{-1}. The hydrocarbon impurities (1430 cm^{-1}) are detected in geothermal sludge sample and the peak is shifted after the end of the process or when the xerogel is formed.

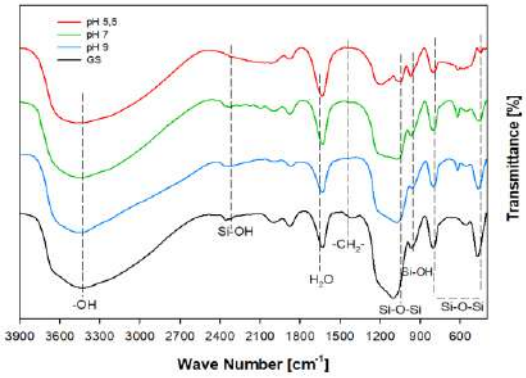


Figure 2. FTIR spectra of geothermal sludge and SiO_2 xerogel synthesized at various pH level

X-ray diffraction patterns of the samples are presented in Figure 3. The XRD pattern showed the amorphous nature of the SiO_2 xerogels. The broad XRD pattern was typical for an amorphous structure [33]. This is indicated by the appearance of a widening peak centered at an area of 2θ around 25˚ where silica with an amorphous structure gives a diffraction that widens at 2θ around 21-25˚. Impurities peaks (19° and 31°) of sodium sulfate (Na_2SO_4) is detected in sample 2 (pH 9) and sample 3 (pH 7). The presence of impurities could be caused by an unfinished cleaning process or salt entrapment in SiO_2 matrices [29]. Aside from the impurity peak, all peaks from all samples are well indexed to JCPDS 7757-82-628.

Surface analysis of each sample is performed by BET method and the result is presented in Table 1. Based on the result, the highest surface area, 400.10 m^2/g, is exhibited by sample synthesized at pH level of 5.5. With the increasing pH, the surface area of the sample is

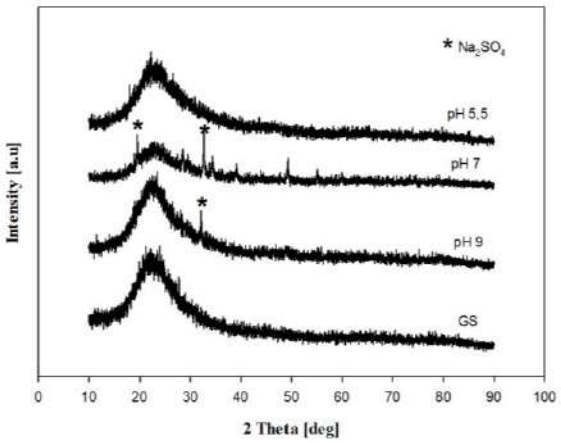


Figure 3. XRD Pattern of geothermal sludge and silica xerogel synthesized at various pH level

TABLE 1. BET analysis result of solid waste and samples synthesized at various pH level

Samples	Specific Surface Area [m^2/g]	Specific Pore Volume [cm^3/g]	Pore Diameter [nm]
GS	100.50	0.11	4.3
pH 9	159.55	0.47	11.8
pH 8	151.74	0.34	8.9
pH 7	223.44	0.48	8.7
pH 6	321.25	0.40	4.9
pH 5.5	400.10	0.45	4.5

decreased. This phenomenon is mainly caused by the slow formation of gel and slow nucleation of SiO_2 occurred at low pH level [29].

The amount of Nitrogen (N_2) adsorbed into the SiO_2 xerogel at various pH level synthesize shown in Figure 4. The isothermal adsorption of N_2 shows the pore volume capacity of SiO_2 xerogel. The highest pore volume is exhibited by the sample synthesized at pH 7 with the value of 0.48 cm^3/g. Based on the pore volume data stated in Table 1, there is no significant effect of pH toward the pore volume of SiO_2 xerogel. In the other hand, the pore diameter is significantly impacted by the pH level of synthesis. Figure 5 shows that the pore diameter is decreased with the decreasing pH level while the specific surface area is increasing. The specific surface area in samples aged at pH level 5.5 increased by 2.5 times compared to samples aged at pH 9 and 4 times compared to raw material (GS). The pore diameter of samples is within the range of 4-12 nm which could be concluded that the porous material has mesoporous size (2-50 nm).

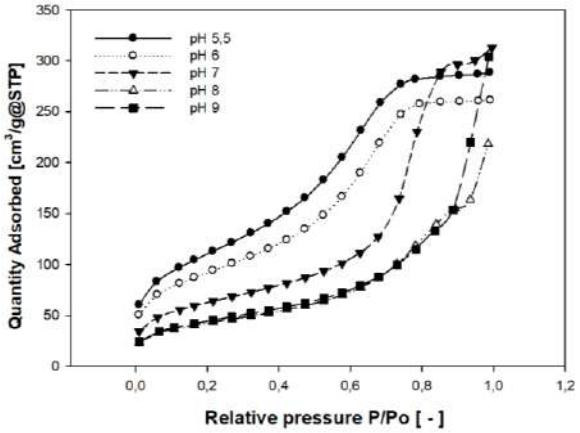


Figure 4. Isothermal adsorption of N_2 in SiO_2 xerogel synthesized at various pH level

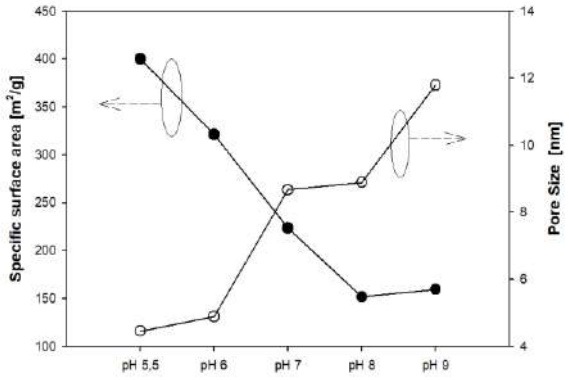


Figure 5. The effect of pH level during synthesis of silica xerogel on the surface area and pore diameter

A side from the pore diameter, the mesoporous SiO_2 xerogel is classified based on its adsorption-desorption profile. Figure 6 shows type-IV-like adsorption-desorption curve which could be applied for porous material hence the samples exhibits mesoporous size (2-50 nm). Type-IV curve initiated with slow adsorption due to the stronger intermolecular interaction between adsorbate than with the adsorbents. Then the curve is slightly bent due to the pore filling by the adsorbate [34–36].

Structure and properties of xerogels are influenced by gelation pH. Because condensation reactions are favored and hydrolysis reactions are restricted under alkali conditions formed particles are fewer in number but larger and denser than particles formed under acidic conditions. Since pH affects the coagulation process, experiments were run at different initial pH 1N H_2SO_4 until the solution reaches various pH level (9, 8, 7, 6 and 5.5). SEM results showed that SiO_2 particles are agglomerated at different pH values. The morphology of materials has distributed with diverse morphological structure. Figure 7f shows the chemical composition of the produced silica xerogel. The sample consist of Si (silicon) and O (oxygen).

TG/DSC was performed to characterize the thermal stability of SiO_2. Figure 8 shows the TG/DSC of SiO_2 prepared at different pH level at temperature from 25 to 1000°C in air atmosphere at heating rate of 10◦C/min.

The weight loss and endothermic peak below 150°C are primarily caused by the desorption of physically adsorbed water [37]. The weight loss and broad

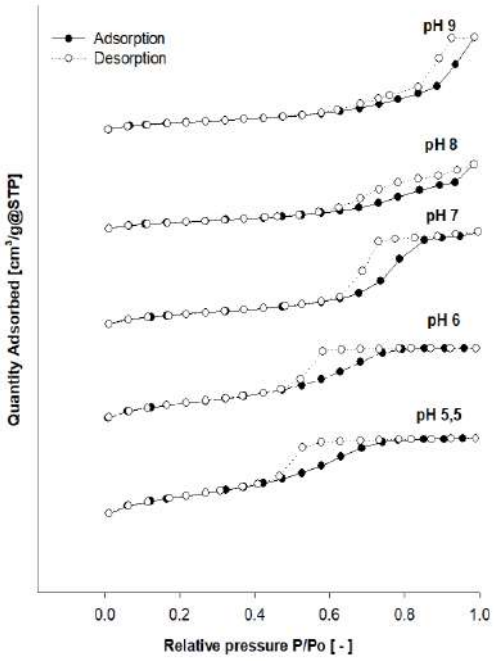


Figure 6. Adsorption-desorption curve of silica xerogel

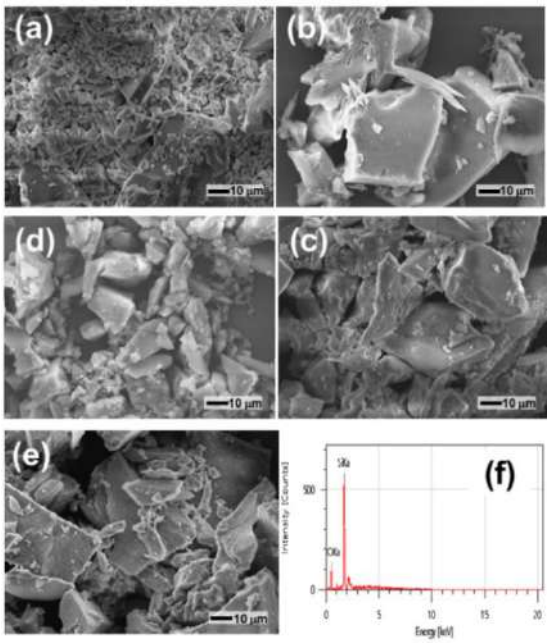


Figure 7. SEM images of SiO_2 particle at different pH values. (a) pH=5.5 ; (b) pH=6; (c) pH=7; (d) pH=8 (e) pH=9, and (f) EDX-spectra of sample

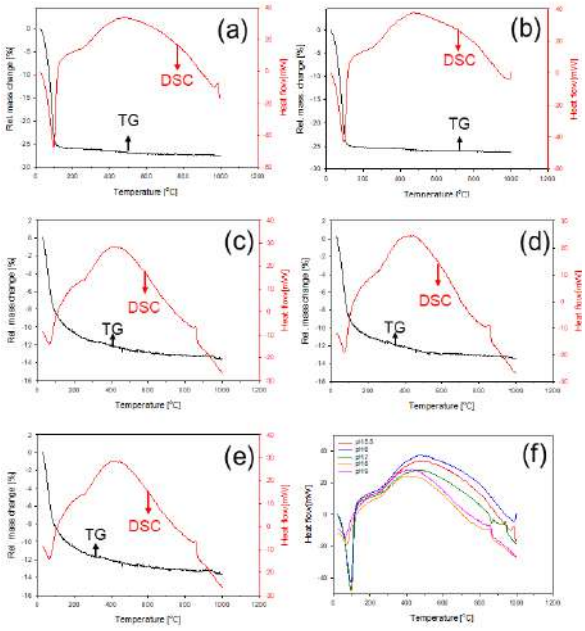


Figure 8. TGA/DSC thermograms of synthesized SiO_2 under different pH values. (a) pH= 5.5; (b) pH = 6; (c) pH=7; (d) pH=8; (e) pH=9, and (f) DSC curve of SiO_2 at different pH level

exothermic peak from 200 to 1000 °C might be attributed to the decomposition of the residual organics and the structural transition from amorf silica to crystalline silica. Obviously, at temperatures below 150, the weight loss in samples aged at pH 5.5 and 6 (acidic condition) was 25% and was greater than that in samples aged at pH 7 and above (neutral and alkaline condition), which was 8%. Figure 8f shows the DSC curve of the samples at various pH levels. Obviously, at temperatures below 150, the heat flow of the endothermic process in the sample aged at acidic conditions (5.5 and 6) was greater than that in the sample aged at pH level 7 and above. Meanwhile, at temperatures between 800-1000, exothermic peaks were seen, where the samples which were aged at acidic pH levels (5.5 and 6) showed exothermic peaks at higher temperatures compared to samples that were aged at neutral and alkaline conditions. This is possible due to the specific surface area and porosity of the sample.

4. CONCLUSIONS

SiO_2 xerogel was successfully synthesized from geothermal sludge by alkaline extraction and acidic precipitation using sulfuric acid. The characteristics of prepared material are strongly affected by the pH level of synthesis. From the FTIR analysis, silanol group and siloxane group is detected in every sample. XRD pattern shows amorphous structure of xerogel. Based on the BET analysis, at a lower pH level of synthesis (pH=5.5), larger surface area was produced (400.10 m^2/g). The pore diameter of each SiO_2 xerogel was categorized as mesoporous (2-50 nm). Pore diameter and specific surface area of the as-prepared silica affected the thermal stability property of the sample. SEM results showed that SiO_2 particles are agglomerated at different pH values.

5. ACKNOWLEDGMENT

This research was funded by Kementrian Riset dan Teknologi/Badan Riset dan Inovasi Nasional, Grant Number: 208/SP2H/AMD/LT/DRPM/2020 and 112/UN27.21/HK/2020.

6. REFERENCES

1. Abdolalipouradl. M, Khalilarya, S, and Jafarmadar. S, "Energy and Exergy Analysis of a New Power, Heating, Oxygen and Hydrogen Cogeneration Cycle Based on the Sabalan Geothermal Wells", ***International Journal of Engineering, Transactions C: Aspects***, Vol. 32, No. 3, (2019), 445-450, doi: 10.5829/ije.2019.32.03c.13.
2. Manzella. A, Bonciani. R, and Allansdottir. A, "Environmental and social aspects of geothermal energy in Italy", ***Geothermics***, Vol. 72, (2017), 232-248, doi: 10.1016/j.geothermics.2017.11.015.
3. Finster. M, Clark. C, Schroeder. J, and Martino. L, "Geothermal produced fluids: Characteristics, treatment technologies, and management options", ***Renewable and Sustainable Energy Reviews***, Vol. 50, (2015), 952-966, doi:

10.1016/j.rser.2015.05.059.

4. Affandi. S, Setyawan. H, Winardi. S, Purwanto. A, and Balqis. R, "A facile method for production of high-purity silica xerogels from bagasse ash", ***Advanced Powder Technology***, Vol. 20, (2015) 468-472, doi: 10.1016/j.apt.2009.03.008.
5. Brinker. C. J, and Scherer. W, "Sol-gel sciences,' in The Processing and the Chemistry of Sol-Gel Processing", ***Academic Press***, Vol. 26, No. 2, (1990), 211-212, doi: 10.1016/0254-0584(90)90039-d.
6. Mathieu. B, Blacher. S, Pirard. R, Pirard. J. P, Sahouli. B, and Brouers. F, "Freeze-dried resorcinol-formaldehyde gels", ***Journal of Non-Crystalline Sollids***, Vol. 212, No. 2-3, (1997), 250-261, doi: 10.1016/S0022-3093(97)00025-2.
7. Klvana. D, Chaouki. J, Repellin-Lacroix. M, and Pajonk. G. M, "A new method of preparation of aerogel-like materials using a freeze-drying process", ***Le Journal de Physique Colloques***, Vol. 50, (1989), C4-29-C4-32, doi: 10.1051/jphyscol.
8. Pajonk. G. M, "Drying Methods Preserving the Textural Properties of Gels", ***Le Journal de Physique Colloques***, Vol. 24, No. C4, (1989), C4-13-C4-22, doi: 10.1051/jphyscol:1989403.
9. Gurav. J. L, Rao. A. V, Rao. A. P, Nadargi. D. Y, and Bhagat. S. D, "Physical properties of sodium silicate based silica aerogels prepared by single step sol-gel process dried at ambient pressure" , ***Journal Alloys and Compounds***, Vol. 476, No. 1-2, (2009), 397-402, doi: 10.1016/j.jallcom.2008.09.029.
10. Shebl. M, Saif. M, Nabeel. A. I, and Shokry. R, "New non-toxic transition metal nanocomplexes and Zn complex-silica xerogel nanohybrid: Synthesis, spectral studies, antibacterial, and antitumor activities", ***Journal of Molecular Structure***, Vol. 1118, (2016), 335-343, doi: 10.1016/j.molstruc.2016.04.037.
11. Capeletti. L. B, do Carmo. M. A. M, Cardoso. M. B, and dos Santos. J. H. Z, "Hybrid silica based catalysts prepared by the encapsulation of zirconocene compound via non-hydrolytic sol-gel method for ethylene polymerization", ***Applied Catalysis A: General***, Vol. 560, (2018), 225-235, doi: 10.1016/j.apcata.2018.03.013.
12. Campos-Molina. M. J, Corral-Pérez. J. J, Mariscal. R, Granados. M. L, "Silica-poly(styrenesulphonic acid) nanocomposites as promising acid catalysts", ***Catalysis Today***, Vol. 279, (2017), 155-163, doi: 10.1016/j.cattod.2016.06.042.
13. Jaiboon. V, Yoosuk. B, and Prasassarakich. P, "Amine modified silica xerogel for H2S removal at low temperature", ***Fuel Processing Technology***, Vol. 128, (2014), 276-282, doi: 10.1016/j.fuproc.2014.07.032.
14. Salehi. S, and Anbia. M, "Investigation of Carbon Dioxide Adsorption on Amino-Functionalized Mesoporous Silica", ***International Journal of Engineering, Transactions C: Aspects***, Vol. 28, No. 6, (2015), 848-854 doi: 10.5829/idosi.ije.2015.28.06c.04.
15. Castillo. X, Pizarro. J, and Ortiz. C, "A cheap mesoporous silica from fly ash as an outstanding adsorbent for sulfate in water", ***Microporous and Mesoporous Materials***, Vol. 272, (2018), 184-192, doi: 10.1016/j.micromeso.2018.06.014.
16. Scherdel. C, and Reichenauer. G, "Highly porous silica xerogels without surface modification", ***The Journal of Supercritical Fluids***, Vol. 106, (2015), 160-166, doi: 10.1016/j.supflu.2015.08.016.
17. Echeverría. J. C, Estella. J, Barbería. V, Musgo. J, and Garrido. J. J, "Synthesis and characterization of ultramicroporous silica xerogels", ***Journal of Non-Crystalline Solids***, Vol. 356, No. 6-8, (2010), 378-382, doi: 10.1016/j.jnoncrysol.2009.11.044.
18. Adhikari. C, Mishra. A, Nayak. D, Chakraborty. A, "Drug delivery system composed of mesoporous silica and hollow mesoporous silica nanospheres for chemotherapeutic drug delivery", ***Journal of Drug Delivery Science and Technology***, Vol. 45, (2018), 303-314, doi: 10.1016/j.jddst.2018.03.020.
19. de Lima. H. H. C, Kupfer. V. L, Moisés. M. P, Rinaldi. J. C, Felisbino. S. L, Rubira. A. F, and Rinaldi. A. W, "Bionanocomposites based on mesoporous silica and alginate for enhanced drug delivery", ***Carbohydrate Polymers***, Vol. 196, (2018), 126-134, doi: 10.1016/j.carbpol.2018.04.107.
20. Vali. K. S, Murugan. B. S, Reddy. S. K, Farsangi. E. N, "Eco-friendly hybrid concrete using pozzolanic binder and glass fibers", ***International Journal of Engineering Transactions A : Basics***, Vol. 33, No. 7, (2020), 1183-1191, doi: 10.5829/ije.2020.33.07a.03.
21. Yakushkin. S. S, Balaev. D. A, Dubrovskiy. A. A, Semenov. S. V, Knyazev. Y. V, Bayukov. O. A, Kirillov. V. L, Ivantsov. R. D, Edelman. I. S, and Martyanova. O. N, "ε-Fe_2O_3 nanoparticles embedded in silica xerogel - Magnetic metamaterial", ***Ceramics International***, Vol. 44, No. 15, (2018), 17852-17857, doi: 10.1016/j.ceramint.2018.06.254.
22. Saputra. R. E, Astuti. Y, and Darmawan. A, "Hydrophobicity of silica thin films: The deconvolution and interpretation by Fourier-transform infrared spectroscopy", ***Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy***, Vol. 199, (2018), 12-20, doi: 10.1016/j.saa.2018.03.037.
23. Nuryono and Narsito, "Pengaruh Konsentrasi Asam Terhadap Karakter Silika Gel Hasil Sintesis dari Natrium Silikat", ***Indonesian Journal of Chemistry***, Vol. 5, No. 1, (2005), 23-30.
24. Guo. Q, Huang. D, Kou. X, Cao. W, Li. L, Ge. L, and Li. J,"Synthesis of disperse amorphous SiO_2 nanoparticles via sol-gel process," ***Ceramics International***, Vol. 43, No. 1, (2017), 192-196, doi: 10.1016/j.ceramint.2016.09.133.
25. Munasir, Triwikantoro, Zainuri. M, and Darminto, "Synthesis of SiO_2 nanopowders containing quartz and cristobalite phases from silica sands" , ***Materials Science- Poland***, Vol. 33, No. 1, (2015), 47-55, doi: 10.1515/msp-2015-0008.
26. Mujiyanti. D. R, Nuryono, and Kunart. E. S, "Sintesis dan karakterisasi silika gel dari abu sekam padi yang diimobilisasi dengan 3-(trimetoksisilil)-1-propantiol," ***Sains dan Terapan Kimia***, Vol. 4, (2010), 150-167
27. Utami. W. S, Herdianita. N. R, and Atmaja. R. W,"The Effect of Temperature and pH on the Formation of Silica Scaling of Dieng Geothermal Field, Central Java, Indonesia," ***Thirty-Ninth Workshop on Geothermal Reservoir Engineering Stanford University***, (2014), 24-26.
28. Lion. M, Maache. M, Lavalley. J. C, Ramis. G, Busca. G, Rossi. P. F, and Lorenzelli. V, "FT-IR study of the Brønsted acidity of phosphated and sulphated silica catalysts", ***Journal of Molecular Structure***, Vol. 218, (1990), 417-422.
29. Muljani. S, Setyawan. H, Wibawa. G and Altway. A, "A facile method for the production of high-surface-area mesoporous silica gels from geothermal sludge", ***Advanced Powder Technology***, Vol. 25, No. 5, (2014), 1593-1599. doi: 10.1016/j.apt.2014.05.012.
30. Jung. I. K, Gurav. J. L, Bangi. U. K. H, Baek. S, Park. H. H, "Silica xerogel films hybridized with carbon nanotubes by single step sol-gel processing", ***Journal of Non-Crystalline Solids***, Vol. 358, No. 3, (2012), 550-556, doi: 10.1016/j.jnoncrysol.2011.11.009.
31. Witoon. T, Tatan. N, Rattanavichian. P, Chareonpanich. M, "Preparation of silica xerogel with high silanol content from sodium silicate and its application as CO_2 adsorbent", ***Ceramics International***, Vol. 37, No. 7, (2011), 2297-2303, doi: 10.1016/j.ceramint.2011.03.020.
32. Lee. J, Kim. J, Lee. B. J, Lee. J, Lee. H. W, Hong. M, Park. H, Shim. D. I, Cho. H. H, Kwon. K, "Characterization of mesoporous silica thin films for application to thermal isolation layer", ***Thin Solid Films***, Vol. 660, (2018), 715-719, doi: 10.1016/j.tsf.2018.04.001.

33. Zhuang. Q, and Miller. J. M, "One-pot sol-gel synthesis of sulfated ZrO2-SiO_2 catalysts for alcohol dehydration," ***Canadian Journal of Chemistry***, Vol. 79, No. 8, (2001), 1220-1223, doi: 10.1139/cjc-79-8-1220.

34. Bleam. W. F, "Adsorption and Surface Chemistry", ***Soil and Environmental Chemistry***, (2012), 371-407, doi: 10.1016/b978-0-12-415797-2.00009-1.

35. Liu. C, and Komarneni. S, "Carbon-silica xerogel and aerogel composites", ***Journal of Porous Materials***, Vol. 1, No. 1, (1995), 75-84, doi: 10.1007/BF00486526.

36. Mosquera. M, J, Santos. D, d. los. M, Valdez-Castro. L, Esquivias. L, "New route for producing crack-free xerogels: Obtaining uniform pore size", ***Journal of Non-Crystalline Solids***, Vol. 354, No. 2-9, (2008), 645-650, doi: 10.1016/j.jnoncrysol.2007.07.095.

37. Yu. H, Tong. Z, Yue. S, Li. X, Su. D, and Ji. H, "Effect of SiO_2 deposition on thermal stability of Al_2O_3-SiO_2 aerogel," ***Journal of the European Ceramic Society***, Vol. 41, No. 1, (2021), 580-589, doi: 10.1016/j.jeurceramsoc.2020.09.015.

Persian Abstract

چکیده

مقدار زیادی لجن توسط نمک زمین گرمایی در نیروگاه زمین گرمایی Dieng تولید می شود که بیش از ۱۶۵ تن در ماه است. این لجن معمولاً استفاده نمی شود ، مگر اینکه در محل دفن زباله استفاده شود. رسوب (لجن) در درجه اول از سیلیس تشکیل شده است. هدف از این تحقیق سنتز سیلیس مزوپور xerogel (SiO2) از لجن زمین گرمایی ((GSو بررسی اثرات pH به عنوان تلاشی برای بالا بردن ارزش اقتصادی لجن از طریق استخراج قلیایی و به دنبال آن اسیدی شدن است. xerogel SiO2 با استخراج GS تهیه شد تا به سیلیکات سدیم تبدیل شود ((Na2SiO3با کمک NaOH پایه و با استفاده از H2SO4 به عنوان عامل ژل سازی رسوب می کند. تجزیه و تحلیل FTIR از xerogel SiO2 گروهی از سیلانول (Si-OH) و سیلوکسان (Si-O-Si) را نشان داد. تجزیه و تحلیل XRD نشان داد که xerogel SiO2 بی شکل است. علاوه بر این ، با استفاده از آزمون BET (Breneur-Emmet-Teller) از آزمون جذب–دفع نیتروژن مشاهده شد که pH کاهش می یابد و به سطح خاص افزایش می یابد و اندازه منافذ کاهش می یابد. بزرگترین سطح اختصاصی مشاهده شده در xerogel SiO2 تهیه شده با pH 5.5 به ۴۰۰.۱۰ مترمربع در گرم با اندازه منافذ ۴.۵ نانومتر رسید. اندازه منافذ برای همه موارد در محدوده ۴ ~ ۱۲ نانومتر بود ، نشان می دهد که xerogels SiO2 مزوپور بودند. اندازه منافذ سیلیس آماده شده خاصیت پایداری حرارتی نمونه را تحت تأثیر قرار می دهد.

IJE TRANSACTIONS A: Basics Vol. 34, No. 7, (July 2021) 1576-1585

International Journal of Engineering

Journal Homepage: www.ije.ir

Spectrophotometric Determination of Naproxen using Chitosan Capped Silver Nanoparticles in Pharmaceutical Formulation

A. Barazandeh[a], G. D. Najafpour[b], A. Alihosseini*[a], S. Kazemi[c], E. Akhondi[a]

[a] *Department of Chemical Engineering, Islamic Azad University, Tehran Central Branch, Tehran, Iran*
[b] *Faculty of Chemical Engineering, Babol Noshirvani University of Technology, Babol, Iran*
[c] *Department of Pharmacology and Physiology, School of Medicine, Babol University of Medical Sciences*

PAPER INFO

Paper history:
Received 31 January 2021
Received in revised form 08 April 2021
Accepted 01 May 2021

Keywords:
Chitosan
Naproxen
Silver Nanoparticle
UV-Vis Spectrophotometric Determination
Characterization

ABSTRACT

Naproxen (NP) is a non-steroidal anti-inflammatory drugs (NSAIDs) are widely used for the treatment of acute to chronic pain. The aim of this work was to develop a new, simple spectrophotometric method for determining the concentration of naproxen in pharmaceutical formulations by chitosan capped silver nanoparticles. The morphology and structure of the chitosan capped silver nanoparticles were determined by ultraviolet spectroscopy (UV), Fourier transform infrared spectroscopy (FT-IR), atomic force microscopy (AFM), dynamic light scattering (DLS) and zeta potential (ZP) under optimal conditions. The results showed that, synthesized chitosan capped silver nanoparticles were approximately with particle size of 100 nm poly dispersity index (PDI) 0.385 and strong anionic (-24.8mV) zeta potential at acidic condition. It was found that chitosan as a chiral selector was able to detect naproxen at optimal conditions. The method was successful to determine naproxen in drug tablets formulation. A relative standard deviation of 1.0% was determined for the analysis of naproxen in real samples. Validation of this method, including limit of detection and limit of quantification were accurately confirmed according to ICH instructions. Based on this method, the limit of detection (LOD) and the limit of quantification (LOQ) of naproxen were calculated 0.022 and 0.066 $molL^{-1}$, respectively. Analysis of statistical data, the reproducibility and accuracy of this method demonstrated that the use of this novel method is valuable and practical for determination of naproxen in pharmaceutical formulations.

***doi**: 10.5829/ije.2021.34.07a.03*

1. INTRODUCTION

Naproxen [(S) -2-(6-methoxy-2-naphthyl) propionic acid (Figure 1) is a chiral and nonsteroidal anti-inflammatory (NSAIDs) drug. Because of the (S)-naproxen isomer has just the favorable therapeutic effect; it is used in purest form for the treatment of diseases [1]. Naproxen is white and odorless powder, that is practically insoluble in water at low pH, but it can be soluble in water at alkaline condition (pH >8) [2]. It is an important member of the family of 2-aryl propionic acid derivatives; which commonly used for reducing moderate to severe pain, fever, inflammation, and stiffness owing to arthritis, gout, injury (such as fractures), tonsillitis and bursitis. It has also recently been reported that naproxen is effectively used in preventing the progression of bladder cancer [3]—Naproxen inhibits COX-1 and COX-2 enzymes can prevent the biosynthesis of prostaglandins [4]. Naproxen (NP) is widely used to treat acute to chronic pain most commonly in veterinary and human medicine because it is available without a prescription. Clinical and pharmacological analysis of this drug requires effective analytical measures for quality control. As well as recent overuse of naproxen has led to the classification of naproxen as an emerging contaminant in wastewater [5]. Therefore, finding a new accurate and fast economic identification method can be the answer to this problem to improve the cognition process. In order to determine naproxen in diagnostic laboratories, several analytical methods have been reported in the scientific

*Corresponding Author Institutional Email: Afsharalihosseini@iauctb.ac.ir (A. Alihosseini)

Please cite this article as: A. Barazandeh ,G. D. Najafpour, A. Alihosseini, S. Kazemi, E. Akhondi, Spectrophotometric Determination of Naproxen using Chitosan Capped Silver Nanoparticles in Pharmaceutical Formulation, International Journal of Engineering, Transactions A: Basics Vol. 34, No. 7, (2021) 1576-1585

literature such as spectrophotometry [6], electrochemical [7], high-performance liquid chromatography (HPLC) [8], etc. It can be mentioned that, HPLC method are highly accurate and sensitive in compared to other methods, but it is too expensive and time-consuming. Among the above methods, spectrophotometer is considered as the usual analytical method, due to its simplicity, availability, fast and low cost [9]. For that reason, a novel UV-spectrophotometric method for determining the amount of naproxen in tablets has been investigated. In this work, we developed a new, simple and economical spectrophotometric method for quantitative determination of concentration of naproxen in pharmaceutical formulations. There are number of investigations on spectrometric method for determination of naproxen. Maheshwari et al. [10] have examined a, simple UV-spectrophotometric method for determining the amount of naproxen tablets using niacinamide as hydrotropic solubilization additive. Mahmood et al. [11] have described a method for detecting the amount of naproxen in commercial tablets. In this method, first naproxen after modulation reaction to hydroxyl analog and then oxidation using potassium permanganate as an oxidant in an acidic environment, in commercial tablets was determined. Sastry et al. [12] have developed a spectrophotometric method to quantify naproxen. In their methods, naproxen was methylated to produce a color oxidative bonding with Gibb reagent in neutral phosphate buffer. Later on, Sastry et al. [13] have described another approach of spectrophotometric method based on the formation of a dye species with MBTH via oxidation of Ce (IV) or Fe (III) to determine naproxen. Pendari et al. [14] have examined the amount of naproxen in human plasma using spectrophotometry; in this approach, the non-ionized form of the drug with pure diethyl ether was extracted having yield of 94.6% and then naproxen was determined by measuring the peak amplitude in the second-order derivative spectrum at 328.2 nm. In this method, the amount of LOQ base on IUPAC was 2.42 μgmL^{-1}. Then, Khan et al. [15] have used 1-naphthylarnin and sodium nitrite to determine naproxen using spectrophotometer at wavelength of 480 nm. Duymus, et al. [16] have examined naproxen-induced complex with tetracyanoethyl ethane (TCNE), 2,3-dichloro-5,6-deciano-p-benzoquinone (DDQ), and p-chloranyl using a spectrophotometric device but its reagent was stable only for duration of five minutes. Alizadeh et al. [17] have described two spectrophotometric methods for determination of naproxen. These methods included the formation of ion pair complex with green bromocresol at 424 nm in the first method and the formation of aqueous bromothymol ion pair complex at 422 nm in the second method. Mahmood et al. [11] had proposed a new spectroscopic method which naproxen changed to hydroxy analog. They proposed a new spectroscopic method based on naproxen to hydroxy analog to determine naproxen. The modified compound was paralyzed using alkaline acid. In this approach, the limit of detection and the limit of quantitation were reported to be 0.0098 $\mu g \cdot mL^{-1}$and 0.03296 $\mu g \cdot mL^{-1}$, respectively. However, this method was much faster and easier than previous methods. In another investigation, Tashkhourian et al. [18] have explored spectrophotometric method based on silver nanoparticles with starch coating to determine naproxen which had limit of detection of 2.7×10^{-6} M. The objective of present work was to develop a fast spectrophotometric method using a new reagent of chitosan -capped silver nanoparticles for the detection of naproxen. Chitosan is a natural active polysaccharide derived from chitin. The excellent properties of chitosan such as biodegradability, biocompatibility, non-toxic, antibacterial and hydrophilic properties make it a good choice for enantiomer detection [19]. The amino groups in chitosan are important sites that can be used to synthesize metal nanoparticles [20]. In this regard, silver and gold nanoparticles are highly regarded by chemists due to their unique spectral and optical properties [21]. Gharibshahian [22] have investigated that the size, shape and structural quality of nanoparticles can affect their properties. These parameters are controlled by growth kinetics. The physical, optical, surface and size of nanomaterials make them attractive case study for exploring a new research field. Chitosan-capped silver nanoparticles are silver-based nanoparticles which is noticed as a group of structural and biological materials [9]. The proposed method is valuable for exact identification and quantification of the naproxen.

2. EXPERIMENTAL

2. 1. Materials Racemic naproxen and S- naproxen were purchased from Alborz Balk and Tehran Daru company, respectively. Silver nitrate ($AgNO_3$), sodium borohydride ($NaBH_4$), acetic acid (CH_3COOH), citric acid, sodium hydroxide and sodium citrate were purchased from Merck (Darmstadt, Germany). Chitosan was supplied by Sigma-Aldrich company (St. Louis, Mo, USA).

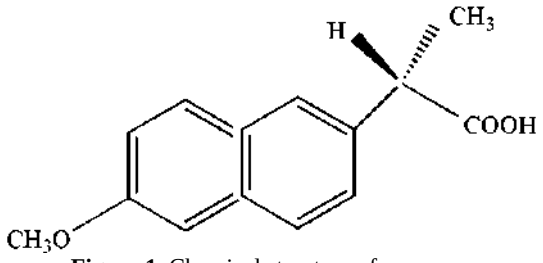


Figure 1. Chemical structure of naproxen

2. 2. Apparatus Chitosan-capped silver nanoparticles were measured by spectrophotometer UV VIS model SPEKOL 1500 Analytik Jena AG; Fourier transform infrared (FT-IR) spectra were analyzed using KBr pellets by a FT-IR spectrometer (WQF - 510A, China). In addition, atomic force microscopy (AFM) model easy scan 2 (Liestal, Switzerland), dynamic light scattering (DLS) and zeta potential (HORIBA - SZ100), were used to characterized nanoparticle.

2. 3. General Procedure To determine naproxen, chitosan-coated silver nanoparticles were first synthesized. Morphology and structure of chitosan-coated silver nanoparticles using ultraviolet (UV) spectroscopy, Fourier transform infrared spectroscopy (FT-IR), atomic force microscopy (AFM), DLS (dynamic light scattering) and ZP (zeta potential) were determined under optimal conditions. Using silver nanoparticles coated with chitosan, the amount of naproxen was determined by spectrophotometry in tablets.

2. 4. Synthesis of Chitosan Capped Silver Nanoparticles Using a method by reducing $NaBH_4$ from $AgNO_3$ in the presence of chitosan, with slight modification according to the previously reported method [23], silver nanoparticles capped with chitosan (Ag NPs) were synthesized. In the first step, 25 mg chitosan was dissolved in 50 mL of 4% aqueous acetic acid and then sonicated for 5 min. In the second stage, 20 mL of the solution was mixed with 50 mL of aqueous $AgNO_3$ solution (6 $mmol.L^{-1}$) and reaction was performed at room temperature under vigorous magnetic stirring for 30 min. In the third stage, a freshly prepared aqueous solution of $NaBH_4$ (1 mL, 58mmol. L^{-1}) was added dropwise to the reaction mixture. After stirring with a magnetic stirrer for 2 hours, observing the color change of the solution to yellowish brown is a sign of synthesis (Ag NPs) in the solution. By increasing the pH of the solution about 12 using NaOH (0.5M), chitosan silver capped nanoparticles were precipitated and separated by centrifugation (12000 rpm in 30 min). After twice washing with deionized water, chitosan silver nanoparticles were dried at 40 °C and stored at dry chamber and room temperature for subsequent experiments.

2. 5. Procedure for Calibration Curves Chitosan capped silver nanoparticles (CS-AgNPs) Method: The equivalent of 100 µL at concentrations of 25-500 µmol mL^{-1} naproxen was transferred to a series of 5 mL bottles. In each bottle, 4.5 mL buffer solution (pH = 4) and 0.5 mL of CS -Ag nanoparticles at concentration of 0.5 $g.L^{-1}$ were added. The prepared standard was placed at room temperature (25 °C) for 8 minutes. The absorption of naproxen in a CS -Ag nanoparticles was measured by UV-vis spectrometer. To obtain the standard calibration curve, the absorption values versus the naproxen concentrations were plotted. Figure 2 illustrates absorption data were perfectly fitted for naproxen concentrations in the range of 25- 500 µmol mL^{-1}.

2. 6. Procedure for The Assay of Tablets Ten samples of 500 mg naproxen tablets from Tehran Daroo Company were used for the analysis of naproxen in tablets by using the proposed methods. The samples were well ground and turned into fine powder. The equivalent of 500 mg of naproxen powdered tablets, after careful weighing, was transferred to a 50 mL glass and dissolved in the 50mL amount of methanol, filtered through a Whatmann paper filter No. 41 and diluted in a 100 mL calibration flask with methanol to a specified volume.

3. RESULTS AND DISCUSSION

3. 1. Validation Parameters of Synthesis Chitosan -Capped Silver Nanoparticles Validation of chitosan-capped silver nanoparticles were characterized by ATM, DLS and ZP, FT-IR, and UV-Vis spectroscopy.

3. 2. UV-Vis Spectroscopy and Mechanism of Action In order to recognize enantiomers component, an interaction in three configuration dependent points are necessary based on researched. The structure of aromatic ring of naproxen and the intermolecular hydrogen bond between naproxen and chitosan-capped silver nanoparticles could possibly provide this "three-point" interaction. It was supposed that intermolecular hydrogen bond between naproxen and chitosan-capped silver nanoparticles and electrostatic interaction could possibly give this recognition. Therefore, this can be used in developing new reagents for determining via colorimetric assays [18]. In this way, the use of UV-vis spectroscopy is a very useful and reliable technique to confirm the initial properties of synthesized nanoparticles. In order to monitor the synthesis and stability of silver nanoparticles

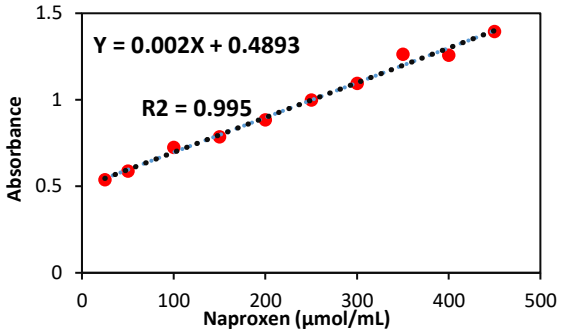


Figure 2. Calibration curve of naproxen with chitosan @Ag nanoparticles

[19] as well as the relationship between the plasmonic properties of silver nanoparticles and their morphology, a fast and easy UV -Vis spectroscopy method was developed for simultaneous monitoring the synthesis [24]. The synthesized chitosan capped silver nanoparticle solution was analyzed by UV-vis spectroscopy in the range of 270-620 nm. The UV-Visible spectrum for chitosan-capped silver nanoparticle is shown in Figure 3. The λmax of chitosan-capped silver nanoparticles was exactly observed at wavelength of 417 nm, which is the characteristic plasmon resonance band (SPR) of silver nanoparticles.

3. 3. FT-IR Spectra of Chitosan and Chitosan-Capped Silver Nanoparticles Figure 4 shows the FT-IR spectrum of standard chitosan and chitosan-capped silver nanoparticles. In the case of chitosan (Figure 4) as shown in this figure, the broad band marked at 3434 cm^{-1} is due to the overlap between the tensile vibrations O-H and N-H, both of them are semi-polysaccharides. In addition, the specified band in 2927 cm^{-1} is aliphatic which indicating of C-H bonds. Also, the tensile vibrations of the amide I, II and III bonds have caused the signals to appear at 1658 and 1600 cm^{-1} which is shown in Figure 4. However, silver ions were reduced to silver nanoparticles when liquid $NaBH_4$ added to $AgNO_3$ for the synthesized chitosan-capped silver nanoparticles. The FTIR spectrum of CS-Ag nanoparticles indicated that the nanoparticles exhibited absorption peaks at 3426, 1595, 1410, 656 cm^{-1} (Figure 4). The tensile bond pressures O - H and N - H in 3434 cm^{-1}originated from the primary hydroxyl and amino groups in chitosan were changed to 3426 cm^{-1}, which is indicating silver ions were chelated with both of amino and hydroxyl groups of chitosan [18]. The band at 1595 cm^{-1} corresponded to amide I due to carbonyl stretch in proteins. In addition, the peak at 1410 cm^{-1} refer to the C–N stretching in CS -Ag nanoparticles [9].

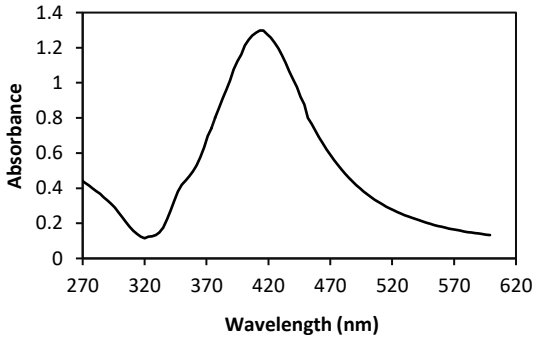


Figure 3. UV-Visible spectrum of chitosan @Ag nanoparticles

3. 4. Atomic Force Microscopy Surface morphology of chitosan capped silver nanoparticle was evaluated by AFM. AFM is provided an accurate information about the distribution of size and the shape of the particles [25]. Figure 5a shows the size distribution analysis of chitosan-coated silver nanoparticles by AFM which is in the range of 3–71.2 nm with an average size of 20±0.15.

It can be concluded that more than 50% of silver nanoparticle is in the range of 20 nm and indicating formation of nanoparticle in desirable size. The surface topography of chitosan capped silver nanoparticle and its properties in 2D and 3D picture are shown in Figure 5b. Also, the spherical shape of nanoparticles is clearly shown in these images which is proper and applicable.

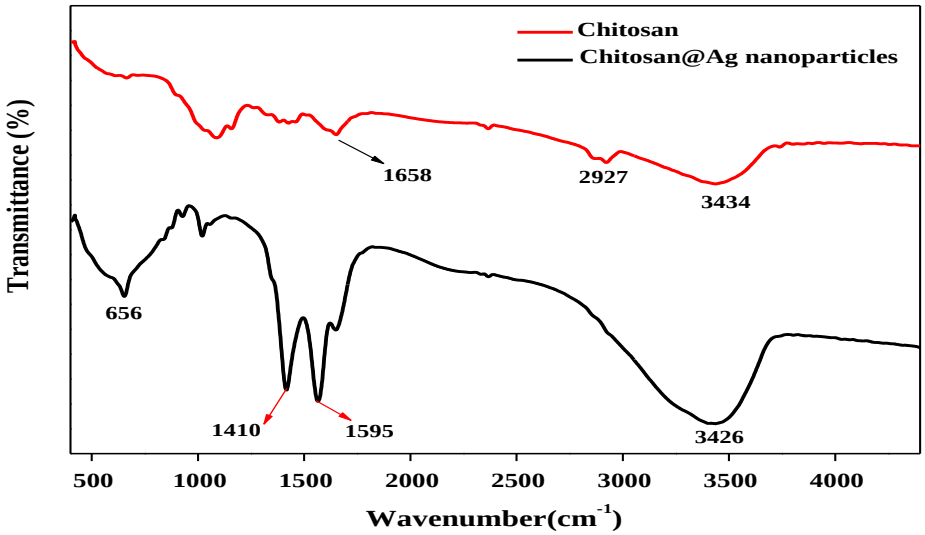


Figure 4. FT-IR spectra of chitosan (red curve) and chitosan @Ag nanoparticles (black curve)

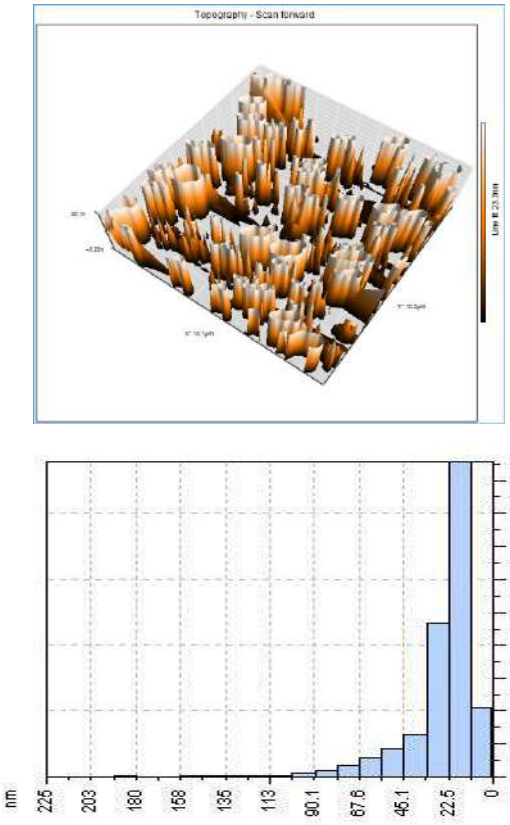


Figure 5. AFM results of chitosan @Ag nanoparticles: Surface topography (Left) and size distribiution (right)

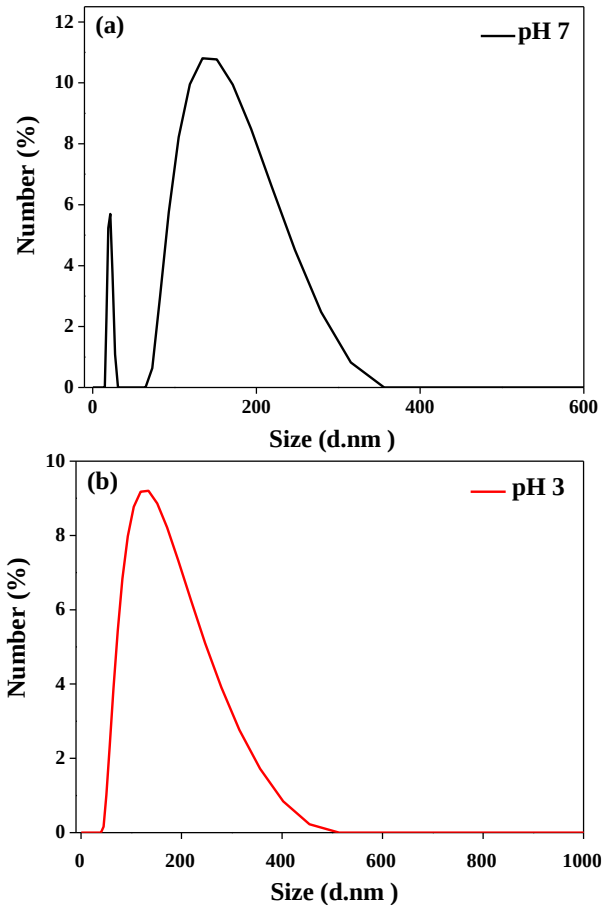


Figure 6. DLS spectra of chitosan @Ag nanoparticles at two pH values of (a) 7 and (b) 3

3. 5. DLS and Zeta Potential In order to investigate the scattering properties of nanoparticles and their hydrodynamic radius characteristics, DLS and zeta potential analysis have been used [19]. Size distribution and poly-dispersity index (PDI) are the two most important factors in the DLS test which are dependent on physio-chemical properties of nanoparticle in solvent. These experiments were carried out at two different pH to examined its effectiveness on nanoparticle size distribution and its poly-dispersity. It can be mentioned that size distribution lower than 150 nm is acceptable as nanoparticle. In addition, based on literature, PDI value in the range of 0.1 to 0.4 is appropriate as poly-dispersed and stable solution [19]. However, Figures 6a and 6b represented the results of DLS analysis for chitosan capped silver (CS-Ag) nanoparticles at pH of 7 and 3, respectively. Because of chitosan insolubility at neutral pH, nanoparticle size was increased (448 nm) and its DPI was 0.618 which is not acceptable.

These results indicate that CS –Ag nanoparticle reagent has resulted in agglomeration at neutral pH, so it can cause limitation of the reagent applications. However, the obtained result at pH 3 showed that the mean diameter of CS- Ag nanoparticles was 116.5 nm and PDI was 0.385, which is less than 0.4. These finding indicated that a homogeneous dispersion of the CS-Ag nanoparticles produced at acidic pH. It can be noted that chitosan polysaccharide is completely soluble at acidic pH. Based on literature, these results can improve nanoparticle size distribution and PDI. In addition, the zeta potential test was used to measure the electrostatic potential at the electrical double layer surrounding a nanoparticle in the solution. Based on literature, nanoparticles with a zeta potential between -10 and +10 mV were categorized in the region of neutral, while nanoparticles having zeta potentials of greater than +30 mV or less than -30 mV were indicated strongly cationic and strongly anionic strengths, respectively. On the other hand, high zeta potential (greater than ±30 mV) of nanoparticles demonstrated that its great stability or its strong resistance against the aggregation, flocculation, and coagulation. Several factors are affected the zeta potential of CS-Ag nanoparticles, for instance particle size and structure, ionic strength, pH of the solution, etc. Therefore, zeta potential of CS-Ag nanoparticles was measured at pH 7 and 3 (see Figures 7). As can be seen,

the zeta potential is approximately neutral (14.4 mV) at pH 7 while it is strong anionic (-24.8mV) at pH 3 which is refer to chitosan surface charge and its functional groups. Therefore, the new proposed reagent exhibited excellent stability at acidic pH at room temperature.

3. 6. Naproxen Determination and Validation Method

3. 6. 1. linearity of the Calibration The first step in the method development is the linearity of the calibration function. The correlation of coefficient is often used as a factor that shows the linearity of the calibration curve. The calibration curve for naproxen absorption with CS-Ag nanoparticles was linear in the concentration range of 25 to 500 µmol mL^{-1} with a correlation coefficient of 0.995. (see Figure 2). The linear equation obtained by least squares regression method were y= 0.4893+0.002X that y is the absorption intensity (in arbitrary units) and x is the concentration of naproxen in (µmol mL^{-1}) which constant obtained values and R square correlation coefficient are equal to B = 0.002, A = 0.4893 and R^2 = 0.995, respectively. It can be concluded that the proposed reagent and analysis method had been correctly functioned.

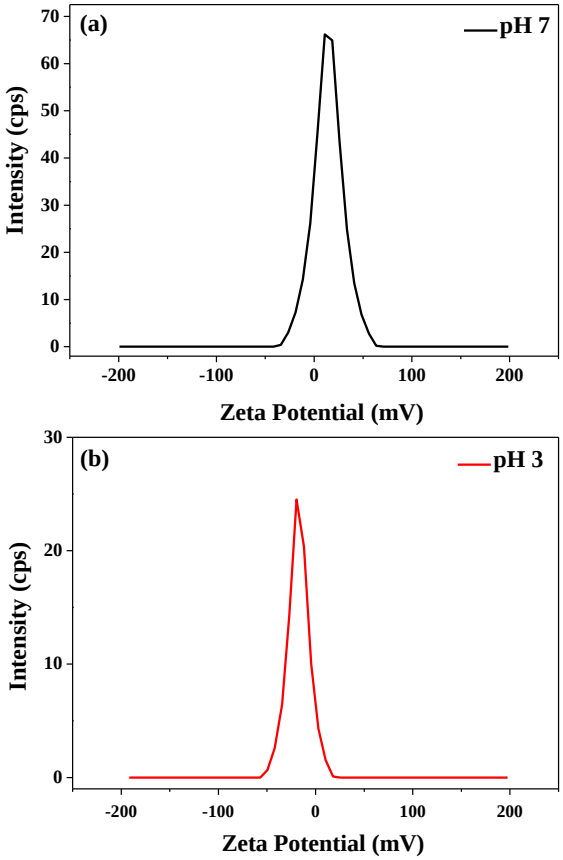


Figure 7. Zeta potential of chitosan @ Ag nanoparticles at two pH values of (a) 7 and (b) 3

3. 6. 2. Absorption Spectrophotometric As shown in Figure 8, the absorption spectra of naproxen and the complex of naproxen and chitosan-capped silver nanoparticles were measured in the range of 230-630 nm. Naproxen at concentration of 50 µL (0.001 M) in methanol showed two absorption peaks which were observed at 262 and 273 nm. The effect of reaction time and pH of the complex of chitosan capped silver nanoparticles and naproxen for finding optimum conditions with maximum stability and sensitivity in absorption was investigated.

3. 7. Optimal Conditions for Detection of Naproxen with Chitosan Silver Nanoparticles

3. 7. 1. Effect of Reaction Time To finding a reagent with minimum reaction time is essential parameter in clinical laboratory. So, this is important for developing a new reagent that reaction between them reach to stable condition in minimum time and stay constant after passing time. Therefore, the effect of the reaction time between chitosan capped silver nanoparticles (reagents) and naproxen on naproxen adsorption was investigated at different time intervals. Initially, the absorption decreased and gradually peaked over time after 8 minutes, then reached to a steady state value after 10 minutes and remained stable (see Figure 9). It can be concluded that the optimum reaction time of 8 min was defined. The reaction time of 8 min was used in the subsequence experiments. This short reaction time is a positive point for a new reagent which reach to stable condition.

3. 7. 2. Effect of pH Chitosan is a cationic polyelectrolyte biopolymer which can be dissolved in acidic pH. The reason is most probably caused by amine groups on chitosan main chain. Therefore, it is soluble in aqueous phase at acidic condition where amine groups

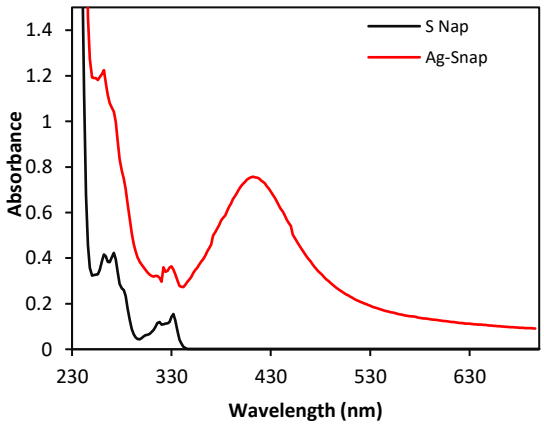


Figure 8. UV-Vis spectrum of pure naproxen and complex of naproxen with chitosan @Ag nanoparticle

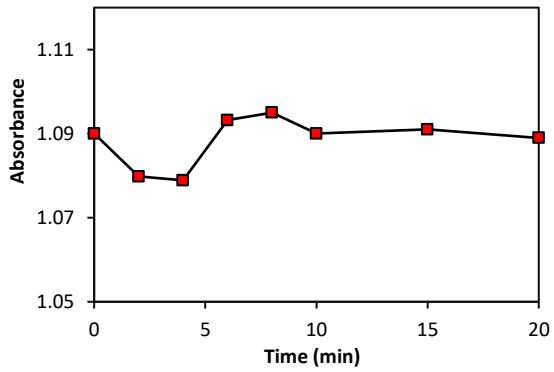


Figure 9. Effect of reaction time on the formation of naproxen- nanoparticles complex

are positively charged. However, chitosan at pK value of 6.5 possess neutral charges. Because of chitosan changeability in various pH, determining appropriate pH is critical matter for silver nanoparticle which capped via chitosan [26, 27]. In addition, based on our experiments chitosan silver nanoparticles are not stable at pH <3.0 and pH> 8.0 [18]. Therefore, the effect of pH in the solution on the stability of chitosan silver nanoparticles (reagent) and naproxen absorption was studied in the range of 3-6. As can be seen the spectrophotometric spectra of samples in Figures 10a and 10b, the maximum intensity of naproxen absorption was observed at pH = 4.0; consequently, the optimum pH for the rest of experiments was chosen to be 4.

3. 7. 3. Determination of Naproxen The existing solution was diluted using methanol at appropriate concentrations in the working range. Then the amount of naproxen in the tablets were measured using the proposed method and the new reagent CS -Ag nanoparticles (see Table 1). The reliability of the amount of naproxen calculated based on the standard calibration curve and reported in Table 1. For examination of the proposed reagent accuracy, different concentration of tablets with 20 μgmL^{-1} variations were prepared and analyzed by the mentioned method. The achieved results and the amount of recovery values are summarized in Table 2, respectively.

3. 7. 4. Detection Limits (LOD) and Quantitative Limit (LOQ) The Limit of detection (LOD) and quantification (LOQ) are the two most important characteristics in validation of a new method. The lowest concentration of analyte in a sample which can be determined with the acceptable accuracy in the defined test conditions is called the detection limit (LOD) [28]. Normally a reagent is not able to exact measure the analyte concentration; although, this claim may have declared in the reagent package. In order to distinguish

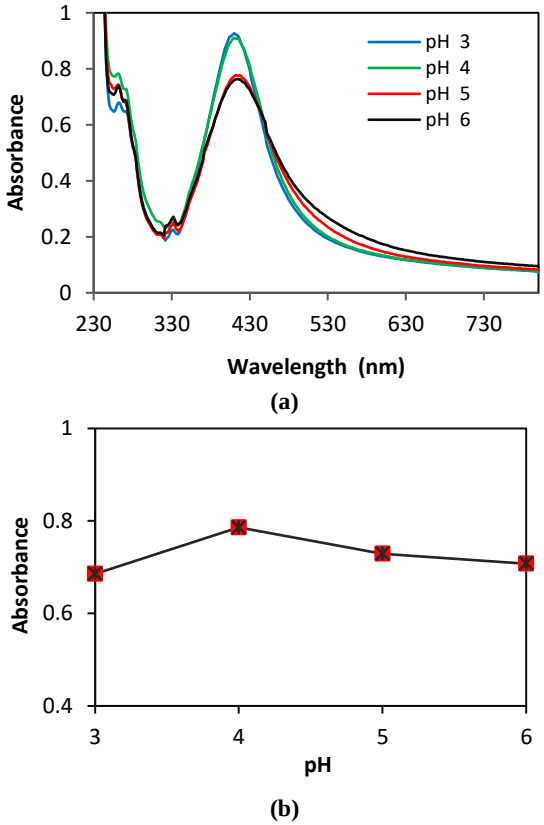


Figure 10. UV-Vis spectrum (a) and absorbance of naproxen at variuos pH values

between the analytical signal and the analytical noise in the absence of analyte, there must be a sufficient concentration of analyte [28]. So, the lowest level of analyte concentration which produced a detectable response is called the detection limit, which is usually three times the noise level. Therefore:

$$LOD = 3\sigma_A/B \quad (1)$$

where the σ_A standard deviation from the fitted regression line is calculated and the B is slope of calibration curve. Standard deviation of the y-intercept of the regression line Sy/x, i.e. standard error of estimate. In the latter method, only the error in the intercept is considered. A better alternative is the equation proposed by Winefordner and Long expressed as follows [2]:

$$LOD = k\,[\sigma^2_{b1} + \sigma^2_A + (A/B)^2\,\sigma^2_B\,]^{1/2} \quad (2)$$

TABLE 1. Determination of naproxen in tablet dosage form using proposed method

Method	Naproxen labeled amount (mg)	Found (mg)	Recovery (%)
chitosan -capped silver nanoparticles	500	503.44	100.688

TABLE 2. Determination of naproxen in tablet dosage form using the standard addition technique

Sample	Taken (µg mL^{-1})	Added (µg mL^{-1})	Found (µg mL^{-1})	Recovery (%)
Naproxen 500 (mg)	100	0	99.67	99.670
		20	119.43	99.525
		40	137.95	99.535
		60	159.56	99.725

TABLE 3. Comparison of the developed method with reported methods

Method	Linearity Coefficient	Relative Error	Ref.
HPLC	0.999	-	[30]
Capillary electrophoresis	0.9969	5	[25]
Spectroscopy	0.9988	3.5	[31]
Spectroscopy	0.999	1.4	[18]
Spectroscopy	0.995	1	This method

Where k is a constant value, which is usually equal to 3, also b_l is the standard deviation of a set of iterations from the empty sample solutions, σ^2_B is the standard deviation of the slope, σ^2_A standard deviation from the fitted regression, the B and A are slope and coefficient of calibration curve respectively. The limit equations (LOD) and quantitative limit (LOQ) according to the International Council for Harmonisation of Technical Requirements for Human Pharmaceuticals Use (ICH) given as follows [29]:

$$LOQ=3.3\sigma/s \quad (3)$$

$$LOQ=10\sigma/s \quad (4)$$

Where the σ standard deviation from the fitted regression line is calculated and the S is slope of calibration curve. In this work, the detection limits (LOD) and quantitative limit (LOQ) for the proposed method based on Equations 3 and 4 were calculated 0.022 $mol.L^{-1}$ and 0.066 $mol.L^{-1}$, respectively. It can be noted that the obtained results of LOD and LOQ are in defined range and so acceptable.

3. 8. Validation Method The characteristics of synthesized chitosan capped silver nanoparticles (reagent) were determined by ultraviolet spectroscopy (UV), Fourier transform infrared spectroscopy (FT-IR). The results were completely in line with the data reported by Tashkhourian et al. [18] and Jafari et al. [9]. In addition, its morphology and size distribution were confirmed by atomic force microscopy (AFM), dynamic light scattering (DLS) and zeta potential (ZP) analyses in optimal conditions. It can be concluded that the proposed reagent is stable in acidic condition which is proper for determination of naproxen. In order to confirm the validity of the new developed method, limits of detection (LOD) and quantitation (LOQ), reproducibility, reproducibility and its accuracy were examined according to ICH guideline. Also, linearity of applied method was investigated in the concentration of 25-500 $\mu mol.mL^{-1}$ naproxen, and correlation coefficients of 0.995 was achieved. In addition, a computational method was developed to define the limits of detection (LOD) and quantitation (LOQ) which calculated 0.022 and 0.066 $mol.L^{-1}$, respectively. The performance of the proposed method was compared with the merit of the developed method with reported methods in literature (see Table 3). {Absalan, 2012 #46}

4. CONCLUSION

This research has demonstrated that chitosan capped silver nanoparticles are a suitable probe for the detection of naproxen using Uv-vis spectrophotometry. Based on the results, the accuracy of this method was very good for determining the amount of naproxen in tablets. Analysis of statistical data, the reproducibility and accuracy of this method demonstrated that the use of this novel method is valuable and practical for determination of naproxen in pharmaceutical formulations. The linear dynamic range and precision, accuracy of developed method chitosan capped silver *nanoparticles* indicate that the figures of merit were comparable with most reported methods for the determination of naproxen. The spectrophotometric method using chitosan capped silver nanoparticles was quite suitable and applicable for the detection of naproxen in bulk and drug formulation.

5. REFERENCES

1. Gilani, S.L., Najafpour, G.D., Heydarzadeh, H.D. and Moghadamnia, A., "Enantioselective synthesis of (s)-naproxen using immobilized lipase on chitosan beads", ***Chirality***, Vol. 29, No. 6, (2017), 304-314. DOI: 10.1002/chir.22689
2. Damiani, P., Bearzotti, M. and Cabezón, M.A., "Spectrofluorometric determination of naproxen in tablets", ***Journal of Pharmaceutical and Biomedical Analysis***, Vol. 29, No. 1-2, (2002), 229-238. DOI:10.1016/S0731-7085(02)00063-8
3. Duggan, K.C., Walters, M.J., Musee, J., Harp, J.M., Kiefer, J.R., Oates, J.A. and Marnett, L.J., "Molecular basis for cyclooxygenase inhibition by the non-steroidal anti-inflammatory drug naproxen", ***Journal of Biological Chemistry***, Vol. 285, No. 45, (2010), 34950-34959. DOI: 10.1074/jbc.M110.162982
4. Gumułka, P., Dąbrowska, M. and Starek, M., "Microanalysis of selected nsaids using the spectrophotometric method", ***Engineering***, Vol. 1, No. 2, (2020), 211-221. DOI: 10.3390/eng1020014
5. Qian, L., Thiruppathi, A.R., Elmahdy, R., van der Zalm, J. and Chen, A., "Graphene-oxide-based electrochemical sensors for the sensitive detection of pharmaceutical drug naproxen", ***Sensors***, Vol. 20, No. 5, (2020), 1252. DOI: 10.3390/s20051252
6. Ahmadi, M., Madrakian, T. and Afkhami, A., "Enantioselective solid phase extraction prior to spectrofluorometric determination: A procedure for the determination of naproxen enantiomers in the presence of each other", ***RSC Advances***, Vol. 5, No. 7, (2015), 5450-5457. DOI: 10.1039/C4RA10405F
7. Guo, H.-S., Kim, J.-M., Chang, S.-M. and Kim, W.-S., "Chiral recognition of mandelic acid by l-phenylalanine-modified sensor using quartz crystal microbalance", ***Biosensors and Bioelectronics***, Vol. 24, No. 9, (2009), 2931-2934. DOI: 10.1016/j.bios.2009.02.002
8. Mikami, E., Goto, T., Ohno, T., Matsumoto, H. and Nishida, M., "Simultaneous analysis of naproxen, nabumetone and its major metabolite 6-methoxy-2-naphthylacetic acid in pharmaceuticals and human urine by high-performance liquid chromatography", ***Journal of Pharmaceutical and Biomedical Analysis***, Vol. 23, No. 5, (2000),917-925. DOI: 10.1016/S0731-7085(00)00365-4
9. Jafari, M., Tashkhourian, J. and Absalan, G., "Chiral recognition of tryptophan enantiomers using chitosan-capped silver nanoparticles: Scanometry and spectrophotometry approaches", ***Talanta***, Vol. 178, (2018), 870-878. DOI: 10.1016/j.talanta.2017.10.005
10. Maheshwari, R., Lakkadwala, S., Vyas, R. and Ghode, P., "Spectrophotometric determination of naproxen tablets using niacinamide as hydrotropic solubilizing additive", ***Journal of Current Pharmaceutical Research***, Vol. 4, (2010), 11-14.
11. Mahmood, H.S. and Al-Sarraj, T.Z., "Two novel spectrophotometric methods for determination of naproxen via a modulation to hydroxy analog", ***Current Applied Science and Technology***, (2020), 295-309. DOI 10.14456/cast.2020.17
12. Sastry, C. and Rao, A.R.M., "Application of folin-ciocalteu reagent for the spectrophotometric determination of some nonsteroidal antiinflammatory agents", ***Journal of Pharmacological Methods***, Vol. 19, No. 2, (1988), 117-125. DOI: 10.1016/0160-5402(88)90032-0
13. Sastry, M., Patil, V. and Sainkar, S., "Electrostatically controlled diffusion of carboxylic acid derivatized silver colloidal particles in thermally evaporated fatty amine films", ***The Journal of Physical Chemistry B***, Vol. 102, No. 8, (1998), 1404-1410. DOI: 10.1021/jp9719873
14. Panderi, I. and Parissi-Poulou, M., "Second-derivative spectrophotometric determination of naproxen in the presence of its metabolite in human plasma", ***Analyst***, Vol. 119,No. 4, (1994), 697-701. DOI: 10.1039/AN9941900697
15. Khan, I.U., Aman, T., Ashraf, A. and Kazi, A.A., "Spectrophotometric determination of naproxen in pure and pharmaceuticalpreparations", ***Analytical letters***, Vol. 32, No. 10., (1999) 2035-2050. DOI: 10.1080/00032719908542950
16. Duymus, H., Arslan, M., Kucukislamoglu, M. and Zengin, M., "Charge transfer complex studies between some non-steroidal anti-inflammatory drugs and π-electron acceptors", ***Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy***, Vol. 65, No. 5, (2006), 1120-1124. DOI: 10.1016/j.saa.2006.01.044
17. Alizadeh, N. and Keyhanian, F., "Simple, sensitive and selective spectrophotometric assay of naproxen in pure, pharmaceutical preparation and human serum samples", ***Acta Poloniae Pharmaceutica-Drug Research***, Vol. 72, No. 5, (2015), 867-875.
18. Tashkhourian, J. and Afsharinejad, M., "Chiral recognition of naproxen enantiomers using starch capped silver nanoparticles", ***Analytical Methods***, Vol. 8, No. 10, (2016), 2251-2258. DOI: 10.1039/C5AY03021H
19. Chen, Q., Jiang, H., Ye, H., Li, J. and Huang, J., "Preparation, antibacterial, and antioxidant activities of silver/chitosan composites", ***Journal of Carbohydrate Chemistry***, Vol. 33, No. 6, (2014), 298-312 DOI: 10.1080/07328303.2014.931962.
20. Amendola, V., Bakr, O.M. and Stellacci, F., "A study of the surface plasmon resonance of silver nanoparticles by the discrete dipole approximation method: Effect of shape, size, structure, and assembly", ***Plasmonics***, Vol. 5, No. 1, (2010), 85-97. DOI: 10.1007/s11468-009-9120-4
21. Sastry, C.S. and Rao, A.R., "Spectrophotometric determination of some analgesic and anti-inflammatory agents with 3-methyl-2-benzothiazolinone hydrazone hydrochloride", ***Microchimica Acta***, Vol. 97, No. 3-4, (1989), 237-244. DOI:10.1007/BF01242470
22. Gharibshahian, E., "The effect of polyvinyl alcohol concentration on the growth kinetics of ktiopo4 nanoparticles synthesized by the co-precipitation method", ***HighTech and Innovation Journal***, Vol. 1, No. 4, (2020), 187-193. DOI: 10.28991/HIJ-2020-01-04-06
23. Demartini, A., Alloisio, M., Cuniberti, C., Dellepiane, G., Jadhav, S.A., Thea, S., Giorgetti, E., Gellini, C. and Muniz-Miranda, M., "Polydiacetylene-functionalized noble metal nanocages", ***The Journal of Physical Chemistry C***, Vol. 113, No. 45, (2009), 19475-19481. DOI: 10.1021/jp905787h
24. Bhattacharjee, S., "Dls and zeta potential–what they are and what they are not?", ***Journal of Controlled Release***, Vol. 235, (2016), 337-351. DOI: 10.1016/j.jconrel.2016.06.017
25. Fillet, M., Fotsing, L., Bonnard, J. and Crommen, J., "Stereoselective determination of s-naproxen in tablets by capillary electrophoresis", ***Journal of Pharmaceutical and Biomedical Analysis***, Vol. 18, No. 4-5, (1998), 799-805. DOI: 10.1016/S0731-7085(98)00218-0
26. Gilani, S.L., Najafpour, G.D., Moghadamnia, A. and Kamaruddin, A.H., "Stability of immobilized porcine pancreas lipase on mesoporous chitosan beads: A comparative study", ***Journal of Molecular Catalysis B: Enzymatic***, Vol. 133, No., (2016), 144-153. DOI: 10.1016/j.molcatb.2016.08.005
27. Heydarzadeh Darzi, H., Gilani, S., Farrokhi, M., Nouri, S. and Karimi, G., "Textural and structural characterizations of mesoporous chitosan beads for immobilization of alpha-amylase: Diffusivity and sustainability of biocatalyst", ***International Journal of Engineering, Transactions B: Applications***, Vol. 32, No. 2, (2019), 207-216. DOI: 10.5829/ije.2019.32.02b.04
28. Shrivastava, A. and Gupta, V.B., "Methods for the determination of limit of detection and limit of quantitation of the analytical methods", ***Chronicles of Young Scientists***, Vol. 2, No. 1, (2011), 21. DOI: 10.4103/2229-5186.7934

29. Thompson, M., Ellison, S.L. and Wood, R., "Harmonized guidelines for single-laboratory validation of methods of analysis (iupac technical report)", ***Pure and Applied Chemistry,*** Vol. 74, No. 5, (2002), 835-855. DOI: 10.1351/pac200274050835

30. Zhang, K., Yuan, Z., Li, L., Shi, X., Cao, L., Du, Y. and Xue, N., "Separation of the two enantiomers of naproxcinod by chiral normal-phase liquid chromatography", ***Journal of Chromatographic Science***, Vol. 49, No. 4, (2011), 272-275. DOI: 10.1093/chrsci/49.4.272

31. Absalan, G., Alipour, Y., Rezaei, Z. and Akhond, M., "Determination of enantiomer compositions of propranolol enantiomers by chiral ionic liquid as a chiral selector and the uv-assisted spectrophotometric method", ***Analytical Methods***,Vol. 4, No. 8, (2012), 2283-2287. DOI: 10.1039/C2AY25161B

Persian Abstract

چکیده

ناپروکسن یک داروی ضد التهاب غیراستروئیدی است که بطور گسترده برای درمان حاد تا مزمن استفاده میشود هدف از این کار توسعه یک روش اسپکتروفتومتری ساده وجدید برای تشخیص ناپروکسن در فرمولاسیون دارویی است که برای اولین بار توسط نانو ذرات نقره با پوشش کیتوسان ایجاد شد ساختار و مرفولوژی نانو ذرات باطیف سنجی مرئی –فرابنفش,(UV-VIS) تکنیک, FTIR میکروسکوپ اتمی, (ATM) تکنیک DSL وآنالیز زتاپتانسیل (ZP) تحت شرایط بهینه تعیین شد نتایج نشان داد که ، نانوذرات نقره پوشانده شده با کیتوزان تقریباً با اندازه ذرات ۱۰۰ نانومتر وشاخص پراکندگی385/0 (PDI)و پتانسیل زتا آنیونی قوی (۲۴/۸–میلی ولت) در شرایط اسیدی بودند. مشخص شد که کیتوزان به عنوان یک انتخاب کایرال قادر به تشخیص ناپروکسن در شرایط بهینه آزمایش است. این روش برای تعیین ناپروکسن در فرمولاسیون قرص های دارویی موفقیت آمیز بود. انحراف استاندارد نسبی ۱.۰٪ برای تجزیه و تحلیل ناپروکسن در نمونه های واقعی تعیین شد. اعتبار سنجی این روش ، از جمله حد تشخیص و حد کمی با توجه به دستورالعمل های ICH به طور دقیق تأیید شد. بر اساس این روش ، حد تشخیص((LOD) و حد کمی(LOQ) ناپروکسن به ترتیب ۰۲۲ / و۰۶۶/ میلی مول بر لیترمحاسبه شد. تجزیه و تحلیل داده های آماری ، قابلیت تکرار و دقت این روش نشان داد که استفاده از این روش جدید برای تعیین ناپروکسن در فرمولاسیون های دارویی ارزشمند و عملی است.

International Journal of Engineering

Journal Homepage: www.ije.ir

Straightforward Prediction for Responses of the Concrete Shear Wall Buildings Subject to Ground Motions Using Machine Learning Algorithms

M. S. Barkhordari*[a], M. S. Es-haghi[b]

[a] *Department of Civil and Environmental Engineering, Amirkabir University of Technology, Tehran, Iran*
[b] *School of Civil Engineering, Khajeh Nasir Toosi University of Technology, Tehran, Iran*

PAPER INFO

Paper history:
Received 16 April 2021
Received in revised form 18 May 2021
Accepted 19 May 2021

Keywords:
Artificial Neural Networks
Regression Model
Tall Buildings
Seismic Response

ABSTRACT

The prediction of responses of the reinforced concrete shear walls subject to strong ground motions is critical in designing, assessing, and deciding the recovery strategies. This study evaluates the ability of regression models and a hybrid technique (ANN-SA model), the Artificial Neural Network (ANN), and Simulated Annealing (SA), to predict responses of the reinforced concrete shear walls subject to strong ground motions. To this end, four buildings (15, 20, 25, and 30-story) with concrete shear walls were analyzed in OpenSees.150 seismic records are used to generate a comprehensive database of input (characteristics of records) and output (responses). The maximum acceleration, maximum velocity, and earthquake characteristics are used as predictors. Different machine learning models are used, and the accuracy of the models in identifying the responses of the shear walls is compared. The sensitivity of input variables to the seismic demand model is investigated. It has been seen from the results that the ANN-SA model has reasonable accuracy in the prediction.

***doi**: 10.5829/ije.2021.34.07a.04*

NOMENCLATURE

T	Control variable	q	Probability of accepting the potential solution
ΔE_c	Changes in the value of the objective function	K_B	Boltzmann constant
y_i	i[th] value of the variable to be predicted	$\bar{y}$	Average of y_i
$\hat{y}_i$	predicted value of y_i		

1. INTRODUCTION

Reinforced concrete (RC) shear walls are efficient members for providing resisting horizontal forces in tall buildings. The non-linear analysis is needed to determine the tall buildings' seismic responses more realistically. However, non-linear modeling is a challenge for practitioner engineers because they should select the proper structural model type, define complex parameters of the materials, elements, and select as well as scale ground motions records. Besides, the ability to predict the structural capacity after an earthquake is essential to inform whether the tall building can be permanently reoccupied or not. Therefore, predicting the structure's response to a new earthquake based on the structure's response to past earthquakes could be an excellent solution to determine the extent of the damage. This is possible using machine learning techniques. In the last few years, research has been done on using machine learning in civil engineering [1-12]. For instance, Thaler et al. [2] developed a machine-learning-enhanced Monte Carlo simulation strategy to predict the structural response in earthquake engineering in which the neural networks are utilized to improve the reliability of the method in the tail end of the distribution. Stoffel et al. [10] developed an Artificial Neural Network accessible to complicated structural deformation under shock-wave loads. They calculated plate deflections by means of

*Corresponding Author Institutional Email: m.s.barkhordari@aut.ac.ir (M. S. Barkhordari)

Please cite this article as: M. S. Barkhordari, M. S. Es-haghi, Straightforward Prediction for Responses of the Concrete Shear Wall Buildings Subject to Ground Motions Using Machine Learning Algorithms, International Journal of Engineering, Transactions B: Applications, Vol. 34, No. 05, (2021) 1586-1601

finite element simulations including a neural network. Mangalathu and Jeon [7] applied machine learning techniques to identify the failure mode of beam-column joints. They also compared various machine learning techniques to estimate the shear strength of beam-column joints using an experimental database. Khaleghi et al. [13] have employed Artificial Neural Networks to predict load-bearing capacity and stiffness of perforated masonry walls. Mangalathu and Jeon [14] conducted a comparative study for failure mode recognition of RC bridge columns using various machine learning models. A clustering algorithm is proposed by Siam et al. [15] for structural performance classifications using a dataset of ninety-seven masonry shear walls. Kiani et al. [16] developed a method for deriving the fragility curves using various machine learning models. They also investigated the effect of training sample size and imbalanced dataset on machine learning models' performance. Gaba et al. [17] classified the damages caused by earthquakes using a previously acquired data set. To establish the best prediction model, they evaluated different machine learning classifier algorithms. Burton et al. [18] described a statistical approach to predict the aftershock collapse vulnerability of buildings. They also mentioned that the Kernel Ridge regression method produces the most accurate and stable predictions. Zhang et al. [19,20] utilized machine learning algorithms to link the capacity of damaged buildings to the response and damage patterns.

As mentioned above, Machine Learning methods to analyze and evaluate the dynamic characteristics of the structure have been studied in the literature. Nevertheless, few articles have focused on tall buildings, which are evaluated in the current paper. In addition, tall buildings have a large number of components and responses that lead to a high dimensional feature space, as opposed to low- or mid-rise buildings. Therefore, there is a need for a simple method to estimate the responses of the tall buildings subject to ground motion. In addition, a hybrid intelligent method, which is the optimization of the parameters in Artificial Neural Network by the revolutionary algorithm of Simulated Annealing, to achieve better performance for predicting the response of tall buildings.

In fact, This study's primary purpose is to evaluate the ability to exist simple machine learning methods and a hybrid technique (ANN-SA model), the Artificial Neural Network (ANN), and Simulated Annealing (SA) to estimate the responses of structures based on earthquake characteristics and the stored responses of a structure subjected to the earthquake. Specifically, the following objectives have been pursued throughout the research: (1) evaluating the performance of various machine learning models (namely linear regression, Ridge regression, Lasso regression, Elastic net regression, Huber regression, RANSAC, and ANN-SA model) in estimating of responses of high-rise-concrete-shear-wall buildings, (2) investigating the effectiveness of using maximum acceleration and maximum speed recorded by the sensor in predicting structural responses (3) identifying the significant input variables which influence the predicting the responses of tall buildings.

2. MATERIALS AND METHODS

The central assumption in this study is that there are sensors in the stories of the tall building so that the maximum acceleration or velocity in the tall building that is subjected to a new earthquake can be captured (these responses are recorded during the earthquake), and the building's responses such as drift, base shear, displacement, the maximum acceleration and velocity under previously recorded ground motions are calculated using any software (Database). In other words, we have the maximum velocity or acceleration in the building and the building's responses, such as the drift caused by previous earthquakes, and only the maximum acceleration and velocity, which created by the new earthquake, are recorded using sensors. Can this information be used to estimate the building's responses under the new earthquake? In order to evaluate this strategy, four buildings (15, 20, 25, and 30-story) with concrete shear walls were analyzed in OpenSees [21,22] to generate a dataset.

A percentage of total data is considered information obtained from the building subjected to the new earthquake (for example, 20%). This means that the obtained acceleration/velocity of this 20 % is considered the sensor's values under the new earthquakes, and the obtained responses (the maximum base shear, maximum drift, and maximum displacement) are considered as unknown variables. Therefore, the maximum acceleration, maximum velocity, and earthquake characteristics are used as predictors in order to estimate the maximum base shear, maximum drift, and maximum displacement for a specific seismic excitement. The characteristics of the earthquake that are considered are the scale factor, significant duration (D5-95 (s)), moment magnitude of the earthquake (magnitude), and Joyner-Boore distance (Rjb (km)).

2. 1. Buildings, Seismic Records, and Modeling

Dual RC (shear wall–frame) high-rise structures are adopted. The dual system buildings have 15, 20, 25, and 30 stories. The story height is equal to 3.5 m. The buildings plan consists of five bays (Figure 1). The gravity framing is considered using the leaning column, which is linked to the main structure. Rigid truss elements are used to connect the shear wall-steel frame and leaning columns and transfer the P-Delta effect. The design dead and live loads are $5kN/m^2$ and $2kN/$

m^2, respectively. The concrete compressive strength is assumed to be 55 MPa. Both longitudinal and transverse reinforcement has a yield strength of 420 MPa. In Appendix A, the fundamental parameters for material properties have been listed. The design was conducted based on ACI [23] and ASCE [24]. The building was also designed based on the modal response spectrum analysis (ASCE [24]), and the first 15 modes were used in the design. Table 1 presents the building site and the design parameters, which and are the maximum considered earthquake (MCE) spectral acceleration at short periods (S_s) and 1-s period (S_1), respectively. Table 2 provides the modal periods of the prototype buildings. Rayleigh damping is assumed. The damping is set as 2% of critical damping proportional to the mass and initial stiffness matrix. The dimensional details of the beams, columns and rebar sections of the concrete shear wall are presented in Appendix A.

The buildings' finite element model is generated by the OpenSees program [21,22] using displacement-based beam-column elements for the RC beams and columns. Concrete02 and Steel02 materials are used to define the material model of concrete and reinforcing steel fibers. The displacement-based beam-column element is a distributed-plasticity-fiber-based element based on Bernoulli's theory. Although different macro elements have been proposed for modeling concrete shear walls [25,26], in this study, RC walls are molded using a state-of-the-art element (SFI_MVLEM- Figure 2) and a nDMaterial FSAM material [26]. The SFI_MVLEM [26] element is a macro element which can simulate the behavior characteristics induced by non-linear shear deformation such as shear– axial/flexural interaction, shear cracking, stiffness deterioration, pinching effect, and strength deterioration. Studies have shown that (1) shear cracking can increase shear deformation of the walls in the plastic hinge region and (2) existing previous models usually underestimate compressive strains at the boundary elements, even for walls that their behavior is dominated by flexure. The confinement parameters of the boundary elements are calibrated according to the model proposed by Mander et al. [27].

TABLE 1. The building site and the design parameters

Latitude (degree)	Longitude (degree)	Design Cat.	Risk Cat.	Soil Cat.	S_s (MCE)	S_1 (MCE)
35.6535	-120.4407	D	I	D (stiff soil)	1.5g	0.6g

TABLE 2. The modal periods of the prototype buildings

Story	15	20	25	30
Modal periods (sec)	2.66	3.17	3.8	4.3

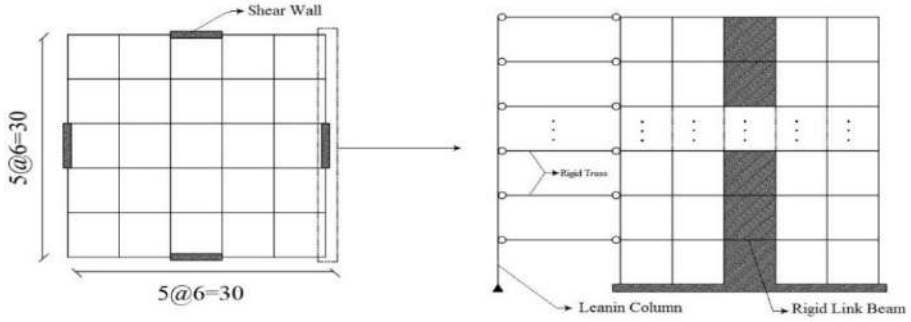


Figure 1. 2D model of the prototype buildings

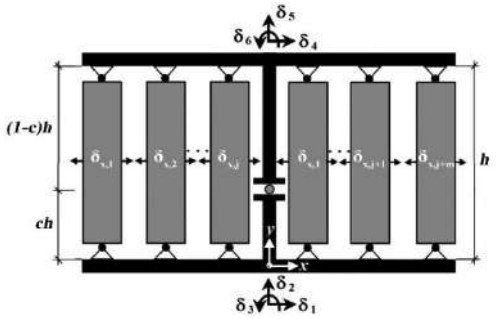


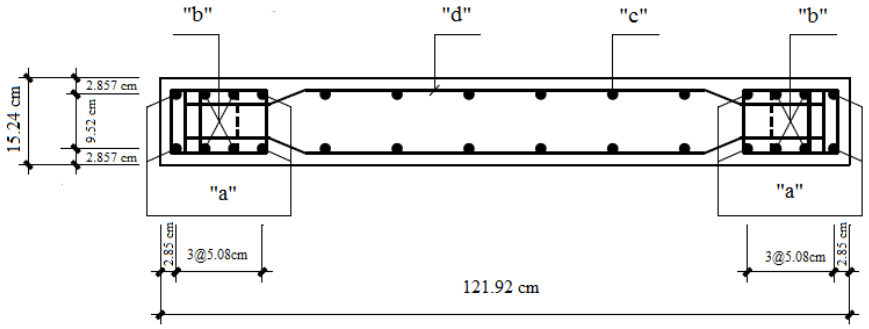
Figure 2. Shear-Flexure Interaction Multiple-Vertical-Line-Element Model (SFI-MVLEM).

The nonlinear time-history analyses are performed for the MCE level. The buildings are subjected to 150 seismic records, resulting in 600 non-linear response history analyses. Earthquake records are selected from the database of the Pacific Earthquake Engineering Research (PEER) center [28]. The key information of these records wall is presented in Appendix A. The minimum magnitude of records is taken as 6.0, and records are within a distance less than 20 km to the fault. Each ground motion is scaled in such a way that its response spectrum equals or exceeds the ASCE [24] spectrum over a determined period range (from 0.2T to

1.5T, where T is the first mode of vibration). All non-linear time-history analyses adopted the Newmark time integration method of constant acceleration. The Newton–Raphson iteration method is utilized to determine how the sequence of steps taken to solve the non-linear equation of motion. The convergence of the algorithm was based on the relative work increment. If a time step failed to converge, the Newton method switches to a modified Newton method with constant stiffness equal to the initial stiffness of the time step.

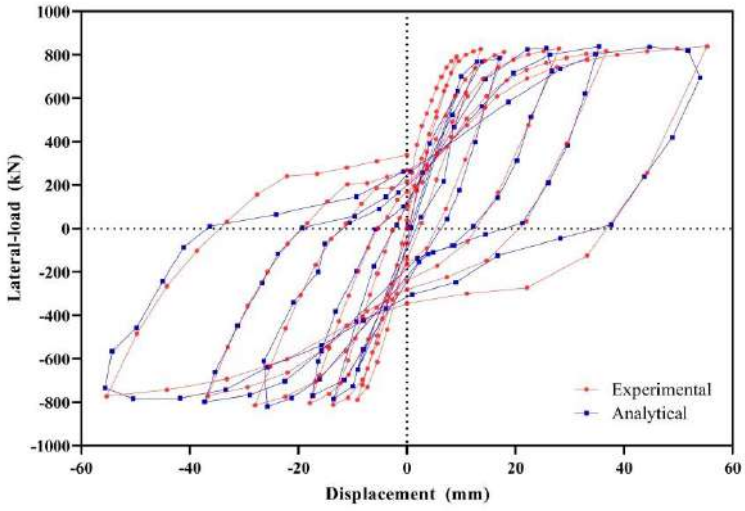
Model calibration is done using experimental results for reverse cyclic loading conducted by Tran and Wallace [29]. As an example, the element's response and related laboratory story test for specimen S78 are shown in Figure 3. Table 3 summarizes specimen information.

2. 2. Supervised Learning Methods One of the simplest supervised machine learning techniques is the family of regression models. Six regression models such as linear regression [30], Ridge regression [31], Lasso



Specimen	a	b	c	d
S78	4#6	4#5	6#3 @ 0.127 m	#3 @ 0.127 m

b) Cross-section and reinforcement distribution



a) Global response

Figure 3. Typical cross-section and response for specimen S78.

TABLE 3. Specimen information

Aspect ratio	Web Reinf.	Boundary Reinf.	Compressive strength of concrete	Yield strengths of Reinf.
1.5	0.0073	0.0606	55 MPa	440-470 MPa

regression [32], Elastic net regression [33, 34], Huber regression [35], and RANSAC [36] are used in this paper. For further information on regression models, interested readers should study corresponding references of each model. In the case of the seismic demand model for the tall buildings, the input vector consists of the scale factor,

significant duration (D5-95 (s)), moment magnitude of the earthquake (magnitude), Joyner-Boore distance (Rjb (km)), and the maximum acceleration as well as velocity (Outputs from the non-linear time history analysis) in the tall building. . Significant duration (D5-95 (s)) is defined as the time needed to build up between 5 and 95 percent of the total Arias intensity for a specific earthquake record. The Joyner-Boore distance is defined as the shortest distance from a seismic station or any other site to the surface projection of the seismic event's rupture surface. Table 4 summarizes the range of parameters used. Other outputs from the non-linear time history analysis (the maximum base shear, maximum drift, and maximum displacement) are considered target variables.

Ordinary Least Square (OLS) regression (or linear regression) is one of the most widely known modeling techniques. The OLS regression is also known as linear regression. The OLS regression assumes that the relationship between the input variable (features vector,X) and the output variable (target vector,Y) is approximately linear (Equation (1)).

$$\hat{Y} = \beta^T X + \beta_0$$
$$\min_{\beta \in \mathbb{R}} \sum_{i=1}^{n} \|\beta^T x_i + \beta_0 - y_i\|^2 \quad (1)$$

where in Equation (1), $\hat{y}$is predicted values vector, $X = (x_1, x_2, \ldots, x_i)$ are the n input variables,$Y = (y_1, y_2, \ldots, y_i)$ are the n output variables, and β^T are the coefficients.

The OLS estimates often are subjected to the drawback of large variance. Previous studies have shown that there is a statistical trade-off between bias and variance. These observations have led to consider biased estimates such as Ridge regression. Ridge regression (Equation 2) introduces some bias by adding a penalty to the sum of the squared errors. Although model efficiency is decreased, the test error is decreased too. The coefficients are shrunk toward 0 as α becomes large.

$$\min_{\beta \in \mathbb{R}} \sum_{i=1}^{n} \|\beta^T x_i + \beta_0 - y_i\|^2 + \alpha \|\beta\|^2 \quad (2)$$

Note that in this case (using Ridge regression) solutions are not equivalent under scaling of the predictors (inputs); therefore, the predictors have to be standardized before using the Ridge regression model. The penalty contains the squared of the L2 norm of β (Equation (2)). The Lasso regression is a shrinkage method like Ridge regression. Lasso regression minimizes a loss function, using the L1 norm which is the sum of absolute values (Equation (3)).

$$\min_{\beta \in \mathbb{R}} \sum_{i=1}^{n} \|\beta^T x_i + \beta_0 - y_i\|^2 + \alpha \|\beta\|_1 \quad (3)$$

The difference between the L1 norm and L2 norm methods is that L1 penalizes coefficients equally but L2 penalizes more very large coefficients. In other words, for some values of α, L1 setting some coefficients equal to 0, and thus the most important variables are kept, this is called feature selection. Elastic Net is similar to Ridge regression and Lasso regression but uses both the L1 norm and L2 norm together (Equation (4)).

$$\min_{\beta \in \mathbb{R}} \sum_{i=1}^{n} \|\beta^T x_i + \beta_0 - y_i\|^2 + \alpha \cdot \eta \cdot \|\beta\|_1 + \alpha \cdot (1 - \eta) \cdot \|\beta\|_2^2 \quad (4)$$

where η is a coefficient that captures the relative amount of L1-penalty. This coefficient (η) is considered 0.5 [33, 34]. α needs to be determined by the analyst in Ridge, Elastic Net, and Lasso. By using the GridSearchCV in python, the value of α that maximizes the R^2 is calculated. The results are discussed further in section 3. In a sample, generally, outliers are considered as an example that differs remarkably from other observations. The models (Ridge, Elastic Net, and Lasso) are presented so far are sensitive to outliers since every single point participates in minimizing the function. To overcome this problem, Robust Regression is proposed. In the following, brief descriptions about Huber Regression and the RANdom SAmple Consensus (RANSAC), which allow the fit of robust regression, are provided. The Huber regression applies a piecewise function (loss) to samples that are classified as outliers. In other words, the loss optimizes either the squared loss or absolute loss for the samples based on a parameter (ε, Equation (5)). The cost function that Huber regression minimizes is given by:

$$\min_{\beta} \sum_{i=1}^{n} L(y_i, f(x_i))$$
$$where \quad (5)$$
$$L(y, f(x)) = \begin{cases} (y - f(x))^2 if |y - f(x)| < \varepsilon \\ 2 \cdot \varepsilon \cdot |y - f(x)| otherwise \end{cases}$$

Hence, the loss function is squared for small prediction errors. RANSAC is a non-deterministic algorithm that divides the complete data set into two different subsets (outlier and inlier). The inlier subset is also known as the hypothetical inliers which are used to fit the model. The basic steps of the RANSAC algorithm are summarized as follows: 1) Select randomly the minimum samples from the original data (the

TABLE 4. The modal periods of the prototype buildings

	Scale Factor	D5-95 (s)	Acc. ($\frac{m}{s^2}$)	Vel. ($\frac{m}{s}$)
Mean	8.43	23.63	7.33	1.37
Std.	9.11	12.24	2.81	0.57
Min.	0.47	7.2	1.52	0.39
25%	2.45	14.62	5.64	0.97
50%	5.68	20.6	6.95	1.27
75%	11.83	28.67	9.14	1.69
Max.	70.15	65.8	14.85	3.72

hypothetical inliers). 2) Fit a model to the selected points. 3) Points from the set of all points are then evaluated against the fitted model by considering a predefined tolerance. If the points fit the computed model well (using loss function), they will be considered as part of the consensus set (CS). 4) Save the estimated model as the best model if the consensus set is large enough ((the number of inliers/ the total number points) > predefined threshold). 5) Otherwise, repeat steps 1 through 4 (a trial and error process).

Although presented regression models provide remarkable feature selection, the prediction performance is limited. The main disadvantage of presented regression models is that they cannot consider non-linearity in the available data. An alternative method of tackling these problems is the use of Artificial Neural Networks. A neural network is a hierarchical organization of neurons which are joined by weighted connections. The structure of Artificial Neural Networks is made of three main components, which are referred to as (1) the input layer, which takes in a numerical representation of the data; (2) the hidden layer, where computations take place; and (3) the output layer. A direct consequence of this approach is an improvement of the estimation of drift, displacement, and base shear of different buildings. The network used to solve the problem in this study consists of three layers (input layer, one hidden layer, and output layer). We determined the number of neurons in the hidden layer and the percentage of the training and test data using a simulated annealing algorithm to reduce the computational time. Simulated Annealing (SA) is a stochastic algorithm for estimating the optimum value of a given function [37]. This method is inspired by the slow cooling of metals. The simulated annealing algorithm randomly selects a new potential solution. The range of the training dataset is 60-80% (X %) of the whole dataset, and the remaining ((100-X)/2 %) is used as the validation and test set. The codes are developed in MATLAB with its toolbox. The dataset (input and output) used in this section is the same as the regression models' dataset. A popular training algorithm (so-called "Trainlm" [37]) that updates weight and bias values according to Levenberg-Marquardt optimization is used. Activation functions for the hidden and output layers are hyperbolic tangent sigmoid (Tansig) and linear transfer function (Purelin), respectively.

2. 3. Simulated Annealing Algorithm Simulated annealing (SA) algorithm is one of the most preferred methods for solving optimization problems developed by Kirkpatrick et al. [39]. The SA algorithm, which is inspired by the slow cooling of metals, is a heuristic method with the basic idea of generating random displacement from any feasible solution. A probability function (Equation 6) is utilized to decide the transition between the current solution and the randomly generated new solution.

$$q = min\{1, e^{-\Delta E/K_B T}\} \quad (6)$$

where T is the control variable, q is the probability of accepting the potential solution, K_B is the Boltzmann constant, and ΔE is changes in the value of the objective function. The SA algorithm has some crucial advantages, including the following: (1) the SA algorithm is relatively easy to code, even for complex problems, and can deal with highly non-linear models, chaotic and noisy data, and many constraints.; (2) most optimization algorithms use the gradient descent, but the SA algorithm does not spend the computational time in calculating it; (3) the SA algorithm can be utilized to identify the minimum of the objective function more efficiently instead of being

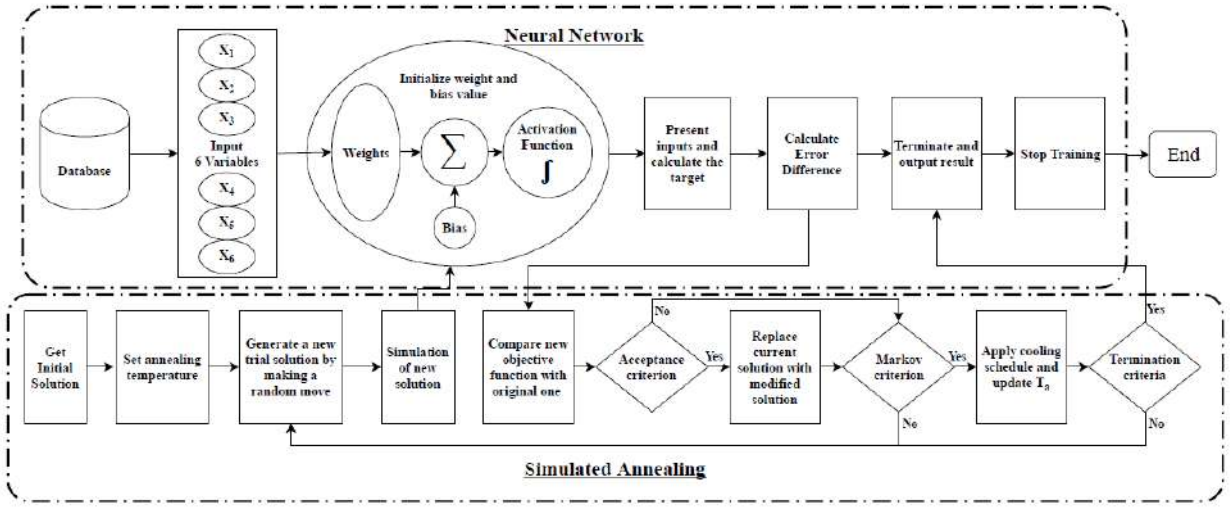


Figure 4. Computation procedure of the number of neurons and the percentage of the training and test data

trapped in a local minimum, and (4) Simulated annealing algorithm is independent of initial conditions [40]. As mentioned earlier, the number of neurons in the hidden layer and the percentage of the training and test data are determined using the simulated annealing algorithm. The proposed computation procedure of the number of neurons in the hidden layer and the percentage of the training data is summarized in the flow chart of Figure 4. The SA algorithm searches in the range 5-30 and 60-90% for the number of neurons in the hidden layer and the training data percentage, respectively.

3. RESULTS AND DISCUSSION

The machine learning techniques explained in the previous section are utilized to predict the high-rise concrete shear wall buildings' responses. The codes (regression models) are developed using a free software machine learning library of the Python programming language, so-called scikit-learn [41]. Observations (targets and features) are (randomly) split into two sets, traditionally called the test set and the training set. In this study, 80% and 20% of the entire dataset are considered for training and testing, respectively. The input variables are centered and scaled (a standard space with 0 mean and unit variance). Generally, the model is fitted on the training data, and the performance of the model is evaluated using unknown (test) data and the R^2 (Equation 7) or residual sum of squares (RSS, Equation 8) or mean square error (MSE, Equation 9) as score metric.

$$R^2 = 1 - \frac{\sum_i (y_i - \hat{y}_i)^2}{\sum_i (y_i - \bar{y})^2} \quad (7)$$

where $y_i$is the ith value of the variable to be predicted, $\bar{y}$is the average ofy_i, and $\hat{y}_i$predicted value ofy_i.

$$RSS = \sum_i (y_i - \hat{y}_i)^2 \quad (8)$$

$$MSE = \frac{1}{n}\sum_i (y_i - \hat{y}_i)^2 \quad (9)$$

The $R^2$is used in this study in order to compare the efficiency of the models in predicting the seismic demand (e.g., Table 5). The$R^2$is utilized because it is easily interpretable and it is a normalized version of the RSS. Besides, the R^2 does not depend on the scale of the data. The $R^2$is computed for the remaining data (test data).

In machine learning, a hyperparameter is a parameter whose value is utilized to control the learning process. As an example, the changes in the performance of the models (Elastic Net regression) against the changes in the hyperparameter α are shown in Figure 5. The performance of the models dramatically decreases as the hyperparameter α gets bigger. Based on these results, an optimum value of the hyperparameter α is chosen for each method and the target variable (displacement, drift, or base shear). The optimum value of the hyperparameter α is given in Appendix B. Here the performance of different methods utilized in this study is compared. Figure 6 (or Table 6) shows theR^2scores from 5 different regression models for displacement, drift, and base shear obtained using a test set for the different tall buildings. Overall Ridge, Lasso, Huber, and Elastic Net regression have very closeR^2scores for displacement, drift, and base shear. On the contrary, there is a difference between the RANSAC and other methods. The RANSAC regression performs the worst among the methods in estimating the base shear. Based on Figure 6, it can be concluded that the regression models have differentR^2scores for almost all the target variables and various buildings.

It is helpful to understand what factors may or may not impact estimating the tall building responses using regression methods. In order to compare regression coefficients, first, the average coefficients for all buildings are calculated for all target variables. Then, average coefficients are normalized by dividing each average coefficient by the sum of all the average coefficients to form a sum of 1.0. As an example, the process for displacement is shown in Figure 7.

TABLE 5. Results of Linear regression

Linear Reg.	Displacement	Drift	Base shear	Structure
R^2	0.65	0.70	0.65	15-stroy
R^2	0.7	0.70	0.65	20-story
R^2	0.6	0.65	0.69	25-story
R^2	0.72	0.75	0.77	30-story
Average	0.66	0.7	0.69	

TABLE 6. Results of Regression Models

Model	Structure		Displacement	Drift	Base-shear
Ridge	15-stroy	R^2	0.66	0.72	0.66
	20-story		0.72	0.72	0.65
	25-story		0.61	0.65	0.70
	30-story		0.73	0.75	0.78
	Average		0.68	0.71	0.69
Lasso	15-stroy	R^2	0.63	0.71	0.63
	20-story		0.76	0.75	0.69
	25-story		0.63	0.68	0.70
	30-story		0.73	0.77	0.75
	Average		0.69	0.73	0.69

Elastic Net	15-stroy	R^2	0.63	0.70	0.63
	20-story		0.74	0.75	0.69
	25-story		0.62	0.68	0.71
	30-story		0.71	0.77	0.76
	Average		0.68	0.73	0.70
Huber	15-stroy	R^2	0.64	0.73	0.65
	20-story		0.73	0.76	0.69
	25-story		0.63	0.68	0.70
	30-story		0.71	0.76	0.75
	Average		0.68	0.73	0.70
RANSAC	15-stroy	R^2	0.58	0.68	0.45
	20-story		0.65	0.72	0.51
	25-story		0.60	0.68	0.65
	30-story		0.73	0.79	0.82
	Average		0.64	0.72	0.61

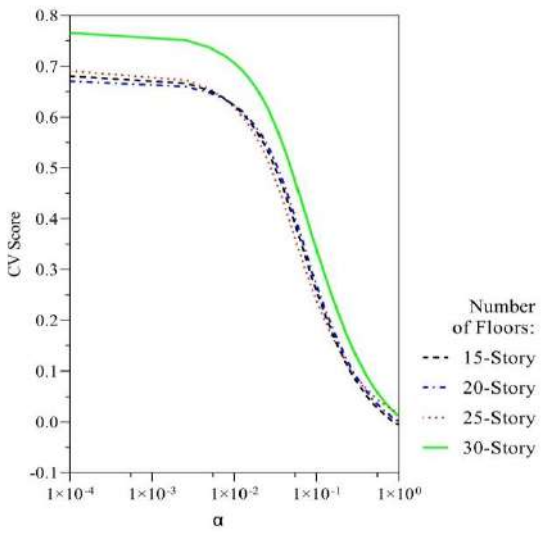


Figure 5. Elastic Net regression performance for predicting base shear of various structures

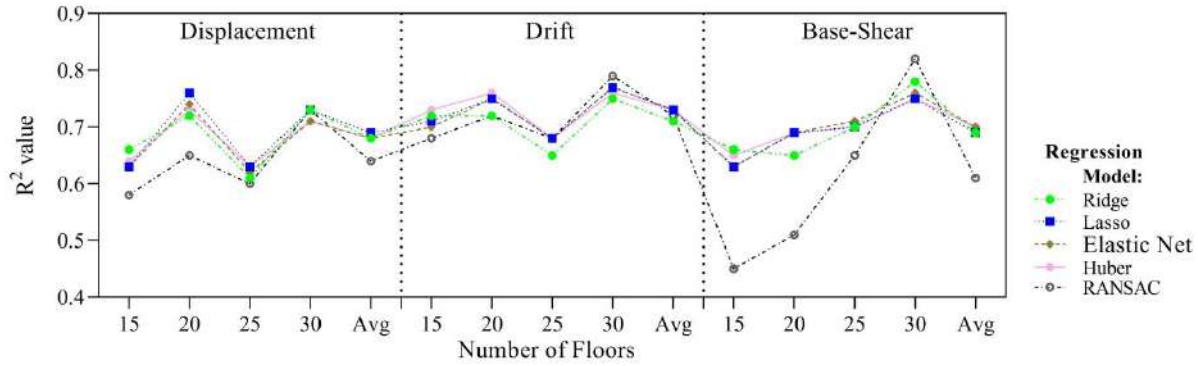


Figure 6. Results of Regression Models

Target	Scale Factor	5-95 Duration	Magnitude	Rjb	Acc.	Vel.	Structure
Displacement	x_{11}	x_{12}	x_{13}	x_{14}	x_{15}	x_{16}	15-story
	x_{21}	x_{22}	x_{23}	x_{24}	x_{25}	x_{26}	20-story
	x_{31}	x_{32}	x_{33}	x_{34}	x_{35}	x_{36}	25-story
	x_{41}	x_{42}	x_{43}	x_{44}	x_{45}	x_{46}	30-story
	$r_1=\frac{\sum_{i=1}^{4} x_{i1}}{4}$	$r_2=\frac{\sum_{i=1}^{4} x_{i2}}{4}$	$r_3=\frac{\sum_{i=1}^{4} x_{i3}}{4}$	$r_4=\frac{\sum_{i=1}^{4} x_{i4}}{4}$	$r_5=\frac{\sum_{i=1}^{4} x_{i5}}{4}$	$r_6=\frac{\sum_{i=1}^{4} x_{i6}}{4}$	average coefficient
	$r_1^N=\frac{r_1}{\sum_{i=1}^{6} r_i}$	$r_2^N=\frac{r_2}{\sum_{i=1}^{6} r_i}$	$r_3^N=\frac{r_3}{\sum_{i=1}^{6} r_i}$	$r_4^N=\frac{r_4}{\sum_{i=1}^{6} r_i}$	$r_5^N=\frac{r_5}{\sum_{i=1}^{6} r_i}$	$r_6^N=\frac{r_6}{\sum_{i=1}^{6} r_i}$	normalized coefficient

Figure 7. Results of Regression Models

Figure 8 shows the average-normalized estimated regression coefficients of various regression models for each target variable (displacement, drift, and base shear). Figure 8, the 0.0 values indicate that the associated features are not significant in predicting target variables. Also, Figure 8 illustrates that:

The crucial parameters to take into account tend to vary from method to method.

As mentioned above, the Elastic Net and Huber regression have the most R^2 scores, but unlike the first method, the second method recognizes more variables as influential input variables,

All regression models identify velocity as a significant input variable,

Huber and RANSAC regressions recognize all the input variables as influential variables,

In the case of displacement, all regression models identify velocity and magnitude as significant input variables,

In the case of base shear, all regression models identify velocity, acceleration, and magnitude as significant input variables,

Lasso, Ridge, and linear methods identify that the Rjb, 5-95 Duration, and scale factor have a minimal effect on predicting the target variable (seismic response).

In this study, the ANN-SA algorithm is utilized as an alternative solution. The Artificial Neural Network parameters are adjusted to maximize the R^2 to 1. Table 7 gives the number of neurons of the neural networks, which are determined using the simulated annealing algorithm for different buildings and target variables. Figure 9 depicts the results obtained from the ANN-SA algorithm. Comparison of the ANN-SA algorithm and regression results (Tables 6 and 7) reveals that the ANN-SA algorithm gives more accurate results for all predicted variables. Surely this could be due to the fact that the non-linearity of the relationship of the responses and features can be captured by an Artificial Neural Network. The above results emphasize the need for a comprehensive evaluation of different models before establishing a machine-learning-based response prediction model. Also, the percentage points of training, validation, and test data are determined using the simulated annealing algorithm. The results (Table 7) indicate that selecting the percentage points of training, validation, and test is an influential parameter.

The sensitivity analysis examines how uncertainty in a model's target variables can be apportioned to different uncertainty sources in the model input parameters. In other words, the sensitivity analysis allows the determination of the model key input factors of an output of interest. In this section, a MATLAB toolbox developed by Vu-Bac et al. [42] is used to carry out the sensitivity analysis. The framework links different steps from generating a sample, constructing the surrogate model, and implementing the sensitivity analysis method. The joint and conditional probability distribution functions of the input parameters are used to generate the

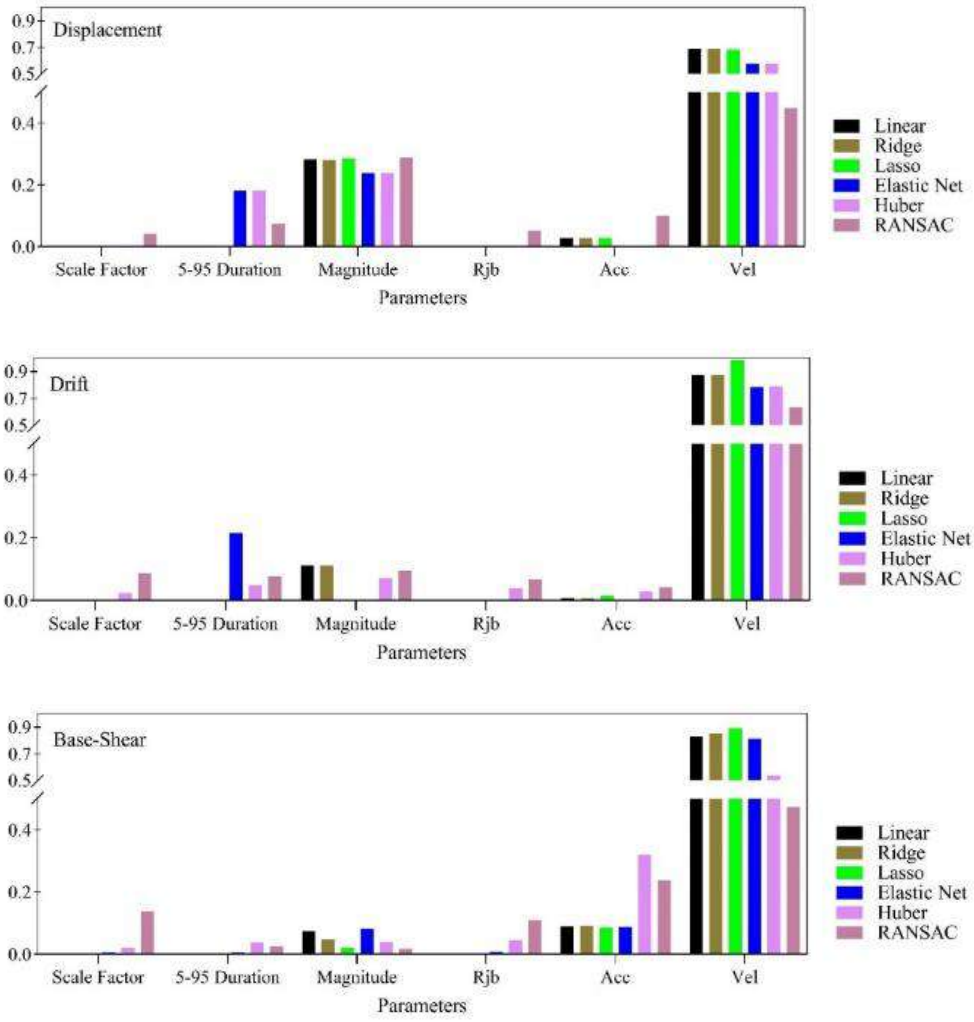
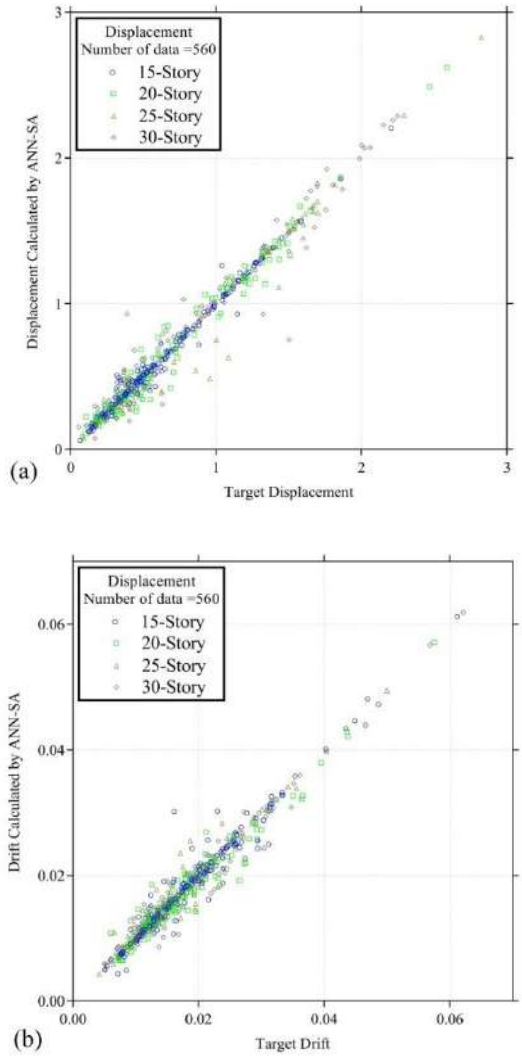


Figure 8. Average estimated coefficients for target variables



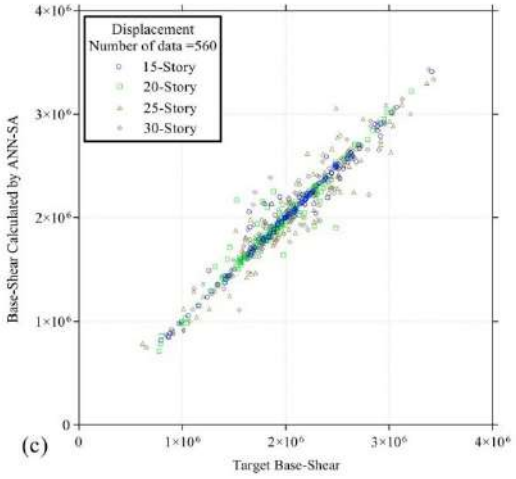


Figure 9. Performance of ANN-SA for estimating (a) Displacement (cm) (b) Drift (c) Base-Shear (N)

4. SENSITIVITY ANALYSIS

sample data since they must account for the input space constraints. The so-called surrogate-based approach is employed as an approximation of the real model for sensitivity analysis. The computation procedure of the sensitivity analysis is summarized in the flow chart of Figure 10. The description of the toolbox has been presented in literature [42]. Table 8 shows the results of the sensitivity analysis for all buildings. For all target variables (displacement, drift, and base shear), the earthquake's magnitude is estimated as the most crucial parameter. The second important parameter varies according to the building and the type of the target variable. Possible reasons for what may have caused this

TABLE 7. Results of Optimized Artificial Neural Networks

Story	Displacement			Drift			Base-shear		
	Neurons Num.	Training Percent	R^2	Neurons Num.	Training Percent	R^2	Neurons Num.	Training Percent	R^2
15	14	70	0.94	28	90	0.96	13	70	0.91
20	16	75	0.93	24	85	0.92	30	75	0.94
25	17	75	0.94	28	85	0.94	30	75	0.9
30	28	75	0.95	26	85	0.96	14	75	0.95
Average			0.94			0.94			0.92

issue can be: (1) structural responses are not identical since earthquake records have a random nature and their content are different from one another [43], and (2) as the building height increases, the effect of the modes (especially higher modes [44, 43]) on the structural response is increased, changing the structural behavior and response under a given earthquake.

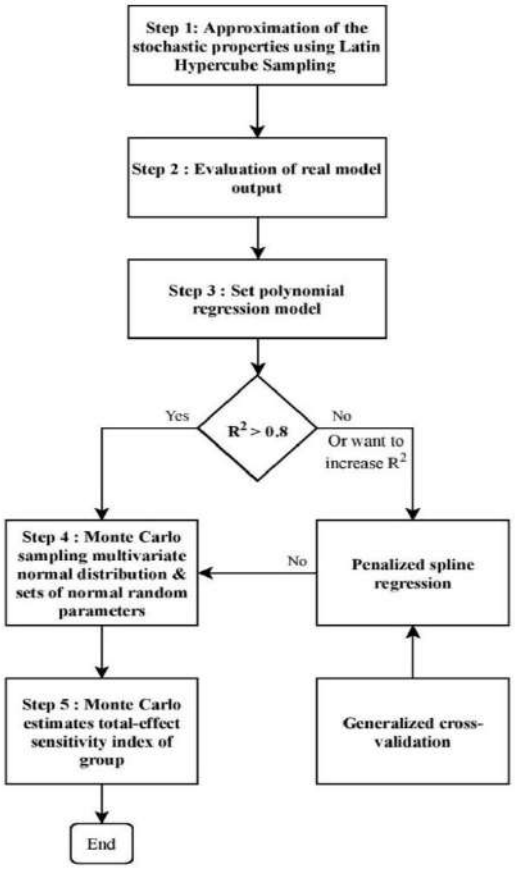


Figure 10. Diagram for sensitivity analysis

TABLE 8. Results of the sensitivity analysis for all buildings

15-story						
Target	**Scale Factor**	**Duration**	**Magnitude**	**Rjb**	**Acc.**	**Vel.**
Disp.	0.0004	0.0003	0.6661	0.0000	0.0012	0.0437
Drift	0.0031	0.0959	0.4580	0.0935	0.0312	0.0154
B. shear	0.0008	0.1668	0.3934	0.0091	0.0385	0.0440
20-story						
Target	**Scale Factor**	**Duration**	**Magnitude**	**Rjb**	**Acc.**	**Vel.**
Disp.	0.0000	0.0001	0.6769	0.0171	0.0020	0.0006
Drift	0.0036	0.0010	0.5450	0.0018	0.0021	0.0502
B. shear	0.0039	0.0009	0.4993	0.0004	0.0387	0.0097
25-story						
Target	**Scale Factor**	**Duration**	**Magnitude**	**Rjb**	**Acc.**	**Vel.**
Disp.	0.0000	0.0004	0.6764	0.0056	0.0001	0.0007
Drift	0.0016	0.0029	0.4438	0.0001	0.0003	0.0208
B. shear	0.0049	0.0379	0.1698	0.0019	0.0092	0.1429
30-story						
Target	**Scale Factor**	**Duration**	**Magnitude**	**Rjb**	**Acc.**	**Vel.**
Disp.	0	0	0.7268	0	0.0004	0.0016
Drift	0.0002	0.0065	0.6074	0.0001	0.0008	0.0304
B. shear	0.0004	0.0001	0.7296	0.0006	0.0153	0.0033

5. CONCLUSIONS

Reinforced concrete shear walls are used in high-rise buildings to resist earthquakes or wind loads. The need for an easy-to-use response estimation method for rapid damage assessment of the high-rise buildings after an earthquake leads to the study of existing simple regression methods and a hybrid technique, the Artificial Neural Network (ANN), and Simulated Annealing (SA) (ANN-SA model), for estimating the response of the structures in this study. In the initial part of this paper, four tall buildings were molded, and non-linear time-history analyses were performed to generate an extensive database. The computer software OpenSees was used to simulate the buildings under 150 earthquakes and calculate the responses. The primary purpose was to compare regression models and a standard Artificial Neural Network in predicting the tall building's response.

Analysis of results showed that if (1) during the earthquake, the maximum velocity created in the structure was stored (which can be done using the sensor) and (2) a database of the structure's responses to past earthquakes was produced (database) using existing software, the ANN-SA algorithm can use this information to estimate structural responses with acceptable accuracy.

Besides, the efficiency of different regression models such as RANSAC, Huber, linear, Ridge, Lasso, and Elastic Net regressions was studied in terms of estimation of structures' response. The training set (Eighty percent of the data) was utilized to fit the models, and the performance of the models was evaluated through the remaining unknown data (the test set). The performance of the regression models was assessed using scores. In general, the Elastic Net and Huber regression had better performance compared to other regression methods. Also, by using Ridge, Lasso, and Elastic Net regressions, the various input variables' relative importance on the estimated responses was identified. From the further exploration of the Elastic Net regression, critical parameters in determining the responses were velocity, acceleration, magnitude, and 5-95-Duration.

In order to evaluate the effect of non-linearity in the available data, the hybrid technique (ANN-SA algorithm) was utilized. The developed model had three-layer structures (input, hidden layer, and output layer). A simulated annealing algorithm was utilized to determine

the optimal number of the Artificial Neural Network neurons and the percentage of data that should be used in the training, validation, and testing set. By comparing the results of the ANN-SA algorithm and regression models, it can be concluded that (1) the effect of the non-linear relationship between data is significant, and considering it increases the accuracy of the model in predicting the target variables, and (2) the Artificial Neural Network outperforms regression models.

In addition, the sensitivity analysis was performed to examine how uncertainty in the target variables of a model could be apportioned to different sources of uncertainty in the model input parameters. The earthquake's magnitude was estimated as the most critical parameter, but the second important parameter varied according to the building and the type of target variable. Although the findings and conclusions are based on the case studies of four concrete shear wall buildings, the methodology has a wealth of applications in functional domains.

According to the literature and the results obtained in this study, it is suggested that researchers follow the process of the current paper for 3D modeling of irregular buildings and investigate the efficiency of Artificial Neural Networks for predicting their responses. Furthermore, the investigation of the Soil-Structure Interaction (SSI) effect can complement this research. Besides, comparing the performance of finite element and Neural Network models with the empirical vulnerability model of the actual seismic damage investigation can be very helpful and practical.

6. REFERENCES

1. Baltacıoğlu, A. K., Öztürk, B. A. K. İ., Civalek, Ö. M. E. R., Akgöz, B. E. K. İ. R. "Is Artificial Neural Network Suitable for Damage Level Determination of RC-Structures?" ***International Journal of Engineering and Applied Sciences***, Vol. 2, No. 3, (2010), 71-81.
2. Thaler, D., Stoffel, M., Markert, B., Bamer, F. "Machine-learning-enhanced tail end prediction of structural response statistics in earthquake engineering." ***Earthquake Engineering & Structural Dynamics***, (2021). https://doi.org/10.1002/eqe.3432
3. Goswami, S., Anitescu, C., Chakraborty, S., Rabczuk, T.: Transfer learning enhanced physics informed neural network for phase-field modeling of fracture, Theoretical and Applied Fracture Mechanics, 2020, 106, Article number 102447
4. H. Guo, X. Zhuang, and T. Rabczuk: A deep collocation method for the bending analysis of Kirchhoff plate. Computers, Materials and Continua, Vol. 59, No. 2, (2019), 433-456.
5. C. Anitescu, E. Atroshchenko, N. Alajlan, and T. Rabczuk: Artificial Neural Network methods for the solution of second order boundary value problems. Computers, Materials and Continua, Vol. 59, No. 1, (2019), 345-359.
6. Nguyen-Thanh, V.M., Zhuang, X., Rabczuk, T.: A deep energy method for finite deformation hyperelasticity, European Journal of Mechanics, A/Solids, 2020, 80, Article number 103874
7. Mangalathu, S., Jeon, J. S. "Classification of failure mode and prediction of shear strength for reinforced concrete beam-column joints using machine learning techniques." ***Engineering Structures***, Vol. 160, (2018), 85-94. https://doi.org/10.1016/j.engstruct.2018.01.008
8. Ebtehaj, I., Bonakdari, H., Es-haghi, M. S. "Design of a hybrid ANFIS–PSO model to estimate sediment transport in open channels." ***Iranian Journal of Science and Technology, Transactions of Civil Engineering***, Vol. 43, No. 4, (2019), 851-857. https://doi.org/10.1007/s40996-018-0218-9
9. Safari, M. J. S., Ebtehaj, I., Bonakdari, H., Es-haghi, M. S. "Sediment transport modeling in rigid boundary open channels using generalize structure of group method of data handling." ***Journal of Hydrology***, Vol. 577, (2019), 123951. https://doi.org/10.1016/j.jhydrol.2019.123951
10. Stoffel, M., Bamer, F., Markert, B. "Artificial Neural Networks and intelligent finite elements in non-linear structural mechanics." ***Thin-Walled Structures***, Vol. 131, (2018), 102-106. https://doi.org/10.1016/j.tws.2018.06.035
11. Sun, H., Burton, H. V., Huang, H. "Machine learning applications for building structural design and performance assessment: state-of-the-art review." ***Journal of Building Engineering***, (2020), 101816. https://doi.org/10.1016/j.jobe.2020.101816
12. Aydin, E., Öztürk, B., Guney, D. "Sensitivity analyses of variations on seismic response via viscous damper placement in planar building structures", ***10th International Conference on Urban Earthquake Engineering***, Tokyo, Japan, (Mar. 1-2, 2013), 2013.
13. Khaleghi, M., Salimi, J., Farhangi, V., Moradi, M. J., Karakouzian, M. "Application of Artificial Neural Network to Predict Load Bearing Capacity and Stiffness of Perforated Masonry Walls." ***Civil Engineering***, Vol. *2*, *No.* 1, (2021), 48-67. https://doi.org/10.3390/civileng2010004
14. Mangalathu, S., Jeon, J. S. "Machine learning–based failure mode recognition of circular reinforced concrete bridge columns: Comparative study." ***Journal of Structural Engineering***, Vol. 145, No. 10, (2019), 04019104. https://doi.org/10.1061/(ASCE)ST.1943-541X.0002402
15. Siam, A., Ezzeldin, M., El-Dakhakhni, W. "Machine learning algorithms for structural performance classifications and predictions: Application to reinforced masonry shear walls." ***Structures***, Vol. 22, (2019), 252-265. https://doi.org/10.1016/j.istruc.2019.06.017
16. Kiani, J., Camp, C., Pezeshk, S. "On the application of machine learning techniques to derive seismic fragility curves." ***Computers & Structures***, Vol. 218, (2019), 108-122. https://doi.org/10.1016/j.compstruc.2019.03.004
17. Gaba, A., Jana, A., Subramaniam, R., Agrawal, Y., Meleet, M. Analysis and Prediction of Earthquake Impact-a Machine Learning approach. 4th International Conference on Computational Systems and Information Technology for Sustainable Solution (CSITSS), Bengaluru, India, (2019). https://doi.org/10.1109/CSITSS47250.2019.9031026
18. Burton, H. V., Sreekumar, S., Sharma, M., Sun, H. "Estimating aftershock collapse vulnerability using mainshock intensity, structural response and physical damage indicators." ***Structural Safety***, Vol. 68, (2017), 85-96. https://doi.org/10.1016/j.strusafe.2017.05.009
19. Zhang, Y., Burton, H. V., Sun, H., Shokrabadi, M. "A machine learning framework for assessing post-earthquake structural safety" ***Structural Safety***, Vol. 72, (2018), 1-16. https://doi.org/10.1016/j.strusafe.2017.12.001
20. Zhang, Y., Burton, H. V. "Pattern recognition approach to assess the residual structural capacity of damaged tall buildings." ***Structural Safety***, Vol. 78, (2019), 12-22. https://doi.org/10.1016/j.strusafe.2018.12.004
21. Mazzoni S., McKenna F., Scott M. H., Fenves G. L. OpenSees command language manual. Pacific Earthquake Engineering Research (PEER) Center. 2006.

22. McKenna, F., Scott, M. H., Fenves, G. L. "Nonlinear finite-element analysis software architecture using object composition." ***Journal of Computing in Civil Engineering***, Vol. 24, No. 1, (2010), 95-107. https://doi.org/10.1061/(ASCE)CP.1943-5487.0000002
23. ACI A. 318-19. Building Code Requirements for Structural Concrete. ACI: Farmington Hills, MI, USA. 2019.
24. American Society for Civil Engineers (ASCE), "Minimum Design Loads and Associated Criteria for Buildings and Other Structures", ASCE/SEI 7-16, 2016. https://doi.org/10.1061/9780784414248
25. Barkhordari, M. S., Tehranizadeh, M., Scott, M. H. "Numerical modelling strategy for predicting the response of reinforced concrete walls using Timoshenko theory." ***Magazine of Concrete Research***, (2021), 1-23. https://doi.org/10.1680/jmacr.19.00542
26. Kolozvari, K., Orakcal, K., Wallace, J. W. "New opensees models for simulating nonlinear flexural and coupled shear-flexural behavior of RC walls and columns." ***Computers & Structures***, Vol. 196, (2018), 246-262. https://doi.org/10.1016/j.compstruc.2017.10.010
27. Mander J. B., Priestley M. J., Park R. "Theoretical stress-strain model for confined concrete." ***Journal of structural engineering***. Vol. 114, No. 8, (1988), 1804-26. https://doi.org/10.1061/(asce)0733-9445(1988)114:8(1804)
28. Ancheta T. D., Darragh R. B., Stewart J. P., Seyhan E., Silva W. J., Chiou B. S., Wooddell K. E., Graves R. W., Kottke A. R., Boore D. M., Kishida T. PEER 2013/03: PEER NGA-West2 Database. Pacific Earthquake Engineering Research. 2013.
29. Tran, T. A. "Experimental and analytical studies of moderate aspect ratio reinforced concrete structural walls." Doctoral dissertation, UCLA, (2012).
30. Hastie T., Tibshirani R., Friedman J. The elements of statistical learning: data mining, inference, and prediction. Springer Science & Business Media, (2009).
31. Hoerl A. E., Kennard R. W. "Ridge regression: Biased estimation for nonorthogonal problems." ***Technometrics***. Vol. 12, No. 1, (1970), 55-67. https://doi.org/10.1080/00401706.1970.10488634
32. Tibshirani R. "Regression shrinkage and selection via the lasso." ***Journal of the Royal Statistical Society: Series B (Methodological)***. Vol. 58, No. 1, (1996), 267-88. https://doi.org/10.1111/j.2517-6161.1996.tb02080.x
33. Zou H., Hastie T. "Regularization and variable selection via the elastic net." ***Journal of the Royal Statistical Society: Series B (Statistical Methodology)***. Vol. 67, No. 2, (2005), 301-20. https://doi.org/10.1111/j.1467-9868.2005.00503.x
34. Giussani A. Applied Machine Learning with Python. EGEA spa; 2020.
35. Huber P.J. Robust Estimation of a Location Parameter. In: Kotz S., Johnson N.L. (eds) Breakthroughs in Statistics. Springer Series in Statistics (Perspectives in Statistics). Springer, New York, NY, (1992). https://doi.org/10.1007/978-1-4612-4380-9_35
36. Fischler M. A., Bolles R.C. "Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography." ***Communications of the ACM***, Vol. 24, No. 6, (1981), 381-395. https://doi.org/10.1145/358669.358692
37. Delgoshaei A., Rabczuk T., Ali A., Ariffin M. K. "An applicable method for modifying over-allocated multi-mode resource constraint schedules in the presence of preemptive resources." ***Annals of Operations Research***, Vol. 259, (2017), 85-117. https://doi.org/10.1007/s10479-016-2336-8
38. Scales L. E. Introduction to non-linear optimization. Macmillan International Higher Education; (1985). https://doi.org/10.1007/978-1-349-17741-7
39. Kirkpatrick S., Gelatt C. D., Vecchi M. P. "Optimization by simulated annealing." ***Science***, Vol. 220, (1983), 671-680. https://doi.org/10.1126/science.220.4598.671
40. Zain A. M, Haron H., Sharif S. "Genetic algorithm and simulated annealing to estimate optimal process parameters of the abrasive waterjet machining." ***Engineering with Computers***, Vol. 27, No. 3, (2011), 251-259. https://doi.org/10.1007/s00366-010-0195-5
41. Pedregosa F., Varoquaux G., Gramfort A., Michel V., Thirion B., Grisel O., Blondel M., Prettenhofer P., Weiss R., Dubourg V., Vanderplas J. "Scikit-learn: Machine learning in Python." ***Journal of Machine Learning Research***, Vol. 12, (2011), 2825-2830.
42. Vu-Bac N., Lahmer T., Zhuang X., Nguyen-Thoi T., Rabczuk T. "A software framework for probabilistic sensitivity analysis for computationally expensive models." ***Advances in Engineering Software***. Vol. 100, (2016), 19-31. https://doi.org/10.1016/j.advengsoft.2016.06.005
43. Chakroborty S., Roy R. "Role of ground motion characteristics on inelastic seismic response of irregular structures." ***Journal of Architectural Engineering***. Vol. 22, No. 1, (2016), B4015007. https://doi.org/10.1061/(asce)ae.1943-5568.0000185
44. Fatemi H., Paultre P., Lamarche C. P. "Experimental Evaluation of Inelastic Higher-Mode Effects on the Seismic Behavior of RC Structural Walls." ***Journal of Structural Engineering***, Vol. 146, No. 4, (2020), 04020016. https://doi.org/10.1061/(asce)st.1943-541x.0002509
45. Barkhordari, M. S., Tehranizadeh, M. "Ranking Passive Seismic Control Systems by Their Effectiveness in Reducing Responses of High-Rise Buildings with Concrete Shear Walls Using Multiple-Criteria Decision Making." ***International Journal of Engineering, Transactions B: Applications***, Vol. 33, No. 8, (2020), 1479-1490. https://doi.org/10.5829/ije.2020.33.08b.06

APPENDIX A

TABLE A1. Parameters of steel material

Yield strength	Initial elastic tangent	Strain-hardening ratio
$420MPa$	$200\ GPa$	0.01

TABLE A2. Parameters of concrete material

Compressive strength (MPa)	Unconf.	55
	Confined	66
Strain at the compressive strength	Unconf.	-0.002
	Confined	-0.005
Strain at the tensile strength	0.00008	
Tensile strength	$1.9MPa$	
Concrete modulus of elasticity	$37GPa$	

TABLE A3. The frame section of the buildings

Building	No. of story	Thickness (cm)	Long. Reinforcement
20-story	1-5	45	Ø25@15cm
	6-10	45	Ø20@20cm
	11-15	35	Ø20@25cm

	16-20	35	Ø20@25cm
25-story	1-5	45	Ø25@15cm
	6-10	45	Ø20@20cm
	11-15	45	Ø20@25cm
	16-20	35	Ø20@25cm
	21-25	35	Ø20@25cm
30-story	1-5	45	Ø25@15cm
	6-10	45	Ø20@20cm
	11-15	45	Ø20@20cm
	16-20	35	Ø20@25cm
	21-25	35	Ø20@25cm
	26-30	35	Ø20@25cm

TABLE A.4. List of ground motions

ID(s)	Name	Year	ID - Station Name
1-2	"Imperial Valley-06"	1979	1-"Brawley Airport" 2-"El Centro Array #10"
3-6	"Loma Prieta"	1989	3- "Gilroy - Historic Bldg." 4-"Gilroy Array #2" | 5-"Gilroy Array #3" | 6- "Saratoga - W Valley Coll." |
7	"Chi-Chi_ Taiwan"	1999	7-"CHY101"
8	"Duzce_ Turkey"	1999	8-"Bolu"
9	"Chuetsu-oki_ Japan"	2007	9- "Joetsu Kakizakiku Kakizaki"
10	"Darfield_ New Zealand"	2010	10- "Riccarton High School "
11-12	"El Mayor-Cucapah_ Mexico"	2010	11- "El Centro Array #12" 12- "Westside Elementary School"
13	"Imperial Valley-06"	1979	13- "El Centro Array #11"
14	"Superstition Hills-02"	1987	14- "Poe Road (temp)"
15	"Superstition Hills-02"	1987	15- "Westmorland Fire Sta"
16	"Northridge-01"	1994	16- "Beverly Hills - 14145 Mulhol"
17	"Kobe_ Japan"	1995	17- "Amagasaki"
18	"Kocaeli_ Turkey"	1999	18- "Duzce"
19	"Iwate_ Japan"	2008	19- "MYG005"
20-23	"El Mayor-Cucapah_ Mexico"	2010	20- "CERRO PRIETO GEOTHERMAL" 21- "MICHOACAN DE OCAMPO" | 22- "RIITO" | 23- "EJIDO SALTILLO" |
24	"Darfield_ New Zealand"	2010	24- "DFHS"
25	"Christchurch_ New Zealand"	2011	25- "Papanui High School "
26	"Northern Calif-03"	1954	26- "Ferndale City Hall"
27	"Coalinga-01"	1983	27- "Parkfield - Fault Zone 14"
28	"Loma Prieta"	1989	28- "Hollister - South & Pine" 29- "Hollister City Hall"
30	"Kobe_ Japan"	1995	30-"Sakai" 31- "Yae"
32-35	"Chi-Chi_ Taiwan"	1999	32- "TCU038" 33-"TCU112" | 34-"TCU117" | 35-"TCU118" |
36	"St Elias_ Alaska"	1979	36- "Icy Bay"
37	"Chi-Chi_ Taiwan-03"	1999	37-"CHY025" 38- "TCU065"
39	"Chuetsu-oki_ Japan"	2007	39- "Joetsu City"
40-42	"Iwate_ Japan"	2008	40- "Nakashinden Town" 41- "Semine Kurihara City" | 42- "Yokote Masuda Tamati Masu" |
43-47	"El Mayor-Cucapah_ Mexico"	2010	43- "TAMAULIPAS" 44-"El Centro - Meloland Geot. Array" | 45- "El Centro - Meloland Geotechnic" | 46-"El Centro Array #7" | 47- "El Centro - Meadows Union School" |
48-50	"Darfield_ New Zealand"	2010	48- "Hulverstone Drive Pumping Station" 49- "NNBS North New Brighton School " 50- "SPFS"
51	"Northwest Calif-02"	1941	"Ferndale City Hall"
52	"Northern Calif-01"	1941	"Ferndale City Hall"
53	"Borrego"	1942	"El Centro Array #9"
54	"Kern County"	1952	"Santa Barbara Courthouse"
55	"Kern County"	1952	"Taft Lincoln School"
56	"Southern Calif"	1952	"San Luis Obispo"
57	"Parkfield"	1966	"San Luis Obispo"
58	"Borrego Mtn"	1968	"El Centro Array #9"
59-75	"San Fernando"	1971	59- "2516 Via Tejon PV" 60-"Carbon Canyon Dam"

	61- "Cedar Springs Pumphouse" | 62- "Cedar Springs_ Allen Ranch" | 63- "Colton - So Cal Edison" | 64- "Fort Tejon" |		
	65-"Gormon - Oso Pump Plant" | 66- "LB - Terminal Island" | 67- "Pearblossom Pump" | 68- "Port Hueneme" | 69-"Puddingstone Dam (Abutment)" | 70- "Santa Anita Dam" | 71- "Tehachapi Pump" | 72- "Upland - San Antonio Dam" | 73- "Wheeler Ridge - Ground" | 74- "Whittier Narrows Dam" | 75- "Wrightwood - 6074 Park Dr" |		
76-78	"Friuli_ Italy-01"	1976	76- "Barcis" 77- "Codroipo" 78-"Conegliano"
79	"Tabas_ Iran"	1978	79-"Ferdows"
80-83	"Imperial Valley-06"	1979	80- "Coachella Canal #4"
	81- "Niland Fire Station" | 82-"Plaster City" | 83- "Victoria" |		
84	"Victoria_ Mexico"	1980	84- "SAHOP Casa Flores"
85-87	"Trinidad"	1980	85- "Rio Dell Overpass - FF"
	86- "Rio Dell Overpass_ E Ground" | 87- "Rio Dell Overpass_ W Ground" |		
88-91	"Irpinia_ Italy-01"	1980	88-"Arienzo"
	89-"Bovino" | 90-"Torre Del Greco" | 91-"Tricarico" |		
92-95	"Irpinia_ Italy-02"	1980	92- "Bovino"
	93-"Brienza" | 94- "Mercato San Severino" | 95- "Tricarico" |		
96-97	"Coalinga-01"	1983	96- "Parkfield - Cholame 12W" 97-"Parkfield - Cholame 1E"
98-126	"Coalinga-01"	1983	98-"Parkfield - Cholame 2E"
	99- "Parkfield - Cholame 2WA" | 100- "Parkfield - Cholame 3E" | 101- "Parkfield - Cholame 3W" | 102- "Parkfield - Cholame 4AW" | 103-"Parkfield - Cholame 4W" | 104- "Parkfield - Cholame 5W" | 105- "Parkfield - Cholame 6W" | 106- "Parkfield - Cholame 8W" | 107- "Parkfield - Fault Zone 1" | 108- "Parkfield - Fault Zone 10" | 109- "Parkfield - Fault Zone 2" | 110- "Parkfield - Fault Zone 3" | 111- "Parkfield - Fault Zone 4" | 112-"Parkfield - Fault Zone 6" |		
	113- "Parkfield - Fault Zone 9" | 114- "Parkfield - Gold Hill 1W" | 115-"Parkfield - Gold Hill 2E" | 116- "Parkfield - Gold Hill 2W" | 117-"Parkfield - Gold Hill 3W" | 118-"Parkfield - Gold Hill 4W" | 119- "Parkfield - Gold Hill 5W" | 120- "Parkfield - Gold Hill 6W" | 121-"Parkfield - Stone Corral 2E" | 122- "Parkfield - Stone Corral 3E" | 123-"Parkfield - Stone Corral 4E" | 124- "Parkfield - Vineyard Cany 3W" | 125- "Parkfield - Vineyard Cany 4W" | 126-"Parkfield - Vineyard Cany 6W" |		
127	"Ierissos_ Greece"	1983	127- "Ierissos"
128-136	"Taiwan SMART1(25)"	1983	128-"SMART1 C00"
	129-"SMART1 E01" | 130- "SMART1 E02" | 131- "SMART1 I01" | 132- "SMART1 I07" | 133- "SMART1 M01" | 134-"SMART1 M06" | 135- "SMART1 O01" |		
	136- "SMART1 O07" |		
137-143	"Borah Peak_ ID-01"	1983	137- "CPP-601" 138-"CPP-610"
	139- "PBF (second bsmt)" | 140-"TAN-719" | 141- "TRA-642 ETR Reactor Bldg(Bsmt)" | 142- "TRA-670 ATR Reactor Bldg(Bsmt)" |		
143-150	"Morgan Hill"	1984	143- "APEEL 1E - Hayward"
	144- "Capitola" | 145- "Foster City - APEEL 1" | 146-"Fremont - Mission San Jose" | 147- "Hollister City Hall" |		
	148- "Los Banos" | 149-"SF Intern. Airport" | 150- "San Justo Dam (L Abut)" |		

APPENDIX B

TABLE B1. Comparison of α (Ridge Reg.)

Ridge Reg.	Displacement	Drift	Base shear	Structure
α	0.001	0.022	0.022	15-stroy
α	0.001	0.278	0.278	20-story
α	0.001	0.022	0.022	25-story
α	0.001	0.002	0.001	30-story

TABLE B2. Comparison of α (Elastic Net Reg.)

Elastic Net Reg.	Displacement	Drift	Base shear	Structure
α	0.01326	0.00130	0.2222	15-stroy
α	0.02811	0.00010	0.2222	20-story
α	0.04941	0.00167	0.0193	25-story
α	0.01900	0.00115	0.1264	30-story

TABLE B3. Comparison of α (Lasso Reg.)

Lasso Reg.	Displacement	Drift	Base shear	Structure
α	0.00052	0.00010	1526.41	15-stroy
α	0.00168	0.00010	2223.00	20-story
α	0.00268	0.00010	3237.45	25-story
α	0.00066	0.00010	3237.46	30-story

Persian Abstract

چکیده

پیش بینی پاسخ‌های دیوارهای برشی بتن آرمه تحت اثر حرکات قوی زمین در طراحی، ارزیابی و تصمیم گیری راهبردهای مقاوم سازی بسیار مهم است. این مطالعه توانایی مدل‌های رگرسیون و یک روش ترکیبی (مدل ANN-SA)، شبکه عصبی مصنوعی (ANN) و الگوریتم تبرید شبیه‌سازی‌شده (SA)، را برای پیش بینی پاسخ دیوارهای برشی بتن آرمه تحت تأثیر حرکت‌های قوی زمین ارزیابی می‌کند. برای این منظور، چهار ساختمان (۱۵ ، ۲۰ ، ۲۵ و ۳۰ طبقه) با دیوارهای برشی بتنی در OpenSees مورد تجزیه و تحلیل قرار گرفته است. ۱۵۰ رکورد لرزه ای برای تولید یک پایگاه داده جامع از ورودی (مشخصات) و خروجی (پاسخ ها) مورد استفاده قرار گرفته است. حداکثر شتاب، حداکثر سرعت و مشخصات زلزله به عنوان پیش بینی کننده (متغییرهای ورودی) استفاده شده است. در این مطالعه دقت مدل های ساده و روش ترکیبی در شناسایی پاسخ‌های دیوارهای برشی مقایسه شده و همچنین حساسیت متغیرهای ورودی به مدل تقاضای لرزه‌ای بررسی شده است. نتایج نشان می‌دهد که مدل ANN-SA از دقت منطقی در پیش بینی برخوردار است.

IJE TRANSACTIONS A: Basics Vol. 34, No. 7, (July 2021) 1602-1610

International Journal of Engineering

Journal Homepage: www.ije.ir

Numerical Evaluation of Slope Stability for Construction and Seismic Loads: Case Study

Z. Alsharifi[a], M. Shakir Mahmood*[a], A. Akhtarpour[b]

[a] *Civil Engineering Department, University of Kufa, Al-Najaf, Iraq*
[b] *Civil Engineering Department, Ferdowsi University of Mashhad, Mashhad, Iran*

PAPER INFO

Paper history:
Received 28 March 2021
Received in revised form 17 May 2021
Accepted 18 May 2021

Keywords:
Slope Stability
Finite Element Method
Strip Footing
Earthquake

ABSTRACT

Al-Najaf, one of Iraq's most important cities, is expected to grow in the coming years. Many buildings located on the slopes of Al-Najaf, which is near the shrine of Imam Ali, possess important economic and tourism values. This paper examines the stability of the city's slope under a variety of conditions, including slope geometry, neighbor structure loading, and earthquake magnitude. A computer-aided 2D Finite Element Method is adopted in the analysis. A set of soil classification and identification tests were carried out in addition to the available required soil parameters in the constitutive modeling. The Mohr-Coulomb model is applied to define the failure state that began in the slope. The results show that the slope is stable due to its weight and geometry, with a minimum factor of safety of 2.6. While under footing loading, this factor of safety decreases to less than 1.8. The most hazardous condition is when the slope has been subjected to seismic loading and the factor of safety is less than unity, for all investigated cases and characteristics, where the slopes would collapse.

***doi**: 10.5829/ije.2021.34.07a.05*

NOMENCLATURE

A	Footing area (m^2)	PI	Plastic index
B	Strip footing width (m)	q_u	Bearing capacity(kPa)
Cc	Coefficient of curvature	X	Distance from slope edge to the footing (m)
Cu	Coefficient of uniformity	**Greek Symbols**	
C	Cohesion (kPa)	σ	Normal stress(kPa)
D	Damping Ratio	σ'_m	Mean stress(kPa)
E	Yong's modulus	ɵ	Slope angle (degree)
FS	Factor of safety	τ_f	Failure shear stress (kPa)
G	Shear modulus(kPa)	ϕ	Internal friction Angle (degree)
G_{max}	Maximum shear modulus(kPa)	γ_{dry}	Dry soil density (g/cm^3)
Gs	Specific gravity	γ_c	Cyclic shear strain
H	Slope height (m)	ω	Natural Moistness content
k	Permeability (m/s.)	υ	Poisson's ratio
P	Aplied load (kN)	ε	Strain
Pa	Atmospheric pressure(kPa)		

1. INTRODUCTION

Flatland shortage and rapid development in hilly terrain with appealing sights are impacting construction on hillsides, and so are many urban excavations in the vicinity of existing structures [1].

Rainfall, earthquakes, erosion, geological characteristics, and construction events are the factors that can cause slope failure. The slope stability analyses are demonstrated on hypotheses, where the

*Corresponding Author Email: mohammedsh.alshakarchi@uokufa.edu.iq (M. Shakir Mahmood)

Please cite this article as: Z. Alsharifi, M. Shakir Mahmood, A. Akhtarpour, Numerical Evaluation of Slope Stability for Construction and Seismic Loads: Case Study, International Journal of Engineering, Transactions A: Basics Vol. 34, No. 7, (2021) 1602-1610

configuration of the slope depends on practice and reliable inspection, (faults, stratification, etc.) [2].

The forms of a slope failure can be classified as, fall, topple slide, spread, and flow [3].

The limit equilibrium methods are suitable for practical purposes where the equilibrium equations are less than the unknowns in slope stability problems, and with assumptions, the problem can be reformed from indeterminate to determinate [4].

The stabilization of footings adjacent to the slope and the slope is obligatory since both bearing capacity and slope stability need to be revealed [5, 6]. Plentiful standards and codes primarily focus on the footings' design in flatlands, giving unsuitable guidelines for the constructions near the slope [5].

Enormous earthquakes are significant in slope collapse [7]. For seismic loading with a decreased soil bearing capacity and an increase in deformation, the security of structures should be assured in recently constructed [8, 9]. The input parameters include characteristics, the earthquake magnitude, frequency, intensity of ground motions, and period of a shake [10].

Iraq is positioned in a seismically active zone and, as a result, is an instability region that has undergone many powerful earthquakes in recent decades. Extreme events have taken numerous lives, damaged several towns, and caused extensive financial harm in this era. The last enormous seismic event was struck at Iran- Iraq frontier in November 2017 [11].

The challenge of a loaded slope with footing has been widely analyzed in a normal loading state, static load. Yet, the dynamic analysis of a loaded slope exposed to seismic force was not thoroughly explored. However, limit equilibrium analysis was employed to inspect the stability of dynamically loaded slope [12].

The soil's shear strength can be enhanced by the long term soaking method [13]. The risk of collapse in all soil groups rises due to a drop in soil matric suction [14-16].

The Mohr-Coulomb model (elastic-perfectly plastic model) is the simplest method for representing soil failure in constitutive models, giving conservative values to all inspected parameters of the footing and dam soils when compared to the Modified Cam-Clay modeling (elastoplastic model) [17-19].

For several years, slope stability analysis that used the finite element approach has been commonly recognized in the literature. The major finite element slope stability approaches used today are the enhanced limit strength method and the strength reduction method. Similarities of the finite element and limit equilibrium approach analyses of slope stability demonstrate the benefits and drawbacks of these approaches for convenient technical challenges [20].

Even though many modern numerical approaches, including the finite element method, have been introduced over the last three decades; the traditional limit equilibrium procedure has remained the most effective in slope stability analysis. Moreover, because of the finite element method's numerous benefits over the limit equilibrium method, However, due to the numerous benefits of the finite element method above the limit equilibrium method, the finite element method now has an increase in applications for stability analysis [21].

The recent paper focuses on the stability analysis of Al-Najaf city's slope under different geometry and loading conditions. The assessment of the slope is important due to the attraction and it is the first numerical model be made for the slopes of Al-Najaf city for the effect of adjacent buildings, earthquakes and for its specific soil characteristics.

2. RESEARCH METHODOLOGY

2. 1. Materials The study area is located within the slope that lay beside Al-Najaf old city and near to the shrine of Imam Ali in south-western of Iraq, as shown in Plate 1. between the longitude lines of 43^o 00' and 44^o 30' and latitudes of 30^o 45' and 32^o 15'. This location has recently developed, and more development is expected as a result of the city's religious attraction. Construction is one of the development aspects, and there is a need to investigate the city's slope instability caused by the extra loading from neighboring construction, earthquakes and rainfall.

The samples were collected from three different sites along the slope with symbols of S1, S2, and S3, as shown in Plate 2.

A set of tests were conducted on the soil samples from the slope site according to ASTM procedures. Figure 1 illustrates the grain size distribution of the different soil samples and reveals that all soil samples are similar in gradation. For this reason, a standard Proctor test is performed for S2, as shown in Figure 2. Table 1 summarizes the results of soil identification and classification tests. The soil samples are classified according to the Unified Soil Classification System (USCS) as' 'SP'. These data are important to select the soil density, angle of internal friction, permeability coefficient, and Yong's modulus.

Many other soil properties are required for the selected constitutive model. Many site investigations and papers are published concerning the soil mechanical properties which are computed through experimental work. Table 2 illustrates the selected input soil properties.

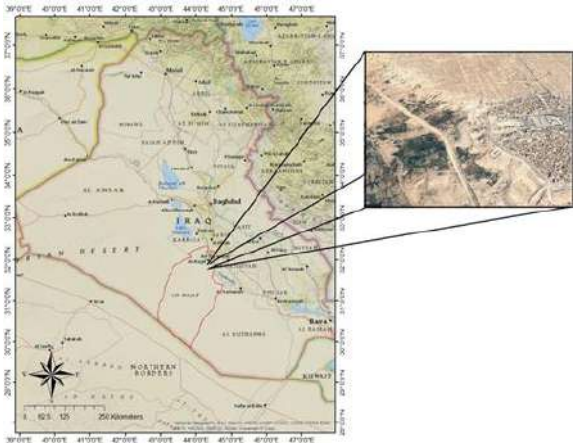


Plate 1. The site of research area

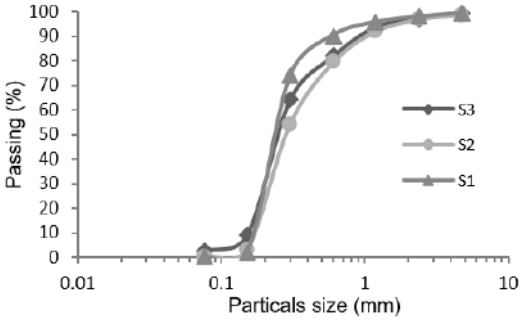


Figure 1. Grain size distribution of the soil samples

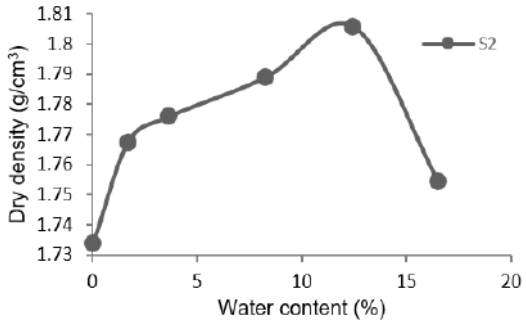


Figure 2. Standard Proctor test results

2. 2. Slope Geometry Using the software titled "Google Earth", the height and angle of the slope are determined. Plate 2 shows the sections of the slope geometry measurements. The height of the slope is ranged between 16 m and 24 m where the slope angle is ranged between 4° and 9° as shown in Table 3.

The slope safety is examined in this paper for the various geometry boundaries where the slope heights (H) are 15, 20 and 30 m and the slope angles (θ) are 5, 7, 10 and 15 degrees.

2. 3. Loading A static load represented by building with strip footing is covered in this study which had a constant applied stress equal to 85 kPa. The study investigates a different footing width (B) and distance from the slope edge (X), as shown in Figure 3.

To investigate the effect of the most dangerous earthquake case that occurred in Iraq, Halabjah earthquake, 2017 returning quake on Kirkuk, North-South direction (with magnitude of 4.9) data (acceleration time history) was adopted from literature [22].

TABLE 1. The results of soil identification tests

Soil identification	S-1	S-2	S-3
Sand, %	99.09	98.66	96.73
Fine, %	0.34	0.33	2.65
D_{10}, mm	0.167	0.170	0.153
D_{30}, mm	0.208	0.229	0.207
D_{60}, mm	0.271	0.368	0.289
Coefficient of uniformity (Cu)	1.623	2.167	1.891
Coefficient of curvature (Cc)	0.961	0.836	0.973
Soil Classification (USCS)	SP	SP	SP
Natural Moistness content (w), %	3.14	3.51	5.05
Maximum dry density(γ_{dry}), g/cm^3	-	18.06	-
Optimum moisture content, %	-	12.47	-
Specific gravity (Gs)	2.623	2.622	2.633
Gypsum content, %	-	1.23	-

TABLE 2. Additional necessary soil properties

Soil properties	Min.	Max.	Avg.
Angle of internal friction (Φ), deg. [23]	30	35	33
Cohesion (C), kPa [23]	0	0	0
Bearing capacity, kPa [23]	70	100	85
Yong's modulus (E), kPa [24]		20125	
Poisson's ratio (υ) [24]		0.35	

Plate 2. The investigation of slope geometry

TABLE 3. The geometry of the city's slopes

Section	Slope height (H), m	Slope angle (θ), deg.
1	24	6.36
2	21	4.92
3	16	4.16
4	29	9.02
5	19	7.61
6	24	7.10
7	16	6.80
8	21	7.72

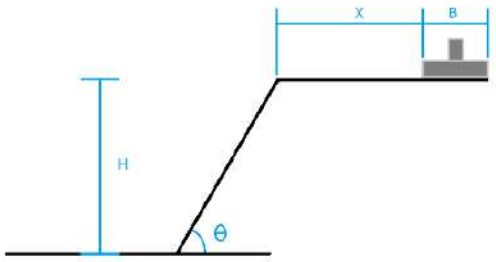


Figure 3. Sketch of slope's investigated parameters

2. 4. Constitutive Model The Finite Element Method (FEM) is performed for numerical analysis. The analysis depended on the principle of limit equilibrium, which combines a finite element method that is advanced, especially for the stability and deformation of slopes and embankment structures [25].

The case study in this study were modeled in two-dimensional finite element of quads and triangle mesh with global elements size equal to 8 m and a 0.24m mesh size for the ground surface of the model. The integration order were 4 nodes for quadrilateral elements and 3 nodes for triangular elements in addition to secondary nodes on the boundary.

The fundamentals of the soil modeling program, using the gathered parameters of the study area (Tables 1 and 2), are explained as the following: The Mohr-Coulomb formula, Equation (1), is applied to define the failure state that began on the slope [26]:

$$\tau_f = C + \sigma \tan \phi \quad (1)$$

The soil has an elastic, perfectly-plastic with regular stress-strain relationships. Stresses are directly proportional to strains until the yield point is reached. Past the yield point, the stress strain curve is faultlessly horizontal (behaves linearly), with two parameters from Hooke's law: Young's modulus (E=20125 kPa) and Poisson's ratio (ν=0.35). As well as two parameters which describe the failure conditions: the internal friction angle (ϕ=35º) and cohesion (C=0) [27].

The soil will be modeled using an equivalent linear method during the dynamic analysis. Equation (2) states the linear relationship between stress and strain using the proportionality factor of Young's modulus. Yet, the variation is that the shear modulus is changed in response to strains. With the defined soil stiffness, identifying the maximum shear strains for each Gauss numerical integration point in each element. The shear modulus is then adjusted in accordance with a predetermined G reduction function, and the step was repeated. This iterative process is repeated until the required G adjustments are made according to shear strain. G is a constant when stepping via the earthquake record, which is an important behavior to understand. G can be changed by each pass via the record, but it always stays constant during a single pass [28].

$$\sigma = E\varepsilon \quad (2)$$

In this study, Equation (3) was provided to determinate shear modulus (G) during the analysis stages by inbuilt function. For medium dense sand (as it appear from seive analysis data and the value of internal friction angle ϕ), K=50, where K is a parameter based on soil type found by Seed and Idriss [29], plastic index (PI)=0 and γ_{dry}= 18 KN/m^3 giving the G_{max} function (Figure 4) for the slopes' soil according to Equation (3).

$$G_{max} = 22K\sqrt{P_a \sigma'_m} \quad (3)$$

A dynamically stressed soil will be "softer" in contrast to cyclic shear stress. This softness is explained as a ratio of Gmax in the Equivalent Linear modeling. This is known as G-reduction function. For the calculation of new G values in each iteration, the determined shear strain along with the function and Gmax is used. Equations (4), (5) and (6) express a formula for calculating the G/G_{max} ratio [28] using plastic index (PI)=0 and γ_{dry}= 18 kN/m^3 gives the G-reduction function (Figure 5) for the slopes' soil according to Equation (4).

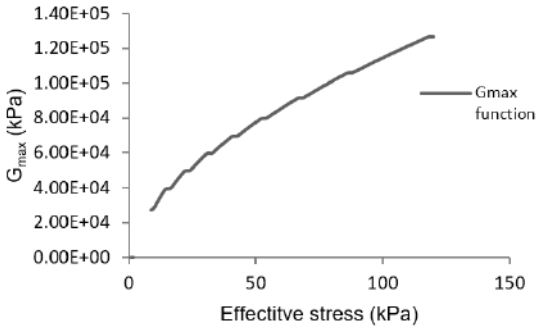


Figure 4. Max. shear modulus (G_{max}) function of the studied soil according to Equaction (3)

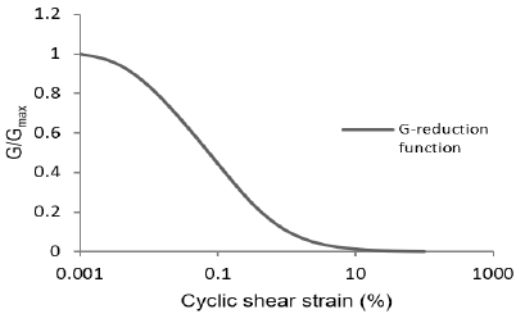


Figure 5. G-reduction function of the studied soil according to Equaction (4)

$$\frac{G}{G_{max}} = K(\gamma, PI)(\sigma'_m)^{m(\gamma,PI)-m_0} \quad (4)$$

$$K(\gamma,PI) = 0.5\left[1+\tanh\left(\ln\left(\frac{0.000102+n(PI)}{\gamma}\right)^{0.492}\right)\right] \quad (5)$$

$$m(\gamma,PI)-m_0 = 0.272\,[1-\tanh(\ln(\frac{0.000556}{\gamma})^{0.4})]\exp(-0.0145\,PI^{1.3}) \quad (6)$$

The damping ratio was also generated from a inbuilt function (Equation (7)) was developed by Kramer [30] using plastic index (PI)=0 and γ_{dry}= 18 kN/m^3 gives the Damping function (Figure 6) for the slopes' soil according to Equation (7).

$$D = 0.333\frac{1+e^{-0.0145PI^{1.3}}}{2}\left(0.586\left(\frac{G}{G_{max}}\right)^2 - 1.547\frac{G}{G_{max}} + 1\right) \quad (7)$$

The validation of the numerical modeling of this research was not done using models for the study area soil and slope properties as this research is an original study, a verification is made by comparison of the results with published limit equilibrium calculations. Three modeles were exploit:

1. A hand calculation based on limit equilibrium methods puplished by GEO-SLOPE International Ltd. (homogeneous soil, 2:1 slope H=10 m, γ = 20 kN/m^3, C'=5 kPa and ϕ' = 30°).

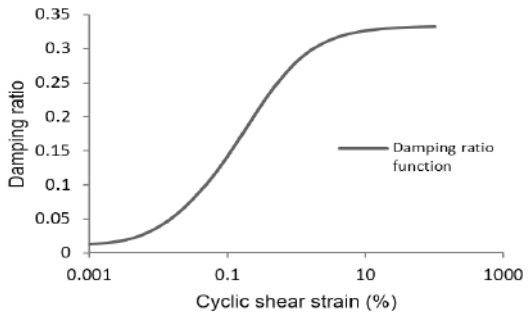


Figure 6. Damping ratio function of the studied soil according to Equaction (7)

2. Verification problem#1in SLIDE verification manual (homogeneous soil, 2:1 slope H=10 m, γ = 20 kN/m^3, C'=3 kPa and ϕ' = 19.6°).
3. Verification problem#14 in SLIDE verification manual (homogeneous soil, 3:2 slope H=20 m, γ = 18.82 kN/m^3, C'=41.65 kPa and ϕ' = 15°).

The results gave a good agreement between the analysis for the inspected models ($1 \geq R^2 > 0.97$).

3. RESULTS AND DISCUSSION

3. 1. Effect of Slope Geometry Figure 7 illustrates an example of slope slip surface and factor of safety in the case of the highest values of slope height (H) and angle (θ). Figure 8 provides the factor of safety values for various investigated slope heights and angles. Noticeably, there is a strong impact from slope angle on the safety where the effect of the slope height is marginal within the investigated ranges (critical values of the city's slope). These results are matched with literature [31], and may be attributed to the significant effect of the slope angle on the force component that is parallel to the slip surface where the failure occurs when the shear force is larger than the shear strength of the soil [32].

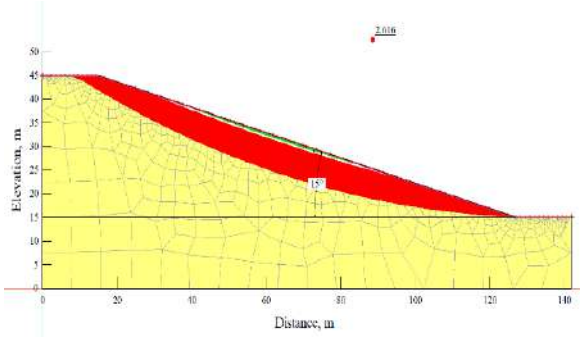


Figure 7. Slope Dermation with Slope height (H) of 30m and Slope angle (θ) of 15degree

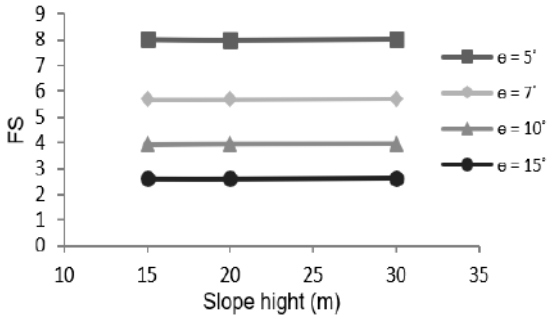


Figure 8. The effect of slope geometry on factor of safety

The minimum value of factor of safety is about 2.6, indicating that the slope in Al-Najaf is safe within the investigated geometry.

3. 2. Effect of Neighbor Construction The effect of strip footing width (B) and its distance from the slope edge (X) are investigated. For all cases with neighbor footing, a fixed applied stress of 85 kPa is given. Figure 9, as an example, proves that the failure in the slope of the city due to the specific footing is locally, i.e., the slope is safe and the failure is due to pressure limit of the footing and it matched with findings of failure modes by Mofidi et. al. [5].

Generally, there is a decrease in the factor of safety (FS) due to the application of the footing with respect to the initial condition (before construction). Figure 10 illustrates the effect of footing width on FS of the slope. There is a slight relationship between footing width and FS where this behavior is reflected in the increase of bearing capacity (q_u) of the footing with the increase of its width. The rise in q_u may be credited to the bigger interaction area between footing and subsoil [33].

Obviously, there is a local failure (due to footing pressure) where the footing is closer to the slope (X=1.8-2m). The maximum FS decreases with the existence of building footing to less than 1.8, yet, the distance between footing and slope crest effects FS with direct proportion.

Figure 11 verifies the effect of footing distance ratio (X/B) on the factor of safety of the slope. The results indicate that the FS increases with increasing X/B. When the footing is moved away from the slope crest (X/B=4), there is a clear increase in FS (about 50%). Nevertheless, the rate of increase in FS decreases with increasing distance beyond X/B=4 where the failure due to bearing capacity is dominated.

The influence of slope is lessened when the footing is sited at an edge distance above four times the width of the footing. This increase in FS with respect to slope crest distance could be return to ground passive resistance from the slope edge [33].

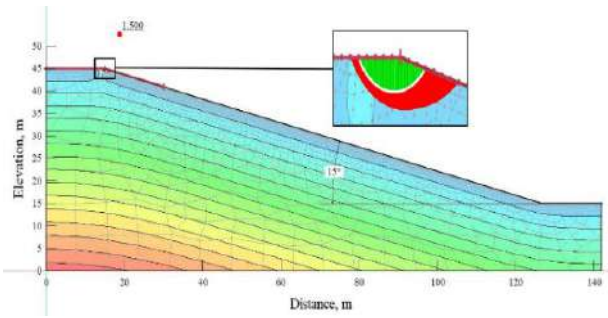


Figure 9. Slope failure with Footing of 0.8m in Width at a distance (X) of 0.8m for Slope of H=30m and θ=15 degree^o

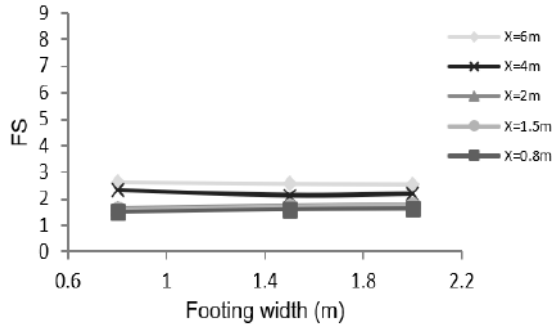


Figure 10. The effect of footing width on slope factor of safety for slope (H=30m and θ=15^o)

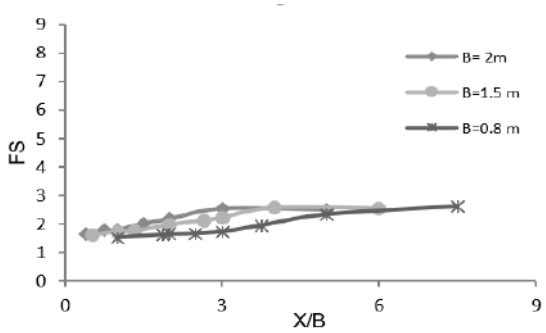


Figure 11. The relationship between factor of safety and X/B for slope (H=30m and θ=15^o)

3. 3. Effect of Seismic Load To investigate the effect of the earthquake on the stability of the slope in Al-Najaf city, an earthquake of 4.9 magnitude is applied to the 30m height (H) and different inclination slope (θ=7^o, 10^o and 15^o) the critical geometry of the city's slope. Figure 12 shows the variation of FS throughout the earthquake period.

The factor of safety swayed with the variation of acceleration along seismic waves initiating instability and risky state on the slope, where the FS reaches a value less than unity. Figure 13 compares the FSs before and after the earthquake. The FS decreased more than 81% from their original FS before the quake. The effect of earthquakes on slope stability can be understood as an equivalent increase in the inclination of joint failure plane [34]. The inspected slope within the study area failed after being exposed to the earthquake for the investigated parameters when it was not even severed from footing loads.

To express the huge damage that the existing constructions in the studied area could suffer, the seismic loading impact was examined with a combination of footing. Figure 14 compares the FSs before and after the seismic event for different footing width (B). After the earthquake, the FSs dropped by 68-78%, with larger B indicating a higher percentage drop.

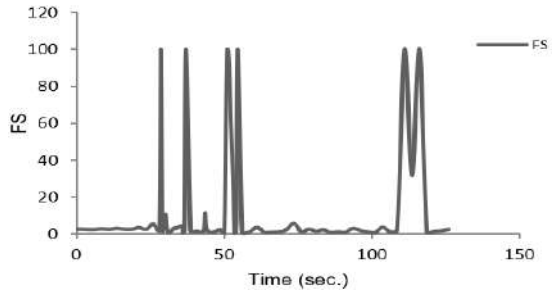


Figure 12. FS change along earthquake period for slope without footing, (H)=30m and (θ)=15 deg

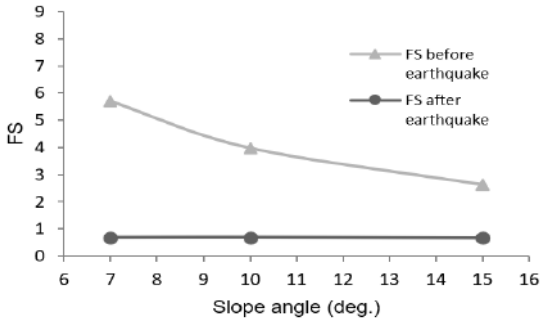


Figure 13. The effect of earthquake on factor of safety on slopes without footing (H=30m)

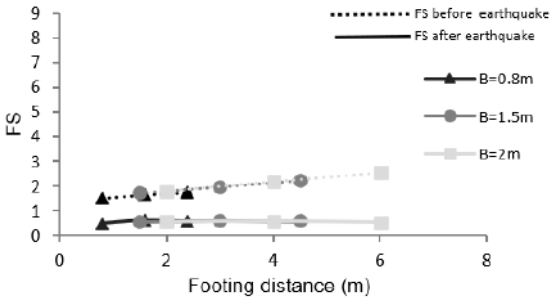


Figure 14. The effect of earthquake on factor of safety on slopes with footing B=(0.8, 1.5 and 2)m for slope (H=30m,θ=15º)

This reduction in FS is caused by the irritation effect of the dynamic load created on the superstructure generating extra forces which are transported to the footings, causing an additional decrease in the bearing capacity (qu). The results gave a minimum and maximum FS equal to 0.48 and 0.58 respectively.

4. CONCLUSIONS

The recent paper investigates the effect of neighboring footing and seismic forces on the stability of Al-Najaf city slopes numerically. The following conclusions can be drawn from the above research:

- The smallest value of the factor of safety is around 2.6, implying that the slope in Al-Najaf is safe in regards to the investigated slope characteristics without loading.
- The slopes' inclined angle (θ) had the greatest influence on safety (in reverse proportion), while the height had only a minor effect under the scenarios of slope geometry and no loading.
- The impact of footing on slope stability in the study area is local (due to bearing capacity failure) concerning the investigated footing width (B), distance (X), and stress (85kPa). As a maximum value, the FS is reduced to less than 1.8.
- The factor of safety increases with the increasing of the footing distance ratio (X/B) until distance beyond (X/B=4) where the increase reduces as the bearing capacity of the footing on slope approaches that of a footing on level ground.
- During the earthquake, the FS of the slope is less than unity with its own weight for all investigated slope geometry. The FSs values are reduced by approximately 81%.
- When neighbour footing and seismic loads are applied, the FS of slope was reduced by about 68-78% compared to the static footing state.
- The current structures built on the slope of Al-Najaf are vulnerable to collapse and destruction in the event of an earthquake similar to the one used in the investigation modelling.

5. REFERENCES

1. Mahmood, M.S., Akhtarpour, A. and Alsharifi, Z., "A review of slopes stability challenges and neighbour buildings", in Journal of Physics: Conference Series, IOP Publishing. Vol. 1895, No. 1, (2021), 012014.
2. Budhu, M., *Soil mechanics and foundation–3rd edition" john wiley & sons inc*. 2010, ISBN 978-040-55684-9.
3. Cruden, D. and Varnes, D., *Landslide types and processes. Landslides: Investigation and mitigation (ed. Turner, ak and schuster, rl), special report 247 of the transport research board, national research council*. 1996, National Academy Press, Washington DC.
4. Samani, H.M. and Meidani, M., "Slope stability analysis using a non-linear optimization technique (research note)", ***International Journal of Engineering***, Vol. 16, No. 2, (2003), 147-156, doi.
5. Mofidi Rouchi, J., Farzaneh, O. and Askari, F., "Bearing capacity of strip footings near slopes using lower bound limit analysis", ***Civil Engineering Infrastructures Journal***, Vol. 47, No. 1, (2014), 89-109, https://dx.doi.org/10.7508/ceij.2014.01.007.

6. Supandi, S., "Simple slope stabilization on quartz sandstone using horizontal drain", ***International Journal of Engineering***, Vol. 34, No. 4, (2021), 1046-1051, doi.
7. Raj, D., Singh, Y. and Kaynia, A.M., "Behavior of slopes under multiple adjacent footings and buildings", ***International Journal of Geomechanics***, Vol. 18, No. 7, (2018), 04018062, https://doi.org/10.1061/(ASCE)GM.1943-5622.0001142.
8. Nadi, B., Askari, F. and Farzaneh, O., "Seismic performance of slopes in pseudo-static designs with different safety factors", ***Iranian Journal of Science Technology. Transactions of Civil Engineering***, Vol. 38, No. C2, (2014), 465, doi.
9. Al-Ameri, A., Jawad, F. and Fattah, M., "Vertical and lateral displacement response of foundation to earthquake loading", ***International Journal of Engineering***, Vol. 33, No. 10, (2020), 1864-1871, doi.
10. Barmenkova, E., "Design of base and foundation for the earthquake-resistant building", in IOP Conference Series: Materials Science and Engineering, IOP Publishing. Vol. 661, No. 1, (2019), 012093. https://doi.org/10.1088/1757-899X/661/1/012093
11. Majdi, A., "Earthquake hazard mitigation in iraq: Recommendations to decision makers", (2020).
12. Loukidis, D., Bandini, P. and Salgado, R., "Stability of seismically loaded slopes using limit analysis", ***Geotechnique***, Vol. 53, No. 5, (2003), 463-479.
13. Mahmood, M.S., "Effect of time-based soaking on shear strength parameters of sand soils", ***Applied Research Journal***, Vol. 3, No. 5, (2017), 142-149.
14. Mahmood, M.S., Al-Baghdadi, W.H., Rabee, A.M. and Almahbobi, S.H., "Reliability of shear-box tests upon soaking process on the sand soil in al-najaf city", in Key Engineering Materials, Trans Tech Publ. Vol. 857, (2020), 212-220. https://doi.org/10.4028/www.scientific.net/kem.857.212
15. Mahmood, M.S., Akhtarpour, A., Almahmodi, R. and Husain, M.M.A., "Settlement assessment of gypseous sand after time-based soaking", in IOP Conference Series: Materials Science and Engineering, IOP Publishing. Vol. 737, No. 1, (2020), 012080. https://doi.org/10.1088/1757-899X/737/1/012080
16. Mahmood, M.S. and Abrahim, M.J., "A review of collapsible soils behavior and prediction", in IOP Conference Series: Materials Science and Engineering, IOP Publishing. Vol. 1094, No. 1, (2021), 012044. DOI: 10.1088/1757-899X/1094/1/012044
17. Sh, M.M., Akhtapour, A. and AA, A.A., "Mechanical behavior of dam-vertically sand drained foundation, case study: Sombar dam", ***Journal of Engineering and Technological Sciences***, Vol. 51, No. 3, (2019), DOI: 10.5614/j.eng.technol.sci.2019.51.3.6
18. Akhtarpour, A., Mahmood, M.S. and Alali, A., "Evaluation of vertical drain with different materials", in IOP Conference Series: Materials Science and Engineering, IOP Publishing. Vol. 584, No. 1, (2019), 012011. DOI:10.1088/1757-899X/584/1/012011
19. Akhtarpour, A., Mahmood, M.S. and Alali, A., "Stability analysis of geosynthically piled foundation earth dam; a case study: Sombar dam", in IOP Conference Series: Materials Science and Engineering, IOP Publishing. Vol. 888, No. 1, (2020), 012003. DOI:10.1088/1757-899X/888/1/012003
20. Liu, S., Shao, L. and Li, H., "Slope stability analysis using the limit equilibrium method and two finite element methods", ***Computers Geotechnics***, Vol. 63, (2015), 291-298, doi.
21. Duncan, J.M., "State of the art: Limit equilibrium and finite-element analysis of slopes", ***Journal of Geotechnical Engineering***, Vol. 122, No. 7, (1996), 577-596, https://doi.org/10.1016/j.compgeo.2014.10.008
22. Al-Taie, A.J. and Albusoda, B.S., "Earthquake hazard on iraqi soil: Halabjah earthquake as a case study", ***Geodesy Geodynamics***, Vol. 10, No. 3, (2019), 196-204, https://doi.org/10.1016/j.geog.2019.03.004.
23. Al-Saoudi, N.K. and Al-Shakerchy, M.S.M., "Statistical analysis of some geotechnical properties of najaf city", in Proceedings of International Geotechnical Conference. Vol. 3, (2010), 1173-1180.
24. Alzabeebee, S. and Ammash, H., "Effect of presence of cavity on behavior of strip foundation rested on al-najaf soil as case study", in the Proceeding of the 2nd international conference on engineering, university of Mosul, Iraq. (2013), 1-11.
25. Devi, D.D.L. and Anbalagan, R., "Study on slope stability of earthen dams by using geostudio software", ***International Journal of Advanced Research, Ideas Innovations in Technology***, Vol. 3, No. 6, (2017), 408-414.
26. Das, B.M., "Principles of geotechnical engineering, Cengage learning, (2021).
27. Ti, K.S., Huat, B.B., Noorzaei, J., Jaafar, M.S. and Sew, G.S., "A review of basic soil constitutive models for geotechnical application", ***Electronic Journal of Geotechnical Engineering***, Vol. 14, (2009), 1-18.
28. Krahn, J., "Dynamic modeling with quake/w: An engineering methodology, GEO-SLOPE Calgary, (2004).
29. Seed, H.B., Idriss, I.M., Lee, K.L. and Makdisi, F.I., "Dynamic analysis of the slide in the lower san fernando dam during the earthquake of february 9, 1971", ***Journal of the Geotechnical Engineering division***, Vol. 101, No. 9, (1975), 889-911.
30. Kramer, S.L., "Geotechnical earthquake engineering, Pearson Education India, (1996).
31. Sazzad, M. and Haque, M., "Effect of surcharge on the stability of slope in a homogeneous soil by fem", in 2nd International Conference on Advances in Civil Engineering. (2014), 315-318.
32. Plummer, C.C., Carlson, D.H. and Hammersley, L., "Physical geology, New York, NY: McGraw-Hill/Education, Inc., (2016).
33. Keskin, M.S. and Laman, M., "Model studies of bearing capacity of strip footing on sand slope", ***Journal of Civil Engineering***, Vol. 17, No. 4, (2013), 699-711, https://doi.org/10.1007/s12205-013-0406-x.
34. Ling, H.I. and Cheng, A.H.-D., "Rock sliding induced by seismic force", ***International Journal of Rock Mechanics Mining Sciences***, Vol. 34, No. 6, (1997), 1021-1029, DOI: 10.1016/S0148-9062(97)00261-1.

Persian Abstract

چکیده

انتظار می رود در سالهای آینده النجف ، یکی از مهمترین شهرهای عراق رشد کند. بسیاری از ساختمانهای واقع در دامنه های النجف ، که نزدیک حرم امام علی است ، دارای ارزشهای اقتصادی و گردشگری مهمی هستند. در این مقاله ثبات شیب شهر تحت شرایط مختلفی از جمله هندسه شیب ، بارگذاری ساختار همسایه و بزرگی زمین لرزه بررسی می شود. در تجزیه و تحلیل یک روش اجزای محدود D۲ با کمک رایانه استفاده شده است. مجموعه ای از آزمایش های طبقه بندی و شناسایی خاک علاوه بر پارامترهای مورد نیاز خاک در مدل سازی سازه انجام می شود. مدل Mohr-Coulomb برای تعریف وضعیت خرابی که از شیب آغاز شده استفاده می شود. نتایج نشان می دهد که شیب به دلیل وزن و هندسه پایدار است و حداقل ضریب ایمنی آن ۲.۶ است. در حالی که تحت بارگذاری قرار می گیرید ، این ضریب ایمنی به کمتر از ۱.۸ کاهش می یابد. خطرناکترین شرایط زمانی است که دامنه تحت بارگذاری لرزه ای قرار گرفته باشد و ضریب ایمنی کمتر از یک واحد باشد ، برای همه موارد و خصوصیات بررسی شده ، جایی که دامنه ها فرو می ریزند.

IJE TRANSACTIONS A: Basics Vol. 34, No. 7, (July 2021) 1611-1622

International Journal of Engineering

Journal Homepage: www.ije.ir

Effect of Foundation Flexibility on the Seismic Performance of a High-rise Structure under Far-field and Near-field Earthquakes

R. Zomorodian, F. Soltani, A. Sivandi-Pour, E. Noroozinejad Farsangi*

Faculty of Civil and Surveying Engineering, Graduate University of Advanced Technology, Kerman, Iran

PAPER INFO

Paper history:
Received 23 March 2021
Received in revised form 15 May 2021
Accepted 17 May 2021

Keywords:
High-rise Structure
Foundation Flexibility
Storey Displacement
Drift Ratio
Layered Soils
Far-field
Near-field

ABSTRACT

In this study, the seismic performance of a 20-storey steel structure with a mat foundation located on layered soil is investigated under an array of strong ground excitations, which includes 6 far-fault and 6 near-fault earthquakes. Eight different modes for soil layering have been considered in the numerical simulation. FLAC 2D nonlinear platform has been used to model the near-realistic behavior. To this end, hundred lines of codes and subroutines have been developed in this platform to perform the analysis. The results of the analyzes include the absolute displacement of the floors, the ratio of the relative displacement of the floors, the shear force, the axial force, and the bending moment of the columns. It was concluded that for a 20-story structure on a mat foundation under both far-field and near-field earthquakes, the most reliable type of soil is the dense sandy soil and the most critical case is the soft clay soil. It was also observed that the near-field strong ground motions have imposed more critical structural responses compared to far-field records.

doi: *10.5829/ije.2021.34.07a.06*

1. INTRODUCTION

Due to the seismicity and the existence of active faults in many parts of the world, the study of the effect of foundation flexibility on the seismic behavior of buildings is of great importance, so extensive studies have been conducted in this field [1-4]. It has been observed that the performance levels of models with flexible foundations, especially in severe earthquakes, may change significantly compared to structures with rigid foundations. However, for moment-resisting frames on soft soils, flexible foundations have a significant effect on displacement and force responses. This indicates the need to pay more attention to the behavior of flexible foundations in modern design in order to achieve more economical and safer structures [5]. Choosing the right type of foundation can significantly affect the response of the structure and the foundation of the building [6]. Although seismic waves probably pass through tens of kilometers of rock and cross only less than 100 meters of soil layer to reach the earth's surface, the greatest impact, and change in the characteristics of earth movements occur within the soil layer. Some limited studies have been conducted on the effect of foundation flexibility and soil types on the seismic performance of structures [7, 8]. The results showed that all types of soil amplify bedrock movements in the soil-structure interface, but with different degrees that the amount of amplification is affected by many factors such as soil type and its characteristics, earthquake frequency content, and building characteristics [9]. Soil modulus has also a significant effect on the natural period of the system and the overall performance of structures [10]. Lateral displacement and internal forces of columns such as shear force, axial force, and moment in columns increase for all building frames when soil type changes from hard to medium and from medium to soft [11]. The dynamic response of the structure to earthquakes is significantly affected by the interaction of the structure and the foundation and the soil beneath the foundation.

*Corresponding Author Institutional Email: noroozinejad@kgut.ac.ir
(E. Noroozinejad Farsangi)

Please cite this article as: R. Zomorodian, F. Soltani, A. Sivandi-Pour, E. Noroozinejad Farsangi, Effect of Foundation Flexibility on the Seismic Performance of a High-rise Structure under Far-field and Near-field Earthquakes, International Journal of Engineering, Transactions A: Basics Vol. 34, No. 7, (2021) 1611-1622

Soil-structure interaction is important for heavy structures, especially for high-rise buildings located on soft soils [12]. Many studies have evaluated the effects of soil-foundation-structure interaction on the seismic response of buildings with different types of foundations on different soils [13, 14]. Interaction affects the amplitude of the displacement of the foundation and depends on the frequency of seismic waves [15]. The results show that the fundamental natural period and base shear of a structure considering the interaction is greater than the non-interaction state, and with increasing the soil shear modulus and the number of stories, the fundamental natural period, the base shear, the maximum lateral displacement, drift ratio and moments in structure would increase [16-18]. Hence, upper stories displacements are more affected with soil-foundation-structure interaction (SFSI) than the lower stories [19]. Therefore, it is essential to consider the effects of SFSI in the seismic design of building frames, especially when located on soft soil [20].

According to the studies described above, various software such as ABAQUS, PLAXIS, SAP2000, ETABS, etc. has been used for modeling such the phenomenon. In this study, FLAC nonlinear platform was used for the simulation. A variety of behavioral models are defined in this program that allows researchers to model and analyze various problems under nonlinear analysis. One of the capabilities of this software is the ability to dynamically analyze geotechnical problems to model harmonic loads, earthquake loads, and explosions [21]. According to the subject of study and as this platform is more reliable especially in damping modeling, this software was used for numerical analysis. In this research, extensive investigations have been performed to study the seismic performance of a high-rise building with a mat foundation on different types of soil layers under various earthquake ground motions. Different engineering demand parameters (EDP) such as floor displacement, drift ratio, and internal forces of the columns have been investigated.

2. MODELING PROCEDURE

The nonlinear environment of FLAC software, which is a finite difference program, was used to model the near-realistic behavior. First, a soil layer was modeled according to the Mohr-Coulomb nonlinear constitutive model using this platform [22]. The Mohr-Coulomb model is one of the most common models that well introduces shear failure in soil and rock. This model has been selected due to its simplicity and reputation compared to other available complicated models used to represent shear failure in soils and rocks. Many of the current research works in the field of SSI are using this model to simulate the failure phenomena in soil materials. Hence, the simulated models would be quite reliable. The general form of the Mohr-Coulomb failure criterion is shown in Figure 1. The failure path from point A to point B is introduced using the Mohr-Coulomb yield function.

$$f^s = \sigma_1 - \sigma_3 N_\varphi + 2c\sqrt{N_\varphi} \quad (1)$$

The tension yield function also occurs in path B to C using the following function.

$$f^t = \sigma^t - \sigma_3 \quad (2)$$

In these relations φ is the angle of internal friction, c is the cohesion, σt is the tensile strength and N_φ is:

$$N_\varphi = \frac{1+sin\varphi}{1-sin\varphi} \quad (3)$$

According to Kuhlemeyer and Lysmer [23] to accurately represent the wave transmission through a model, the spatial element size, Δl, should be smaller than approximately one-tenth to one-eighth of the input wavelength associated with the highest frequency of the input wave.

$$\Delta l \leq \frac{\lambda}{10} \quad (4)$$

To have a computationally efficient model, mesh sensitivity analysis has been conducted to find the optimal mesh size based on the above formulation removing frequency values greater than 10 Hz.

After modeling the soil layers, the foundation and then the structure were modeled. Due to the fact that in this study, a high-rise structure is used and the superstructure loads are relatively large, and also different modes of soil layering are used, in some cases several types of soil are used and the subgrade is variable, and in some cases the subgrade is weak, so in order to minimize non-uniform settlements and deal with local complications and case weaknesses in the subgrade, mat foundation has been used for all cases. Different methods are common for the mat foundation analysis, the two main groups of which are rigid and nonrigid. In this

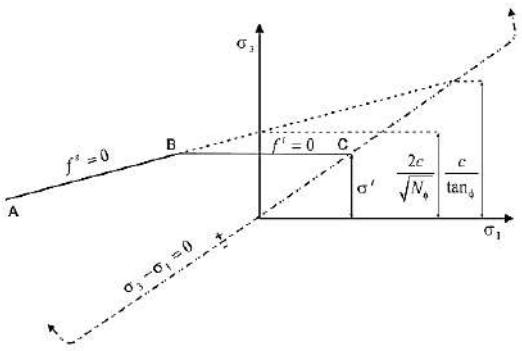


Figure 1. Mohar-Coulomb failure criterion [21]

study, the rigid solution is used for the mat foundation, which is known as the conventional method of static equilibrium. In this method, it is assumed that the foundation is much stiffer and harder than the subgrade [24]. After completing the modeling steps, earthquake acceleration records were applied to the lowest soil layer. All the mentioned steps including modeling, dynamic analysis, etc. have been done by developing the codes and sub-routines in FLAC 2D. As the investigated models are symmetrical both in plan and elevation, 2D models provide satisfactory results and are reliable. On the other hand, 2D models have fewer degrees of freedom (DOF) compared to 3D models and are more computationally efficient. It should be noted that in this study, nonlinear properties of the soils and steel materials have been used and the stress-strain curves of steel and concrete are shown in Figures 2 and 3, respectively.

2. 1. Modeling Approach Validation To validate the numerical modeling approach used in the present study, a 4-storey building on a strip foundation on a soil layer was modeled and its seismic performance was investigated. FLAC 2D nonlinear environment was used to model the near-reality behavior. The results of modeling were compared with the simplified nonlinear model of Tahghighi and Rabiee's research results [5]. The modeling steps are as follows: first, the soil layers were modeled according to Mohr-Coulomb behavioral model

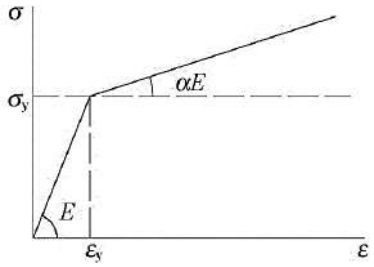


Figure 2. Bilinear stress-strain curve used for steel material [5]

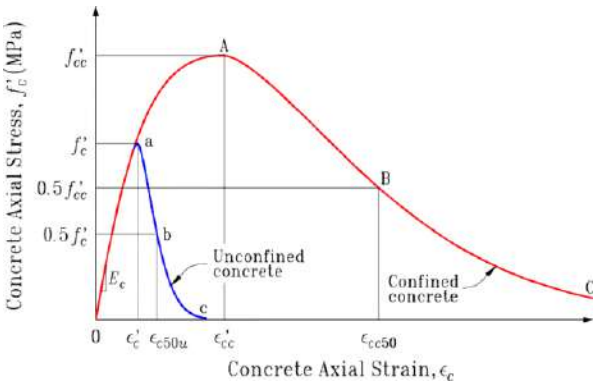


Figure 3. Curvilinear stress-strain relationship for confined concrete material [25]

using FLAC software, and then the strip foundation and the main structure were modeled. Finally, the obtained model is shown in Figure 4. After completing the modeling steps, the selected earthquakes were applied to the lowest layer of soil, which is the bedrock. All geotechnical, structural, and earthquake characteristics used are given in Tables 1-4.

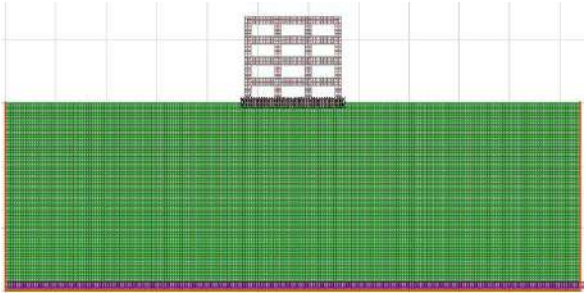


Figure 4. Building, soil layer and foundation modeled in FLAC software

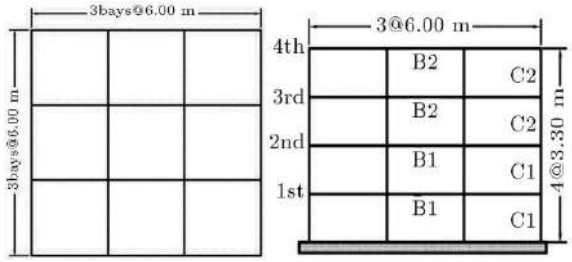


Figure 5. Plan and Elevation of the studied structure for verification [5]

TABLE 1. Sections of beams and columns [5]

Section tag	Web		Flange	
	Depth (cm)	Thickness (cm)	Width (cm)	Thickness (cm)
C1	31.96	2.11	39.88	3.33
C2	32.10	1.64	37.34	2.62
B1	27.84	1.09	30.48	1.70
B2	27.65	0.99	30.48	1.54

TABLE 2. Details of soil parameters used [5]

Φ (degree)	C (kg/cm^2)	υ	G(kgf/cm^2)	Vs(m/s)
30	0.15	0.35	6707	560

TABLE 3. Specifications of the foundation used [5]

Footing type	B (m)	L (m)	H (m)	q_{all} (kgf/cm^2)	K_s (kgf/cm^3)
strip	1	19	0.65	2	2.40

TABLE 4. Specifications of used earthquakes for verification [5]

No.	Earthquake	Year	Station	Mw	D (km)	PGA (g)	PGV (cm/s)	PGD (cm)
1	Northridge, USA	1994	Old Ridge Route	6.7	22.6	0.57	52.1	4.2
2	Cape Mendocino, USA	1992	Rio Dell Overpass	7.1	18.5	0.55	42.1	18.6
3	Chi-Chi, Taiwan	1999	TCU045	7.6	24.0	0.51	39.0	14.3
4	Loma Prieta, USA	1989	Gilroy Gavilan Coll.	6.9	12.0	0.36	28.6	6.3
5	San Fernando, USA	1971	Lake Hughes	6.6	20.3	0.37	17.0	1.6
6	Victoria, Mexico	1980	Cerro Prieto	6.1	17.0	0.62	31.6	13.2
7	Whittier Narrows, USA	1987	LA-116th St School	6.0	22.5	0.39	21.0	1.8

After completing the dynamic analysis, the time history diagrams of the displacement of the floors in FLAC software were examined and were compared with the results of Tahghighi and Rabiee's research for DBE and MCE hazard levels in Figure 6. As can be seen, this comparison indicates that the results are almost identical to each other and the modeling approach in the current study is verified. The model proposed in the current study is more comprehensive compared to Tahghighi and Rabiee's research [5], as the soil-structure-interaction (SSI) have been implemented with more details and near the realistic behavior. The purpose has not been to develop a model for practical engineering application as it has been a scientific investigation considering as much complexity and uncertainties in models as possible.

3. STRUCTURAL AND GEOTECHNICAL SPECIFICATIONS OF THE MODEL

In this study, a 20-storey steel moment-resisting frame building located on 8 soil layering modes and a mat foundation has been investigated. This building model has a floor plan of 30.48× 36.58 m^2 and its total height is 80.77 m. The bays are 6.1 m, as shown in Figure 7. Floors are made of composite slabs and the building's lateral load-resisting system consists of steel perimeter moment-

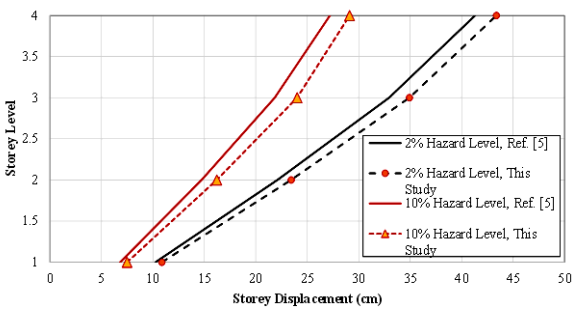


Figure 6. Model validation at two hazard levels [5]

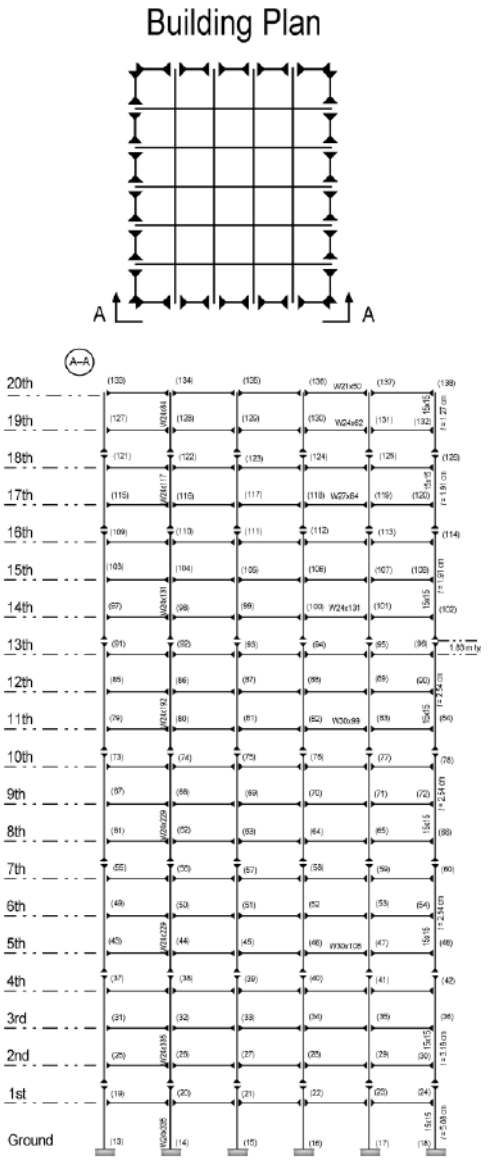


Figure 7. Specifications of the investigated 20-storey structure [22]

resisting frames (MRFs). The type of steel selected for the beams and columns of the structure is of St-37 type. The middle columns of the MRF are wide flange but the corner columns are BOX type. The typical heights from floor to floor are 3.96 m and only the height of the first floor is 5.49 m. The building has been selected from pre-designed SAC project models [22]. The amount of live load is considered as 200 kg/m^2, while the dead load is 500 kg/m^2, and the live load reduction coefficient is considered as 20%. The characteristics of the investigated structure and it's plan are shown in Figure 1.

The complete characteristics of the soils used in this study are presented in Tables 5 and 6. The soils are layered and named in 8 different categories.

TABLE 5. Specifications of soil layers used in this study [3,4,24]

Soil type	G (Pa)	K (Pa)	E (Pa)	ν	C (Pa)	Φ (deg)	γ (kg/m^3)
Soft clay	8.40×10^6	5.70×10^7	2.4×10^7	0.43	5×10^4	20	1575
Medium stiff clay	15.84×10^6	9.37×10^7	4.5×10^7	0.42	9×10^4	24	1595
Stiff clay	19.23×10^6	4.16×10^7	5.0×10^7	0.30	12×10^4	26	1500
Loose sand	23.08×10^6	5.00×10^7	6.0×10^7	0.30	0	28	1500
Medium density sand	35.18×10^6	10.55×10^7	9.5×10^7	0.35	0	33	1900
Dense sand	57.90×10^6	$24.39\times\times10^7$	16.1×10^7	0.39	0	38	2000

TABLE 6. Soil categories used in this study

Type / Layer Number	Soil 1	Soil 2	Soil 3	Soil 4	Soil 5	Soil 6	Soil 7	Soil 8
1st Layer	Loose sand	Medium density sand	Dense sand	Loose sand	Soft clay	Medium stiff clay	stiff clay	Soft clay
2nd Layer	Loose sand	Medium density sand	Dense sand	Medium density sand	Soft clay	Medium stiff clay	stiff clay	Medium stiff clay
3rd Layer	Loose sand	Medium density sand	Dense sand	Dense sand	Soft clay	Medium stiff clay	stiff clay	stiff clay

The concrete used in the design of the foundation has a modulus of elasticity E_c = 2.2×1010 (Pa) compressive strength f_c = 2.1×107 (Pa), Poisson ratio υ = 0.2. The foundation is designed as a Mat foundation with the moment of inertia (I) of 1.302 m^4 and the cross-section area (A) of 2.5 m^2 for a unit width. The schematic presentation of the developed model in FALC is depicted in Figure 8.

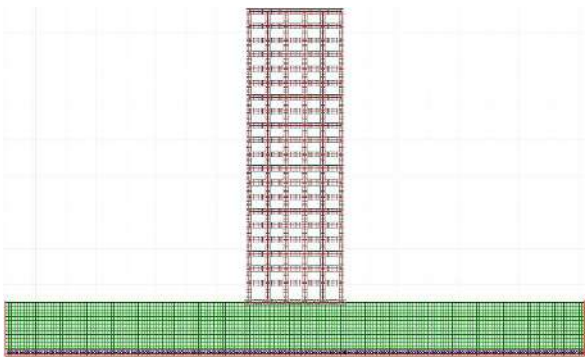


Figure 8. Building, soil layer and foundation modeled in FLAC nonlinear platform

4. STRONG GROUND MOTIONS USED

Twelve earthquake records were applied to the models, 6 of which are far-fault earthquakes and the other 6 are near-fault earthquakes, and all are taken from the FAMA P695 database [26]. To make sure that the residual velocity or displacement will not affect the final outcome, baseline filtering has been applied to them [21]. To match the design response spectra, ASCE 7 procedure has been used for the scaling procedure. A 5% Rayleigh damping is considered as the inherent damping of the investigated structure. To get the most accurate results, the excitations have been applied to the bottom of the lowest layer, representing the bedrock. Earthquake records' characteristics are given in Tables 7 and 8.

5. RESULTS AND DISCUSSION

In this part, the seismic performance of a SAC project high-rise steel structure on layered soils is investigated and the results are discussed. Nonlinear time history analyzes were performed using FLAC 2D software and

the results of various EDPs were obtained, which are compared and discussed below.

5. 1. Time History Analysis Nonlinear dynamic time history analyzes have been performed using strong ground motion records proposed in Tables 7 and 8. A total of 96 time-history analyzes were performed to study the seismic performance of this structure, as shown in Figures 9-17. Floor displacement, drift ratio, maximum axial force, shear, and moment in columns have been selected as the most important response parameters to estimate the seismic behavior of the structure in this study.

5. 1. 1. Absolute Displacement The maximum absolute displacements of all structural floors on eight soil types were obtained under all earthquake records. Finally, the average displacement of each floor in all soil types was calculated. Then, the diagrams of the average floor displacement for the investigated structure were obtained, which are shown in Figures 9 and 10. The horizontal axis shows the average displacement of the floors and the vertical axis shows the level of floors. As shown in Figures 9 and 10, the displacement of floors increases with increasing floor height, but the rate of increase in displacement in the lower floors is more significant than in the upper floors. Figure 9 shows the average displacement for all soil types under far-fault and near-fault earthquakes. As can be seen, the displacement of floors in the case of clay soils is more than the sandy soils, and the highest displacement is related to type 5 soil (soft clay), while the lowest displacement is related to type 3 soil (dense sand). As shown in Figure 9(a), the displacement range of the floors for the 20-story structure under far-field earthquakes is 0.025-0.371 m, whilst in Figure 9(b) the displacement range for near-field earthquakes is 0.045-0.706 m, which is much greater than that of the far-field case.

Figure 10 shows the average floor displacement for all earthquakes in two categories. In Figure 10, the average floor displacement range for all far-fault earthquakes has been 0.023-0.482 m; while the range for the near-field earthquakes has been 0.025-0.944m. According to the mentioned ranges, the average floor displacement under near-fault records is significantly higher than the far-fault ones. Imperial Valley – 06 El Centro and San Fernando earthquakes in the far-fault category, and Imperial Valley - 06 El Centro and Superstition Hills - 02 earthquakes in the near-fault category have caused the highest storey displacements.

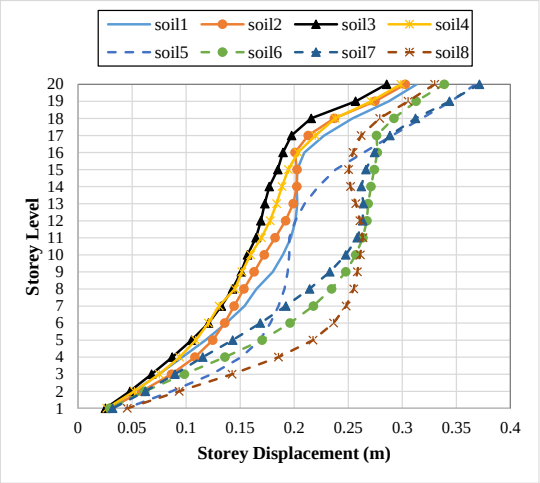
Figure 11 shows the average displacement of floors for all types of soils under different earthquake categories. In Figure 11, the range of floor displacement for all soil types in all far-fault earthquakes has been 0.032-0.325 m, while the range for the near-fault earthquakes has been 0.052-0.563 m.

TABLE 7. Characteristics of selected far-field earthquakes [25]

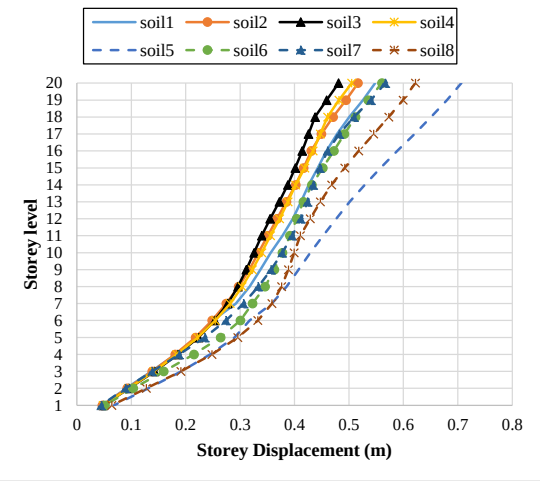
Record Number	Earthquake Name	Station Name	Magnitude (MW)	PGAmax (g)	PGVmax (cm/s)
1	San Fernando	LA-Hollywood Store FF	6.6	0.21	**19**
2	Imperial Valley-06	Delta	6.5	0.35	**33**
3	Imperial Valley - 06	El Centro Arrey #11	6.5	0.38	**42**
4	Northridge - 01	Beverly Hills-Mulhol	6.7	0.52	**63**
5	Loma Prieta	Capitola	6.9	0.53	**35**
7	Landers	Coolwater	7.3	0.42	**42**

TABLE 8. Characteristics of selected near-field earthquakes [25]

Record Number	Earthquake Name	Station Name	Magnitude (MW)	PGAmax (g)	PGVmax (cm/s)
26	Landers	Lucerne	7.3	0.79	**140.3**
24	Cape Mendocino	Petrolia	7.0	0.63	**82.1**
32	Northridge- 01	Rinaldi Receiving Sta	6.7	0.87	**167.3**
8	Imperial Valley - 06	El Centro Arrey 06	6.5	0.44	**111.9**
14	Superstition Hills - 02	Parachute Test Site	6.5	0.42	**106.8**
20	Loma Prieta	Saratoga – Alloha Ave	6.9	0.38	**55.6**

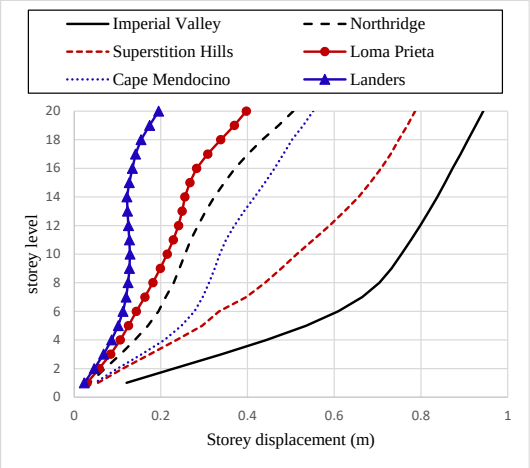


(a)



(b)

Figure 9. Average displacement of the building under (a) far-fault, and (b) near-fault



(b)

Figure 10. Average floor displacement for earthquakes: (a) far-fault, and (b) near-fault

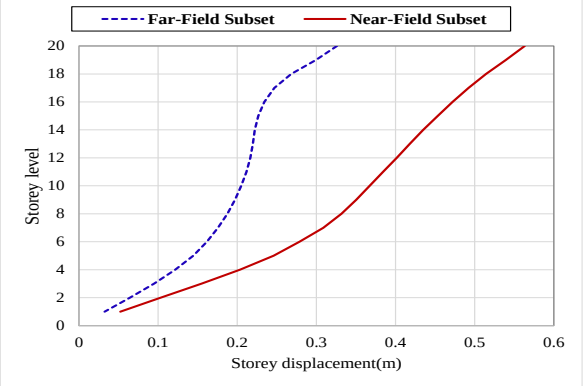
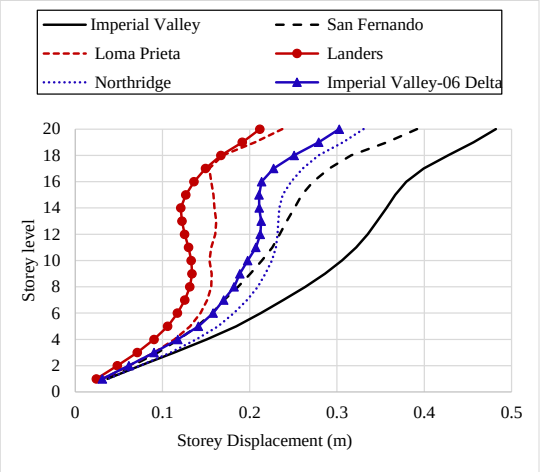


Figure 11. Average floor displacement in all earthquakes and all soils



(a)

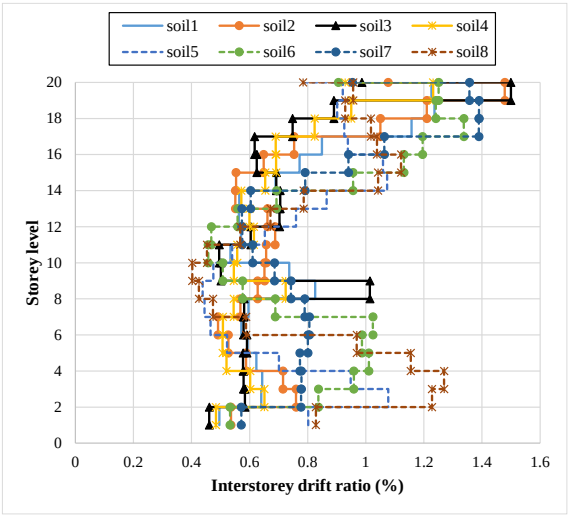
5. 1. 2. Drift Ratio Drift Ratio is the most commonly used EDP to determine damage level in a structure. Figure 12 shows the average drift ratio of structural floors for all types of soils under near-fault and far-fault earthquakes. As shown in this figure, the drift ratios are higher in the first and last floors of the structure, and in these areas, the drift ratio has an increasing trend. The maximum drift ratio of floors for all soils under far-fault earthquakes according to Figure 12(a) is related to dense sand soil and medium-density sand, which are 1.5% and 1.48%, respectively. According to Figure 12(b), the maximum drift ratio for all soils under near-fault earthquakes is related to soft clay with a value of 1.83%. Therefore, the maximum average ratio of structural floors of near-fault earthquakes is higher than the earthquakes far from faults.

Figure 13 shows the graphs of the average floor drift ratio for all earthquakes in areas far from the fault and

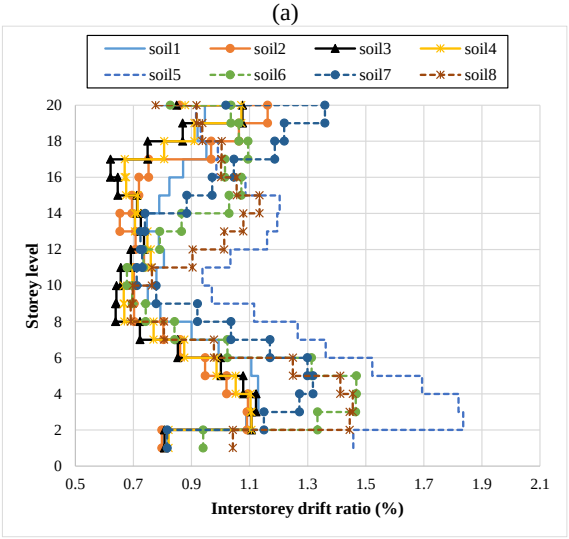
near the fault. In Figure 13(a), the San Fernando earthquake has caused the highest floor drift ratio of 1.6%. In Figure 13(b), the Imperial Valley-06 El Centro Earthquake has induced a drift ratio of 2.8, which has been the highest drift ratio.

Figure 14 shows the average drift ratio of structural floors. As can be seen in this figure, the maximum average drift ratio has been 1.24% under earthquakes far from faults and the highest drift ratio has been about 1.31% in near-fault earthquakes. According to the diagram, it can be seen that the drift ratio is higher in the first and last floors.

5. 1. 3. Internal Forces As shown in the diagrams above, the highest displacement and drift ratio is related to soft clay in the Imperial Valley-06 El Centro earthquake in the near-fault areas, which was determined as the most critical case. Thus, soft and stiff clay were selected under the Imperial Valley-06 El Centro earthquake in the area near the fault to make a comparison between the results of the internal forces of the columns in terms of soil stiffness. The internal forces of the columns include the shear force, bending moment,

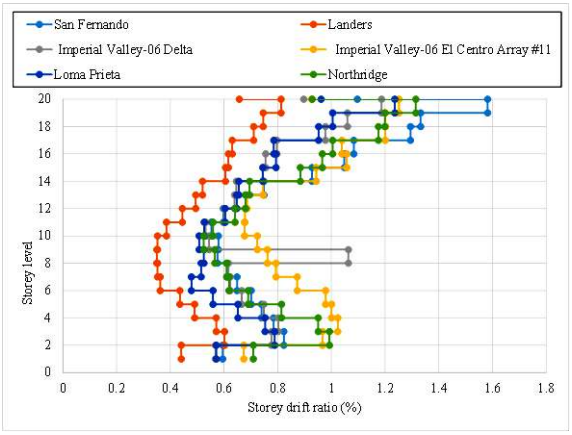


(a)

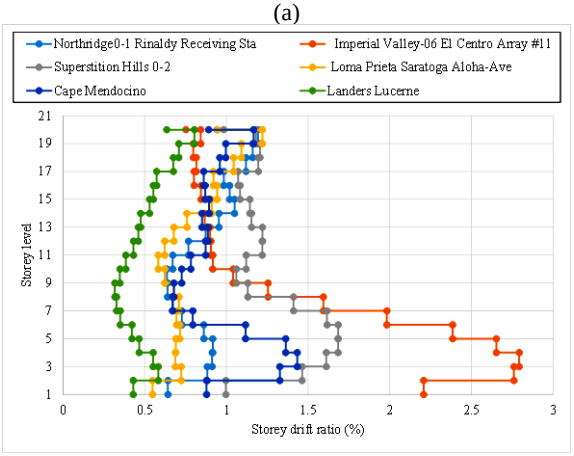


(b)

Figure 12. Average ratio of relative displacement of structural floors: (a) far-fault, and (b) near-fault



(a)



(b)

Figure 13. The average ratio of relative displacement of structural floors (a) far-fault, and (b) near-fault

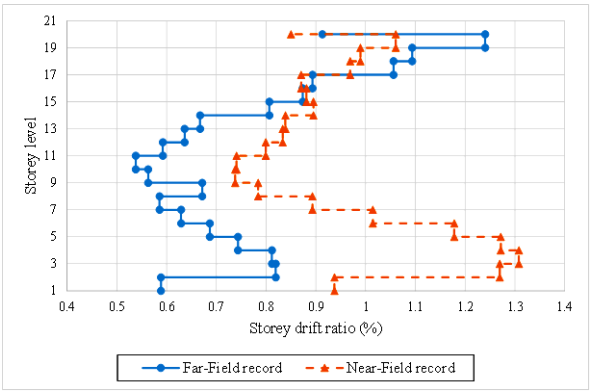


Figure 14. Average ratio of relative displacement of structural floors in all earthquakes and all soils

and axial force of the middle and side columns. Figure 15 shows the shear force of the middle and lateral columns of the structure vs. the structural floors for stiff and soft clay soils under Imperial Valley-06 El Centro earthquake. As can be seen, with increasing the number of floors and the height of the structure, the values of shear forces decrease, so the diagrams have a downward trend. According to the diagram, the shear force of the middle columns is more than the shear force of the side columns. Also, the amount of shear forces in soft clay is more than stiff clay.

Figure 16 shows the bending moment values of the middle and lateral columns of the structure vs. the structural floors for soft and stiff clay soils in Imperial Valley-06 El Centro earthquake. As can be seen in the figure, the bending moment values decrease with the increasing number of floors and have a decreasing trend. The amount of bending moment for the middle columns is more than the side columns. Also, the shear forces for soft clay are more than for stiff clay.

Figure 17 shows the axial force of the middle and lateral columns of the structure versus the number of structural floors for soft and stiff clay soils in Imperial Valley-06 El Centro earthquake. According to the figure below, it can be seen that the axial force in the side columns is more than the middle columns and the values for soft clay are more than those for stiff clay, while in the middle columns the axial force values for these columns on soft clay are less than the stiff clay soil. Also, according to this diagram, it can be seen that the axial force of the columns decreases with the increasing number of floors and has a downward trend.

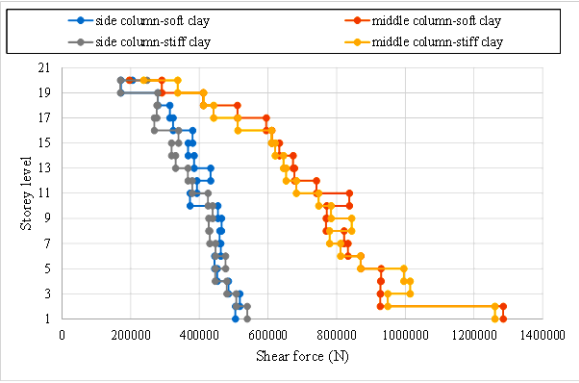


Figure 15. Shear force of side and middle columns of the structure in the Imperial Valley-06 El Centro earthquake in the area near the fault for soft clay and stiff clay

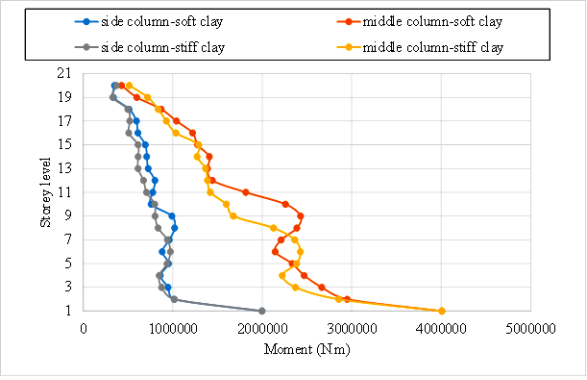


Figure 16. Bending moment of side and middle columns of the structure in the Imperial Valley-06 El Centro earthquake in the area near the fault for soft clay and stiff clay

5. 1.6. Maximum Principal Stress The following relation is known as Mohr-Coulomb relation and the criterion of Mohr-Coulomb rupture is based on literature [27]. The circle of Mohr and rupture envelope is shown in Figure 18.

$$\tau_f = c + \sigma_f \tan\varphi \quad (5)$$

In the above formulation, σ_f depends on the load and is a variable of shear strength. The value of σ_1 is increased by vertical loading until the Mohr circle touches the rupture line and the soil element is ruptured. In this case, σ_1, for which the Mohr circle is tangent to the rupture line, is the minimum principal maximum stress that

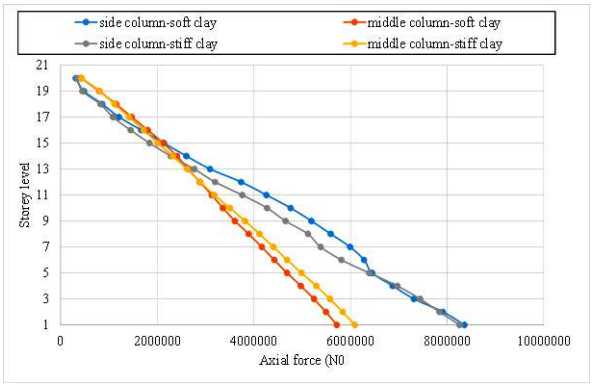


Figure 17. Axial force of side and middle columns of the structure in the Imperial Valley-06 El Centro earthquake in the area near the fault for soft clay and stiff clay

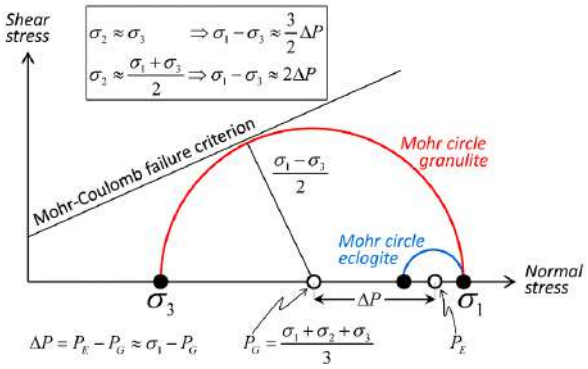


Figure 18. Mohr circle and rupture envelope [28]

causes a rupture in the θ direction in the soil element. The value of the rupture plane angle with the principal stress plane is given in Equation (6) [27].

$$\theta = 45 + \frac{\varphi}{2} \quad (6)$$

Equation (7) shows the relationship between the principal stresses at the moment of rupture, which is called the rupture relationship of the principal stresses [27]. Due to the importance of the principal maximum stress value of the σ_1 stress contour in the soil layers for more critical soils and earthquakes, it is shown in the following figures. Figure 19 shows the maximum principal stress in soft clay layers. Figure 20 shows the maximum principal stress for stiff clay layers. Figures 19 and 20 are related to the Imperial earthquake in the area near the fault.

$$\sigma_1 = \sigma_3 \times tan^2\left(45 + \frac{\varphi}{2}\right) + 2c \times \tan\left(45 + \frac{\varphi}{2}\right) \quad (7)$$

Figure 21 shows the maximum principal stress of the soft clay layer in the Imperial Valley-06 El Centro earthquake far from the fault. For the sake of brevity, only these figures are presented for the case of stress conditions in the soil types, while the rest has shown a similar trend.

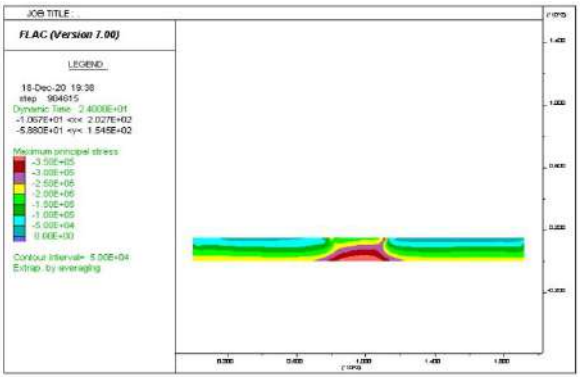


Figure 19. maximum principal stress in soft clay layers in Imperial Valley-06 El Centro earthquake

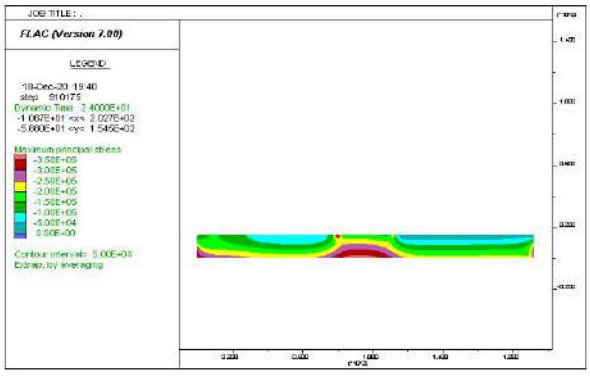


Figure 20. maximum principal stress in stiff clay layers in Imperial Valley-06 El Centro earthquake

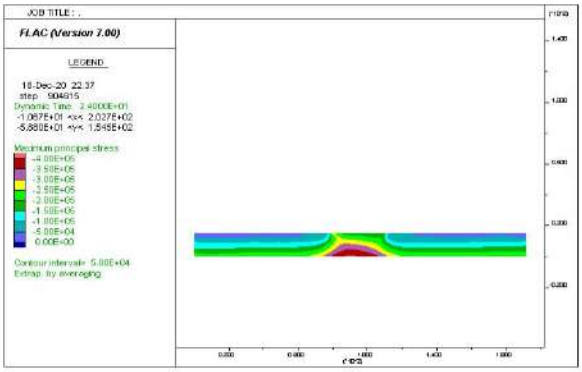


Figure 21. maximum principal stress in soft clay layers in Imperial Valley-06 El Centro earthquake

6. CONCLUSION

In this study, the seismic performance of a 20-story steel structure with a mat foundation on layered soils has been investigated. Nonlinear dynamics time-history analyzes were performed using strong ground motion records in FLAC platform. The results of the analysis include absolute floor displacement, floor drift ratio, the shear force, axial force, and bending moment of columns, which are described below:

The displacement of layers in clay soils is more than sandy soils so that the highest displacement is related to soft clay and also the least displacement is related to dense sand, so the most suitable soil for a 20-storey structure on a mat foundation under far- and near-fault earthquakes is dense sand and the most critical soil type would be the soft clay.

The results indicated that the maximum average drift ratio has been 1.24% under earthquakes far from faults and the highest drift ratio has been about 1.31% in near-fault earthquakes. It was also observed that the drift ratio is higher in the first and last floors.

The values of shear forces and bending moment in the middle columns are more than the side columns of the structure and the amount of shear forces and bending moment in soft clay is more than stiff clay, so the more flexible the soil, the higher the shear force and bending moment of columns. Internal forces decrease with increasing number of structural floors and have a downward trend.

7. RECOMMENDATIONS

The above results can be useful to help design and practicing engineers to consider the effects of soil and earthquake type on the seismic design of buildings. However, this study still needs to be extended for other types of structures with more floors with other structural systems. Other structural behavioral indicators, behavior

coefficient, ductility coefficient, and resistance reduction coefficient can be investigated in future studies. It is also suggested to use incremental and pushover dynamic analysis for future studies. In this study, the behavioral model of Mohr-Coulomb for soil has been used. It is suggested that other behavioral models be investigated as well. In this study, mat foundation is considered as rigid, which can be further investigated as semi-rigid or non-rigid in future studies.

8. REFERENCES

1. Jin, L. and Liang, J., "The effect of foundation flexibility variation on system response of dynamic soil–structure interaction: An analytical solution", ***Bulletin of Earthquake Engineering***, Vol. 16, No. 1, (2018), 113-127, https://doi.org/10.1007/s10518-017-0212-9.
2. Ponbunyanon, P., Limkatanyu, S., Prachasaree, W. and Damrongwiriyanupap, N., "Seismic assessments of 3-storey rc frame buildings including effects of pile-foundation flexibility".
3. Wulandari, P.S. and Tjandra, D., "Analysis of piled raft foundation on soft soil using plaxis 2d", ***Procedia Engineering***, Vol. 125, (2015), 363-367, https://doi.org/10.1016/j.proeng.2015.11.083.
4. Johnson, R.T., Varghese, R.M. and Joseph, J., Parametric study on the behavior of combined pile raft foundation founded on multi-layered soil using plaxis 3d, in Soil dynamics and earthquake geotechnical engineering. 2019, Springer.217-225. https://doi.org/10.1007/978-981-13-0562-7_24
5. Tahghighi, H. and Rabiee, M., "Influence of foundation flexibility on the seismic response of low-to-mid-rise moment-resisting frame buildings", ***Scientia Iranica***, Vol. 24, No. 3, (2017), 979-992, https://doi.org/10.24200/SCI.2017.4081
6. Qi, S. and Knappett, J., "Influence of foundation type on seismic response of low-rise structures in liquefiable soil", ***Soil Dynamics Earthquake Engineering***, Vol. 128, (2020), 105786, https://doi.org/10.1016/j.soildyn.2019.105786.
7. Abd-Elhamed, A. and Mahmoud, S., "Seismic response evaluation of structures on improved liquefiable soil", ***European Journal of Environmental Civil Engineering***, Vol., No., (2019), 1-23, https://doi.org/10.1080/19648189.2019.1595738.
8. Olarte, J., Dashti, S. and Liel, A.B., "Can ground densification improve seismic performance of the soil-foundation-structure system on liquefiable soils?", ***Earthquake Engineering Structural Dynamics***, Vol. 47, No. 5, (2018), 1193-1211, https://doi.org/10.1002/eqe.3012.
9. Matinmanesh, H. and Asheghabadi, M.S., "Seismic analysis on soil-structure interaction of buildings over sandy soil", ***Procedia Engineering***, Vol. 14, (2011), 1737-1743, https://doi.org/10.1016/j.proeng.2011.07.218.
10. Ismail, A., "Effect of soil flexibility on seismic performance of 3-d frames", ***Journal of Mechanical Civil Engineering***, Vol. 11, No. 4, (2014), 135-143, https://doi.org/10.9790/1684-1142135143.
11. Anand, N., Mightraj, C. and Prince Arulraj, G., "Seismic behaviour of rcc shear wall under different soil conditions", in Indian geotechnical conference. (2010), 119-120.
12. Karthika, A.P. and Gayathri, V., "Literature review on effect of soil structure interaction on dynamic behaviour of buildings", ***International Research Journal of Engineering Technology***, Vol. 5, No. 04, (2018), 2522-2525, doi.
13. Pitilakis, D. and Karatzetzou, A., "Performance-based design of soil–foundation–structure systems", in Proceedings of the 15th world conference on earthquake engineering, Lisbon, Portugal. (2012).
14. Hosseinzadeh, N., Davoodi, M. and Roknabadi, E.R., "Shake table study of soil structure interaction effects in surface and embedded foundations", in 15th World Conference on Earthquake Engineering (15WCEE). (2012).
15. Elwi, M., Muhammed, B. and Alhussiny, N., "Evaluation of soil-structure interaction for structures subjected to earthquake loading with different types of foundation", in MATEC Web of Conferences, EDP Sciences. Vol. 162, (2018), 04026. https://doi.org/10.1051/matecconf/201816204026
16. Priyanka, R.J., Anand, N. and Justin, D.S., "Studies on soil structure interaction of multi storeyed buildings with rigid and flexible foundation", ***International Journal of Emerging Technology Advanced Engineering***, Vol. 2, No. 12, (2012), 111-118, doi.
17. Kuladeepu, M., Narayana, G. and Narendra, B., "Soil structure interaction effect on dynamic behavior of 3d building frames with raft footing", ***International Journal of Research in Engineering and Technology***, (2015).
18. Zhu, G. and Lee, V.W., "Three-dimensional (3d) soil structure interaction with normal-plane p-wave incidence: Rigid foundation", ***Soil Dynamics Earthquake Engineering Structural Dynamics***, Vol. 105, (2018), 11-21, https://doi.org/10.1016/j.soildyn.2017.11.016.
19. Khazaei, J., Amiri, A. and Khalilpour, M., "Seismic evaluation of soil-foundation-structure interaction: Direct and cone model", ***Earthquakes Structures***, Vol. 12, No. 2, (2017), 251-262, http://dx.doi.org/10.12989/eas.2017.12.2.251.
20. Raheem, S.E.A., Ahmed, M.M. and Alazrak, T.M., "Evaluation of soil–foundation–structure interaction effects on seismic response demands of multi-story mrf buildings on raft foundations", ***International Journal of Advanced Structural Engineering***, Vol. 7, No. 1, (2015), 11-30, https://doi.org/10.1007/s40091-014-0078-x.
21. Itasca, F., "Fast lagrangian analysis of continua", ***Itasca Consulting Group Inc., Minneapolis, Minn***, (2000).
22. Ohtori, Y., Christenson, R., Spencer Jr, B. and Dyke, S., "Benchmark control problems for seismically excited nonlinear buildings", ***Journal of engineering mechanics***, Vol. 130, No. 4, (2004), 366-385, https://doi.org/10.1061/(ASCE)0733-9399(2004)130:4(366).
23. Kuhlemeyer, R.L. and Lysmer, J., "Finite element method accuracy for wave propagation problems", ***Journal of the Soil Mechanics Foundations Division***, Vol. 99, No. 5, (1973), 421-427, https://doi.org/10.1061/JSFEAQ.0001885.
24. Eslami, A., "Foundation engineering design and construction", Building Housing Research Center, BHRC. No. B-437, Tehran, Iran, (2006),.
25. ROY, N., PAULTRE, P. and PROULX, J., "A bridge ductility study for seismic assessment and rehabilitation", in 13th World Conference on Earthquake Engineering, Vancouver, Canada. (2004).
26. Council, A.T. and Agency, U.S.F.E.M., "Quantification of building seismic performance factors, US Department of Homeland Security, FEMA, (2009).
27. Das, B.M. and Sivakugan, N., "Fundamentals of geotechnical engineering, Cengage Learning, (2016).
28. Jamtveit, B., Moulas, E., Andersen, T.B., Austrheim, H., Corfu, F., Petley-Ragan, A. and Schmalholz, S.M., "High pressure metamorphism caused by fluid induced weakening of deep continental crust", ***Scientific reports***, Vol. 8, No. 1, (2018), 1-8, https://doi.org/10.1038/s41598-018-35200-1

Persian Abstract

چکیده

در این پژوهش، عملکرد لرزه ای سک ساختمان ۲۰ طبقه فلزی با فونداسیون گسترده روی خاک لایه ای تحت گروهی از شتاب نگاشت های حرکت شدید زمین شامل ۶ رکورد حوزه دور و ۶ رکورد حوزه نزدیک مورد مطالعه قرار گرفته است. ۸ لایه بندی مختلف خاک برای شبیه سازی عددی در نظر گرفته شده است. نرم افزار غیرخطی FALC 2D به منظور مدلسازی نزدیک به واقعیت مد نظر قرار گرفته است. برای این منظور صدها خط کد نویسی جهت انجام تحلیل ها صورت گرفته است. نتایج مستخرج شامل جابجایی مطلق کف ها، جابجایی نسبی طبقات و تلاش های داخلی در ستونها بوده است. نتیجه گیری شد برای یک ساختمان ۲۰ طبقه برروی فونداسیون گسترده تحت زلزله های حوزه دور و نزدیک، خاک ماسه ای متراکم و خاک رسی نرم به ترتیب بیشترین و کمترین قابلیت اعتماد را برای سازه ایجاد می نمایند. همچنین مشاهده گردید اثرات زلزله های حوزه نزدیک به مراتب بحرانی تر از زلزله های حوزه دور بوده است.

IJE TRANSACTIONS A: Basics Vol. 34, No. 7, (July 2021) 1623-1631

International Journal of Engineering

Journal Homepage: www.ije.ir

Optimization of Calcined Bentonite Caly Utilization in Cement Mortar using Response Surface Methodology

S. Sahith Reddy, M. Achyutha Kumar Reddy*

Department of Civil Engineering, Koneru Lakshmaiah Education Foundation, Vaddeswaram, Guntur, India

PAPER INFO

Paper history:
Received 07 April 2021
Received in revised form 27 April 2021
Accepted 30 April 2021

Keywords:
Calcined Bentonite
Workability
Compressive Strength
Response Surface Methodology
Optimization

ABSTRACT

Discovery of alternative to the pozzolanic materials generated from industrial wastes was needed because of its unavailability when the industries was shutdown permanently. This paper deals the optimization of calcined bentonite clay utilization in cement mortar using response surface methodology (RSM). The variables were taken as three levels of calcination temperature (room temperature, 700^{0}C and 800^{0}C) and seven levels of calcined bentonite (0%, 5%, 10%, 15%, 20%, 25% and 30%). The compressive strength, workability, strength activity index and sorpitivity were taken as responses. The fresh and hardened properties of all determined for all mixes. Design Expert 11.0 version was utilized to carried out modelling and optimization using RSM. Workability was decreased upon increasing the calcination temperature and bentonite content in cement mortar. This attributed to high water absorption capacity of bentonite. The peak compressive strength was displayed by 20% replaced bentonite calcined at 800^{0}C cement mortar after 28 days curing. Strength activity was improved upon increasing the percentage of bentonite calcined at 800^{0}C. The sorpitivity of cement mortar was improved by incorporation of bentonite calcined at 800^{0}C. The generated models from RSM were significance in all the factors considered. Optimum performance of the responses was observed at 15.25 % bentonite substitution calcined at 800^{0}C

***doi**: 10.5829/ije.2021.34.07a.07*

NOMENCLATURE

Degree Celcius	^{0}C	*Mega Pascal*	*MPa*
Kilo Newton	kN	Analysis of Variance	ANOVA

1. INTRODUCTION

Utilization of pozzolanic materials will enhance the fresh and hardened properties of cement mortar and concrete [1]. Presently, fly ash, GGBS, silica fume, metakaolin etc. are utilizing as a pozzolanic materials in concrete for better performance as well as reduction of CO_2 emissions to the environment [2]. Fly ash is one of the discarded materials of thermal power plants. It is one of the widely using pozzolanic materials in construction industry [3]. The performance of fresh and hardened properties of concrete will enhance by incorporation of fly ash in concrete [4]. Fly ash will not be generated once the thermal power plants is permanently shutdown. So, the discovery of alternative to the pozzolonic materials generated from industrial wastes was needed.

Bentonite is a natural pozzolonic material, which can be used instead of industrial wastes in concrete [5, 6]. A few research was available on the evaluation of the properties of bentonite incorporated cement mortar [6-8] as well as concrete [9, 10]. The fresh and hardened properties of concrete was increased by incorporation of bentonite in concrete [11, 12]. Few examinations were done to assess the combination of bentonite and other pozzolonic materials [13, 14]. The ambiguity in the effect of bentonite was observed based on it's source of collection and process of treatment [15, 16]. therefore, investigation on process of treatment of bentonite is needed to estimate the influence on bentonite.

*Corresponding Author Institutional Email: achyuthakumar@kluniversity.in (M. Achyutha Kumar Reddy)

Please cite this article as: S. Sahith Reddy, M. Achyutha Kumar Reddy, Optimization of Calcined Bentonite Caly Utilization in Cement Mortar using Response Surface Methodology, International Journal of Engineering, Transactions A: Basics Vol. 34, No. 7, (2021) 1623-1631

Calcination is a method of heating a material under regulated temperature and in a monitored environment. A few investigations were made to evaluate the effect of calcined bentonite on strength and durability of cement mortar and concrete as well as self-compacting concrete which is available in Pakistan [17-19].

Mirza et. Al. [20] initiated the research on the effect of calcined bentonite on cement mortar and concrete by considering 150^0C, 250^0C, 500^0C, 750^0C and 950^0C as calcination temperatures. Apart from the effect of bentonite calcined at 150^0C on cement mortar and concrete, no results or discussion was made on the effect of the remaining calcination temperatures. The research on the same topic was extended by Ahmad et. al. [21]. They considered the calcination temperatures as 500^0C and 900^0C, evaluated the fresh and hardened properties of cement mortar and concrete. It was observed almost the same strength activity for raw bentonite and bentonite calcined at 500^0C. The strength activity was drastically reduced by incorporation of bentonite calcined at 900^0C. Moreover, no optimal solution was provided in terms of bentonite substitution or calcination temperature of bentonite [21].

Optimization included the utilization of Response Surface Methodology (RSM) employing theories of mathematical and statistical analysis techniques between variables and responses [22]. RSM was utilized for the optimization of the wanted set of objectives, either independent variables or responses, to implement multi-objective optimization in different concrete materials [23, 24]. Mohammed et al. [25] performed multi-objective optimization to accomplish a connection between variables and responses of the properties of roller-compacted concrete by keeping fly ash content as constant, considering the combined effect of both crumb rubber and nano-silica [26]. This research aimed at the optimization of calcined bentonite usage in cement mortar by using RSM. The bentonite substitution (0%, 5%, 10%, 15%, 20%, 25% and 30%) and calcination temperatures (RT, 700^0C and 800^0C) were taken as variables. Compressive strength, workability, strength activity index and sorpitivity were considered as responses

2. EXPERIMENTAL PROGRAM.

2. 1. Materials OPC 53 grade cement was used in experimental work, the properties are within the permissible limits as per the standard procedure IS12269. Stand sand was used in the determination of compressive strength, the properties are within the permissible limits as per the standard procedure IS650. Bentonite was collected form Tandur (17°14′27″N and 77°35′14″E), southern region of India. The calcination of the bentonite was done with muffle furnace having maximum temperature of 1500^0C and working temperature of 1200^0C. The chemical composition of bentonite was displayed in Table 1. The XRD analysis was performed by using X-Pert X-ray Diffractometer of PANalytical, model number: PW 3040/00. Five mineral crystalline structures was observed upon analysis, Figure 1 exhibits the different phases. Scanning electron microscopy (SEM) images were obtained by using Nova Nano SEM/ FEI for further pursuit. It was observed that the crystalline structure by studying the images displayed in Figure 2.

2. 2. Testing of Specimens Normal consistency, initial setting time and final setting time tests for all mixes was conducted as per the standard procedure IS4031. The workability of mixes was tested as per the standard procedure IS 4031 Part-7.

TABLE 1. Chemical analysis of bentonite sample

S. No.	Component	% Mass
1	SiO_2	51.11
2	Al_2O_3	16.38
3	CaO	7.12
4	MgO	7.57
6	Fe_2O_3	7.65
7	K_2O	1.34
8	Na_2O	0.29
9	P_2O_5	0.29
10	MnO	0.14
11	V_2O_5	0.07
12	TiO_2	1.29
13	LoI*	6.75

* Loss on Ignition

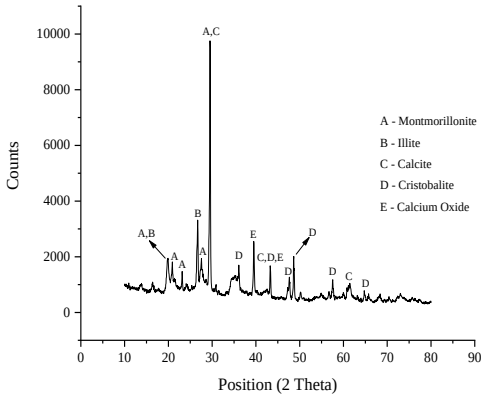


Figure 1. XRD Analysis of Bentonite

A total number of 189 cement mortar cubes were cast and tested to failure after 3,7 and 28 days curing as per the standard procedure IS 4031 Part-6. Testing of all specimens was done with compression testing machine of maximum capacity 3000 kN and rate of loading as1kN/s. The strength activity index test was conducted as per the standard procedure ASTM C311. The sorpitivity test was conducted as per the standard procedure ASTM C1585 – 20.

The sorpitivity test was conducted as per the standard procedure ASTM C1585 – 20. The proportions of all mixes were calculated as per the standard procedure IS 4031 Part-6, displayed in Table 2

3. RESULTS AND DISCUSSION

Analysis was performed for the values taken after testing the specimens of all mixes as per the standards.

3. 1. Normal Consistency, Initial & Final Setting The normal consistency of all mixes is shown in Figure 3. Rise in the normal consistency was observed upon increasing calcined bentonite addition. Initial and final setting times of all mixes was displayed in Figures 4. and 5. Increase in the initial and final setting times were observed by addition of bentonite. This attribute the high-water absorption capacity of bentonite and removal of moisture in bentonite upon calcination.

3. 2. Workability The workability of calcined bentonite mortar was summarized, displayed in Figure 6. The workability was decreased upon increasing the percentage of calcined bentonite addition. This may be the result of high-water demand of bentonite, it will increase by calcination process.

3. 3. Compressive Strength The compressive strength of all mixes was calculated, the results were summarized in Figures 7, 8 and 9 for 3, 7 and 28 days of curing, respectively. Less compressive strength was noticed at 3 days curing calcined bentonite blended cement mortar mixes. Compressive strength was improved at 7 days curing for calcined bentonite at 800^0C among all mixes. The cement mortar exhibits highest compressive strength for 20 % bentonite substitution and calcined at 800^0C temperature. This attribute the pore filling effect between cement particles by bentonite addition, the particle size of calcined bentonite was less than cement. The secondary C-S-H gel formation at lateral ages of cement mortar because bentonite is a pozzolanic material. Calcination at 800^0C eliminates the impurities of bentonite, leads to enhancement of compressive strength at lateral ages.

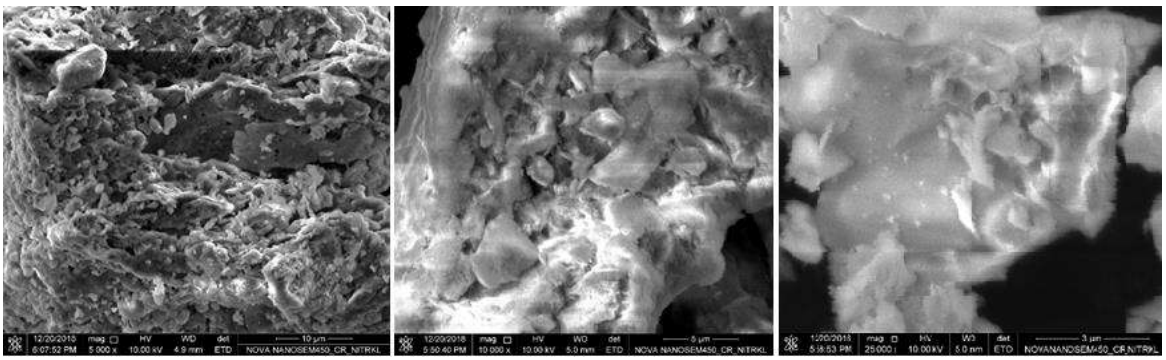


Figure 3. SEM Analysis of Bentonite

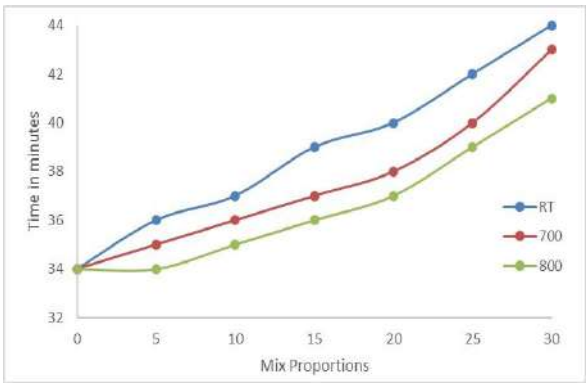


Figure 3. Normal Consistency setting time of all mixes

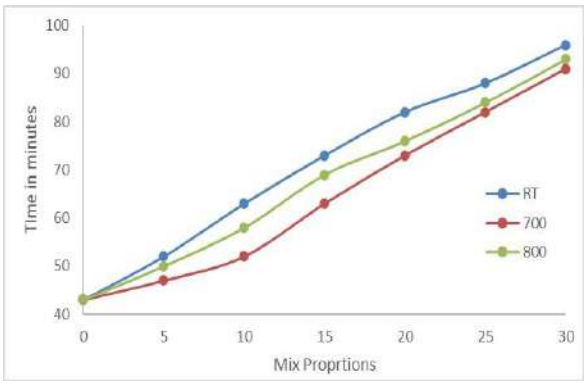


Figure 4. Initial setting time of all mixes

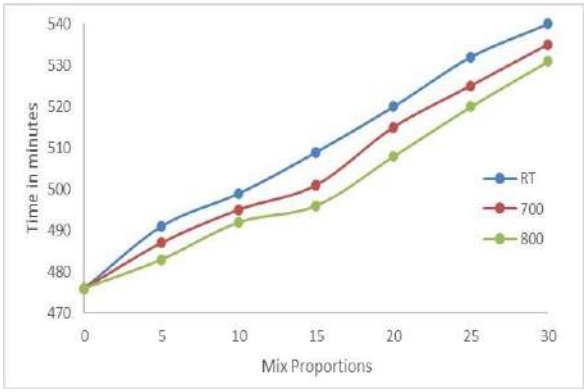


Figure 5. Final setting time of all mixes

3. 4. Strength Activity Index Strength activity index of all specimens was calculated at 28 days curing. The summary of results was displayed in Figure 11. More strength activity index was seen for the cement mortar replaced by 20 % bentonite calcined at 800^{0}C. This may be because of pozzolonic activity occurrence upon addition of calcined bentonite.

3. 4. Strength Activity Index Strength activity index of all specimens was calculated at 28 days curing. The summary of results was displayed in Figure 11. More strength activity index was seen for the cement mortar

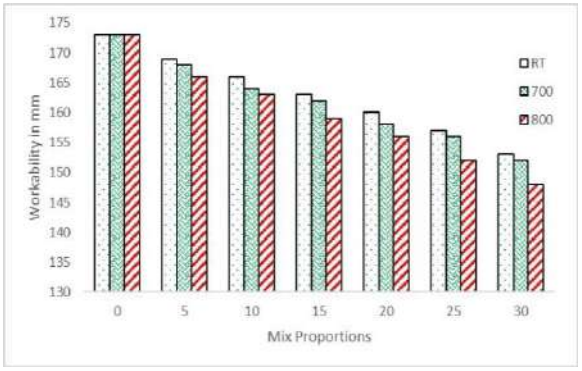


Figure 6. Workability of all mixes

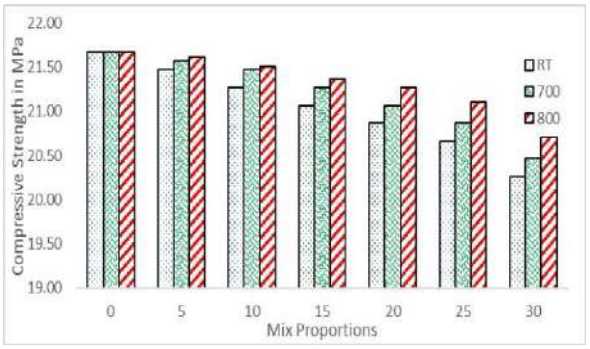


Figure 7. Compressive strength of calcined bentonite mixes after 3 days curing

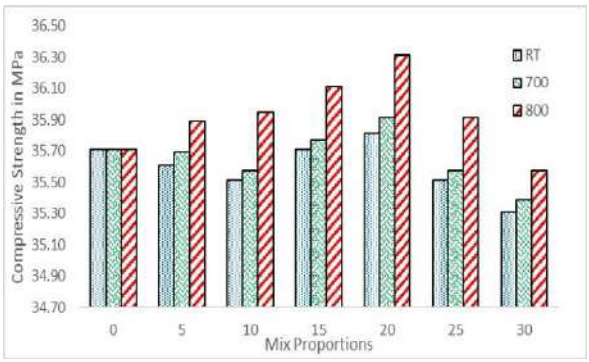


Figure 8. Compressive strength of calcined bentonite mixes after 7 days curing

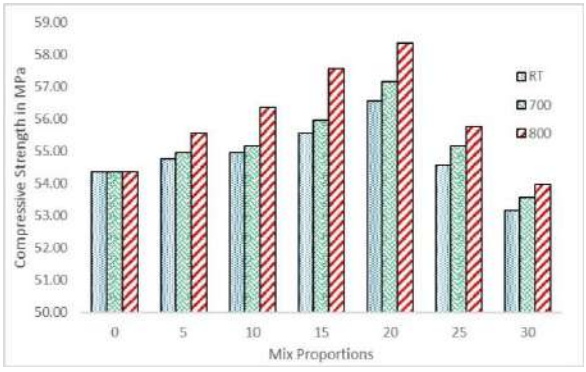


Figure 9. Compressive strength of calcined bentonite mixes after 28 days curing

replaced by 20 % bentonite calcined at 800^{0}C. This may be because of pozzolonic activity occurrence upon addition of calcined bentonite.

3. 5. Sorpitivity Figure 11 displays the sorpitivity for all mixes after 28 days curing. Drastic decrease in the sorpitivity was observed upon addition of bentonite as well as calcined bentonite. This may be as a result of high-water absorption capacity of bentonite and calcined bentonite.

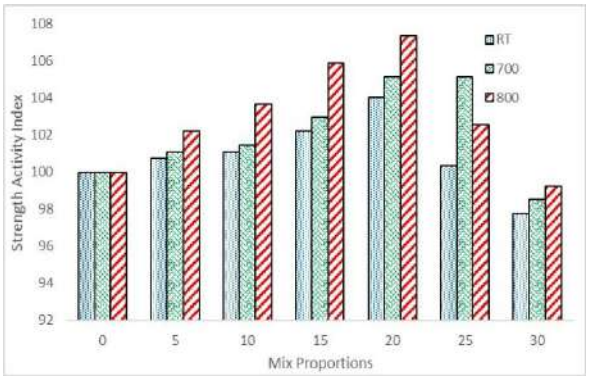


Figure 10. Strength Activity Index of calcined bentonite mixes after 28 days curing

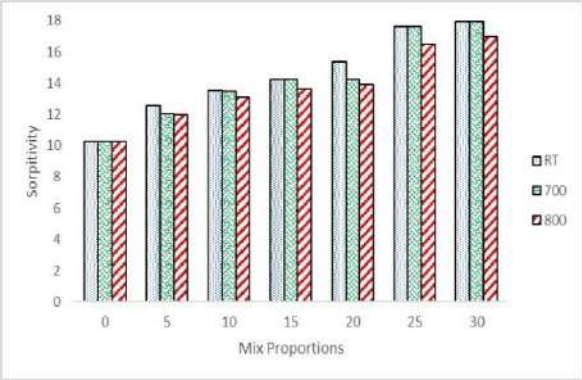


Figure 11. Sorpitivity of calcined bentonite mixes

4. RSM MODELLING

4. 1. Development of Model Various types of models are available like Central Composite, Box Behnken and optimal (custom) in a randomized design in RSM. Selecting a suitable model depends on the type of available data and levels for each factor. The modelling and optimization were done after performance of laboratory experiments using central composite design method of RSM. CCC models are the unique form of the central composite design. The star points are at some α distance from the center based on the properties required for the design and the number of factors in the design. To maintain rotatability, the value of α depends on the number of trial runs in the factorial portion of the CCC. However, the factorial portion can also be a fractional factorial design of resolution V. The value of α also varies on whether the design is orthogonally blocked.

4. 2. Mix Matrix Design In this investigation, Design Expert 11.0 version was used. The design of experiments was created by a composite design technique based on two variables (calcination temperature and bentonite replacement). Three levels of calcination temperature (Room Temperature (RT), 700^0C and 800^0C) and six levels % of bentonite replacement (0%, 5%, 10%, 15%, 20%, 25% and 30%) were used. 21 combinations of mixtures were developed in RSM. Table 3 represents the details of all combination of variables. The responses of calcined bentonite blended concrete (workability, compressive strength, strength activity index and sorpitivity) were determined for all mixtures, considered for RSM analysis and optimization.

4. 3. Analysis of Variance The summaries of ANOVA for the responses were analyzed, displayed in Tables 5-8. F-values are 9.60, 260.48, 7.10 and 155.56 for workability, compressive strength, strength activity index and sorpitivity respectively, pointing all models to be significant. The quadratic model was used for workability, compressive strength and strength activity index. For sorpitivity, linear model was used. In models of workability and sorpitivity, all terms are significant. For compressive strength and strength activity index, AB and A^2 and was the insignificant term and interaction in the quadratic model. The final models for workability, compressive strength, strength activity index and sorpitivity of all mixes consist of all the terms are given in Equations (1)-(4).

$$\text{Workability}=+172.50305+ 0.020037A - 0.741675B- 0.000129AB - 0.000027\ A^2 + 0.003651\ B^2 \quad (1)$$

$$\text{Compressive strength}=+53.83245 - 0.006978A + 0.331324B + 0.000030AB + 9.69751E\text{-}06\ A^2 – 0.011656\ B^2 \quad (2)$$

$$\text{Strength activity index}=+98.80342 - 0.006795A + 0.612087B + 0.000084AB + 0.000010\ A^2 –0.021419\ B^2 \quad (3)$$

$$\text{Sorpitivity} = +10.94914 - 0.000724A + 0.238712B \quad (4)$$

3-dimensional (3D) response surface plot was used to illustrate the relationship between responses and independent variables. Figures 12-15 show the 3D response surface plots illustrating the relationship between responses (workability, compressive strength, strength activity index and sorpitivity) and independent variables (RT, 700^0C and 800^0C) for all mixes.

4. 4. Optimization Optimization was employed with the aim of getting an optimized variable of all bentonite mixes. Optimization was performed by using Design Expert Software. Table 8 shows optimization results of bentonite mixes. The optimized mix was achieved at 15.25 % of substitution of bentonite calcined at 800^0C. The optimum values of responses were 57.17 MPa compressive strength, 159.01 mm workability, 105.14 strength activity index and 14.03 sorpitivity. The optimized performance was proposed by considering different performance indexes.

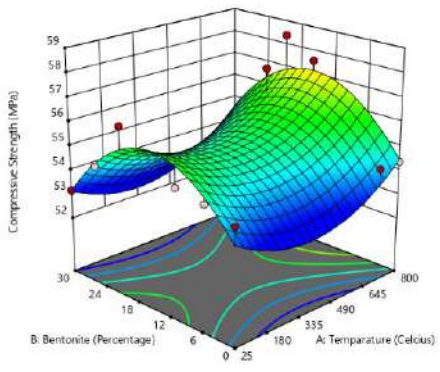


Figure 12. 3D surface model for compressive strength

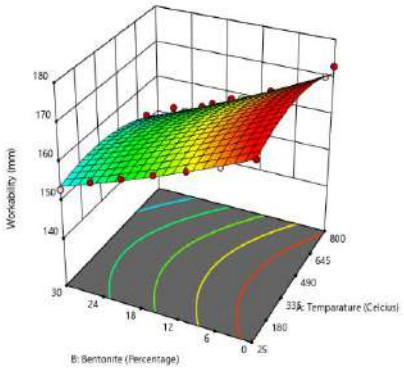


Figure 13. 3D surface model for workability

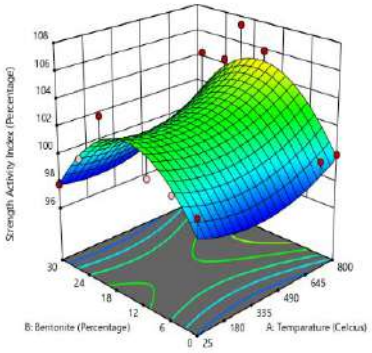


Figure 14. 3Dsurface model for Strength activity Index

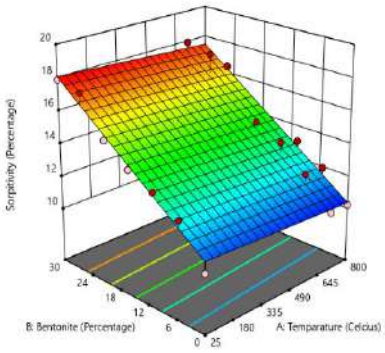


Figure 15. 3D surface model for Sorpitivity

TABLE 3. details of all combination of variables

	F:1	F:2	R:1	R:2	R:3	R:4
S. No	**Calcination Temparture**	**Bentonite Percentage**	**Compressive strength**	**Workability**	**Strength Activity Index**	**Sorpitivity**
1	25	0	54.37	173	100	10.23
2	25	5	54.77	169	100.73	12.54
3	25	10	54.97	166	101.1	13.523
4	25	15	55.57	163	102.2	14.208
5	25	20	56.58	160	104.06	15.345
6	25	25	54.57	157	100.36	17.619
7	25	30	53.17	153	97.79	17.92
8	700	0	54.37	173	100	10.23
9	700	5	54.97	168	101.1	12.01
10	700	10	55.17	164	101.47	13.45
11	700	15	55.98	162	102.96	14.208
12	700	20	57.18	158	105.16	14.208
13	700	25	55.17	156	105.15	17.619
14	700	30	53.57	152	98.52	17.965
15	800	0	54.37	173	100	10.23
16	800	5	55.57	166	102.2	11.98
17	800	10	56.38	163	103.69	13.12
18	800	15	57.58	159	105.9	13.64
19	800	20	58.38	156	107.37	13.894
20	800	25	55.77	152	102.57	16.482
21	800	30	53.97	148	99.26	16.98

TABLE 4. ANOVA for quadratic model, R1: Compressive strength

Source	Sum of Squares	df	Mean Square	F-value	p-value	
Model	26.49	5	5.30	9.60	0.0003	significant
A-Temperature	4.60	1	4.60	8.33	0.0113	
B-Bentonite	0.0652	1	0.0652	0.1181	0.7359	
AB	0.2306	1	0.2306	0.4177	0.5279	
A^2	1.69	1	1.69	3.06	0.1006	
B^2	21.40	1	21.40	38.76	< 0.0001	
Residual	8.28	15	0.5520			
Cor Total	34.77	20				

TABLE 5. ANOVA for quadratic model, R2: Workability

Source	Sum of Squares	df	Mean Square	F-value	p-value	
Model	1070.90	5	214.18	260.48	< 0.0001	significant
A-Temperature	41.14	1	41.14	50.04	< 0.0001	
B-Bentonite	915.22	1	915.22	1113.07	< 0.0001	
AB	4.13	1	4.13	5.02	0.0406	
A^2	13.40	1	13.40	16.29	0.0011	
B^2	2.10	1	2.10	2.55	0.1309	
Residual	12.33	15	0.8222			
Cor Total	1083.24	20				

TABLE 6. ANOVA for quadratic model, R3: Strength Activity Index

Source	Sum of Squares	df	Mean Square	F-value	p-value	
Model	89.92	5	17.98	7.10	0.0014	significant
A-Temperature	15.54	1	15.54	6.13	0.0256	
B-Bentonite	0.0341	1	0.0341	0.0135	0.9092	
AB	1.76	1	1.76	0.6937	0.4180	
A^2	1.80	1	1.80	0.7098	0.4127	
B^2	72.26	1	72.26	28.53	< 0.0001	
Residual	38.00	15	2.53			
Cor Total	127.92	20				

TABLE 7. ANOVA for linear model, R4: Compressive strength

Source	Sum of Squares	df	Mean Square	F-value	p-value	
Model	120.97	2	60.48	155.56	< 0.0001	significant
A-Temperature	1.30	1	1.30	3.35	0.0838	
B-Bentonite	119.67	1	119.67	307.76	< 0.0001	
Residual	7.00	18	0.3888			
Cor Total	127.97	20				

TABLE 8. Optimized calcined bentonite mixes

Bentonite percentage	Calcination temperature	Compressive strength	Workability	Strength activity Index	Sorpitivity
15.26	800	57.17	159.01	105.14	14.03
21.34	538.78	55.00	159.72	102.32	15.65
19.37	792.08	56.90	156.26	104.80	15.00
17.91	310.96	54.96	163.26	102.21	15.00
29.11	800	54.93	149.55	101.39	17.31

5. CONCLUSIONS

The following conclusions were made based on experimental and statistical analysis,

- Initial and final setting times were increased upon addition of calcined bentonite.
- Workability was decreased by increasing the percentage of calcined bentonite addition. This attributed to the higher water absorption capacity of the bentonite
- The maximum compressive strength was observed for 20 % bentonite substitution and calcined at 800^0C temperature due to pore filling effect since the particle size of bentonite is lesser than cement.
- Strength activity was improved; best strength activity was observed for the mix made with 20 % bentonite substitution calcined at 800^0C because bentonite obeys pozzolanic properties.
- The sorpitivity of cement mortar decreased significantly upon increasing the addition of calcined bentonite due to the lesser particle size of bentonite comparatively with cement.
- The model was prepared, and optimization was done by using RSM. The optimized values of variables were determined as 15.25 % bentonite substitution and 800^0C calcination temperature.
- The optimum values of responses were 57.17 MPa compressive strength, 159.01 mm workability, 105.14 strength activity index and 14.03 sorpitivity.

6. REFERENCES

1. Siddique, R.J.P.E., "Utilization of industrial by-products in concrete", Vol. 95, (2014), 335-347, doi: 10.1016/j.proeng.2014.12.192.
2. Latawiec, R., Woyciechowski, P. and Kowalski, K.J.J.E., "Sustainable concrete performance—CO_2-emission", Vol. 5, No. 2, (2018), 27, doi: 10.3390/environments5020027.
3. Zeng, Q., Li, K., Fen-chong, T., Dangla, P.J.C. and Materials, B., "Determination of cement hydration and pozzolanic reaction extents for fly-ash cement pastes", Vol. 27, No. 1, (2012), 560-569, doi: 10.1016/j.conbuildmat.2011.07.007.
4. Ahad, M.Z., Ashraf, M., Kumar, R. and Ullah, M.J.M., "Thermal, physico-chemical, and mechanical behaviour of mass concrete with hybrid blends of bentonite and fly ash", Vol. 12, No. 1, (2019), 60, doi: 10.3390/ma12010060.
5. Memon, S.A., Arsalan, R., Khan, S., Lo, T.Y.J.C. and materials, b., "Utilization of pakistani bentonite as partial replacement of cement in concrete", Vol. 30, (2012), 237-242, doi: 10.1016/j.conbuildmat.2011.11.021.
6. Ghonaim, S.A. and Morsy, R.A.J.J.o.A.-A.U.E.S., "Study of bentonite usage in environmentally friendly concrete", Vol. 15, No. 57, (2020), 1012-1024, doi: 10.21608/auej.2020. 120366.
7. Liu, M., Hu, Y., Lai, Z., Yan, T., He, X., Wu, J., Lu, Z., Lv, S.J.C. and Materials, B., "Influence of various bentonites on the mechanical properties and impermeability of cement mortars", Vol. 241, (2020), 118015, doi: 10.1016/j.conbuildmat.2020.118015.
8. Karunarathne, V.K., Paul, S.C. and Šavija, B.J.M., "Development of nano-sio2 and bentonite-based mortars for corrosion protection of reinforcing steel", Vol. 12, No. 16, (2019), 2622, doi: 10.3390/ma12162622.
9. Man, X., Haque, M.A., Chen, B.J.C. and Materials, B., "Engineering properties and microstructure analysis of magnesium phosphate cement mortar containing bentonite clay", Vol. 227, (2019), 116656, doi: 10.1016/j.conbuildmat.2019.08.037.
10. Garcia-Lodeiro, I., Boudissa, N., Fernández-Jiménez, A. and Palomo, A.J.M.L., "Use of clays in alkaline hybrid cement preparation. The role of bentonites", Vol. 233, (2018), 134-137, doi: 10.1016/j.matlet.2018.08.098.
11. Darweesh, H. and Nagieb, Z., "Hydration of calcined bentonite portland blended cement pastes", Vol. 14, No. 3, (2007), 301-307.
12. Wei, J., Gencturk, B.J.C. and Research, C., "Hydration of ternary portland cement blends containing metakaolin and sodium bentonite", Vol. 123, (2019), 105772, doi: 10.1016/j.cemconres.2019.05.017.
13. Kumar, B.P., Rao, V.R. and Reddy, M.A.K.J.T., "Effect on strength properties of concrete by partial replacement of cement with calcium bentonite and fly ash", Vol. 8, No. 4, (2017), 450-455, doi.
14. Joel, S.J.C.E.J., "Compressive strength of concrete using fly ash and rice husk ash: A review", Vol. 6, No. 7, (2020), doi.
15. Reddy, M.A.K. and Rao, V.R., "Utilization of bentonite in concrete: A", in International Conference on Advances in Civil Engineering (ICACE-2019). Vol. 21, (2019), 23.
16. Kaufhold, S., Dohrmann, R. and Ufer, K.J.A.C.S., "Determining the extent of bentonite alteration at the bentonite/cement interface", Vol. 186, (2020), 105446, doi: 10.1016/j.clay.2020.105446.
17. Laidani, Z.E.-A., Benabed, B., Abousnina, R., Gueddouda, M.K., Kadri, E.-H.J.C. and Materials, B., "Experimental investigation on effects of calcined bentonite on fresh, strength and durability properties of sustainable self-compacting concrete", Vol. 230, (2020), 117062, doi: 10.1016/j.conbuildmat.2019.117062.

18. Balamuralikrishnan, R. and Saravanan, J.J.E.S.J., "Effect of addition of alccofine on the compressive strength of cement mortar cubes", Vol. 5, No. 2, (2021), 155-170, doi: 10.28991/esj-2021-01265.
19. Abd El-Motaal, A.M., Abdel-Reheem, A. and Mahdy, M.J.C.E.J., "Effect of low mixing speed on the properties of prolonged mixed concrete", Vol. 6, No. 8, (2020), 1581-1592, doi: 10.28991/cej-2020-03091568.
20. Mirza, J., Riaz, M., Naseer, A., Rehman, F., Khan, A. and Ali, Q.J.A.C.S., "Pakistani bentonite in mortars and concrete as low cost construction material", Vol. 45, No. 4, (2009), 220-226, doi: 10.1016/j.clay.2009.06.011.
21. Ahmad, S., Barbhuiya, S., Elahi, A. and Iqbal, J.J.C.M., "Effect of pakistani bentonite on properties of mortar and concrete", Vol. 46, No. 1, (2011), 85-92.
22. Şimşek, B., İç, Y.T., Şimşek, E.H., Güvenç, A.B.J.C. and Systems, I.L., "Development of a graphical user interface for determining the optimal mixture parameters of normal weight concretes: A response surface methodology based quadratic programming approach", Vol. 136, (2014), 1-9, doi: 10.1016/j.chemolab.2014.05.001.
23. Ghafari, E., Costa, H., Júlio, E.J.C. and Materials, B., "Rsm-based model to predict the performance of self-compacting uhpc reinforced with hybrid steel micro-fibers", Vol. 66, (2014), 375-383, doi: 10.1016/j.conbuildmat.2014.05.064.
24. Rajaee, M., Hosseinipour, S. and Jamshidi Aval, H.J.I.J.o.E., "Multi-objective optimization of hmgf process parameters for manufacturing aa6063 stepped tubes using fem-rsm", Vol. 34, No. 5, (2021), 1305-1312, doi: 10.5829/ije.2021.34.05b.25.
25. Mohammed, B.S., Achara, B.E., Liew, M.S., Alaloul, W., Khed, V.C.J.C. and Materials, B., "Effects of elevated temperature on the tensile properties of ns-modified self-consolidating engineered cementitious composites and property optimization using response surface methodology (RSM)", Vol. 206, (2019), 449-469, doi: 10.1016/j.conbuildmat.2019.02.033.
26. Khorami, M., Houseinpour, M., Younesi, H. and Najafpour, G.J.I.J.o.E., "Lipase production in solid state fermentation using aspergillus niger: Response surface methodology", Vol. 25, No. 3, (2012), 151-160.

Persian Abstract

چکیده

کشف جایگزین برای مواد پوزولانی تولید شده از ضایعات صنعتی به دلیل در دسترس نبودن آن هنگام خاموشی دائمی صنایع مورد نیاز بود. در این مقاله بهینه سازی استفاده از رس بنتونیت کلسینه در ملات سیمان با استفاده از روش سطح پاسخ (RSM) پرداخته شده است. متغیرها به عنوان سه سطح دمای کلسینه (دمای اتاق ، ۷۰۰C° و ۸۰۰ C° و هفت سطح بنتونیت کلسینه شده (۰٪ ، ۵٪ ، ۱۰٪ ، ۱۵٪ ، ۲۰٪ ، ۲۵٪ و ۳۰٪) گرفته شدند. مقاومت فشاری ، کارایی ، شاخص فعالیت قدرت و میزان جذب به عنوان پاسخ در نظر گرفته شد. خواص تازه و سخت شده همه برای همه مخلوط ها تعیین می شود. برای انجام مدل سازی و بهینه سازی با استفاده از RSM از نسخه Design Expert 11.0 استفاده شده است. با افزایش دمای کلسیم و مقدار بنتونیت در ملات سیمان ، میزان کارایی آن کاهش یافت. این امر به ظرفیت جذب بالای بنتونیت نسبت می دهد. حداکثر مقاومت فشاری پس از ۲۸ روز پخت با ۲۰٪ جایگزین شده بنتونیت کلسینه شده در ملات سیمان C° ۸۰۰ نشان داده شد. با افزایش درصد بنتونیت کلسینه شده در C° ۸۰۰ ، فعالیت قدرت بهبود یافت. میزان جذب ملات سیمان با ترکیب بنتونیت کلسینه شده در C° ۸۰۰ بهبود یافت. مدلهای تولید شده از RSM در کلیه فاکتورهای مورد بررسی اهمیت داشتند. عملکرد مطلوب پاسخها در ۱۵/۲۵ درصد تعویض بنتونیت که در ۸۰۰ درجه سانتیگراد کلسینه شده مشاهده گردید.

IJE TRANSACTIONS A: Basics Vol. 34, No. 7, (July 2021) 1632-1643

International Journal of Engineering

Journal Homepage: www.ije.ir

Performance of Reinforced Concrete Shear Wall Equipped with an Innovative Hybrid Damper

M. Abedi, M. Jalali*, J. Shafaei

Faculty of Civil engineering, Shahrood University of Technology, Shahrood, Iran

PAPER INFO

Paper history:
Received 26 April 2021
Received in revised form 18 May 2021
Accepted 19 May 2021

Keywords:
Residual Displacement
Shape Memory Alloy
Concrete Shear Wall
Hybrid Damper
Numerical Modelling
Super-elastic Behaviour

ABSTRACT

The performance of simultaneous application of steel cantilever damper and Shape Memory Alloy (SMA) rods in the reinforced concrete (RC) shear wall was investigated. In this regard, the critical numerical validation of three full-scale experimental models were distinctly performed and the results were analyzed. Various aspects of numerical modelling, including material modelling assumptions, behavioural models, elements, and solution methods were compared with experimental results. Specimens considering SMA rods as well as steel cantilever damper were numerically investigated. The results illustrated that with increasing the SMA rod angle, the maximum force was decreased, and the residual displacement and dissipated energy was improved. Also, comparing the specimen results without the SMA rods and the specimen with the SMA rods showed that despite the positive effect of the SMA rods, which leads to an increase in maximum force and reduction of residual displacement, the dissipated energy was decreased.

doi: *10.5829/ije.2021.34.07a.08*

1. INTRODUCTION

The use of steel dampers is one of the passive control methods against earthquakes, which has expanded due to economic and production advantages. In this method, by damper inelastic behaviour, energy is dissipated, and by concentrating the damage in it, damage to other members is prevented. It is also easy to replace this type of damper [1]. In addition to structural damage that leads to the unusability of the structure, residual displacement also causes residents' insecurity.

Due to shaped memory alloy features, researchers have studied their performance in the structure in recent years. The first known example of using a shaped memory alloy in a structure dates back to the repair of the bell tower of the Church of San Georgia in the Trignano region of Italy. The tower was damaged by a 4.8 magnitude earthquake in 1996. To repair, four vertical prestressed steel bars with SMA were placed in the inner corners to increase the structure's flexural strength. SMA machine was made of 60 wires with a diameter of 1 mm and a length of 300 mm. In 2000, the structure was hit by a 4.5 magnitude earthquake; observations showed that the structure was not damaged. In another similar project, Croci retrofit a building damaged by the 1997 earthquake in Assisi, Italy, with super-elastic SMAs [2]. DesRoches et al. [3] Studied the properties of wires and rods made of shape memory alloys composed of nickel and titanium alloys to determine the effects of rod size, loading history, and loading rate on the amount of energy dissipation, self-centering ability, and stress of shift phase. Sayyaadi and Zakerzadeh [4] also examined SMA wires. Kim et al. [5] suggested a type of steel damper to improve the seismic of existing structures. The damper was fixed at one end and free at another end, resulting in behaviour like a cantilever. Lu et al. [6] Examined three systems of structures resistant to lateral forces. They conducted their research on a self-centering concrete frame exposed to a vibrating table, quasi-static loading on a concrete shear wall equipped with self-centering coupler beams, and a concrete shear wall equipped with replaceable members at the foot of the wall. All three structural systems performed effectively against lateral force. Kim et al. [7] examined steel cantilever dampers

*Corresponding Author Institutional Email: mjalali@sharoodut.ac.ir
(M. Jalali)

Please cite this article as: M. Abedi, M. Jalali, J. Shafaei Performance of Reinforced Concrete Shear Wall Equipped with an Innovative Hybrid Damper, International Journal of Engineering, Transactions A: Basics Vol. 34, No. 7, (2021) 1632-1643

under different conditions. The results showed that the damper in a small displacement yielded, and the behaviour of stable hysteresis and the loops' shape is close to a parallelogram, which indicates its high energy dissipation capacity. Ahn et al. [8] tested a concrete shear wall equipped with a steel cantilever damper and an isolator under quasi-static loading. In this study, four specimens with different loading conditions were used. One of the specimens was drifted at 2%, and the other three specimens were loaded in two stages. The results showed that with increasing the initial load drift, the deformation of the steel cantilever damper plastic increases, and the total energy dissipation in the second stage is significantly reduced; however, no severe damage was observed in the wall. Naeem et al. [9] built a hybrid energy dissipator by combining a memory alloy rod with a slotted steel plate. The results showed that the maximum drift between the floor and the displacement of the roof of the model structure equipped with a bar made of memory alloy is significantly reduced. Puentes and Palermo [10] examined braced concrete shear walls with and without steel bracing and SMA. The results showed that in the shear wall model equipped with bracing, resistance, energy dissipation, and displacement recovery increased, stiffness and strength degradation decreased. Liu and Jiang [11] tried to focus possible damage on the replaceable members at the corners of the concrete shear wall. The results showed that lateral load-bearing capacity, ductility, and energy dissipation capacity increased. Liu and Jiang [12] modelled a concrete shear wall with replaceable members at the corners of the wall's foot with different compressive axial force ratios in ABAQUS software. The results showed that walls equipped with replaceable members with a larger axial load ratio, larger load capacity, and larger deformation. Chen et al. [13] Examined a concrete shear wall equipped with a coupler beam and replaceable members at the wall's foot using numerical modelling. In general, shear walls equipped with replaceable members have been shown to dissipate energy better. Puentes and Palermo [14] developed a bracing system consisting of a nickel-titanium super-elastic memory alloy under tensile force to improve the fat shear concrete wall. This study focused on 1.3-scale walls representing concrete shear walls before the 1970s that are prone to shear slippage and oblique cracking. The results showed that walls equipped with shaped memory alloy bracing improved seismic performance, including lateral resistance capacity, ductility, energy dissipation, and displacement recovery. Wang and Zhu [15] explored the possibility of using super-elastic memory alloy bars to access the self-centering reinforced concrete walls. In this study, modelling and nonlinear analysis were performed using the OpenSees finite element program and compared with laboratory results. The results show that although the self-centered reinforced concrete walls dissipation relatively less energy through hysteresis loops, almost no residual deformation remains after cyclic loading with a maximum drift of 2.5%. NourEldin et al. [16] used a slotted steel damper equipped with a shaped memory rod in a steel frame with eccentric bracing and analyzed the fragility and cost-effectiveness of the life cycle. The results showed that the frame equipped with a hybrid damper had a lower seismic response than the frame with a slotted damper due to the increase in seismic performance due to the extra stiffness, energy dissipation, and self-centering ability provided by the SMA rod. The results also showed that the life cycle cost of frames equipped with hybrid dampers was lower compared to frames without slotted dampers, although the initial costs of hybrid dampers were higher than those of slotted dampers. Wang et al. [17] investigated the connection of a beam to a steel column using a shaped memory alloy. It was observed that the hysteresis diagram is stable and has good ductility and energy dissipation. Falahian et al. [18] investigated the seismic performance of a steel frame equipped with a self-centering damper. The results showed that the proposed damper limits the residual drifts. Issa and Alam [19] evaluate steel frames equipped with Buckling Restrained Bracing (BRB), Piston Based Self Centering (PBSC) bracing, and Friction Spring Based Piston Bracing (SBPB). The results showed that frames equipped with SBPB and PBSC performed better than frames equipped with BRB. Bogdanovic et al. [20] evaluated steel structures with and without prestressed viscous dampers. The results showed that using this damper, the structural responses are reduced by 10 to 70%. Kamaludin et al. [21] evaluated three concrete frame structures equipped with three types of viscoelastic, friction, and BRB dampers. It was found that viscoelastic dampers perform better than the other two dampers. Alavi et al. [22] developed and presented a combined framework of control-structural optimization. Fathizadeh et al. [23] proposed a new system called " curved damper truss moment frame", and it was found that the proposed system satisfies the requirements of the FEMA P695 code. Aydin et al. [24] investigated the effect of soil-structure interaction on viscous dampers. Barkhordari and Tehranizadeh [25] evaluated the effect of tuned mass damper (TMD), viscous damper, friction damper, and lead core rubber bearing in damage control and seismic response of high-rise structures equipped with a concrete shear wall. Hosseinnejad et al. [26] studied the load-bearing capacity of the post-tensioned tapered steel beams by shaped memory alloy (SMA) tendons. Heydari and Gerami [27] investigated the approach of moment frames with conventional welded connections using a reversible system. Pourzangbar et al. [28] investigated the effect of different viscous damper configurations on the performance of steel frames.

Shojaeifar et al. [1] evaluated the performance of triangular added damping and stiffness (TADAS) dampers in combination with curved dampers (Curved-TADAS damper) in moment resisting steel frame (MRSF).The combination of nickel and titanium shapes memory alloys with two unique behaviours: super-elastic and shape memory. This behaviour results in them being able to withstand large strains of about 8% without creating residual strains. Also, nitinol alloy has excellent corrosion and fatigue resistance, which means that it does not need to be replaced under cyclic loads such as earthquakes.

In this study, due to the model's complexity, three reference papers [3, 7, 8] were utilized to evaluate the results of numerical modelling of SMA rods, steel dampers, and concrete shear walls equipped with steel dampers and isolator. Various numerical modelling, including material modelling, assume, behavioural models, elements used, and solution methods, different experimental results were discussed and interpreted with a numerical model. Also, the proposed system of design steel cantilever damper and SMA rod with different angles were examined.

2. SHAPE MEMORY ALLOY AND CANTILEVER DAMPER

Shaped-memory alloys are known as intelligent materials due to their unique properties. With the combination of different materials, this alloy can be produced that NiTi is one of the most widely used compounds due to its ability to withstand large strains of about 8%. The memory alloy is composed of two crystalline structures, austenite and martensite. The austenite phase is stable at high temperature, and low stress, which leads to super-elastic behaviour, and the martensite phase is stable at low temperature and high stress, which produces a shape memory behaviour. Due to its ability to change from one phase to another by applying temperature and stress, this alloy can change the residual shapes to zero.

Cantilever dampers are a type of slotted damper with one end fixed and the other end free. These dampers have deformation inside the plate and high elastic stiffness, and their geometry reduces the strain concentration due to the reduction of width to the free end. As shown in Figure 1, due to its optimized geometry, it is much more economical than other types of dampers [7].

3. NUMERICAL MODELLING AND INTERPRETATION OF RESULTS

3. 1. Numerical Modelling of Sma Rods, Validation with Experimental Results

The experimental research of DesRoches et al. [3] was selected for validation on the numerical modelling. For boundary conditions, all degrees of freedom of the two ends, except for the displacement in the axial direction for the left end, was restrained. The loading was applied in a cyclic pattern in the tensile direction. Also, mesh with dimensions of 1, 2, 3, and 4 mm (corresponds to 1-SMA, 2-SMA, 3-SMA and 4-SMA numerical models) was employed. Figures 2 show the loading protocol of the SMA rod for simulation.

As shown in Figure 3, the SMA rod's stress decreases at the same strains for larger mesh models than for smaller mesh, although the results have converged in the 3-SMA and 4-SMA models. Table 1 compares the maximum numerical stress of SMA rods with experimental. In addition to the closeness of results, the solution time is also an important issue. So, the 2-SMA model with 331 seconds of solving time was selected.

3. 2. Numerical Modelling of Steel Damper, Verification with Experimental Results

Experimental research of Kim et al. [7] was chosen to validate the performance of steel dampers. In the experimental specimen, the damper is connected to the rigid frame by two strong members on the left and right. For numerical modelling, boundary conditions were applied directly to the plates. Loading was applied to a reference point which restrained to the left plate by the coupling constraint. For the right plate, all degrees of freedom were fully constrained. Two types of isotropic

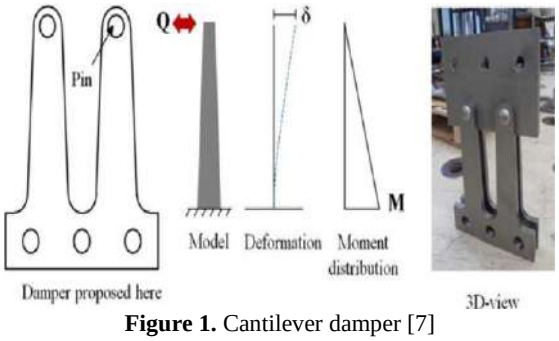


Figure 1. Cantilever damper [7]

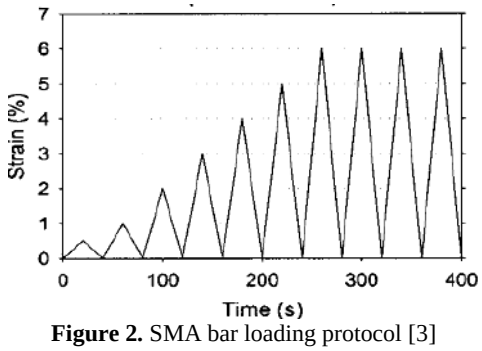


Figure 2. SMA bar loading protocol [3]

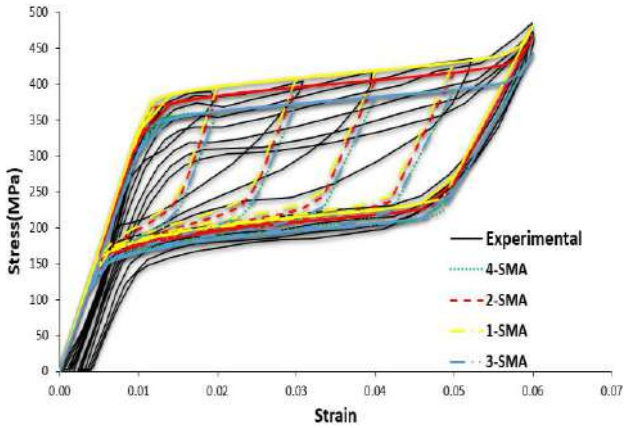


Figure 3. Comparison of stress-strain diagram of SMA rod of the experimental specimen with numerical modelling

TABLE 1. Comparison of experimental results and numerical modelling of SMA rod

Specimen		Experimental model	1-SMA	2-SMA	3-SMA	4-SMA
Maximum stress(MPa)	Drift(%2)	389	397	384	362	362
	Drift(%3)	405	409	396	373	373
	Drift(%4)	417	419	405	382	382
	Drift(%5)	436	430	416	392	392
	Drift(%6)	486	481	467	443	445
Solving time(second)		-	1650	331	117	123

and combined hardening were employed for numerical modelling. The bolts at the damper's right end were neglected in modelling and attached to the tie plate. The pins were rigidly modelled for simplification. Figures 4 and 5 show the boundary conditions and damper finite element mesh.

As shown in Figure 6, in isotropic stiffening specimens, larger hysteresis loops have been formed than in the experimental specimen. In the inelastic region, specimens constrained in the Y direction, the force-displacement diagram has revealed a steeper slope than

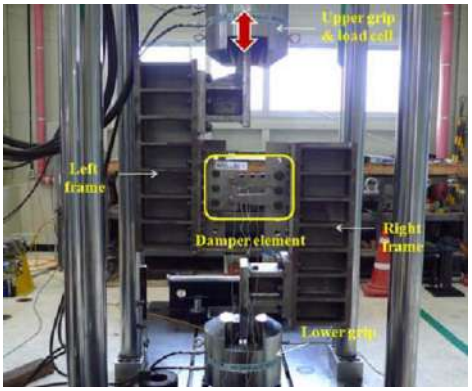


Figure 4. Steel damper boundary conditions [7]

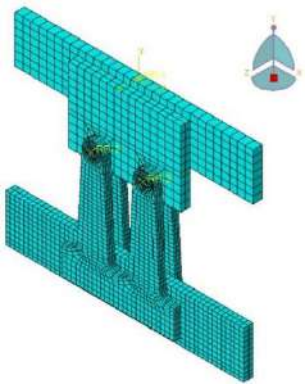


Figure 5. Damper finite element mesh

the experimental model. Nevertheless, the model combine stiffening and accessible in the Y direction fits well with the experimental findings. In Table 2, the numerical results with combined stiffening and free in the Y direction have been compared with the experimental results.

Figure 7 shows the Mises stress contours of the damper. In models released in the Y direction, stress concentration is observed at the bottom of the damper, indicating a rupture formed in this position.

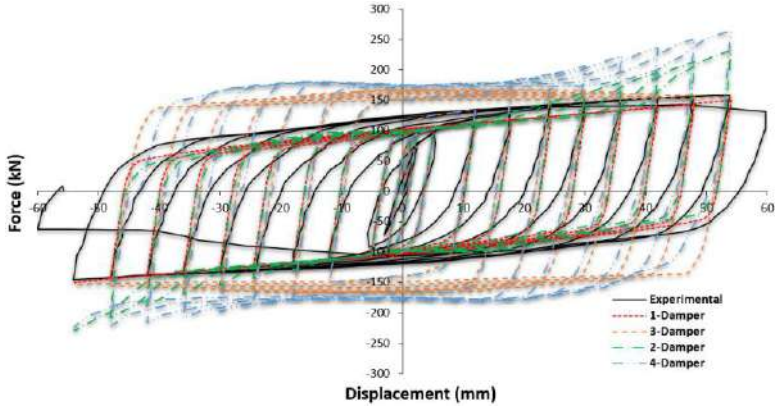
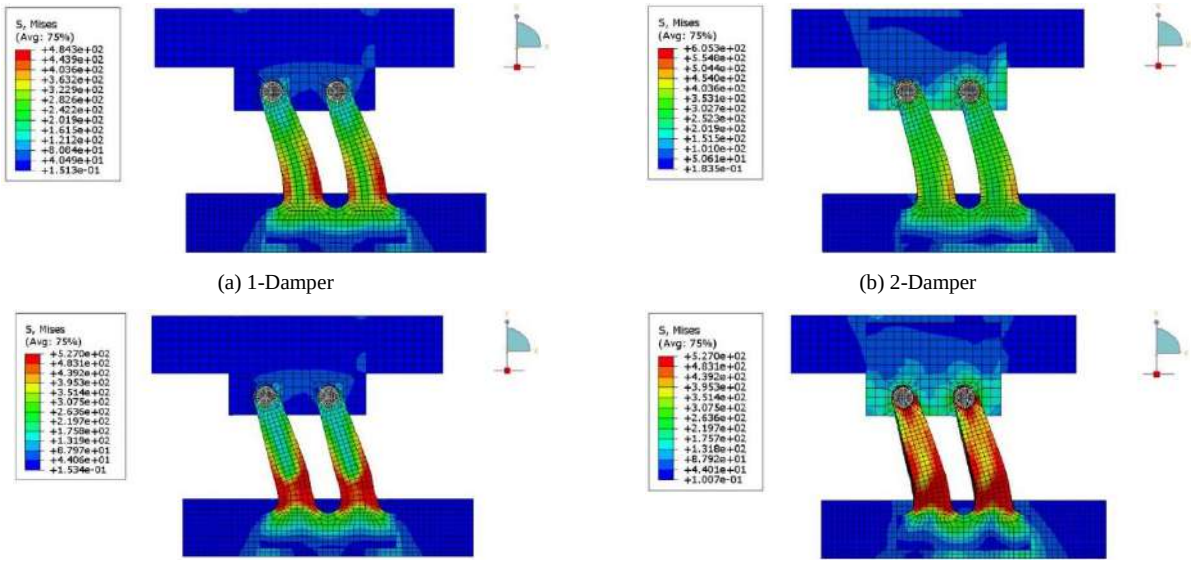


Figure 6. Comparison of force-displacement diagram of steel damper, experimental via numerical results

TABLE 2. Comparison of 1-Damper specimen with the experimental data

Specimen	1-Damper	Experimental	Difference
Yield displacement (mm)	2.5	2.63	-4.94
Yield force (kN)	96	108.40	-11.44
Initial stiffness (kN/mm)	38.4	41.22	-6.84
Second stiffness (kN)	149.68	151.30	-1.07



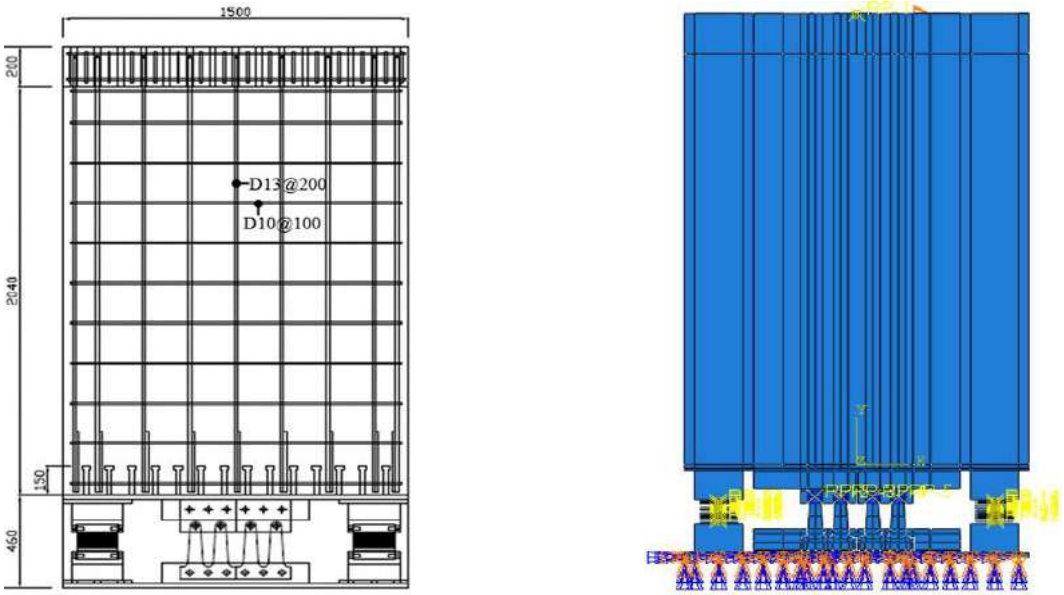
(a) 1-Damper (b) 2-Damper

(c) 3-Damper (d) 4-Damper

Figure 7. Mises stress contours in damper -54 mm displacement of (in MPa)

3. 3. Numerical Modelling of Shear Wall Equipped with Steel Damper and Seismic Isolator, Verification with Experimental Results

Experimental research of Ahn et al. [8] was selected to verify the numerical modelling of concrete shear wall equipped with an isolator and steel damper. Figure 8 shows the configuration of the wall equipped with a damper and seismic separator. Isolator rubber with different values was modelled to validate the Poisson ratio. The concrete was modelled elastic. The effect of cracking on the stiffness of the structure was considered using a cracking coefficient of 0.5. Tables 3 and 4 show the specifications of concrete and rubber materials.



(a) Experimental modelling [8] (b) Numerical modelling

Figure 8. configuration of the wall equipped with damping and seismic separator

TABLE 3. Specifications of concrete materials

Density (kg/m^3)	2400
Specified strength (MPa)	30
Modulus of elasticity (MPa)	26154
Modulus of crack elasticity (MPa)	13077
poisson's ratio	0.15

TABLE 4. Specifications of rubber materials

Poisson's ratio	**K_H (N/mm)**	**G (N/mm^2)**	**K (N/mm^2)**	**C_{10} (N/mm^2)**	**D_1 (mm^2/N)**
0.4990	58	0.4124	206.0555	0.2062	0.0097
0.4993	58	0.4124	294.4239	0.2062	0.0068
0.4995	58	0.4124	412.2484	0.2062	0.0049

As shown in Figure 9 and Table 5, the maximum force and dissipated energy decrease by decreasing the Poisson ratio, and the residual displacement increases. The model with Poisson's ratio of 0.4993 is in good agreement with the experimental results; therefore, Poisson's ratio of 0.4993 was used to model the rubbers.

Experimental diagrams were compared with the numerical model in Figures 10 to 13 as well as Tables 6 to 8. The absence of the columns on both sides of the wall in the numerical model is responsible for differences in results of numerical models compared to experiments. There is also a distance between the hole and the pins connecting the column to the frame. Such detail has not been simulated in numerical modelling and causes the negligible pinching phenomenon in cyclic behaviour.

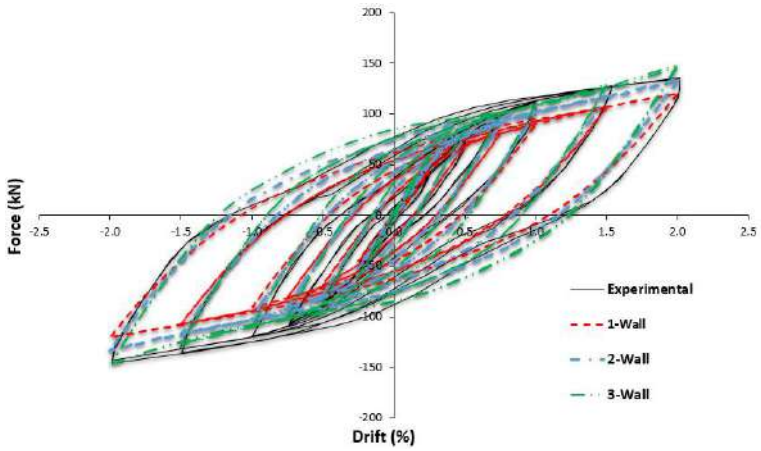


Figure 9. Comparison of force-drift diagram, experimental result vs. numerical finding

TABLE 5. Comparison of models with different Poisson ratios vs. the experimental results

Specimen	Maximum force (kN)	Maximum force difference (%)	Dissipated energy (kN.m)	Dissipated energy difference (%)	Residual displacement (mm)	Residual displacement difference (%)
Experimental	136	0	49	0	-31.59	0
1-Wall	121	-11	41.43	-15.45	-29.29	-7.28
2-Wall	134	-1.47	48.78	-0.45	-31.46	-0.41
3-Wall	149	9.5	56.84	16	-33.5	6.04

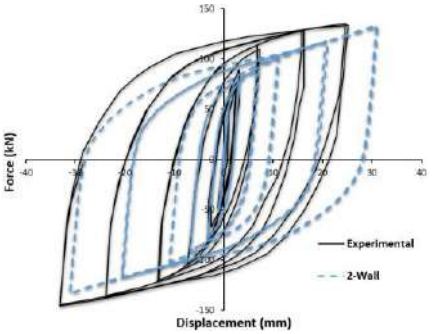


Figure 10. Dampers force-displacement diagram in 2-Wall specimen, Experimental via numerical results

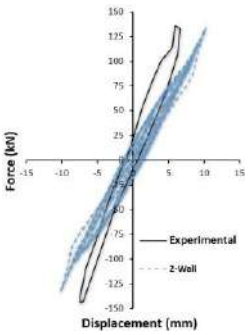


Figure 11. Wall force-displacement diagram in 2-Wall specimen, Experimental via numerical results force-displacement diagram in 2-Wall specimen, Experimental via numerical results

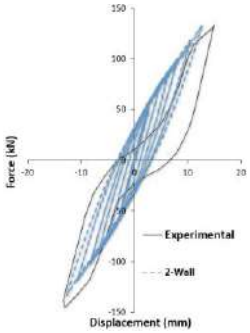


Figure 12. Wall's rigid body rotation diagram in 2-Wall specimen force-displacement diagram in 2-Wall specimen, Experimental via numerical results

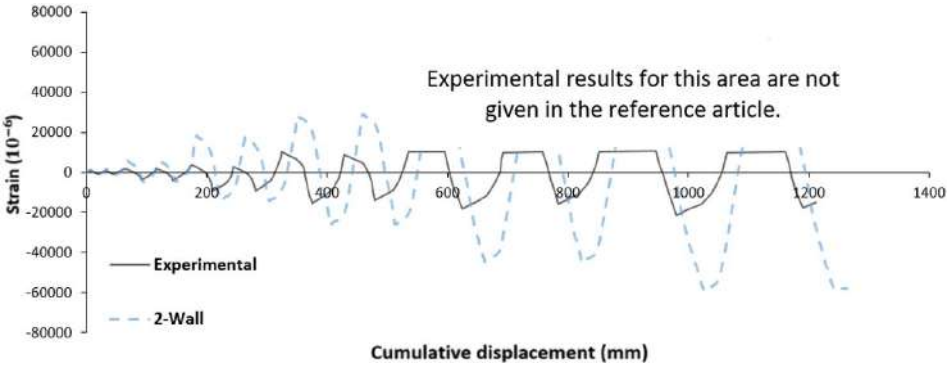


Figure 13. Comparison of damper cumulative displacement-strain diagram with numerical results in 2-Wall specimen

TABLE 6. Comparison of force-displacement results of 2-Wall specimen with the experimental model

Specimen	Maximum displacement in right side (mm)	Right maximum displacement difference (%)	Maximum displacement in left side (mm)	Left maximum displacement difference (%)
Experimental	25	0	-33	0
2-Wall	31	24	-31	-6.06

TABLE 7. Comparison of 2-Wall specimen wall deformation results with the experimental model

Specimen	Maximum displacement in right side (mm)	Right maximum displacement difference (%)	Maximum displacement in left side (mm)	Left maximum displacement difference (%)
Experimental	6.62	0	-7.38	0
2-Wall	10.32	56.89	-10.21	38.35

TABLE 8. Comparison of 2-Wall rigid body rotation result with the experimental model

Specimen	Maximum displacement in right side (mm)	Right maximum displacement difference (%)	Maximum displacement in left side (mm)	Left maximum displacement difference (%)
Experimental	14.90	0	-13.40	0
2-Wall	12.72	-14.63	-12.66	-5.52

Earthquake foreshock was taken into account in two stages. For the first step, 1%, 1.5%, and 2% of drift were applied to the specimens. 2% of drift was considered for the second stage as the main earthquake. Force-drift diagrams are presented in Figures 14 to 16.

The diagrams' difference is due to the seismic isolator's performance in the numerical model. However, the poor performance of the seismic isolator is maybe due to simplification in the simulation of elastic concrete behaviour and absence of columns in the numerical model; it has not been considered on both sides of the wall. Propagation of cracks in the concrete has caused pinching phenomena in cyclic response. There is also a difference between the hole and the pins connecting the column to the frame, which has led to differences in experimental results and numerical modelling. Figure 17 shows the seismic isolator deformation for 2% drift. Figure 18 shows the damper deformation compared to the experimental results in different drifts. As can be seen, in drift 2%, numerical deformation matches the experimental specimen. However, with increasing drift, there is a difference between the deformation of numerical models and corresponding experimental results due to high seismic isolator deformation in the numerical model compared to the experimental.

3. 4. Performance Evaluation of Shear Wall Equipped with a Damper, Seismic Isolator and Sma Rod In this study, the effect of SMA rod with angles of 30, 45, and 60 degrees on the performance of concrete shear wall equipped with steel cantilever damper and seismic isolator was also evaluated. Due to the buckling in the slender members, the rod was attached to the damper so that it was only stretched. Figure 19 reveals the SMA rod assembly with a 30-degree angle to the damper.

Figure 20 presents the force-drift diagram. Table 9 shows the comparison of the models. In the model with 2 rods at an angle of 30 degrees, the displacement of residual has the most significant decrease, and in the model with 2 rods at an angle of 60 degrees, the dissipated energy has the highest increase. The results also prove that by increasing the angle of the rod, the maximum force decreases, and the dissipated energy as well as the residual displacement increases.

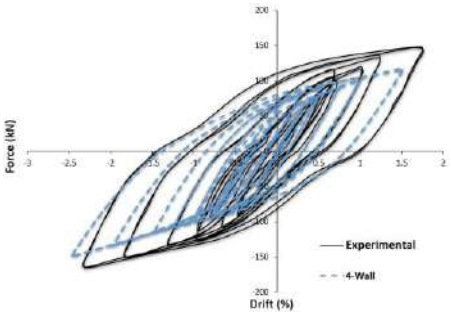


Figure 14. Comparison of force-drift diagram - experimental - via numerical model - 1% and 2% drift

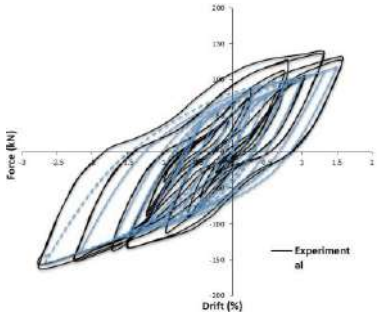


Figure 15. Comparison of force-drift diagram -experimental via numerical model - 1.5% and 2% drift

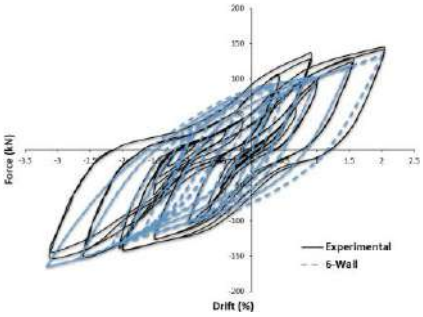


Figure 16. Force-drift diagram of the experimental via numerical model - 2% and 2% drift

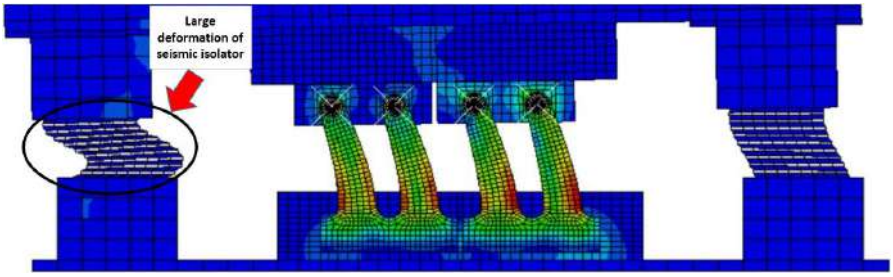


Figure 17. Deformation of the seismic isolator in the 6-Wall specimen

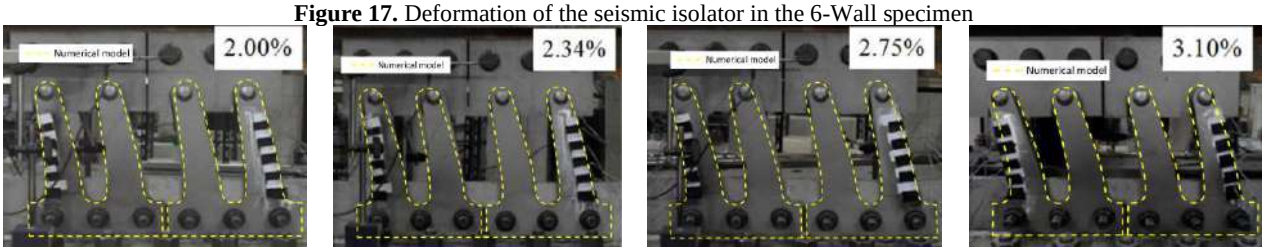


Figure 18. Comparison of experimental specimen deformation with numerical in 2% drift

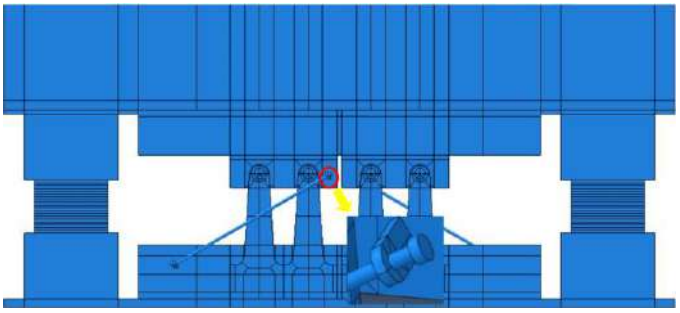
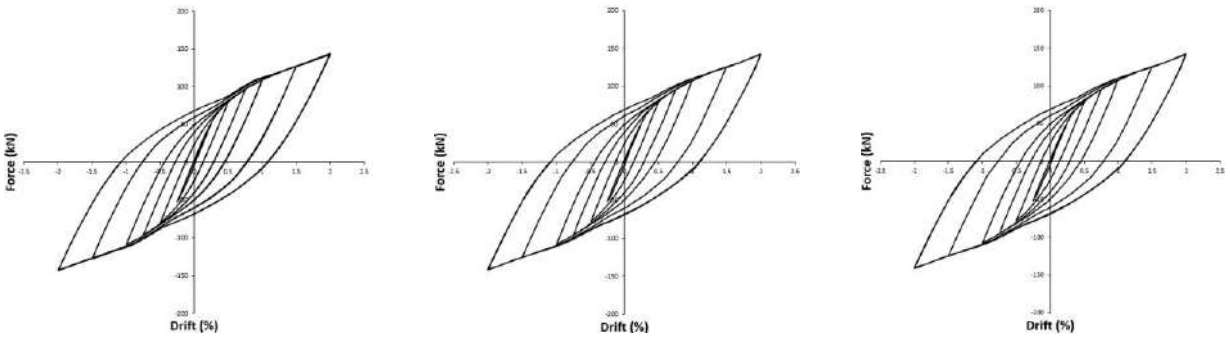


Figure 19. Specimen assembly 1-Wall-SMA



(a) specimen 1-Wall-SMA (b) specimen 2-Wall-SMA (c) specimen 3-Wall-SMA

Figure 20. Force-drift diagram of wall model equipped with steel plan damper, seismic isolator, and SMA rod

TABLE 9. Comparison of specimens with SMA rods

Specimen	Maximum force (kN)	Dissipated energy (kN.m)	Residual displacement (mm)
1-Wall-SMA	143.853	46.57	-29.28
2-Wall-SMA	142.342	47.06	-29.67
3-Wall-SMA	139.921	47.68	-30.22

Figures 21 to 23 compare a specimen's results with 2 SMA rods and a specimen without SMA rods. In the specimen with 2 SMA rods at an angle of 30 degrees compared to the specimen without SMA rods, the maximum force increases by 7.35%, the dissipated energy decreases by 4.53%, and the residual displacement decreases by 6.93%. In the specimen with 2 SMA rods at an angle of 45 degrees compared to the specimen without SMA rods, the maximum force increases by 6.23%, the dissipated energy decreases by 3.53%, and the residual displacement decreases by 5.69%. In the specimen with 2 SMA rods at an angle of 60 degrees compared to the specimen without SMA rods, the maximum force increases by 4.42%, the dissipated energy decreases by 2.26%, and the residual displacement decreases by 3.94%. The results exhibit that despite the SMA rod's positive effect, which leads to an increase in maximum force and a decrease in residual displacement, the dissipated energy decreases.

The effect of 4 SMA bars with an angle of 30 degrees on the structural system's behaviour was also examined. Figure 24 offers the assembly of 4 SMA bars with an angle of 30 degrees to the steel damper. Figure 25 demonstrates the hysteresis diagram for the specimen with 4 SMA bars, and in Figure 26 to 28**,** the results of the model without SMA bars and with 2 and 4 SMA bars are compared. Based on the findings, the maximum force

in the specimen with 4 SMA rods compared to the specimen with 2 SMA rods and without SMA rods increases by 6.38% and 14.20%, respectively, and the dissipated energy by 4.4% and 8.73%, respectively, and the residual displacement by It decreases by 7.31% and 13.73%.

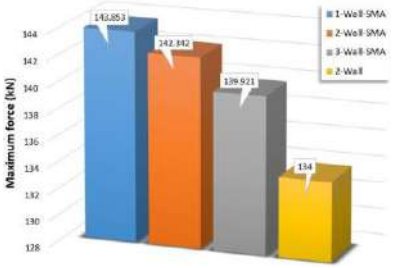


Figure 21. Comparison of maximum force for the specimen with SMA rod and without SMA rod

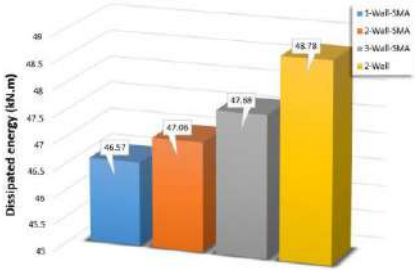


Figure 22. Comparison of dissipated energy for the specimen with SMA rod and without SMA rod

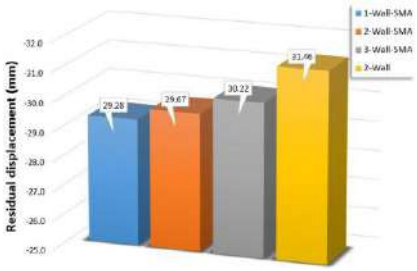


Figure 23. Comparison of residual displacement for the specimen with SMA rod and without SMA rod

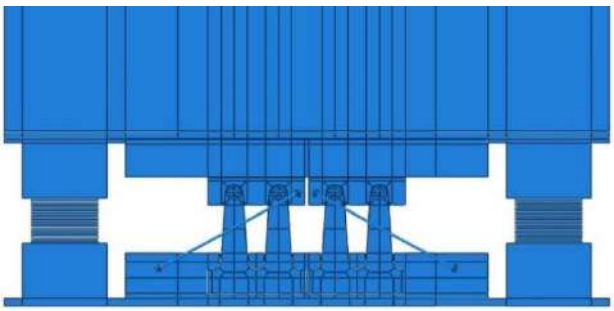


Figure 24. Specimen assembly 4-Wall-SMA

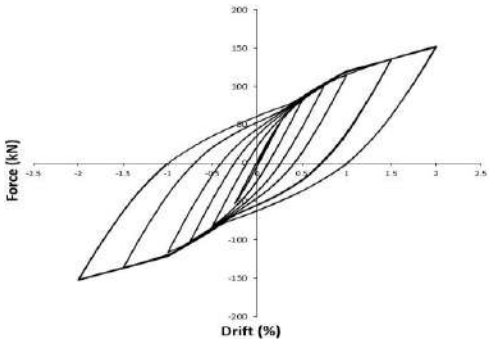


Figure 25. Force-drift diagram - 4-Wall-SMA

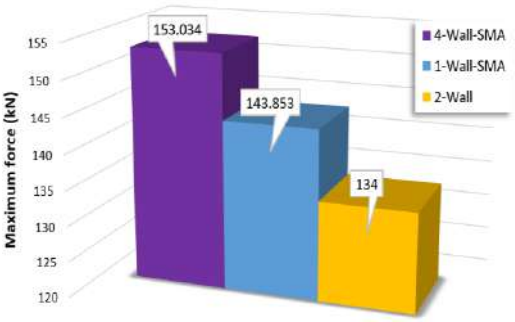


Figure 26. Comparison of maximum force for specimens with 2 and 4 SMA bars and without SMA bars

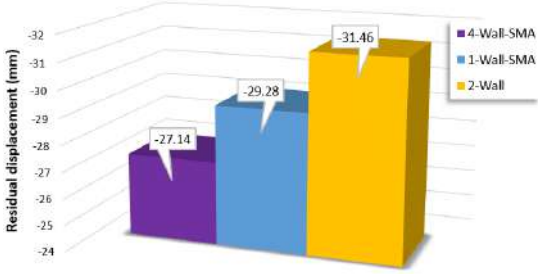


Figure 27. Comparison of residual displacements for specimens with 2 and 4 SMA bars and without SMA bars

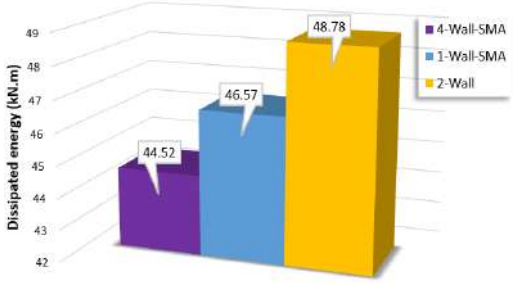


Figure 28. Comparison of dissipated energy for specimens with 2 and 4 SMA rods and without SMA rods

4. CONCLUSION

The most important findings of the present study are summarised as follows.

1. Steel dampers were modelled with isotropic, combined hardness, restrain and free in Y coordinate, respectively. Models with isotropic hardening revealed larger hysteresis loops than the experimental specimen. Moreover, in the specimens restrained in the Y direction, the force-displacement diagram had a steeper slope in the inelastic region than the experimental specimen. The model with combined hardening and release in the Y direction fits well with the experimental data.
2. Seismic isolator rubbers were modelled with Poisson's ratios of 0.4990, 0.4993, and 0.4995. The model with Poisson's ratio of 0.4993 compared well to the experimental specimen in maximum force and dissipated energy had a difference of -1.47 and -0.45%, respectively.
3. The effect of earthquake foreshock loading was simulated in two stages, including 1%, 1.5%, and 2% drift for the first and 2% drift for the second stage. The results demonstrate that the proposed model could not simulate these types of loads due to the poor performance of the seismic separator.
4. Three specimens were modelled with 30, 45, and 60-degree of SMA rod angles to investigate the SMA rod's efficiency. The results illustrated that with increasing rod angle, the maximum force decreases, and the dissipated energy and the residual displacement increased.
5. Comparing the results of specimens with SMA rods and without SMA rods showed that despite the positive effect of SMA rods to increase in the maximum force and a decrease in residual displacement, the dissipated energy had been decreased. The results also showed that in the specimen with four SMA rods compared to the sample with two SMA rods and without SMA rods, the maximum force increases by 6.38 and 14.20%, respectively, the dissipated energy by 4.4 and 8.73%, respectively, and the residual displacement decrease by 7.31 and 13.73%, respectively.

5. REFERENCES

1. Shojaeifar H., Maleki A., and Lotfollahi-Yaghin M.A., "Performance evaluation of curved-TADAS damper on seismic response of moment resisting steel frame", ***International Journal of Engineering, Transactions A: Basics***, Vol. 33, No. 1, (2020), 55-67, DOI: 10.5829/IJE.2020.33.01A.07.
2. DesRoches R. and Smith B., "Shape memory alloys in seismic resistant design and retrofit: a critical review of their potential and limitations", ***Journal of Earthquake Engineering***, Vol. 8, No. 3, (2004), 415-429, DOI: https://doi.org/10.1080/13632460409350495.
3. DesRoches R., McCormick J., and Delemont M., "Cyclic Properties of Superelastic Shape Memory Alloy Wires and Bars", ***Journal of Structural Engineering***, Vol. 130, No. 1, (2004), 38-46, DOI: 10.1061/(ASCE)0733-9445(2004)130:1(38).
4. Sayyaadi H. and Zakerzadeh M., "Nonlinear Analysis of a Flexible Beam Actuated by a Couple of Active SMA Wire Actuators", ***International Journal of Engineering, Transactions A: Basics***, Vol. 25, No. 3, (2012), 249-264, DOI: 10.5829/idosi.ije.2012.25.03a.07.
5. Kim Y., Ahn T., Kim H., and Jang D., "Development of new steel damper for seismic retrofit of existing structures", ***in Proceedings of 15th world conference on earthquake engineering***, (2012).
6. Lu X., Mao Y., Chen Y., Liu J., and Zhou Y., "New structural system for earthquake resilient design", ***Journal of Earthquake and Tsunami***, Vol. 7, No. 03, (2013), 1350013, DOI: https://doi.org/10.1142/S1793431113500139.
7. Kim Y.J., Ahn T.-S., Bae J.-H., and Oh S.-H., "Experimental study of using cantilever type steel plates for passive energy dissipation", ***International Journal of Steel Structures***, Vol. 16, No. 3, (2016), 959-974, DOI: https://doi.org/10.1007/s13296-014-0121-6.
8. Ahn H.-J., Kim Y.-J., Bae J.-H., and Jung I.-Y., "Cyclic loading test of wall damping system with steel dampers", ***Advances in Structural Engineering***, Vol. 19, No. 8, (2016), 1262-1274, DOI: 10.1177/1369433216642068.
9. Naeem A., Eldin M.N., Kim J., and Kim J., "Seismic performance evaluation of a structure retrofitted using steel slit dampers with shape memory alloy bars", ***International Journal of Steel Structures***, Vol. 17, No. 4, (2017), 1627-1638, DOI: 10.1007/s13296-017-1227-4.
10. Cortés-Puentes W.L. and Palermo D., "SMA tension brace for retrofitting concrete shear walls", ***Engineering Structures,*** Vol. 140, (2017), 177-188, DOI: https://doi.org/10.1016/j.engstruct.2017.02.045.
11. Liu Q. and Jiang H., "Experimental study on a new type of earthquake resilient shear wall", ***Earthquake Engineering & Structural Dynamics***, Vol. 46, No. 14, (2017), 2479-2497, DOI: 10.1002/eqe.2914.
12. Liu Q. and Jiang H., "Study on a new type of earthquake resilient shear wall", ***in Proceedings of the 16th World Conference on Earthquake Engineering***, (2017).
13. Chen C., Xiao R., Lu X., and Chen Y., "Study on the shear wall structure with combined form of replaceable devices", ***Advances in Structural Engineering***, Vol. 21, No. 9, (2018), 1327-1348, DOI: 10.1177/1369433217742525.
14. Cortés-Puentes W.L. and Palermo D., "Seismic Retrofit of Concrete Shear Walls with SMA Tension Braces", ***Journal of Structural Engineering***, Vol. 144, No. 2, (2018), 04017200, DOI: doi:10.1061/(ASCE)ST.1943-541X.0001936.
15. Wang B. and Zhu S., "Seismic behavior of self-centering reinforced concrete wall enabled by superelastic shape memory alloy bars", ***Bulletin of Earthquake Engineering***, Vol. 16, No. 1, (2018), 479-502, DOI: 10.1007/s10518-017-0213-8.
16. NourEldin M., Naeem A., and Kim J., "Life-cycle cost evaluation of steel structures retrofitted with steel slit damper and shape memory alloy–based hybrid damper", ***Advances in Structural Engineering***, Vol. 22, No. 1, (2019), 3-16, DOI: 10.1177/1369433218773487.
17. Wang W., Fang C., Feng W., Ricles J., Sause R., and Chen Y., "SMA-based low-damage solution for self-centering steel and composite beam-to-column connections", ***Journal of Structural Engineering***, Vol. 146, No. 6, (2020), 04020092, DOI: https://doi.org/10.1061/(ASCE)ST.1943-541X.0002649.

18. Falahian A., Asadi P., Riahi H.T., and Haghy M., "Seismic performance assessment of steel frames with shape-memory alloy wire-based dampers", ***The Structural Design of Tall and Special Buildings***, Vol. 29, No. 16, (2020), e1797, DOI: https://doi.org/10.1002/tal.1797.
19. Issa A. and Alam M.S., "Comparative seismic fragility assessment of buckling restrained and self-centering (friction spring and SMA) braced frames", ***Smart Materials and Structures***, Vol. 29, No. 5, (2020), 055029, DOI: https://doi.org/10.1088/1361-665X/ab7858.
20. Bogdanovic A., Rakicevic Z., and Noroozinejad Farsangi E., "Shake table tests and numerical investigation of a resilient damping device for seismic response control of building structures", ***Structural Control and Health Monitoring***, Vol. 26, No. 11, (2019), e2443, DOI: https://doi.org/10.1002/stc.2443.
21. Kamaludin P.N.C., Kassem M.M., Farsangi E.N., Nazri F.M., and Yamaguchi E., "Seismic resilience evaluation of RC-MRFs equipped with passive damping devices", ***Earthquakes and Structures***, Vol. 18, No. 3, (2020), 391-405, DOI: https://doi.org/10.12989/eas.2020.18.3.391
22. Alavi A., Dolatabadi M., Mashhadi J., and Farsangi E.N., "Simultaneous optimization approach for combined control–structural design versus the conventional sequential optimization method", ***Structural and Multidisciplinary Optimization***, Vol. 63, No. 3, (2021), 1367-1383, DOI: https://doi.org/10.1007/s00158-020-02765-3.
23. Fathizadeh S.F., Dehghani S., Yang T.Y., Noroozinejad Farsangi E., Vosoughi A.R., Hajirasouliha I., Takewaki I., Málaga-Chuquitaype C., and Varum H., "Trade-off Pareto optimum design of an innovative curved damper truss moment frame considering structural and non-structural objectives", ***Structures***, Vol. 28, (2020), 1338-1353, DOI: https://doi.org/10.1016/j.istruc.2020.09.060.
24. Aydin E., Ozturk B., Bogdanovic A., and Farsangi E.N., "Influence of soil-structure interaction (SSI) on optimal design of passive damping devices", ***Structures***, Vol. 28, (2020), 847-862, DOI: https://doi.org/10.1016/j.istruc.2020.09.028.
25. Barkhordari M.S. and Tehranizadeh M., "Ranking Passive Seismic Control Systems by Their Effectiveness in Reducing Responses of High-Rise Buildings with Concrete Shear Walls Using Multiple-Criteria Decision Making", ***International Journal of Engineering, Transactions B: Applications***, Vol. 33, No. 8, (2020), 1479-1490, DOI: 10.5829/ije.2020.33.08b.06.
26. Hosseinnejad H., Lotfollahi -Yaghin M., Hosseinzadeh Y., and Maleki A., "Numerical Investigation of Response of the Post-Tensioned Tapered Steel Beams with Shape Memory Alloy Tendons", ***International Journal of Engineering, Transactions A: Basics***, Vol. 34, No. 4, (2021), 782-792, DOI: 10.5829/ije.2021.34.04a.04.
27. Heydari. T A.M. and Gerami M., "Multi-stage Performance Upgrade of Steel Moment Frames by Post-tension Connections", ***International Journal of Engineering, Transactions B: Applications***, Vol. 34, No. 5, (2021), 1132-1144, DOI: 10.5829/ije.2021.34.05b.07.
28. Pourzangbar A., Vaezi M., Mousavi S.M., and Saber A., "Effects of Brace-viscous Damper System on the Dynamic Response of Steel Frames", ***International Journal of Engineering, Transactions B: Applications***, Vol. 33, No. 5, (2020), 720-731, DOI: 10.5829/ije.2020.33.05b.02.

Persian Abstract

چکیده

آلیاژ حافظه‌دار شکلی(SMA) دارای قابلیت منحصر به فرد رفتار ابر کشسان و حافظه شکلی است؛ این رفتار باعث می‌شود تا در کرنش‌های بزرگ، آلیاژ حافظه‌دار شکلی کرنش پسماندی در حدود صفر داشته باشد. در کنار این مزیت، رفتار تنش- کرنش آلیاژ حافظه‌دار شکلی در رفتار ابرکشسان شکل پرچم گونه دارد و قابلیت استهلاک انرژی زیادی نداشته و قابل جایگزین شدن با میراگر فلزی نمی باشد. برای جبران ضعف ذکر شده در میراگرهای فلزی، در تحقیق حاضر عملکرد کاربرد هم زمان این دو عضو در دیوار برشی بتن مسلح مورد بررسی قرار گرفته است و عملکرد میراگر ترکیبی (میراگر طره‌ای فولادی و میله SMA) به عنوان یک سیستم لرزه‌ای نوآورانه مورد بررسی قرار گرفته است. مدلسازی عددی در نرم افزار آباکوس انجام شده است. در این خصوص به صحت سنجی کامل ۳ مدل آزمایشگاهی پرداخته شده است. به منظور داشتن درک جامع از صحت نتایج مدلسازی عددی، در قسمت صحت سنجی تنها به مقایسه رفتار نیرو-تغییرمکان اکتفا نشده است وتاریخچه کرنشها و تغییر مکان های قسمت های مختلف سازه حاصل از مدل آزمایشگاهی با مدل عددی مورد مقایسه انتقادی قرار گرفته است. همچنین جنبه‌های گوناگون مدلسازی عددی اعم از فرضیات مدلسازی مصالح، مدل‌های رفتاری، المان‌های مورد استفاده و روش‌های حل در مقایسه با نتایج آزمایشگاهی مورد بحث و تفسیر قرار گرفته است. نتایج نشان دادند با افزایش زاویه میله SMA نیروی حداکثر کاهش و جابه‌جایی پسماند و انرژی مستهلک شده افزایش می‌یابد. علاوه براین، مقایسه نتایج نمونه بدون میله SMA و نمونه با میله SMA نشان دادند برخلاف تاثیر مثبت میله SMA که منجر به افزایش نیرو حداکثر و کاهش جابه‌جایی پسماند می‌شود، انرژی مستهلک شده کاهش می‌یابد.

IJE TRANSACTIONS A: Basics Vol. 34, No. 7, (July 2021) 1632-1643

International Journal of Engineering

Journal Homepage: www.ije.ir

Performance of Reinforced Concrete Shear Wall Equipped with an Innovative Hybrid Damper

M. Abedi, M. Jalali*, J. Shafaei

Faculty of Civil engineering, Shahrood University of Technology, Shahrood, Iran

PAPER INFO

Paper history:
Received 26 April 2021
Received in revised form 18 May 2021
Accepted 19 May 2021

Keywords:
Residual Displacement
Shape Memory Alloy
Concrete Shear Wall
Hybrid Damper
Numerical Modelling
Super-elastic Behaviour

ABSTRACT

The performance of simultaneous application of steel cantilever damper and Shape Memory Alloy (SMA) rods in the reinforced concrete (RC) shear wall was investigated. In this regard, the critical numerical validation of three full-scale experimental models were distinctly performed and the results were analyzed. Various aspects of numerical modelling, including material modelling assumptions, behavioural models, elements, and solution methods were compared with experimental results. Specimens considering SMA rods as well as steel cantilever damper were numerically investigated. The results illustrated that with increasing the SMA rod angle, the maximum force was decreased, and the residual displacement and dissipated energy was improved. Also, comparing the specimen results without the SMA rods and the specimen with the SMA rods showed that despite the positive effect of the SMA rods, which leads to an increase in maximum force and reduction of residual displacement, the dissipated energy was decreased.

doi: *10.5829/ije.2021.34.07a.08*

1. INTRODUCTION

The use of steel dampers is one of the passive control methods against earthquakes, which has expanded due to economic and production advantages. In this method, by damper inelastic behaviour, energy is dissipated, and by concentrating the damage in it, damage to other members is prevented. It is also easy to replace this type of damper [1]. In addition to structural damage that leads to the unusability of the structure, residual displacement also causes residents' insecurity.

Due to shaped memory alloy features, researchers have studied their performance in the structure in recent years. The first known example of using a shaped memory alloy in a structure dates back to the repair of the bell tower of the Church of San Georgia in the Trignano region of Italy. The tower was damaged by a 4.8 magnitude earthquake in 1996. To repair, four vertical prestressed steel bars with SMA were placed in the inner corners to increase the structure's flexural strength. SMA machine was made of 60 wires with a diameter of 1 mm and a length of 300 mm. In 2000, the structure was hit by a 4.5 magnitude earthquake; observations showed that the structure was not damaged. In another similar project, Croci retrofit a building damaged by the 1997 earthquake in Assisi, Italy, with super-elastic SMAs [2]. DesRoches et al. [3] Studied the properties of wires and rods made of shape memory alloys composed of nickel and titanium alloys to determine the effects of rod size, loading history, and loading rate on the amount of energy dissipation, self-centering ability, and stress of shift phase. Sayyaadi and Zakerzadeh [4] also examined SMA wires. Kim et al. [5] suggested a type of steel damper to improve the seismic of existing structures. The damper was fixed at one end and free at another end, resulting in behaviour like a cantilever. Lu et al. [6] Examined three systems of structures resistant to lateral forces. They conducted their research on a self-centering concrete frame exposed to a vibrating table, quasi-static loading on a concrete shear wall equipped with self-centering coupler beams, and a concrete shear wall equipped with replaceable members at the foot of the wall. All three structural systems performed effectively against lateral force. Kim et al. [7] examined steel cantilever dampers

*Corresponding Author Institutional Email: mjalali@sharoodut.ac.ir
(M. Jalali)

Please cite this article as: M. Abedi, M. Jalali, J. Shafaei Performance of Reinforced Concrete Shear Wall Equipped with an Innovative Hybrid Damper, International Journal of Engineering, Transactions A: Basics Vol. 34, No. 7, (2021) 1632-1643

under different conditions. The results showed that the damper in a small displacement yielded, and the behaviour of stable hysteresis and the loops' shape is close to a parallelogram, which indicates its high energy dissipation capacity. Ahn et al. [8] tested a concrete shear wall equipped with a steel cantilever damper and an isolator under quasi-static loading. In this study, four specimens with different loading conditions were used. One of the specimens was drifted at 2%, and the other three specimens were loaded in two stages. The results showed that with increasing the initial load drift, the deformation of the steel cantilever damper plastic increases, and the total energy dissipation in the second stage is significantly reduced; however, no severe damage was observed in the wall. Naeem et al. [9] built a hybrid energy dissipator by combining a memory alloy rod with a slotted steel plate. The results showed that the maximum drift between the floor and the displacement of the roof of the model structure equipped with a bar made of memory alloy is significantly reduced. Puentes and Palermo [10] examined braced concrete shear walls with and without steel bracing and SMA. The results showed that in the shear wall model equipped with bracing, resistance, energy dissipation, and displacement recovery increased, stiffness and strength degradation decreased. Liu and Jiang [11] tried to focus possible damage on the replaceable members at the corners of the concrete shear wall. The results showed that lateral load-bearing capacity, ductility, and energy dissipation capacity increased. Liu and Jiang [12] modelled a concrete shear wall with replaceable members at the corners of the wall's foot with different compressive axial force ratios in ABAQUS software. The results showed that walls equipped with replaceable members with a larger axial load ratio, larger load capacity, and larger deformation. Chen et al. [13] Examined a concrete shear wall equipped with a coupler beam and replaceable members at the wall's foot using numerical modelling. In general, shear walls equipped with replaceable members have been shown to dissipate energy better. Puentes and Palermo [14] developed a bracing system consisting of a nickel-titanium super-elastic memory alloy under tensile force to improve the fat shear concrete wall. This study focused on 1.3-scale walls representing concrete shear walls before the 1970s that are prone to shear slippage and oblique cracking. The results showed that walls equipped with shaped memory alloy bracing improved seismic performance, including lateral resistance capacity, ductility, energy dissipation, and displacement recovery. Wang and Zhu [15] explored the possibility of using super-elastic memory alloy bars to access the self-centering reinforced concrete walls. In this study, modelling and nonlinear analysis were performed using the OpenSees finite element program and compared with laboratory results. The results show that although the self-centered reinforced concrete walls dissipation relatively less energy through hysteresis loops, almost no residual deformation remains after cyclic loading with a maximum drift of 2.5%. NourEldin et al. [16] used a slotted steel damper equipped with a shaped memory rod in a steel frame with eccentric bracing and analyzed the fragility and cost-effectiveness of the life cycle. The results showed that the frame equipped with a hybrid damper had a lower seismic response than the frame with a slotted damper due to the increase in seismic performance due to the extra stiffness, energy dissipation, and self-centering ability provided by the SMA rod. The results also showed that the life cycle cost of frames equipped with hybrid dampers was lower compared to frames without slotted dampers, although the initial costs of hybrid dampers were higher than those of slotted dampers. Wang et al. [17] investigated the connection of a beam to a steel column using a shaped memory alloy. It was observed that the hysteresis diagram is stable and has good ductility and energy dissipation. Falahian et al. [18] investigated the seismic performance of a steel frame equipped with a self-centering damper. The results showed that the proposed damper limits the residual drifts. Issa and Alam [19] evaluate steel frames equipped with Buckling Restrained Bracing (BRB), Piston Based Self Centering (PBSC) bracing, and Friction Spring Based Piston Bracing (SBPB). The results showed that frames equipped with SBPB and PBSC performed better than frames equipped with BRB. Bogdanovic et al. [20] evaluated steel structures with and without prestressed viscous dampers. The results showed that using this damper, the structural responses are reduced by 10 to 70%. Kamaludin et al. [21] evaluated three concrete frame structures equipped with three types of viscoelastic, friction, and BRB dampers. It was found that viscoelastic dampers perform better than the other two dampers. Alavi et al. [22] developed and presented a combined framework of control-structural optimization. Fathizadeh et al. [23] proposed a new system called " curved damper truss moment frame", and it was found that the proposed system satisfies the requirements of the FEMA P695 code. Aydin et al. [24] investigated the effect of soil-structure interaction on viscous dampers. Barkhordari and Tehranizadeh [25] evaluated the effect of tuned mass damper (TMD), viscous damper, friction damper, and lead core rubber bearing in damage control and seismic response of high-rise structures equipped with a concrete shear wall. Hosseinnejad et al. [26] studied the load-bearing capacity of the post-tensioned tapered steel beams by shaped memory alloy (SMA) tendons. Heydari and Gerami [27] investigated the approach of moment frames with conventional welded connections using a reversible system. Pourzangbar et al. [28] investigated the effect of different viscous damper configurations on the performance of steel frames.

Shojaeifar et al. [1] evaluated the performance of triangular added damping and stiffness (TADAS) dampers in combination with curved dampers (Curved-TADAS damper) in moment resisting steel frame (MRSF).The combination of nickel and titanium shapes memory alloys with two unique behaviours: super-elastic and shape memory. This behaviour results in them being able to withstand large strains of about 8% without creating residual strains. Also, nitinol alloy has excellent corrosion and fatigue resistance, which means that it does not need to be replaced under cyclic loads such as earthquakes.

In this study, due to the model's complexity, three reference papers [3, 7, 8] were utilized to evaluate the results of numerical modelling of SMA rods, steel dampers, and concrete shear walls equipped with steel dampers and isolator. Various numerical modelling, including material modelling, assume, behavioural models, elements used, and solution methods, different experimental results were discussed and interpreted with a numerical model. Also, the proposed system of design steel cantilever damper and SMA rod with different angles were examined.

2. SHAPE MEMORY ALLOY AND CANTILEVER DAMPER

Shaped-memory alloys are known as intelligent materials due to their unique properties. With the combination of different materials, this alloy can be produced that NiTi is one of the most widely used compounds due to its ability to withstand large strains of about 8%. The memory alloy is composed of two crystalline structures, austenite and martensite. The austenite phase is stable at high temperature, and low stress, which leads to super-elastic behaviour, and the martensite phase is stable at low temperature and high stress, which produces a shape memory behaviour. Due to its ability to change from one phase to another by applying temperature and stress, this alloy can change the residual shapes to zero.

Cantilever dampers are a type of slotted damper with one end fixed and the other end free. These dampers have deformation inside the plate and high elastic stiffness, and their geometry reduces the strain concentration due to the reduction of width to the free end. As shown in Figure 1, due to its optimized geometry, it is much more economical than other types of dampers [7].

3. NUMERICAL MODELLING AND INTERPRETATION OF RESULTS

3. 1. Numerical Modelling of Sma Rods, Validation with Experimental Results

The experimental research of DesRoches et al. [3] was selected for validation on the numerical modelling. For boundary conditions, all degrees of freedom of the two ends, except for the displacement in the axial direction for the left end, was restrained. The loading was applied in a cyclic pattern in the tensile direction. Also, mesh with dimensions of 1, 2, 3, and 4 mm (corresponds to 1-SMA, 2-SMA, 3-SMA and 4-SMA numerical models) was employed. Figures 2 show the loading protocol of the SMA rod for simulation.

As shown in Figure 3, the SMA rod's stress decreases at the same strains for larger mesh models than for smaller mesh, although the results have converged in the 3-SMA and 4-SMA models. Table 1 compares the maximum numerical stress of SMA rods with experimental. In addition to the closeness of results, the solution time is also an important issue. So, the 2-SMA model with 331 seconds of solving time was selected.

3. 2. Numerical Modelling of Steel Damper, Verification with Experimental Results

Experimental research of Kim et al. [7] was chosen to validate the performance of steel dampers. In the experimental specimen, the damper is connected to the rigid frame by two strong members on the left and right. For numerical modelling, boundary conditions were applied directly to the plates. Loading was applied to a reference point which restrained to the left plate by the coupling constraint. For the right plate, all degrees of freedom were fully constrained. Two types of isotropic

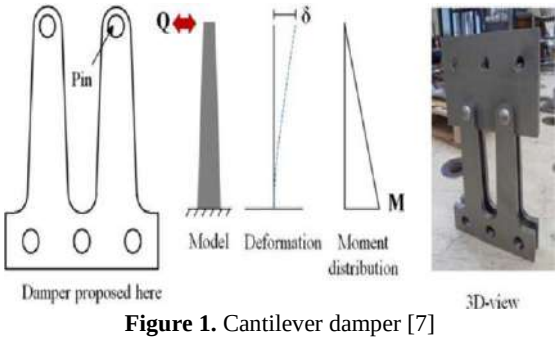


Figure 1. Cantilever damper [7]

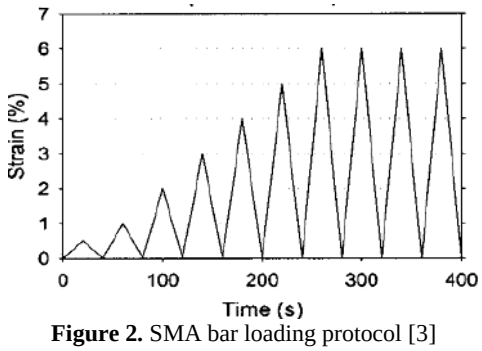


Figure 2. SMA bar loading protocol [3]

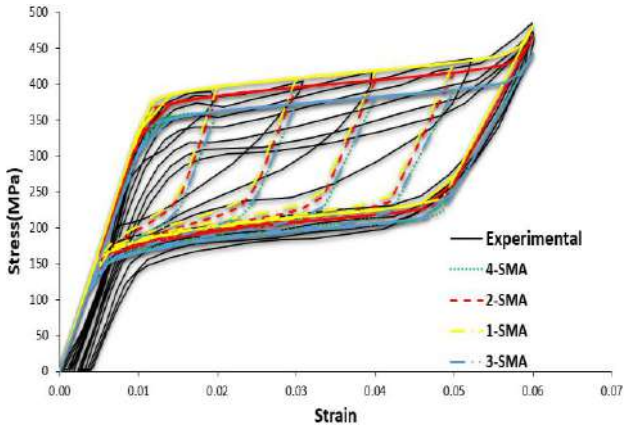


Figure 3. Comparison of stress-strain diagram of SMA rod of the experimental specimen with numerical modelling

TABLE 1. Comparison of experimental results and numerical modelling of SMA rod

Specimen		Experimental model	1-SMA	2-SMA	3-SMA	4-SMA
Maximum stress(MPa)	Drift(%2)	389	397	384	362	362
	Drift(%3)	405	409	396	373	373
	Drift(%4)	417	419	405	382	382
	Drift(%5)	436	430	416	392	392
	Drift(%6)	486	481	467	443	445
Solving time(second)		-	1650	331	117	123

and combined hardening were employed for numerical modelling. The bolts at the damper's right end were neglected in modelling and attached to the tie plate. The pins were rigidly modelled for simplification. Figures 4 and 5 show the boundary conditions and damper finite element mesh.

As shown in Figure 6, in isotropic stiffening specimens, larger hysteresis loops have been formed than in the experimental specimen. In the inelastic region, specimens constrained in the Y direction, the force-displacement diagram has revealed a steeper slope than

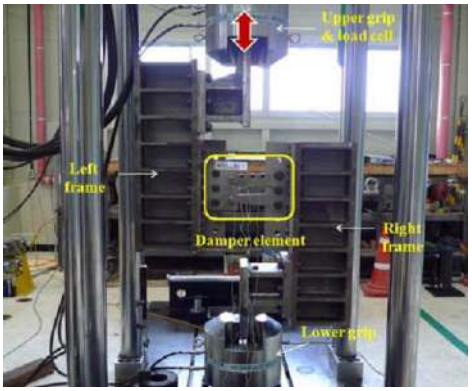


Figure 4. Steel damper boundary conditions [7]

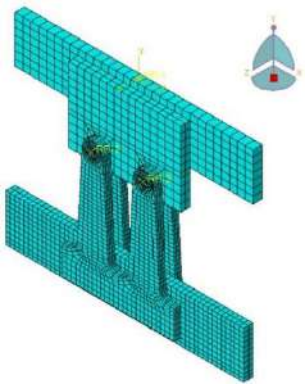


Figure 5. Damper finite element mesh

the experimental model. Nevertheless, the model combine stiffening and accessible in the Y direction fits well with the experimental findings. In Table 2, the numerical results with combined stiffening and free in the Y direction have been compared with the experimental results.

Figure 7 shows the Mises stress contours of the damper. In models released in the Y direction, stress concentration is observed at the bottom of the damper, indicating a rupture formed in this position.

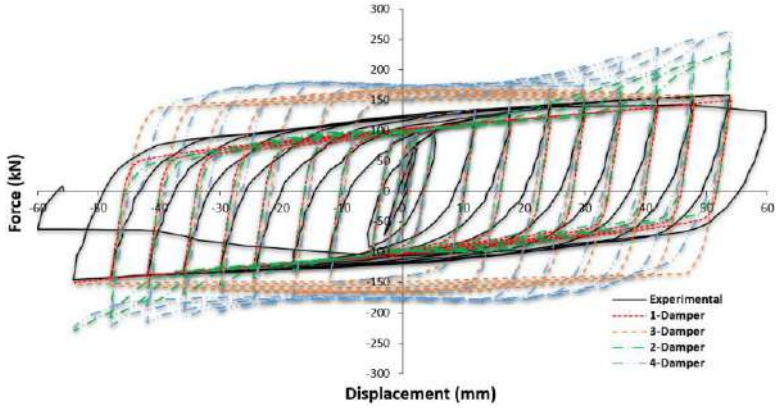
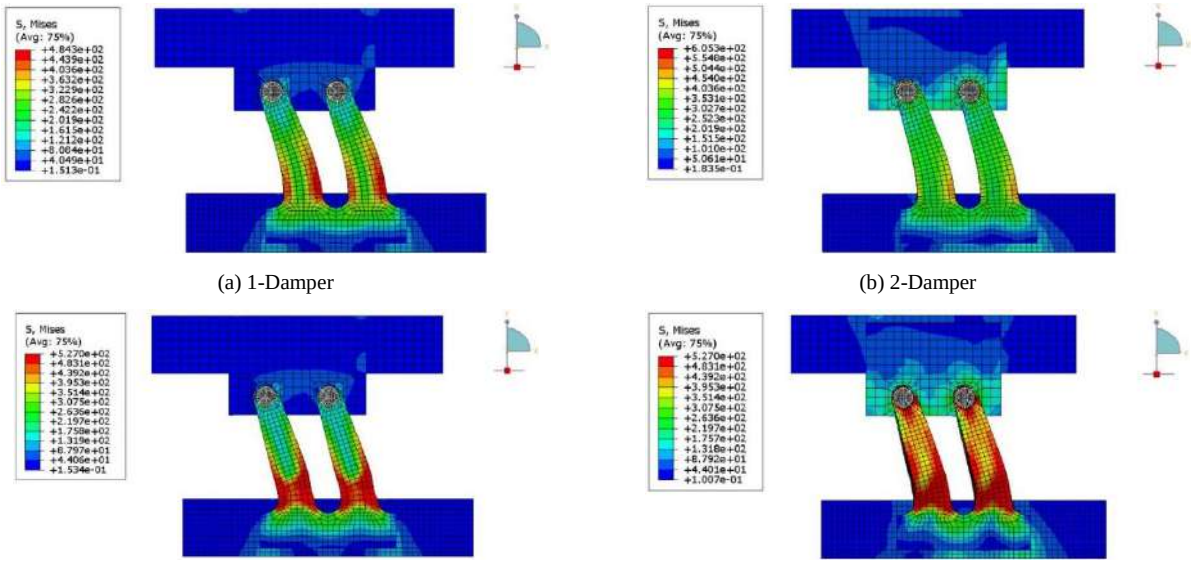


Figure 6. Comparison of force-displacement diagram of steel damper, experimental via numerical results

TABLE 2. Comparison of 1-Damper specimen with the experimental data

Specimen	1-Damper	Experimental	Difference
Yield displacement (mm)	2.5	2.63	-4.94
Yield force (kN)	96	108.40	-11.44
Initial stiffness (kN/mm)	38.4	41.22	-6.84
Second stiffness (kN)	149.68	151.30	-1.07



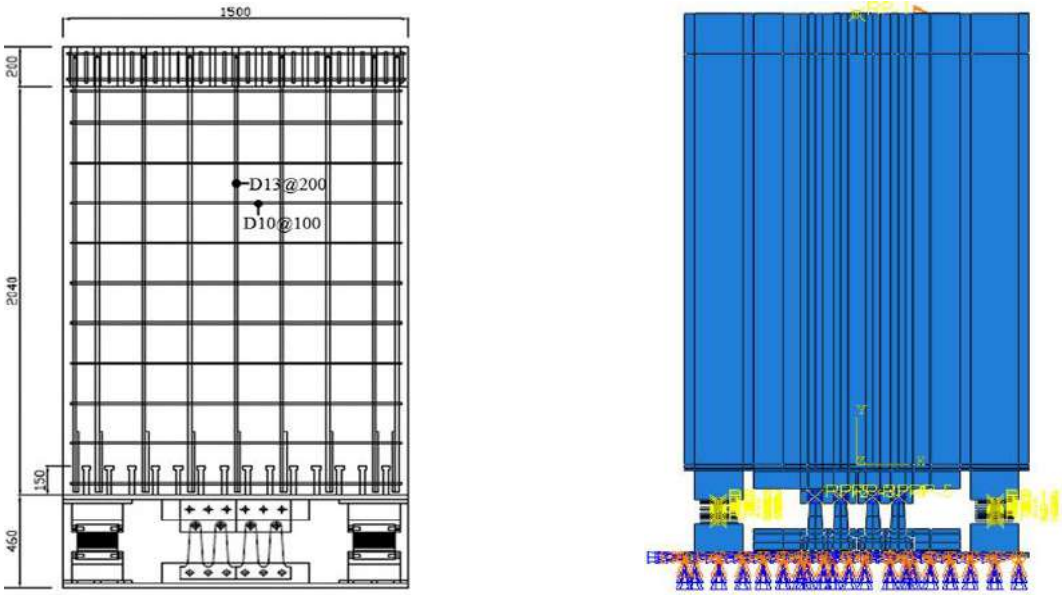
(a) 1-Damper (b) 2-Damper

(c) 3-Damper (d) 4-Damper

Figure 7. Mises stress contours in damper -54 mm displacement of (in MPa)

3. 3. Numerical Modelling of Shear Wall Equipped with Steel Damper and Seismic Isolator, Verification with Experimental Results

Experimental research of Ahn et al. [8] was selected to verify the numerical modelling of concrete shear wall equipped with an isolator and steel damper. Figure 8 shows the configuration of the wall equipped with a damper and seismic separator. Isolator rubber with different values was modelled to validate the Poisson ratio. The concrete was modelled elastic. The effect of cracking on the stiffness of the structure was considered using a cracking coefficient of 0.5. Tables 3 and 4 show the specifications of concrete and rubber materials.



(a) Experimental modelling [8] (b) Numerical modelling

Figure 8. configuration of the wall equipped with damping and seismic separator

TABLE 3. Specifications of concrete materials

Density (kg/m^3)	2400
Specified strength (MPa)	30
Modulus of elasticity (MPa)	26154
Modulus of crack elasticity (MPa)	13077
poisson's ratio	0.15

TABLE 4. Specifications of rubber materials

Poisson's ratio	**K_H (N/mm)**	**G (N/mm^2)**	**K (N/mm^2)**	**C_{10} (N/mm^2)**	**D_1 (mm^2/N)**
0.4990	58	0.4124	206.0555	0.2062	0.0097
0.4993	58	0.4124	294.4239	0.2062	0.0068
0.4995	58	0.4124	412.2484	0.2062	0.0049

As shown in Figure 9 and Table 5, the maximum force and dissipated energy decrease by decreasing the Poisson ratio, and the residual displacement increases. The model with Poisson's ratio of 0.4993 is in good agreement with the experimental results; therefore, Poisson's ratio of 0.4993 was used to model the rubbers.

Experimental diagrams were compared with the numerical model in Figures 10 to 13 as well as Tables 6 to 8. The absence of the columns on both sides of the wall in the numerical model is responsible for differences in results of numerical models compared to experiments. There is also a distance between the hole and the pins connecting the column to the frame. Such detail has not been simulated in numerical modelling and causes the negligible pinching phenomenon in cyclic behaviour.

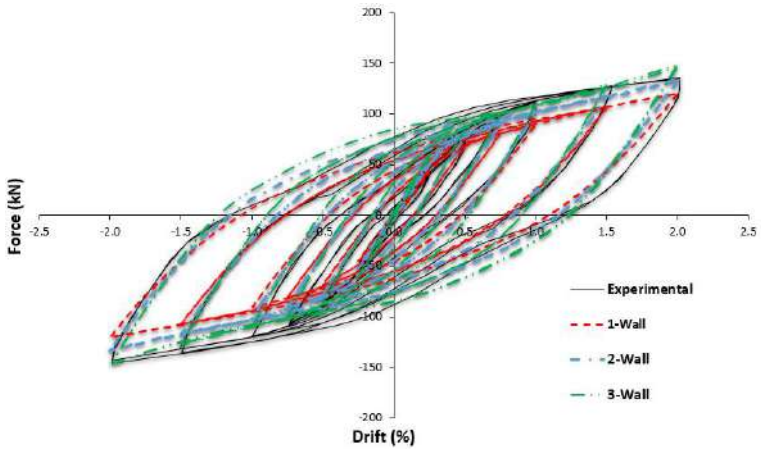


Figure 9. Comparison of force-drift diagram, experimental result vs. numerical finding

TABLE 5. Comparison of models with different Poisson ratios vs. the experimental results

Specimen	Maximum force (kN)	Maximum force difference (%)	Dissipated energy (kN.m)	Dissipated energy difference (%)	Residual displacement (mm)	Residual displacement difference (%)
Experimental	136	0	49	0	-31.59	0
1-Wall	121	-11	41.43	-15.45	-29.29	-7.28
2-Wall	134	-1.47	48.78	-0.45	-31.46	-0.41
3-Wall	149	9.5	56.84	16	-33.5	6.04

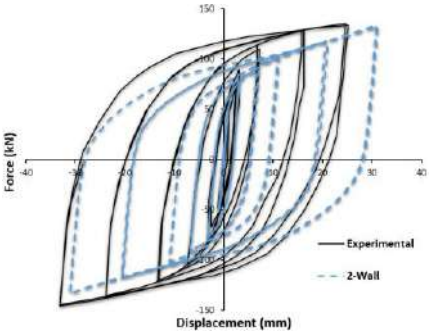


Figure 10. Dampers force-displacement diagram in 2-Wall specimen, Experimental via numerical results

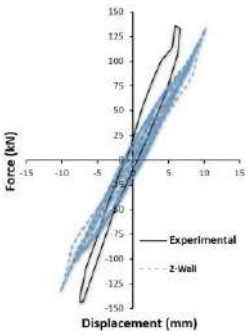


Figure 11. Wall force-displacement diagram in 2-Wall specimen, Experimental via numerical results force-displacement diagram in 2-Wall specimen, Experimental via numerical results

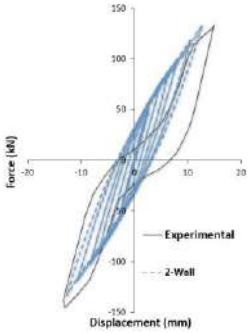


Figure 12. Wall's rigid body rotation diagram in 2-Wall specimen force-displacement diagram in 2-Wall specimen, Experimental via numerical results

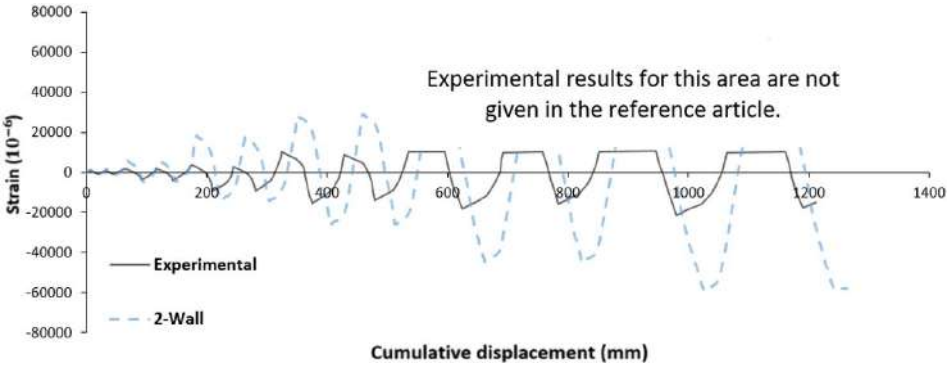


Figure 13. Comparison of damper cumulative displacement-strain diagram with numerical results in 2-Wall specimen

TABLE 6. Comparison of force-displacement results of 2-Wall specimen with the experimental model

Specimen	Maximum displacement in right side (mm)	Right maximum displacement difference (%)	Maximum displacement in left side (mm)	Left maximum displacement difference (%)
Experimental	25	0	-33	0
2-Wall	31	24	-31	-6.06

TABLE 7. Comparison of 2-Wall specimen wall deformation results with the experimental model

Specimen	Maximum displacement in right side (mm)	Right maximum displacement difference (%)	Maximum displacement in left side (mm)	Left maximum displacement difference (%)
Experimental	6.62	0	-7.38	0
2-Wall	10.32	56.89	-10.21	38.35

TABLE 8. Comparison of 2-Wall rigid body rotation result with the experimental model

Specimen	Maximum displacement in right side (mm)	Right maximum displacement difference (%)	Maximum displacement in left side (mm)	Left maximum displacement difference (%)
Experimental	14.90	0	-13.40	0
2-Wall	12.72	-14.63	-12.66	-5.52

Earthquake foreshock was taken into account in two stages. For the first step, 1%, 1.5%, and 2% of drift were applied to the specimens. 2% of drift was considered for the second stage as the main earthquake. Force-drift diagrams are presented in Figures 14 to 16.

The diagrams' difference is due to the seismic isolator's performance in the numerical model. However, the poor performance of the seismic isolator is maybe due to simplification in the simulation of elastic concrete behaviour and absence of columns in the numerical model; it has not been considered on both sides of the wall. Propagation of cracks in the concrete has caused pinching phenomena in cyclic response. There is also a difference between the hole and the pins connecting the column to the frame, which has led to differences in experimental results and numerical modelling. Figure 17 shows the seismic isolator deformation for 2% drift. Figure 18 shows the damper deformation compared to the experimental results in different drifts. As can be seen, in drift 2%, numerical deformation matches the experimental specimen. However, with increasing drift, there is a difference between the deformation of numerical models and corresponding experimental results due to high seismic isolator deformation in the numerical model compared to the experimental.

3. 4. Performance Evaluation of Shear Wall Equipped with a Damper, Seismic Isolator and Sma Rod In this study, the effect of SMA rod with angles of 30, 45, and 60 degrees on the performance of concrete shear wall equipped with steel cantilever damper and seismic isolator was also evaluated. Due to the buckling in the slender members, the rod was attached to the damper so that it was only stretched. Figure 19 reveals the SMA rod assembly with a 30-degree angle to the damper.

Figure 20 presents the force-drift diagram. Table 9 shows the comparison of the models. In the model with 2 rods at an angle of 30 degrees, the displacement of residual has the most significant decrease, and in the model with 2 rods at an angle of 60 degrees, the dissipated energy has the highest increase. The results also prove that by increasing the angle of the rod, the maximum force decreases, and the dissipated energy as well as the residual displacement increases.

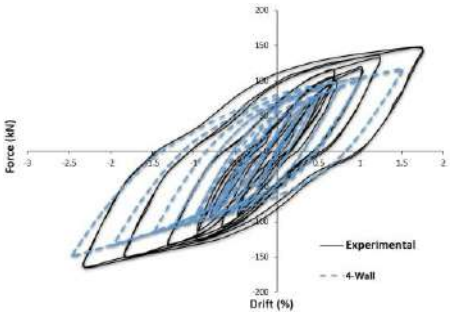


Figure 14. Comparison of force-drift diagram - experimental - via numerical model - 1% and 2% drift

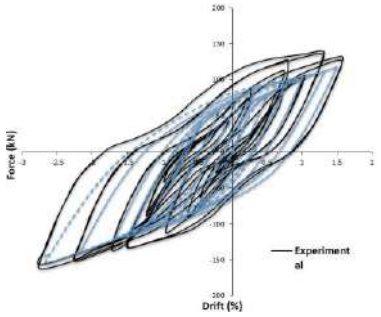


Figure 15. Comparison of force-drift diagram -experimental via numerical model - 1.5% and 2% drift

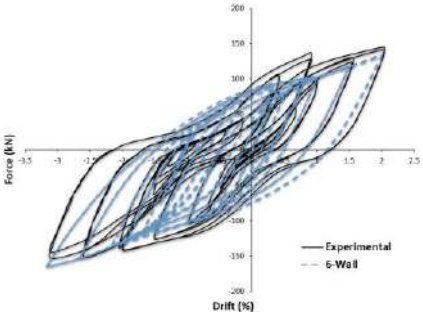


Figure 16. Force-drift diagram of the experimental via numerical model - 2% and 2% drift

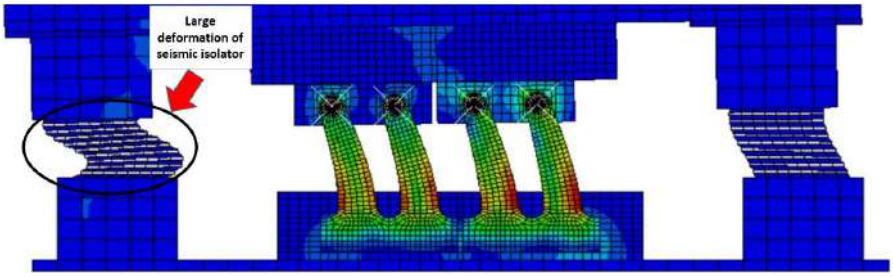


Figure 17. Deformation of the seismic isolator in the 6-Wall specimen

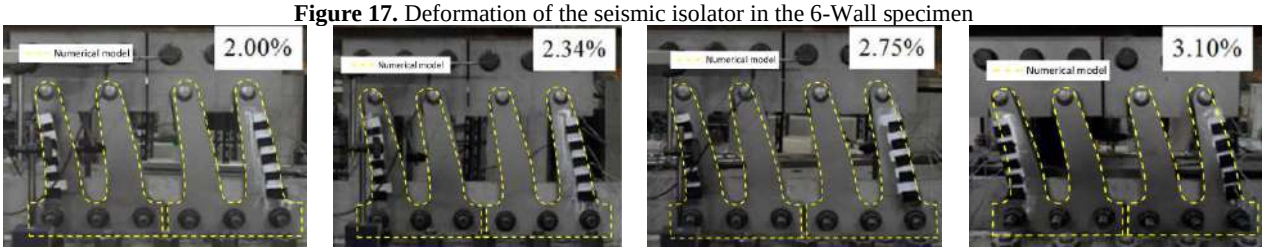


Figure 18. Comparison of experimental specimen deformation with numerical in 2% drift

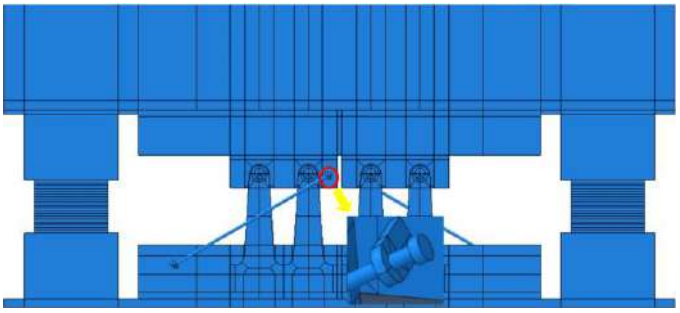
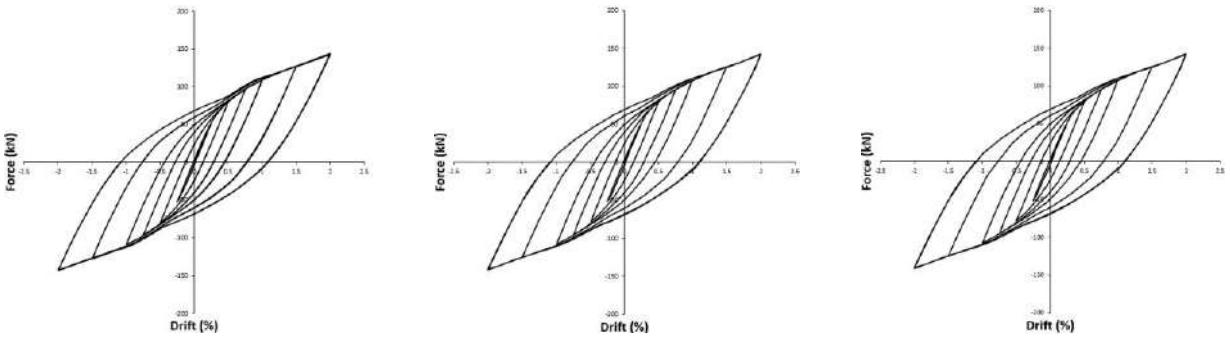


Figure 19. Specimen assembly 1-Wall-SMA



(a) specimen 1-Wall-SMA (b) specimen 2-Wall-SMA (c) specimen 3-Wall-SMA

Figure 20. Force-drift diagram of wall model equipped with steel plan damper, seismic isolator, and SMA rod

TABLE 9. Comparison of specimens with SMA rods

Specimen	Maximum force (kN)	Dissipated energy (kN.m)	Residual displacement (mm)
1-Wall-SMA	143.853	46.57	-29.28
2-Wall-SMA	142.342	47.06	-29.67
3-Wall-SMA	139.921	47.68	-30.22

Figures 21 to 23 compare a specimen's results with 2 SMA rods and a specimen without SMA rods. In the specimen with 2 SMA rods at an angle of 30 degrees compared to the specimen without SMA rods, the maximum force increases by 7.35%, the dissipated energy decreases by 4.53%, and the residual displacement decreases by 6.93%. In the specimen with 2 SMA rods at an angle of 45 degrees compared to the specimen without SMA rods, the maximum force increases by 6.23%, the dissipated energy decreases by 3.53%, and the residual displacement decreases by 5.69%. In the specimen with 2 SMA rods at an angle of 60 degrees compared to the specimen without SMA rods, the maximum force increases by 4.42%, the dissipated energy decreases by 2.26%, and the residual displacement decreases by 3.94%. The results exhibit that despite the SMA rod's positive effect, which leads to an increase in maximum force and a decrease in residual displacement, the dissipated energy decreases.

The effect of 4 SMA bars with an angle of 30 degrees on the structural system's behaviour was also examined. Figure 24 offers the assembly of 4 SMA bars with an angle of 30 degrees to the steel damper. Figure 25 demonstrates the hysteresis diagram for the specimen with 4 SMA bars, and in Figure 26 to 28**,** the results of the model without SMA bars and with 2 and 4 SMA bars are compared. Based on the findings, the maximum force

in the specimen with 4 SMA rods compared to the specimen with 2 SMA rods and without SMA rods increases by 6.38% and 14.20%, respectively, and the dissipated energy by 4.4% and 8.73%, respectively, and the residual displacement by It decreases by 7.31% and 13.73%.

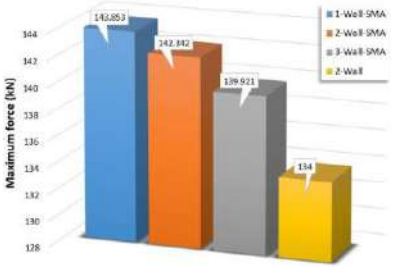


Figure 21. Comparison of maximum force for the specimen with SMA rod and without SMA rod

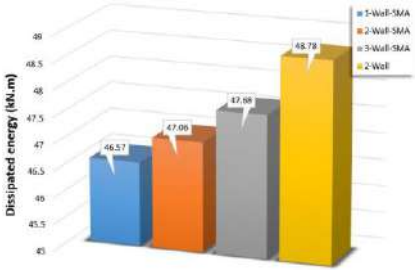


Figure 22. Comparison of dissipated energy for the specimen with SMA rod and without SMA rod

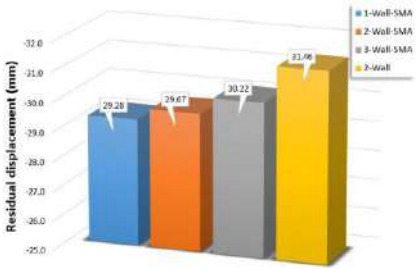


Figure 23. Comparison of residual displacement for the specimen with SMA rod and without SMA rod

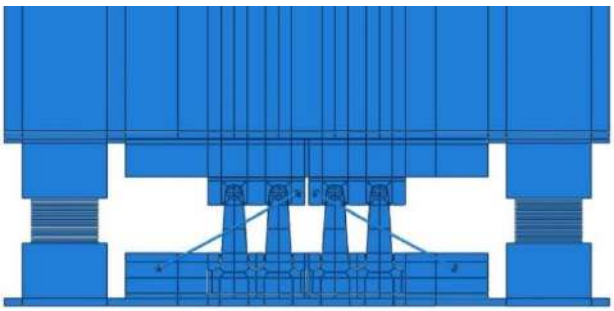


Figure 24. Specimen assembly 4-Wall-SMA

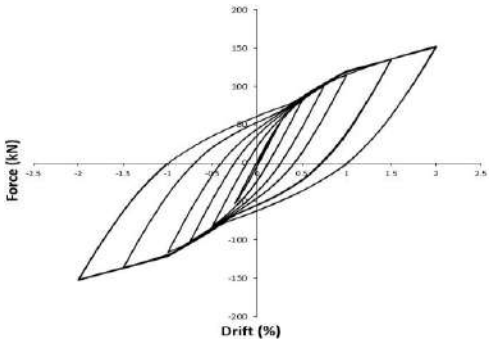


Figure 25. Force-drift diagram - 4-Wall-SMA

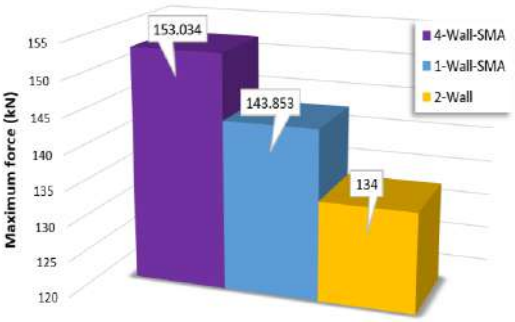


Figure 26. Comparison of maximum force for specimens with 2 and 4 SMA bars and without SMA bars

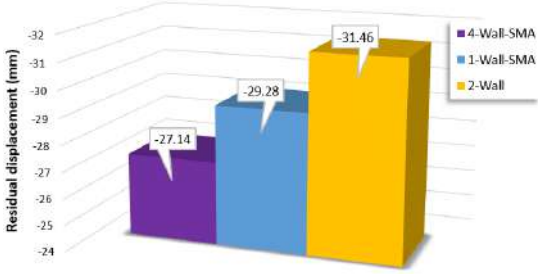


Figure 27. Comparison of residual displacements for specimens with 2 and 4 SMA bars and without SMA bars

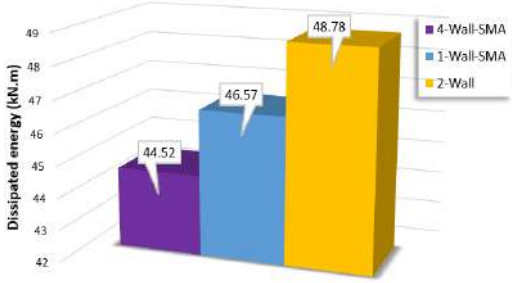


Figure 28. Comparison of dissipated energy for specimens with 2 and 4 SMA rods and without SMA rods

4. CONCLUSION

The most important findings of the present study are summarised as follows.

1. Steel dampers were modelled with isotropic, combined hardness, restrain and free in Y coordinate, respectively. Models with isotropic hardening revealed larger hysteresis loops than the experimental specimen. Moreover, in the specimens restrained in the Y direction, the force-displacement diagram had a steeper slope in the inelastic region than the experimental specimen. The model with combined hardening and release in the Y direction fits well with the experimental data.
2. Seismic isolator rubbers were modelled with Poisson's ratios of 0.4990, 0.4993, and 0.4995. The model with Poisson's ratio of 0.4993 compared well to the experimental specimen in maximum force and dissipated energy had a difference of -1.47 and -0.45%, respectively.
3. The effect of earthquake foreshock loading was simulated in two stages, including 1%, 1.5%, and 2% drift for the first and 2% drift for the second stage. The results demonstrate that the proposed model could not simulate these types of loads due to the poor performance of the seismic separator.
4. Three specimens were modelled with 30, 45, and 60-degree of SMA rod angles to investigate the SMA rod's efficiency. The results illustrated that with increasing rod angle, the maximum force decreases, and the dissipated energy and the residual displacement increased.
5. Comparing the results of specimens with SMA rods and without SMA rods showed that despite the positive effect of SMA rods to increase in the maximum force and a decrease in residual displacement, the dissipated energy had been decreased. The results also showed that in the specimen with four SMA rods compared to the sample with two SMA rods and without SMA rods, the maximum force increases by 6.38 and 14.20%, respectively, the dissipated energy by 4.4 and 8.73%, respectively, and the residual displacement decrease by 7.31 and 13.73%, respectively.

5. REFERENCES

1. Shojaeifar H., Maleki A., and Lotfollahi-Yaghin M.A., "Performance evaluation of curved-TADAS damper on seismic response of moment resisting steel frame", ***International Journal of Engineering, Transactions A: Basics***, Vol. 33, No. 1, (2020), 55-67, DOI: 10.5829/IJE.2020.33.01A.07.
2. DesRoches R. and Smith B., "Shape memory alloys in seismic resistant design and retrofit: a critical review of their potential and limitations", ***Journal of Earthquake Engineering***, Vol. 8, No. 3, (2004), 415-429, DOI: https://doi.org/10.1080/13632460409350495.
3. DesRoches R., McCormick J., and Delemont M., "Cyclic Properties of Superelastic Shape Memory Alloy Wires and Bars", ***Journal of Structural Engineering***, Vol. 130, No. 1, (2004), 38-46, DOI: 10.1061/(ASCE)0733-9445(2004)130:1(38).
4. Sayyaadi H. and Zakerzadeh M., "Nonlinear Analysis of a Flexible Beam Actuated by a Couple of Active SMA Wire Actuators", ***International Journal of Engineering, Transactions A: Basics***, Vol. 25, No. 3, (2012), 249-264, DOI: 10.5829/idosi.ije.2012.25.03a.07.
5. Kim Y., Ahn T., Kim H., and Jang D., "Development of new steel damper for seismic retrofit of existing structures", ***in Proceedings of 15th world conference on earthquake engineering***, (2012).
6. Lu X., Mao Y., Chen Y., Liu J., and Zhou Y., "New structural system for earthquake resilient design", ***Journal of Earthquake and Tsunami***, Vol. 7, No. 03, (2013), 1350013, DOI: https://doi.org/10.1142/S1793431113500139.
7. Kim Y.J., Ahn T.-S., Bae J.-H., and Oh S.-H., "Experimental study of using cantilever type steel plates for passive energy dissipation", ***International Journal of Steel Structures***, Vol. 16, No. 3, (2016), 959-974, DOI: https://doi.org/10.1007/s13296-014-0121-6.
8. Ahn H.-J., Kim Y.-J., Bae J.-H., and Jung I.-Y., "Cyclic loading test of wall damping system with steel dampers", ***Advances in Structural Engineering***, Vol. 19, No. 8, (2016), 1262-1274, DOI: 10.1177/1369433216642068.
9. Naeem A., Eldin M.N., Kim J., and Kim J., "Seismic performance evaluation of a structure retrofitted using steel slit dampers with shape memory alloy bars", ***International Journal of Steel Structures***, Vol. 17, No. 4, (2017), 1627-1638, DOI: 10.1007/s13296-017-1227-4.
10. Cortés-Puentes W.L. and Palermo D., "SMA tension brace for retrofitting concrete shear walls", ***Engineering Structures,*** Vol. 140, (2017), 177-188, DOI: https://doi.org/10.1016/j.engstruct.2017.02.045.
11. Liu Q. and Jiang H., "Experimental study on a new type of earthquake resilient shear wall", ***Earthquake Engineering & Structural Dynamics***, Vol. 46, No. 14, (2017), 2479-2497, DOI: 10.1002/eqe.2914.
12. Liu Q. and Jiang H., "Study on a new type of earthquake resilient shear wall", ***in Proceedings of the 16th World Conference on Earthquake Engineering***, (2017).
13. Chen C., Xiao R., Lu X., and Chen Y., "Study on the shear wall structure with combined form of replaceable devices", ***Advances in Structural Engineering***, Vol. 21, No. 9, (2018), 1327-1348, DOI: 10.1177/1369433217742525.
14. Cortés-Puentes W.L. and Palermo D., "Seismic Retrofit of Concrete Shear Walls with SMA Tension Braces", ***Journal of Structural Engineering***, Vol. 144, No. 2, (2018), 04017200, DOI: doi:10.1061/(ASCE)ST.1943-541X.0001936.
15. Wang B. and Zhu S., "Seismic behavior of self-centering reinforced concrete wall enabled by superelastic shape memory alloy bars", ***Bulletin of Earthquake Engineering***, Vol. 16, No. 1, (2018), 479-502, DOI: 10.1007/s10518-017-0213-8.
16. NourEldin M., Naeem A., and Kim J., "Life-cycle cost evaluation of steel structures retrofitted with steel slit damper and shape memory alloy–based hybrid damper", ***Advances in Structural Engineering***, Vol. 22, No. 1, (2019), 3-16, DOI: 10.1177/1369433218773487.
17. Wang W., Fang C., Feng W., Ricles J., Sause R., and Chen Y., "SMA-based low-damage solution for self-centering steel and composite beam-to-column connections", ***Journal of Structural Engineering***, Vol. 146, No. 6, (2020), 04020092, DOI: https://doi.org/10.1061/(ASCE)ST.1943-541X.0002649.

18. Falahian A., Asadi P., Riahi H.T., and Haghy M., "Seismic performance assessment of steel frames with shape-memory alloy wire-based dampers", ***The Structural Design of Tall and Special Buildings***, Vol. 29, No. 16, (2020), e1797, DOI: https://doi.org/10.1002/tal.1797.
19. Issa A. and Alam M.S., "Comparative seismic fragility assessment of buckling restrained and self-centering (friction spring and SMA) braced frames", ***Smart Materials and Structures***, Vol. 29, No. 5, (2020), 055029, DOI: https://doi.org/10.1088/1361-665X/ab7858.
20. Bogdanovic A., Rakicevic Z., and Noroozinejad Farsangi E., "Shake table tests and numerical investigation of a resilient damping device for seismic response control of building structures", ***Structural Control and Health Monitoring***, Vol. 26, No. 11, (2019), e2443, DOI: https://doi.org/10.1002/stc.2443.
21. Kamaludin P.N.C., Kassem M.M., Farsangi E.N., Nazri F.M., and Yamaguchi E., "Seismic resilience evaluation of RC-MRFs equipped with passive damping devices", ***Earthquakes and Structures***, Vol. 18, No. 3, (2020), 391-405, DOI: https://doi.org/10.12989/eas.2020.18.3.391
22. Alavi A., Dolatabadi M., Mashhadi J., and Farsangi E.N., "Simultaneous optimization approach for combined control–structural design versus the conventional sequential optimization method", ***Structural and Multidisciplinary Optimization***, Vol. 63, No. 3, (2021), 1367-1383, DOI: https://doi.org/10.1007/s00158-020-02765-3.
23. Fathizadeh S.F., Dehghani S., Yang T.Y., Noroozinejad Farsangi E., Vosoughi A.R., Hajirasouliha I., Takewaki I., Málaga-Chuquitaype C., and Varum H., "Trade-off Pareto optimum design of an innovative curved damper truss moment frame considering structural and non-structural objectives", ***Structures***, Vol. 28, (2020), 1338-1353, DOI: https://doi.org/10.1016/j.istruc.2020.09.060.
24. Aydin E., Ozturk B., Bogdanovic A., and Farsangi E.N., "Influence of soil-structure interaction (SSI) on optimal design of passive damping devices", ***Structures***, Vol. 28, (2020), 847-862, DOI: https://doi.org/10.1016/j.istruc.2020.09.028.
25. Barkhordari M.S. and Tehranizadeh M., "Ranking Passive Seismic Control Systems by Their Effectiveness in Reducing Responses of High-Rise Buildings with Concrete Shear Walls Using Multiple-Criteria Decision Making", ***International Journal of Engineering, Transactions B: Applications***, Vol. 33, No. 8, (2020), 1479-1490, DOI: 10.5829/ije.2020.33.08b.06.
26. Hosseinnejad H., Lotfollahi -Yaghin M., Hosseinzadeh Y., and Maleki A., "Numerical Investigation of Response of the Post-Tensioned Tapered Steel Beams with Shape Memory Alloy Tendons", ***International Journal of Engineering, Transactions A: Basics***, Vol. 34, No. 4, (2021), 782-792, DOI: 10.5829/ije.2021.34.04a.04.
27. Heydari. T A.M. and Gerami M., "Multi-stage Performance Upgrade of Steel Moment Frames by Post-tension Connections", ***International Journal of Engineering, Transactions B: Applications***, Vol. 34, No. 5, (2021), 1132-1144, DOI: 10.5829/ije.2021.34.05b.07.
28. Pourzangbar A., Vaezi M., Mousavi S.M., and Saber A., "Effects of Brace-viscous Damper System on the Dynamic Response of Steel Frames", ***International Journal of Engineering, Transactions B: Applications***, Vol. 33, No. 5, (2020), 720-731, DOI: 10.5829/ije.2020.33.05b.02.

Persian Abstract

چکیده

آلیاژ حافظه‌دار شکلی(SMA) دارای قابلیت منحصر به فرد رفتار ابر کشسان و حافظه شکلی است؛ این رفتار باعث می‌شود تا در کرنش‌های بزرگ، آلیاژ حافظه‌دار شکلی کرنش پسماندی در حدود صفر داشته باشد. در کنار این مزیت، رفتار تنش- کرنش آلیاژ حافظه‌دار شکلی در رفتار ابرکشسان شکل پرچم گونه دارد و قابلیت استهلاک انرژی زیادی نداشته و قابل جایگزین شدن با میراگر فلزی نمی باشد. برای جبران ضعف ذکر شده در میراگرهای فلزی، در تحقیق حاضر عملکرد کاربرد هم زمان این دو عضو در دیوار برشی بتن مسلح مورد بررسی قرار گرفته است و عملکرد میراگر ترکیبی (میراگر طره‌ای فولادی و میله SMA) به عنوان یک سیستم لرزه‌ای نوآورانه مورد بررسی قرار گرفته است. مدلسازی عددی در نرم افزار آباکوس انجام شده است. در این خصوص به صحت سنجی کامل ۳ مدل آزمایشگاهی پرداخته شده است. به منظور داشتن درک جامع از صحت نتایج مدلسازی عددی، در قسمت صحت سنجی تنها به مقایسه رفتار نیرو-تغییرمکان اکتفا نشده است وتاریخچه کرنشها و تغییر مکان های قسمت های مختلف سازه حاصل از مدل آزمایشگاهی با مدل عددی مورد مقایسه انتقادی قرار گرفته است. همچنین جنبه‌های گوناگون مدلسازی عددی اعم از فرضیات مدلسازی مصالح، مدل‌های رفتاری، المان‌های مورد استفاده و روش‌های حل در مقایسه با نتایج آزمایشگاهی مورد بحث و تفسیر قرار گرفته است. نتایج نشان دادند با افزایش زاویه میله SMA نیروی حداکثر کاهش و جابه‌جایی پسماند و انرژی مستهلک شده افزایش می‌یابد. علاوه براین، مقایسه نتایج نمونه بدون میله SMA و نمونه با میله SMA نشان دادند برخلاف تاثیر مثبت میله SMA که منجر به افزایش نیرو حداکثر و کاهش جابه‌جایی پسماند می‌شود، انرژی مستهلک شده کاهش می‌یابد.

IJE TRANSACTIONS A: Basics Vol. 34, No. 7, (July 2021) 1644-1652

International Journal of Engineering

Journal Homepage: www.ije.ir

Optimization of Dam Reservoir Operation Using Grey Wolf Optimization and Genetic Algorithms: A Case Study of Taleghan Dam

A. Davani Motlagh[a], M. S. Sadeghian*[a], A. H. Javid[b], M. Asgari[c]

[a] *Department of Civil Engineering, Central Tehran Branch, Islamic Azad University Tehran, Iran*
[b] *Department of Environmental Engineering, Science and Research Branch, Islamic Azad University, Tehran, Iran*
[c] *Department of Mathematics, Central Tehran Branch, Islamic Azad University, Tehran, Iran*

PAPER INFO

Paper history:
Received 25 January 2021
Received in revised form 10 April 2021
Accepted 12 April 2021

Keywords:
Optimization
Taleghan Dam
Grey Wolf Optimization Algorithm
Genetic Algorithm
WEAP Software

ABSTRACT

With the growth of population, shortage of water, and severe lack of water resources, optimization of reservoirs operation is a principal step in water resource planning and management. Therefore, in the present study, water was optimally allocated for a period from 2010 to 2020 using two simulation-optimization models based on Grey Wolf Optimization algorithm (GWO) and Genetic Algorithm (GA) and WEAP model. System operational indices including volumetric reliability, temporal reliability, vulnerability, and sustainability were used to evaluated the perforemance of optimization algorithms as well as WEAP model. The objective function of resources allocation problem was minimizing sum of the squared relative deficiencies for each month and maximizing reliability over the entire 11-year period. The results showed that optimal allocation solution found by the GWO algorithm with volumetric reliability, vulnerability, and sustainability indices which were 86.93, 0.29, and 21.48%, respectively was better and more suitable than the optimal allocation solution found by GA algorithm (which were 87.12, 0.41, and 21.34%, respectively). Finally, given an increase in the water demands , it is possible to obviate the needs of beneficiaries to an acceptable level and prevent severe draught in dry months through optimizing the use of available resources. According to the calculated indices for the WEAP model, in which volumetric reliability, vulnerability, and sustainability were equal to 87.46, 0.92, and 1.03%, respectively. It can be concluded that the use of optimization algorithm in optimal operation of the dam is more reliable than WEAP model.

***doi**: 10.5829/ije.2021.34.07a.09*

1. INTRODUCTION

Currently, given growing needs for water resources as a result of population growth, industry, and agriculture, it is not possible to plan only with variable and uncertain water resources. Therefore, construction of reservoirs to obviate shortage of water is an inevitable and definite matter. The uncontrolled increase in consumption and limited water resources will cause crises in the country in the very near future. In addition, Drought is an inevitable part of the world's climate. It occurs in wet as well as in dry regions. Therefore, planning for drought and mitigating its impacts is essential [1]. As a result, in addition to construction of the dam, operation of the reservoir dams should be done in such a way that the least deficiency occurs during operation period according to inflow to the dam, reservoir geometry, weather conditions, and type of consumption. Following construction of the dam, agricultural, industrial, and urban development programs and finally, structure of the basin water system will change. Changes in structure of the basin water system due to changes in the water supply system or water demand lead to changes in temporal and spatial conditions of the water system. Therefore, optimal operation of dam reservoirs is one of the essential issues in water resources management, especially surface water. Optimization allows for accurate mathematical modeling in a process, and as a result, we will be able to optimize our models using mathematical programming methods. Recently, approximation algorithms have shown

*Corresponding Author Institutional Email: m.sadeghian@iauctb.ac.ir (M. S. Sadeghian)

Please cite this article as: A. Davani Motlagh, M. S. Sadeghian, A. H. Javid, M. Asgari, Optimization of Dam Reservoir Operation Using Grey Wolf Optimization and Genetic Algorithms (A Case Study of Taleghan Dam), International Journal of Engineering, Transactions A: Basics Vol. 34, No. 7, (2021) 1644-1652

considerable ability to achieve optimal operation of dam reservoirs [2]. Sattari et al. [3] investigated efficiency of Alavian Dam reservoir system during three phases. They defined the objective function as maximizing the total water output required for agriculture usage. They showed that the calculated capacity was relatively correct during the preliminary studies and operation was relatively satisfactory during the study period. Lack of environmental flow of the river can be evaluated as a major weakness of this model [3]. Kougias and Theodossiou [4] investigated application of harmony search (HS) algorithm in planning of a four-reservoir dam system for irrigation and hydroelectric purposes. Their objective function was maximizing daily gain of the reservoir system for 2-hour period. They showed that the HS algorithm has the potential to optimize complex problems by comparing the results obtained from this method with other methods [4]. Mehta et al. [5] compared three scenarios including changing cultivation pattern according to economic factors, changing cultivation pattern for less water consumption and a combination of changes in the irrigation system, and changing cultivation pattern in the basin in California using water evaluation and planning (WEAP) model, to provide a solution for overcoming the effects of climate change. Guo [6] used the non-dominated sorting particle swarm optimizer (NSPSO) algorithm as a modified version of particle swarm optimization (PSO) to optimize utilization of multi-purpose reservoirs. They demonstrated good performance of the algorithm.

Pradhan and Tripathy [7] developed a model for optimal multi-purpose operation of the hydraulic reservoir in India, based on GA. Comparing the results of GA with the current policy showed the ability and effectiveness of GA [7]. Garousi-Nejad et al. [8] examined the efficiency of the FA algorithm relative to GA in reservoir operation for agricultural water supply and hydropower generation. The results show that the convergence velocity in FA is better than GA to reach the global optimum point and the value of the objective function. Sonaliy and Suryanarayana [9] used the GA for optimal utilization of Ukai reservoir in India. They showed that GA can fully satisfy the needs of downstream irrigation and minimize release of water leading to significant savings of water. Jian-Xia et al. [10] used genetic algorithm (GA) for optimal allocation of water from the reservoir. They investigated probabilistic sensitivity of GA operators, such as intersection and mutation. The results of GA showed that it could act as a suitable option in optimization problems. Ghadami et al. [11] developed a plan for optimal use of the multi-reservoir system in north of Khorasan for agricultural usage using GA. In this model, the most appropriate algorithm was determined for dam reservoir operation based on certain values of state variables including flow volume at the beginning of the year and river's water during agricultural season. Hamlat et al. [12] in a study using WEAP model, they examined and analyzed the existing water balance and the expected scenarios of water resources management in the western Algerian watershed in the future and considered the various policies in place and the parameters that might affect future demand by 2030, and showed that the needs of the domestic sector could be met by considering the expected scenarios Dehghan [13] in a study investigated allocation of water resources under management scenarios in the Gorganrood basin using the WEAP model. They showed that in the new planning of water resources allocation for the Gorganrood basin, the needs of Voshmgir Damṡ margin industry can be met by 9.5 million cubic meters by accepting 5% reduction of system reliability. Asadi et al. [14] presented a method using a multi-objective structure and utilizing new formulations, in which, instead of meeting 100% of the needs in some months, regardless of dry months, some water of the high-water months or seasons is stored in the reservoir to be used in low-water months for adjusting the failure rate. For this purpose, the multi-objective particle swarm optimization (MOPSO) algorithm was connected to the WEAP simulator model. Finally, the results were evaluated in three scenarios: status quo, land development, and system optimization scenarios. In the status quo scenario, optimal situation was reported in the whole period except for several months. In the land development scenario, in many dry years and in all the last six years of planning in most uses, percentage of supply was zero in 3-8 consecutive dry months and it was less than 5% in other low-water years. But, percentage of supply reached by 28-60 % in these months through implementation of the optimizer model [14].

2. METHODOLOGY

2. 1. Case Study Taleghan Dam was built in 135 km northwest of Tehran with a longitude of 50° 37′ to 51° 10′ and a latitude of 36° 5′ to36° 25′ (Figure 1). It is located on Taleghan River in Sefidrud catchment. This earth dam is made of pebbles with clay core and a crown of 1,111 m long and 109 m high from the foundation and has a useful volume of 320 million cubic meters and a dead volume of 91 million cubic meters. Tables 1 and 2 represent distribution of water inflow into the reservoir dam and downstream based on monthly needs, respectively. In this study, an 11-year period was used. For estimating volume of water evaporated from the lake as well as the volume of rainfall, the Equation (2) was used as follows [6]:

$$\mathrm{Loss_t} = \mathrm{A_t} \times (\mathrm{Ev_t} - \mathrm{R_t}) \quad (1)$$

$$A_t = a \times e^{(b \times S_t)} + c \times e^{(d \times S_t)} \quad (2)$$

where, A, S , Ev, and R are the lake area (km2) , the volume of water stored in the dam (MCM), evaporation in meters, and precipitation in meters respectively, and t denoted the simulation time step..a, b, c, and d are regression coeficients of volume-suface curve and equal to 8908000 and 10-10 × 8.79 and -8621000 and -9-10 × 6.026, respectively. Equation (1) was used to estimate the evaporation from lake surface. In Equation 1, we need the surface of lake which is a variable parameter. Using Equation (2) we determined the lake surface as a function of lake volume. It is worth to mention that the lake volume was determined by water mass balance in the lake. The only question is that what is the relation between volume and surface of the lake. Using available volume-elevation-surface data and Matlab curve-fitting toolbox, we invetigate a variety of model. We used least-square method to determine the coefficient of each regression model. The best model (minimm error) was found by the Equation (2). In other word, Equation (2) is a relationship between the the lake surface (as the dependent variable) and lake volume (as the independent variable).

Based on Tables 1 and 2, sum of annual water demand for drinking, agriculture, artificial recharge, and environment is equal to 456. 49 MCM.

2. 2. Grey Wolf Optimization Algorithm (GWO)

GWO [15] was inspired by life of grey wolves. They have a special interest in social hierarchy. The leaders of males and females are called as alpha. Alpha is mainly responsible for deciding on hunting, where to sleep, when to wake up, and so on. Alphaṡ decisions are dictated to the rest of the community showing that organization and discipline in an association are more important thanpower. Beta is at the second level in the grey wolf hierarchy. Beta wolves are follower wolves helping the alpha in decision-making or other association activities. The lowest rank in grey wolves belongs to the omega grey wolf. The omega plays the role of a sacrificial sheep. The omega wolf must always serve other dominant wolves. They are the last wolves in the community that are allowed to eat. If the wolf is neither alpha, nor beta, and omega, so it is called as delta. They must serve the alpha and beta wolves. Social hierarchy in the group of grey wolves is shown in Figure 2. The most important hunting phases in the grey wolveṡ association include the followings: 1. Pursuing 2. Reaching the hunt 3. Surrounding 4. Forming the attack position 5. Attacking for mathematically modeling this process, the Equations (3)-(6) are used as follow:

$$\vec{X}_{(t+1)} = \overrightarrow{Xp}_{(t)} - (\vec{A} \times \overrightarrow{D_t}) \quad (3)$$

where t is the number of iterations, $\vec{A}$ and $\vec{D}$ are the coefficient vectors, $\overrightarrow{Xp}$ is the position of the prey, and $\vec{X}$ is the position of the gray wolves.

$$\vec{D} = \left|(\vec{C} \times \overrightarrow{Xp}_t) - \vec{X}_t\right| \quad (4)$$

$$\vec{A} = a(2r_1 - 1) \quad (5)$$

$$\vec{C} = 2r_2 \quad (6)$$

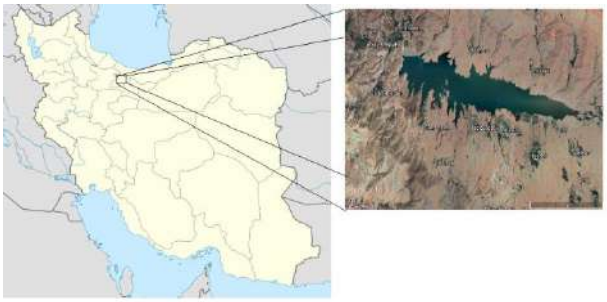


Figure 1. Location of the Taleghan Dam

TABLE 1. The average monthly water flow to Taleghan Dam (during 2010-2020) MCM

Statisticial Summary of inflow	April	May	June	July	August	September	October	November	December	February	January	March
Mean	76.80	102.10	73.62	29.50	13.71	9.89	8.66	16.25	13.05	11.49	13.73	30.17
Standard Deviation	25.10	35.46	37.58	12.42	4.75	4.36	1.86	8.28	5.11	3.70	4.59	9.56
Min	41.11	64.10	32.22	15.74	6.94	4.01	3.91	7.68	5.81	4.12	8.30	20.33
Max	129.21	178.81	174.41	57.48	21.41	18.94	10.82	33.92	22.14	16.53	23.96	46.19

TABLE 2. Water demand in different months in Taleghan Dam

Demand	April	May	June	July	August	September	October	November	December	February	January	March
Drinking	14.88	16.98	19.77	22.75	22.40	21.35	18.72	16.45	14.52	14.00	13.65	14.52
Agriculture	7.5	28.3	27.5	24.8	23.3	14.1	23.1	1.4	0	0	0	0
Artificial recharge	1.1	0	0	0		0	0	0	4.7	5.6	4.3	4.3
Environment	7.9	12	9.3	5.4	6.4	4.3	4.7	5.1	3.9	3.9	4.1	9
Sum	31.38	57.78	56.57	52.95	52.1	39.75	46.52	22.95	23.12	23.5	22.05	27.82

where, component "a" is linearly decreased from 2 to 0 during the repetition period, and r1 and $r2$ are the random vectors in the range of [0 and 1]. For mathematically modeling hunting behavior of the grey wolf, α (the best candidate solution), β (the second best candidate solution), and δ (the third best candidate solution) are used assuming that they have the best knowledge about prey position. So, the three best solutions obtained so far are kept and other search agents like omega are forced to update their position according to position of the best search agents. Equation (7) is used to update the position of the wolves:

$$\vec{X}_{(t+1)} = \frac{\overrightarrow{X1}_t + \overrightarrow{X2}_t + \overrightarrow{X3}_t}{3} \quad (7)$$

Here, X is the position of any wolf at iteration t+1.

X1 is the position of the alpha at iteration t, X2 is the position of the beta at iteration t, X3 is the position of the omega at iteration t

where, $\overrightarrow{X1}$, $\overrightarrow{X2}$, and $\overrightarrow{X3}$ were defined as follow:

$$\overrightarrow{X1}_t = \left|\overrightarrow{X\alpha}_t - (\overrightarrow{A1} \times \overrightarrow{D\alpha}_t)\right| \quad (8)$$

$$\overrightarrow{X2}_t = \left|\overrightarrow{X\beta}_t - (\overrightarrow{A2} \times \overrightarrow{D\beta}_t)\right| \quad (9)$$

$$\overrightarrow{X3}_t = \left|\overrightarrow{X\delta}_t - (\overrightarrow{A3} \times \overrightarrow{D\delta}_t)\right| \quad (10)$$

where, $\overrightarrow{X\alpha}$, $\overrightarrow{X\beta}$, and $\overrightarrow{X\delta}$ are the positions of the first three best solutions initeration t. $\overrightarrow{A1}$, $\overrightarrow{A2}$, and $\overrightarrow{A3}$ are introduced before (Equation (5)) and $\overrightarrow{D\alpha}$, $\overrightarrow{D\beta}$ and $\overrightarrow{D\delta}$ are defined as follow, respectively.

$$\overrightarrow{D\alpha}_t = \left|(\overrightarrow{C1} \times \overrightarrow{X\alpha}) - \vec{X}\right| \quad (11)$$

$$\overrightarrow{D\beta}_t = \left|(\overrightarrow{C2} \times \overrightarrow{X\beta}) - \vec{X}\right| \quad (12)$$

$$\overrightarrow{D\delta}_t = \left|(\overrightarrow{C3} \times \overrightarrow{X\delta}) - \vec{X}\right| \quad (13)$$

where $\overrightarrow{C1}$, $\overrightarrow{C2}$, and $\overrightarrow{C3}$ are introduced before (Equation (6)).

2. 3. Genetic Algorithm (GA) This algorithm was first proposed by Holland [16] and then, it was developed as a powerful optimization tool. This method is based on the Darwin's theory and conflict of survival, stating that always the creatures who are the most stable can survive. The GA starts from a set of initial random solutions called as population. Each population is made up of a set of chromosomes, each of which is a solution to the problem, and each chromosome is made up of a set of genes, or indeed problem-solving variables. Size of the population influences performance of GA so that, if the population is too small, due to not searching all the solution space, the algorithm may not converge to the desired solution, and if it is too large, although more space is searched, but convergence speed for optimal solution will be slow and execution time of the program will be longer. There are two types of operators in the GA: evolutionary operators, such as selection and genetic operators, such as crosses and mutations. Selection process is based on the degree of suitability of the objective functions corresponding to each chromosome in each generation, and the criterion for selecting chromosomes is based on their suitability.

2. 4. WEAP Software WEAP is a software tool used for integrated planning of water resources providing a comprehensive, flexible, and user-friendly framework for policy planning and analysis [17]. Many areas are exposed to heavy challenges of freshwater management. Allocation of the limited water resources has raised concerns about environmental quality, climate diversity planning, and uncertainty. In addition, the necessity for developing and implementing sustainable water usage strategies has increasingly imposed pressure on water resources policy-makers. The need node in WEAP depends on issues, such as water consumption patterns, equipment efficiency, reuse strategy, costs, and water allocation schemes. Furthermore, the supply side refers to issues, such as surface runoff, groundwater resources, reservoirs, and water transfer. WEAP is distinguished from natural simulation (e.g., need for evapotranspiration; runoff, base flow) and engineering components of the water system (e.g., reservoirs, groundwater pumping) by a comprehensive approach. This allows the designer to have access on a more comprehensive view of the wide range of factors that must be considered in water resources management for the current and future usage. The results obtained from

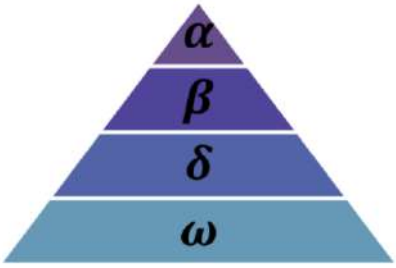


Figure 2. Social hierarchy in the group of grey wolves

WEAP are a very useful tool for assessing different options for water management and development. In this study, WEAP software was used to allocate the available water resource among different demand sites. In other words, the amount of supplied water to each demand site by SOP algorithm was calculated by WEAP software. Inputs data including river inflow, precipitation, evaporation, demands, and dam physical properties were collected and used as inputs of WEAP software. The time horizon of investigation was set to 2010-2020. The results of supplied water to each demand site were then exported to Excel and the reservoir performance indicators of SOP algorithm were calculated.

2. 5. Indicators of Reservoir Performance For checking performance of GA and GWO algorithms, indicators of reliability, vulnerability, and sustainability were used. The index of reliability offers the possibility that the system has normal operation (no failure) during its performance period. This index can be defined in two volume and time forms as follows:

2. 6. Volumetric Reliability Volumetric Reliability is referred to the ratio of volume of water released during the whole period t (Re_t) to the amount of water needed in downstream of the reservoir (D_{et}) and is obtained based on the Equation (14) as follows:

$$\delta_v = 100 \times (\frac{Re_t}{De_t}) \quad (14)$$

2. 7. Time Reliability Time Reliability is referred to a percent of time, at which the reservoir is able to meet the demand (facing no failure) and calculated as follows (Equation (15)):

$$\delta_t = 100 \times \left[1 - \frac{NDef}{T}\right] \qquad NDef = Number\ (\alpha \times Det > Ret) \quad (15)$$

where, N_{Def} and T are the number of periods facing the failure and the total operating periods, respectively. α is the confident coefficient indicating how much failure is acceptable in reliability calculation. In this study, α=0.9 was used. It means all deficits lower than 10 percent are considered as full allocation. Generally, the confident coefficient for domestic use is 0.95, for environmental needs is 0.9, and for agriculture is 0.85. In this study, because we investigate the total allocation as a whole, we used α=0.9 which can be considered as an average value for α.

2. 8. Vulnerability Index Vulnerability Index represents the extent of system failures and is obtained using the Equation (16) as follows [18]:

$$\eta = \max_{t=1,2,\dots,T} (\frac{De_t - Re_t}{De_t}) \quad (16)$$

where, De_t and Re_t represent the required and released water volumes in t-th period, respectively and T shows the entire number of periods of operation.

2. 9. Sustainability Loucks [19] introduced sustainability index, φ, as follows (Equation (17)):

$$\Phi = \delta\ \gamma\ (1 - \eta) \quad (17)$$

where, δ, η, and γ are Reliability and Vulnerability, and Resilience, respectively. Resilience is calculated by the Equation (18) stated as follows:

$$\gamma = 1/f/fs \quad (18)$$

where, fs is the number of failure periods continually, and f is the number of entire time periods.

2. 10. Mathematical Model of Optimal Operation of Reservoir In this study, The objective function of resources allocation problem was minimizing sum of the squared relative deficiencies for each month (Modified vulnerability index) and maximizing reliability.

the objective function and constraints are defined as follows:

$$min\ F = \frac{1}{T}\sum_{t=1}^{T}\left(\frac{R_t - D_t}{D_t}\right)^2 + \frac{1}{\delta_t} + P \quad (19)$$

Subject to:

$$S_{t+1} = S_t + Q_t + Loss_t + Re_t + Spill_t \quad (20)$$

$$R_{min} \le R_t \le R_{max} \quad (21)$$

$$S_{min} \le S_t \le S_{max} \quad (22)$$

$$P = \begin{cases} \sum_{t=1}^{T} \frac{S_{min} - S_t}{S_{min}} & if\ (S_t < S_{min}) \\ \sum_{t=1}^{T} \frac{S_t - S_{max}}{S_{max}} & If\ (S_t < S_{max}) \end{cases} \quad (23)$$

where, R_t is the volume of the released water, D_t is the amount of the required water, δ_t is the time reliability, S_{t+1} is the volume of the stored water in the reservoir in the next period, S_t is the volume of the stored water in the reservoir in the current period, and R_{min} and R_{max}

represent the minimum and maximum released water from reservoir, respectively. S_{min} and S_{max} are the minimum and maximum volumes of the stored water, respectively; P is the penalty function related to the reservoir volume, Q_t is the volume of water inflow to the reservoir. $Loss_t$ refers to the amount of reservoir loss. $Spill_t$ presents the volume of water overflow, and t is the number of period.

3. RESULTS AND DISCUSSION

In order to obtain a reliable reults, two population size (number of wolves), were used to optimize the problem. Different performance indices were tested with population size of 200 - 500 with 2500 iterations. The results are shown in Tables 3 and 4.

For considering different modes that can be used in GA operator of roulette wheel selection, gene mutation operator with probability between 0.2 - 0.3 and cross – over operator with probability between 0.8-0.9 were evaluated in program execution.Values of the objective function and reservoir performance indicators are given in Table 4 for the whole operation period (132 months).

Values of the objective function and reservoir performance indices are given in Table 5 for the whole statistical period (132 months) and the iteration of 2500.

TABLE 3. Results of GWO algorithm

GWO algorithm		
Objective function	**Iteration**	**Population**
0.040879	2500	200
0.03306	2500	500

TABLE 4. The results of applying GA

GA algorithm					Number
Objective function	**pc**	**pm**	**Iteration**	**Population**	
0.041035	0.8	0.2	2500	200	1
0.040612	0.9	0.3	2500	200	2
0.0392	0.8	0.2	2500	500	3
0.03886	0.9	0.3	2500	500	4
0.040641	0.8	0.3	2500	200	5
0.039974	0.8	0.3	2500	500	6

According to this table, the volumetric reliability of the best solution found by GWO with 86.93 is close to the best GA solution with 87.12 while GWO minimizes the objective function with 0.03306 better than GA with 0.03886. According to the objective function (minimize vulnerability index and maximize time reliability index), the vulnerability index of the best solution found by GWO with 0.2903 is highly lower than best solution found by GA with 0.4131 which means GWO is highly better than GA. In addition, time reliability of the best solution found by GWO with 68.93 is better than the best solution found by GA with 66.66 which means GWO is better than GA. also sustainability index (which is an overall index, see Equation (17)) of GWO with 21.48 is higher than GA with 21.34 and confirms the better performance of the GWO.

In Figure 3, as can be seen, although performance of both methods was acceptable and they were able to meet the required demand in downstream of dam with good accuracy, GWO has better performance than GA. As can be seen in Figure 3, the optimum solution found by GA, has more severe droughts than the optimum solution found by GWO.

Figure 4 shows the average annual water shortage (demand minus release) of Taleghan Dam obtained from GWO and GA algorithms. According to Figure 4, severity of shortages in GA is higher than GWO. The maximum amount of dam volume shortage in GA and GWO is equal to 7.89 and 6.77 MCM, respectively. Low amount of shortages in GWO indicates good performance of this algorithm. According to Table 6, it can be concluded that although WEAP model has more suitable performance than GWO algorithm regarding its time reliability index, volumetric reliability index, and shortage, but higher value of the objective function and more vulnerability and less sustainability than GWO algorithm are reasons for poor performance of WEAP.

Figure 5 shows the average monthly water release in two models. Results of applying models showed that during the 11-year period and according to the two models of WEAP and GWO, the reservoir can adjust 399.14 and 391.64 million meter cubic water annually for release on average. Based on Figure 5, the average release of water in April, May, June, and July is more in the WEAP model than GWO model. The condition is completely opposite in August, September, and October, revealing that the GWO model provides more supplies in

TABLE 5. Objective function values and reservoir performance indices in operation of the Taleghan Dam reservoir using the GWO and GA algorithms

Sustainability (%)	Vulnerability(%)	Time reliability α(%)=0.9	Volumetric reliability(%)	Total Deficit (MCM)	Objective function	Model
21.48	0.2903	68.93	86.93	655.92	0.03306	GWO
21.34	0.4131	66.66	87.12	646.377	0.03886	GA

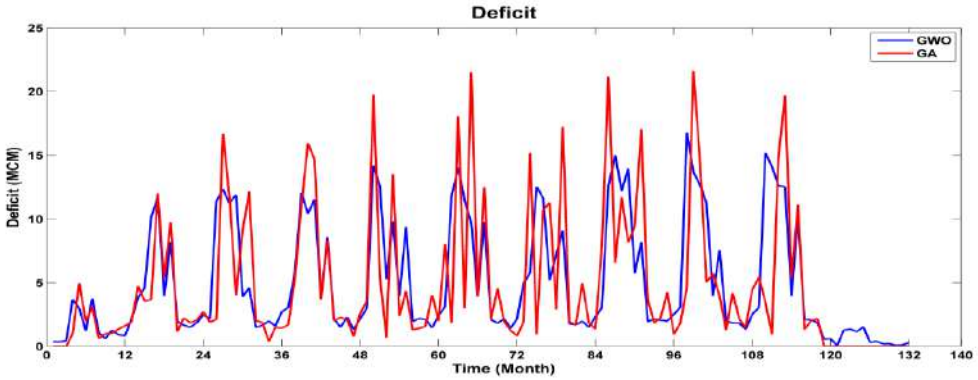


Figure 3. Release value through GWO and GA algorithms

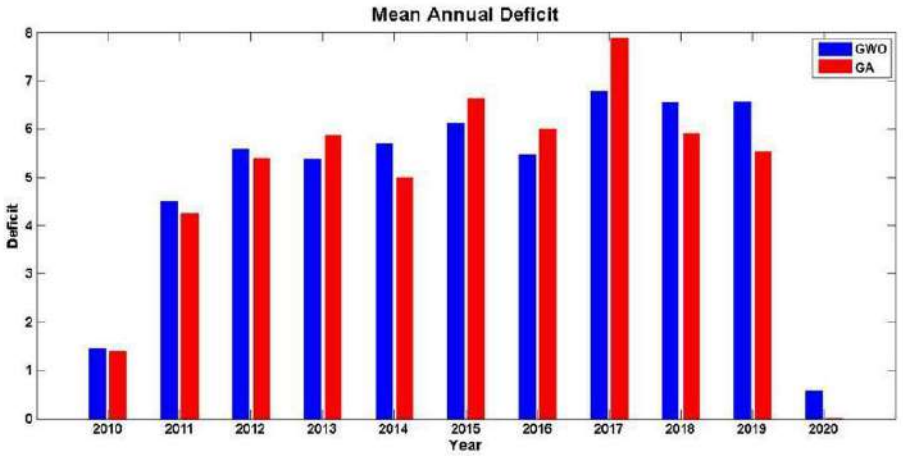


Figure4. Mean values of annual water deficit obtained from GWO and GA algorithms for Taleghan Dam (2010-2020)

TABLE 6. Objective function values and reservoir performance indices in operation of the Taleghan Dam reservoir using the GWO algorithm and WEAP model

Sustainabilit (%)	Vulnerability(%)	Time reliability α=0.9 (%)	Volumetric reliability (%)	Shortage (MCM)	Objective function	Model
21.48	0.29	68.93	86.93	655.92	0.03306	GWO
1.03	0.92	71.9	87.46	627.97	0.09	WEAP

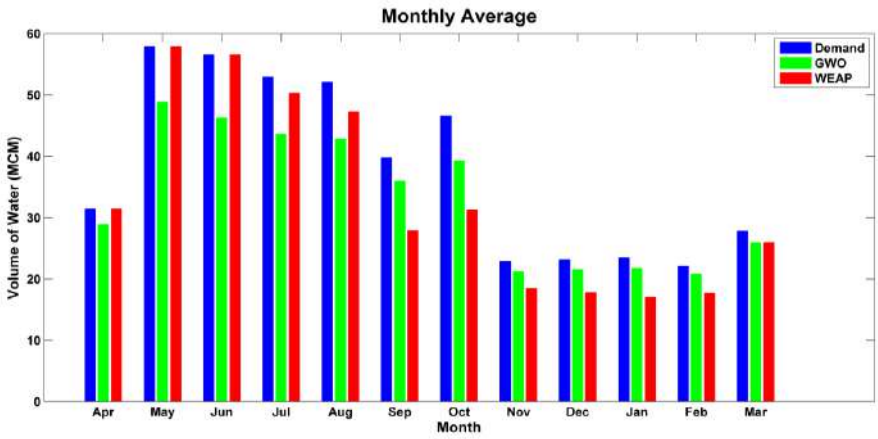


Figure 5. Average monthly release in WEAP and GWO and monthly demand of the dam

August, September, and October by saving water in April, May, June, and July. Using the operating procedure of less release of water in these months, GWO causes more water reservation in the reservoir, which in turn causes uniform distribution of water deficit in these months and decreases severity of water deficit in the critical months. In WEAP model (SOP), the aim is to supply 100% of monthly demands. Based on this policy in dry periods, by consumption of the available water in the reservoir, the elevation of the reservoir decreases and as a result, the amount of evaporation is decreased in comparison by optimal solution found by GWO. Therefore, water deficit in WEAP model decreases in comparison by GWO, and consequently, the volumetric reliability obtained by WEAP model is better than the GWO's. On the other hand, due to the less water available in the reservoir, the obtained vulnerability of the system in WEAP model is considerably more than GWO's.

4. CONCLUSION

Meta-heuristic optimization algorithms have been extensively used by scientists and engineers to optimize water resource system during last two decades. In the present study, first, performance of the GWO algorithm was investigated in compared to the GA and the GWO algorithm was selected as appropriate method. Then, the GWO algorithm was compared with WEAP model and results showed that in the GWO algorithm, always a desirable amount of water would be stored in the reservoir and used in each period faces acceptable amount of water deficit and as a result, preventing severe shortages in drought events. The results of the present study are consistent with the results of other study. Vulnerability index was obtained as 0.29 and 0.92 in GWO algorithm and WEAP model, respectively. Also, sustainability of the system was higher in the GWO algorithm with 21.48 than the WEAP model with 1.03. Optimal usage of water in condition where the recent droughts have caused water deficit in the country naturally and also, irregular usage of underground waters has caused many worries; accordingly, it is suggested to use the water resources in the reservoirs scientifically.

5. REFERENCES

1. Dariane, A.B., "Reservoir operation during droughts", ***International Journal of Engineering, Transactions B: Applications***, Vol. 16, No. 3, (2003), doi: 10.1201/9781315226774.
2. Othman, F., Sadeghian, M. and Azad, I., "Investigate the potential and limitations of meta-heuristics algorithms applied in reservoir operation systems", in 6th International Symposium on Advance Science and Technology, Kuala Lumpur, Malaysia. (2012).
3. Sattari, M.T., Apaydin, H. and Ozturk, F., "Operation analysis of eleviyan irrigation reservoir dam by optimization and stochastic simulation", ***Stochastic Environmental Research Risk Assessment***, Vol. 23, No. 8, (2009), 1187-1201, doi: 10.1007/s00477-008-0292-9
4. Kougias, I.P. and Theodossiou, N., "Application of the harmony search optimization algorithm for the solution of the multiple dam system scheduling", ***Optimization Engineering***, Vol. 14, No. 2, (2013), 331-344, doi: 10.1007/s11081-011-9183-x.
5. Mehta, V.K., Haden, V.R., Joyce, B.A., Purkey, D.R. and Jackson, L.E., "Irrigation demand and supply, given projections of climate and land-use change, in yolo county, california", ***Agricultural Water Management***, Vol. 117, (2013), 70-82, doi. http://dx.doi.org/10.1016/j.agwat.2012.10.021
6. Guo, X., Hu, T., Wu, C., Zhang, T. and Lv, Y., "Multi-objective optimization of the proposed multi-reservoir operating policy using improved nspso", ***Water Resources Management***, Vol. 27, No. 7, (2013), 2137-2153, doi: 10.1029/2010WR009166
7. Pradhan, N. and Tripathy, K., "Optimization of the operating policy of the multipurpose hirakud reservoir by genetic algorithm", ***American Journal of Engineering Research***, Vol. 2, No. 11, (2013), 260-266, doi.
8. Garousi-Nejad, I., Bozorg-Haddad, O., Loáiciga, H.A. and Mariño, M.A., "Application of the firefly algorithm to optimal operation of reservoirs with the purpose of irrigation supply and hydropower production", ***Journal of Irrigation Drainage Engineering***, Vol. 142, No. 10, (2016), 04016041, doi: 10.1061/(ASCE)IR.1943-4774.0001064.
9. Sonaliya, S. and Suryanarayana, T., "Optimal reservoir operation using genetic algorithm: A case study of ukai reservoir project", ***International Journal of Innovative Research in Science, Engineering Technology***, Vol. 3, No. 6, (2014).
10. Jian-Xia, C., Qiang, H. and Yi-min, W., "Genetic algorithms for optimal reservoir dispatching", ***Water resources management***, Vol. 19, No. 4, (2005), 321-331, doi: 10.1007/s11269-005-3018-5.
11. Ghadami, S., Ghahraman, B., Sharifi, M. and RAJABI, M.H., "Optimization of multireservoir water resources systems operation using genetic algorithm", ***Iran-Water Resources Research Journal,*** (2009), doi: 10.3390/su10051660
12. Hamlat, A., Errih, M. and Guidoum, A., "Simulation of water resources management scenarios in western algeria watersheds using weap model", ***Arabian Journal of Geosciences***, Vol. 6, No. 7, (2013), 2225-2236, doi: 10.1007/s12517-012-0539-0.
13. Dehghan, Z., Delbari, M. and Mohammadrezapour, O., "Planning water resources allocation under various managerial scenarios in gorganroud basin", ***Water Soil Science***, Vol. 25, No. 3, (2015), 117-132.
14. Asadi, M., Akhondali, A. and Azari, A., "Optimal operation of water resources systems by using mopso multi-objective algorithm", ***Iranian Journal of Soil Water Research***, Vol. 48, No. 4, (2017), 701-714.
15. Mirjalili, S., Mirjalili, S.M. and Lewis, A., "Grey wolf optimizer", ***Advances in Engineering Software***, Vol. 69, (2014), 46-61, doi.
16. Holland John, H., "Adaptation in natural and artificial systems", Ann Arbor: University of Michigan Press, (1975).
17. Sieber, J., "Weap water evaluation and planning system", Vol., (2006).
18. Hashimoto, T., Stedinger, J.R. and Loucks, D.P., "Reliability, resiliency, and vulnerability criteria for water resource system performance evaluation", ***Water Resources Research***, Vol. 18, No. 1, (1982), 14-20.
19. Loucks, D.P., "Quantifying trends in system sustainability", ***Hydrological Sciences Journal***, Vol. 42, No. 4, (1997), 513-530.

Persian Abstract

چکیده

با افزایش جمعیت، کمبود آب و کمبود شدید منابع آبی، بهینه سازی عملکرد مخازن یک گام اصلی در برنامه ریزی و مدیریت منابع آب است. در مطالعه حاضر، با استفاده از دو مدل بهینه سازی – شبیه سازی بر اساس الگوریتم بهینه سازی گرگ خاکستری (GWO) و الگوریتم ژنتیک (GA) و مدل WEAP ، منابع آب برای یک دوره آماری از سال ۲۰۱۰ تا سال ۲۰۲۰ تخصیص داده شد. از شاخص های عملکرد سیستم شامل قابلیت اطمینان حجمی، قابلیت اطمینان زمانی، آسیب پذیری و پایداری برای ارزیابی الگوریتم های بهینه سازی و همچنین مدل WEAP استفاده شد. تابع هدف مساله تخصیص منابع آب، به حداقل رساندن مجموع کمبودهای نسبی مربع برای هر ماه و به حداکثر رساندن قابلیت اطمینان در کل دوره ۱۱ ساله بود. نتایج نشان داد که راه حل تخصیص بهینه توسط الگوریتم GWO با شاخص های قابلیت اطمینان حجمی ، آسیب پذیری و پایداری به ترتیب با ۸۶.۹۳ ، ۰.۲۹ ، و ۲۱.۴۸ بهتر از راه حل تخصیص بهینه توسط الگوریتم GA با ۸۷.۱۲ ، ۰/۴۱ و ۲۱/۳۴ درصد می باشد. در نهایت، با توجه به افزایش تقاضای آب، می توان نیازهای ذینفعان را تا حد قابل قبولی کاهش داد و از طریق بهینه سازی استفاده از منابع موجود ، از خشکسالی شدید در ماه های خشک جلوگیری کرد. با توجه به شاخص های محاسبه شده برای مدل WEAP، که در آن قابلیت اطمینان حجمی، آسیب پذیری و پایداری به ترتیب برابر با ۸۷/۴۶ ، ۰/۹۲ و ۱/۰۳ درصد می باشد، می توان نتیجه گرفت که استفاده از الگوریتم بهینه سازی در عملکرد بهینه سد مناسب تر است از مدل WEAP می باشد.

IJE TRANSACTIONS A: Basics Vol. 34, No. 7, (July 2021) 1653-1666

International Journal of Engineering

Journal Homepage: www.ije.ir

Evaluation of Combined Use of Waste Paper Sludge Ash and Nanomaterials on Mechanical Properties and Durability of High Strength Concretes

F. Delaram[a], Y. Mohammadi*[b], M. R. Adlparvar[c]

[a] *Department of Civil Engineering, Qeshm Branch, Islamic Azad University, Qeshm, Iran*
[b] *Department of Civil Engineering, University of Mohaghegh Ardabili, Ardabil, Iran*
[c] *Department of Civil Engineering, University of Qom, Qom, Iran*

PAPER INFO

Paper history:
Received 03 April 2021
Received in revised form 12 May 2021
Accepted 15 May 2021

Keywords:
Aluminum Oxide Nanoparticles
Durability Properties
High Strength Concrete
Mechanical Properties
Silica Nanoparticles
Waste Paper Sludge Ash

ABSTRACT

The paper industry burns or buries a significant amount of waste paper sludge. This issue is not suitable from environmental and economic aspects. In this study, the mechanical and durability properties of high-strength concrete containing waste paper sludge ash (WPSA) were evaluated. The variables were WPSA (0, 5, 10, and 15% by weight of cement), silica nanoparticles (0 and 2.5 by weight of cement), and aluminum oxide nanoparticles (0 and 2.5 by weight of cement). Compressive strength, splitting tensile strength, flexural strength, and ultrasonic pulse velocity tests were conducted to evaluate the mechanical properties. The durability properties were also investigated using water penetration depth, water absorption, and electrical resistivity tests. The microstructure of the specimens was analyzed by preparing electron microscopic images. The combined effect of WPSA and nanoparticles on improving the mechanical and durability properties of high-strength concrete are better than using each of them alone. WPSA and nanoparticles react with calcium hydroxide formed due to cement's hydration, and silica produces hydrated calcium, which is the hard material that makes concrete strength. Consumption of calcium hydroxide and production of more hydrated calcium silicate in the presence of nanoparticles and WPSA are among the reasons for water absorption reduction, increased electrical resistance, and water penetration depth reduction in concrete specimens. By replacing part of the cement with WPSA, silica nanoparticles, and aluminum oxide nanoparticles, the transition zone between the aggregates strengthens, and the tensile and flexural strengths increased.

***doi**: 10.5829/ije.2021.34.07a.10*

1. INTRODUCTION

Extensive research showed that some paper industry waste can be used as raw materials in the construction industry [1-4]. Paper waste is obtained from an initial mechanical separation process and turns ash when burned [5-8]. In other words, waste paper sludge ash (WPSA) is a waste produced by the paper industry. These materials are produced when ink and pulp are burned to reduce their volume and produce energy [9]. According to EU standards, the combustion process usually occurs on a molten bed at 850 to 1100 degrees Celsius. The UK has 125,000 tonnes of waste paper ash per year, 70% is used for useless purposes such as spreading it on land to be recycled, and the remaining 30% is routinely transported to landfills [10]. Proper resource efficiency, reduced landfilling, and beneficial reuse are among the reasons why WPSA research has been promoted. WPSA has cement properties and reacts with water. However, due to its high porosity, it needs a lot of water. However, the combination of WPSA paper with pozzolanic materials can improve building materials' properties (including concrete). This combination can reduce harmful effects, such as reducing concrete strength [11, 12].

Past research on WPSA has focused on hydraulic

*Corresponding Author Email: Yaghoubm@uma.ac.ir (Y. Mohammadi)

Please cite this article as: F. Delaram, Y. Mohammadi, M. R. Adlparvar, Evaluation of the Combined Use of Waste Paper Sludge Ash and Nanomaterials on Mechanical Properties and Durability of High Strength Concretes, International Journal of Engineering, Transactions A: Basics Vol. 34, No. 7, (2021) 1653-1666

properties, reactivity and their potential use as cementitious materials. For example, Pera and Amrouz [13] showed that WPSA could be converted to sticky calcareous at 700 to 750 ° C. Later, other studies confirmed such finding [11, 13]. Sutcu and Akkurt [14] investigated porous ceramics' production from a mixture of waste from the paper-making process and three different clays. The results showed that clay soils, aluminum silicate, alkaline clay, and refractory clay could produce refractory bricks [14]. Ismail et al. [15] investigated the manufacture of bricks from paper sludge and fuel oil ash. Although the strength of the prepared brick sample was slightly higher than the minimum required according to BS 6073, the brick sample density prepared from paper sludge and fuel oil ash showed a reduction of 26.10% compared to conventional bricks, which is an important issue [15].

Fava et al. [8] used WPSA as a supplement for cementitious materials and evaluated its application in concrete and building materials as adhesive materials along with cement. It was concluded that if WPSA is used in 10% of Portland cement, it will positively affect mortars' mechanical performance [8]. Martinez et al. [16] investigated bricks' production using two types of waste from the paper industry, including WPSA and waste from cleaning pulp. According to the results, increasing the porosity of the samples reduced the compressive strength [16]. Ahmad et al. [17] examined concrete containing WPSA as a substitute for cement. Mechanical properties, water absorption, and dry density at 28 days were examined. WPSA at a dose of 5% could be effective [17]. Wong et al. [9] examined hydrophobic concrete containing WPSA. The results showed that using 12% WPSA instead of cement without harmful effects on strength, hydration, and concrete density reduces water absorption, water absorption rate, and concrete electrical conductivity by 84, 86, and 85%, respectively [9].

Pourazar et al. [18] investigated concretes' properties containing waste paper ash on concretes containing silica fume. The results showed that the ratio of compressive strength to the tensile strength of concretes containing waste paper ash is 13.55 to 17.79, which is approximately ten times the ratio of ordinary concrete [18]. Marvroulidou and Shah [19] investigated the combined use of WPSA and active alkaline cement in concrete. The dose of activators, processing conditions, and processing time was studied. The results showed that mixing with WPSA in the activating system causes early concrete strength at ambient temperature and sufficient strength [19].

Chemicals penetrate the steel through the concrete's pores, causing corrosion, cracking, and eventually concrete failure. Chemical reactions of concrete take place at the nanoscale. The nanostructure of concrete deserves more attention to make concrete with dense nanostructure, minimum size, and pores, which affects concrete durability [20-22]. Recently, nanotechnology is used to reduce the permeability and concrete seepage [23-25]. In this research, the effect of WPSA in combination with silica nanoparticles and aluminum oxide nanoparticles is investigated.

Faez et al. [26] evaluated fresh and hardened self-compacting concrete properties containing aluminum oxide nanoparticles and silica fume. The results showed that using 2.5% aluminum oxide nanoparticles as a substitute for cement increased the compressive strength of 90 days by 86% [26]. Ghanbari et al. [27] evaluated the effect of nanosilica on the attributes of lightweight concrete containing zeolite and glass fibers. The results showed that the combined use of silica nanoparticles and glass fibers significantly affects self-compacting lightweight concrete containing zeolite and scoria aggregates; Depending on the dose of glass fibers and silica nanoparticles, the tensile strength has increased from about 3 to 56% [27]. Heidarzad et al. [28] evaluated the effects of aluminum oxide nanoparticles on the performance of concrete containing fibers. The results showed that the addition of aluminum oxide nanoparticles to specimens with and without glass fibers significantly increased the electrical resistance. Aluminum oxide nanoparticles prevented more ions in the concrete and reduced concrete corrosion by filling small cavities and increasing the concrete density [28]. Mostafa et al. [29] evaluated the effect of nano-silica and recycled nanomaterials (nano-glass waste, nano-rice husk ash, and nano-metakaolin) on the properties of high-performance concrete. According to the results, the microstructure of concrete and compressive strength is among the parameters that can affect these concretes' resistance against corrosion [29]. Fahmy et al. [30] investigated nanoparticles' effect on the behavior of high-strength concrete beams. Eight rectangular beams with the same geometric and steel specifications were made. The variables included type of nanomaterials (nanosilica, nano-titanium, and the combination of nanotitanium and nanosilica) and their doses. The dual-use of the hybrid mixture (nanosilica and nanotitanium) has significantly improved the compressive and tensile strength compared to using the same percentage of a single type of nanomaterials [30].

As mentioned above, WPSA and pozzolanic materials can protect the concrete against external degradation agents and prevent water penetration in concrete and reduce rebar corrosion. Water penetration prevention in concrete increases the efficiency of concrete structures. This prevents harmful substances from penetrating the concrete and steel. One of the benefits of hydrophobic concretes made from WPSA, and pozzolanic materials are filling voids and cavities in the cement matrix. However, the results of studies showed that although WPSA can improve the durability properties of concrete, but has little effect on improving

the mechanical properties of concrete; Nanoparticles can overcome this weakness and improve the mechanical properties of concrete. The combined effect of WPSA and nanoparticles in high-strength concretes has not been evaluated in previous studies. Therefore, in the present study, the combined use of WPSA, silica nanoparticles, and aluminum oxide nanoparticles on high-strength concrete's mechanical and durability properties are investigated.

2. LABORATORY PROGRAM

Variables included WPSA (0, 5, 10, and 15% by weight of cement), type of nanoparticles (silica nanoparticles and aluminum oxide nanoparticles), and amount of nanoparticles (0 and 2.5% by weight of cement) respectively. The variables and the number of samples were selected according to the literature review experimental studies. The range of silica nanoparticles and aluminum oxide nanoparticles was considered based on studies about the use of nanoparticles. In most studies, the use of 2.5% nanoparticles was introduced as the optimal value [23-27]. The use of different amounts of nanoparticles can be considered as an important parameter. However, in this study, the main focus was on the mechanical properties, durability, and microstructure of high-strength concrete containing WPSA. The most important question that the authors sought to answer was whether the use of WPSA in combined with nanoparticles could improve engineering properties? Instead of using different amounts of nanoparticles, the authors increased the number of experiments to examine more specifications of these concretes. The experiments were conducted in four different sections. In the first part, the fresh properties of concrete were determined by slump test. In the second part, concrete's mechanical properties were evaluated using compressive strength, splitting tensile strength, flexural strength, and ultrasonic pulse velocity (UPV) tests. In the third part, the durability properties of concrete were assessed by performing water absorption, water penetration depth, and electrical resistivity. In the fourth section, the microstructure of concrete using scanning electron microscope (SEM) imaging was evaluated. According to the variables and experiments, 324 specimens were made according to Table 1. A flowchart of the study process is shown in Figure 1.

2. 1. Materials The used materials included coarse aggregates (gravel), fine aggregates (sand), cement, water, superplasticizer, WPSA, silica nanoparticles, and aluminum oxide nanoparticles.

2. 1. 1. Aggregates Aggregates in concrete make up about 75% of its volume, so their quality is important. In addition to being effective in concrete strength, aggregates are also effective in concrete durability and stability. Aggregates used in concrete include coarse aggregates about 63 to 73% of the total aggregates (maximum size is 23 mm), and fine aggregates are about 33 to 43% of the total aggregates (ranging in size from 3 to 76 mm). In this research, sand produced in Joban mine (Roudbar, Iran) has been used. The specific weight of sand and gravel is 2600 and 2650 kg/m^3. Grading of sand and gravel is in accordance with ASTM C330 [31] and is shown in Figure 2. The coarse and fine aggregates are in the range of ASTM C330. The dimensions of aggregates are selected in such a way that they are between the upper and lower boundaries of ASTM C330, so the standard requirements for the materials have been observed.

2. 1. 2. Cement Cement in concrete plays the role of bonding the components together. Portland type II cement was used as the primary adhesive in all samples and was prepared from the Hegmatan cement factory. This cement's chemical and physical characteristics based on ASTM C150 [32] are presented in Table 2. Specific gravity and pH of cement are 1.95 g/cm^3 and 9, respectively.

2. 1. 3. Silica Nanoparticles Silica nanoparticles are synthetic materials composed of very fine SiO_2 particles like cement and have high pozzolanic properties. The powder of this material is 99.9% silica, and its density is 0.2 g/cm^3. The specific surface area of silica nanoparticles is 50 to 100 m^2/g. Its solid particle diameter is about 50 to 100 nanometers. The

TABLE 1. Number of total specimens

Test	Number of mixture design	Number of specimens for each mixture	Age (days)	Number of total specimens
Compressive strength	12	3	7, 28 and, 90	108
Splitting tensile strength	12	3	28	36
Flexural strength	12	3	28	36
UPV	12	3	28	36
Water absorption	12	3	28	36
Water penetration depth	12	3	28	36
Electrical resistivity	12	3	28	36
Total				324

characteristics of silica nanoparticles are presented in Table 2.

2. 1. 4. Aluminum Oxide Nanoparticles These materials are white synthetic material composed of very small Al_2O_3 particles. They are used to improve ceramics' properties, solve their brittleness problem, increase the erosion resistance of coatings, and increase heat resistance [33]. The alumina in the nano-aluminum powder used in this research is more than 99%, its density is 3.89 g/cm^{3,} and the specific surface area of these nanoparticles is more than 138 m^2/g. Its solid particle diameter is about 20 nanometers.

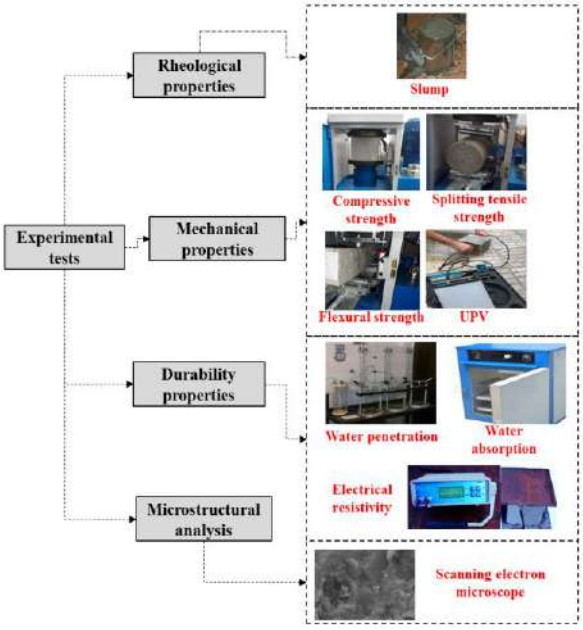


Figure 1. Geometric properties and steel reinforcement arrangement of original beams

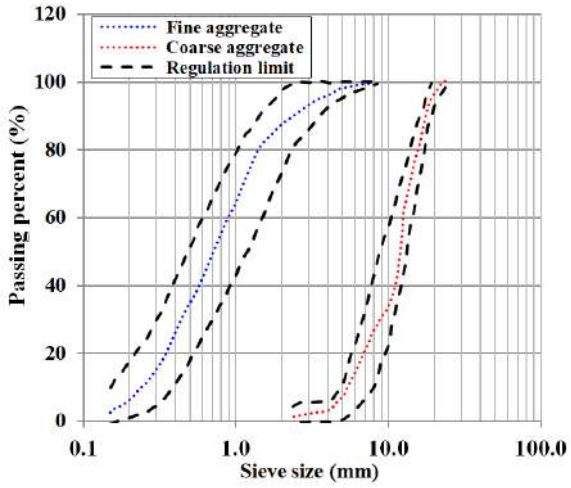


Figure 2. Garding curves of sand and gravel

2. 1. 5. WPSA The pulp production process begins with converting wood or lignocellulosic raw materials into flexible fibers suitable for making paper. During converting wood into pulp and paper, some wood particles and cellulose fibers enter the wastewater treatment system of factories. These cellulosic wastes, along with some minerals, are separated from the effluent treatment unit in the form of dry sludge and buried in the pits around the factories. One of the problems faced by the paper industry's production units is the significant volume of the waste remaining from the production process. These companies now bury all the sludge produced, and this is costly process.

In this study investigates the effect of addition of paper WPSA in the production of high-strength concretes. Waste paper sludge of Chouka Plulp & Paper factory (Rezvan-Shar, Iran) was used to make the specimens. A lot of waste paper sludge from the paper production process is accumulated in this factory in recent decades, the ashes of which can be used to produce concrete. The waste received from the factory was stored in the open air for seven days and then burned at 700 ° C. The chemical composition and mineral oxides in WPSA were evaluated using X-ray fluorescence (XRF) Spectrometer (Table 2). The paper sludge of Chuka factory was a solid waste consisting of fiber and ash residues produced by the paper-making and pulping process. This sludge was prepared, and its ash was used as a substitute for a part of cement. The potential of this

TABLE 2. Chemical characteristics of the binder materials

Components	Cement	AlNPs	SNPs	WPSA
Al_2O_3	4.95٪	99≥٪	-	6.83
CaO	62.95٪	25ppm ≤	-	25.43
SiO_2	21.27٪	-	99.98≥	10.79
Fe_2O_3	4.03٪	80ppm ≤	-	0.46
MgO	1.55٪	-	-	0.87
SiO_3	2.26٪	-	-	0.32
K_2O	0.65٪	-	-	0.22
Na_2O	0.49٪	70ppm ≤	-	0.15
MnO_2	-	3ppm ≤	-	-
Cr	-	4ppm ≤	-	-
Co	-	2ppm ≤	-	-
LOI	-	-	1.00 ≤	54.34
Specific gravity (g/cm^3)	0.75	3.89	0.5	2.81
Specific surface area (m^2/g)	326	138	200	168

recycled material for concrete production was evaluated based on the concept of biorefining. The used binder materials are presented in Figure 3.

2. 1. 6. Superplasticizer Excellent concrete water reducer, suitable for producing super-smooth concrete and self-compacting without additional water and aggregates separation and compatible with various types of cement are the main features of this superplasticizer. This product is manufactured in accordance with ASTM C494 [34]. The specific weight of this superplasticizer is 1.1 g/cm^2.

2. 1. 7. Water Tap water was used to make concrete specimens. According to ASTM C1602 [35], drinking water is suitable for concrete and does not need to be tested [36]. The water's pH was 7.5, the total hardness is 241 mg/l, the amount of chloride is 21.6 mg/l, and the amount of sulfate is 84.75 mg/l, which are within the permissible values mentioned.

2. 2. Experimental Tests Slump test was conducted to measure the workability of concrete in accordance with ASTM C143 [37].

The compressive strength test was performed according to ASTM C39 [38]. This test is one of the most important factors in the quality control of concrete and expresses the sample's tolerance in compressive strength. Cube molds with dimensions of 150×150×150 mm were used to make the specimens. The molds were kept in the open air for 24 hours after filling. After this period, the molds were opened, and the specimens were placed in the water tank and tested at the desired ages (7, 28, and 90 days).

The splitting tensile strength test was performed according to ASTM C496 [39]. In this experiment, standard cylinders of 150×300 mm were used. The conditions for preparing and storing the samples were the same as for compressive strength specimens. Specimens were placed horizontally in the machine. The load gradually increased, and due to the pressure in the direction perpendicular to the pressure, tension was created, and the specimen was broken. The load increase was uniform and loaded at a rate of 0.8 to 1.2 MPa per minute until the specimen was broken. The maximum load was read and recorded by the device. Splitting tensile strength was obtained by Equation (1). In this regard, T is the splitting tensile strength, P is the fracture load, L is the sample length, and D is the sample diameter. Failure of specimens in splitting tensile strength test was performed at 28 days of age.

$$T = \frac{2p}{\pi L d} \quad (1)$$

The flexural strength test was performed according to ASTM C293 [40]. The specimens had dimensions of 100×100×500 mm and were placed symmetrically on two simple supports. They were then loaded evenly and without sudden changes. Flexural strength was calculated by Equation (2). In this experiment, specimens were broken at 28 days of age.

$$S_f = \frac{3pl}{2bh^2} \quad (2)$$

In this equation, S_f is the flexural strength, p is the applied force in the middle of the span, b is the width of the section, h is the section's height, and l is the distance between the two supports.

The ultrasonic pulse velocity test was performed according to ASTM C597 [41]. The ultrasonic pulse velocity method involves measuring the travel time of an ultrasonic pulse passing through a concrete specimen. An ultrasonic device consists of an electronic circuit to generate a pulse and a transmitter to transmit these electronic pulses to mechanical energy, which have vibrational frequencies in the range of 15 to 50 kHz. The travel time between pulses is measured electronically. The pulse propagation's average velocity is obtained by dividing the path length between the transmitters by travel time. There are three methods for measuring pulse velocity in concrete: direct transfer, indirect transfer, and surface transfer. In the present study, the direct method is used by placing the transmitters between two surfaces in front of concrete samples. Pulse velocity is affected by several factors without taking into account the properties of concrete, including the smoothness of the contact surface of the specimen under test, the effect of path length on pulse velocity, test temperature, concrete moisture conditions, the presence of steel rebars, pulse velocity methods for evaluating concrete structures. Attempts have been made to correlate pulse velocity with strength and other properties of concrete. UPV test was performed at 28 days of age.

The initial (30 minutes) and final (72 hours)

Figure 3. Photo images of binder materials used

volumetric water absorption test was performed according to ASTM C642 [42]. Equation (3) was used to determine the water absorption percentage.

$$\text{Water absorption percentage}=\frac{m-m_0}{m_0}\times 100 \quad (3)$$

m is the wet weight of the specimen, and m_0 is the dry weight of the specimen.

The test for determining the water penetration depth in concrete was performed according to DIN1048 [43]. This experiment aims to assess the penetration of pressurized water in hardened concrete that has been treated in water. In this way, water is applied by pressure to the surface of hardened concrete, and then the specimen is divided into two halves, and the penetration depth related to water advancement is measured. 150×150×150 mm cube specimens were used. The sample was placed in the machine, and water pressure of 450 to 550 kPa was applied for 72 hours. During the test, the surfaces of the specimens that are not exposed to water pressure are continuously checked so that water does not leak. After applying pressure for a specified period of time, the sample is removed from the device, and excess surface water droplets are removed from the sample on which water pressure is applied. The sample is then split in half perpendicular to the surface exposed to water pressure. When the sample was divided into halves, the halved sample's surface exposed to water pressure was placed down, and the leading water sample was marked on it as soon as it was broken. The progress of water penetration is clearly visible on the surface. The maximum penetration depth was measured and recorded in millimeters.

Corrosion is an electrochemical phenomenon in which rebars act as anodes and concrete as cathodes, creating an electric current between the rebars and the concrete surface. In this case, the mobility of ions is visible. The more and easier this movement is, the lower the resistance to ion mobility and the greater the concrete's electrical conductivity. Therefore, one of the simplest ways to test concrete durability is to determine its electrical resistance. Moisture reduces electrical resistance. The presence of small cracks that fill with water also reduces electrical resistance. A concrete electrical resistance measuring device was used to determining the electrical resistivity. Specimens were tested at 28 days of age. An average of three readings was reported for each mixture. Two copper plates were placed in complete contact with two concrete surfaces. Equation (4) was used for determining the electrical resistivity.

$$\rho = \frac{R.L}{A} \quad (4)$$

where ρ is the specific electrical resistance of concrete in ohm-meters, R is the electrical resistance read by the device in ohms-meters. A is the area on concrete in square meters, and L is the distance between the positive and negative poles in meters. The estimation of corrosion probability of reinforcements buried in concrete based on electrical resistance is presented in Table 3.

2. 3. Mixture Design The ratios of the concrete mix components were determined using the absolute volume method in accordance with ACI 211 [47]. The water to cement ratio was considered 0.45 after reducing the aggregate moisture. The materials for one cubic meter of concrete are presented in Table 4. WPSA in the amount of 0, 5, 10, and 15% by weight of cement, silica nanoparticles at 0 and 2.5% by weight of cement, and aluminum oxide nanoparticles at 0 and 2.5% by weight of cement were used. The reason for choosing 2.5% nanoparticles is that this value has been introduced as the optimal value in many studies [26-28]. On the other hand, the main variable studied in the present study is WPSA, and four different values were considered for it.

The combined use of WPSA and two different types of nanoparticles in high-strength concrete is one of the

TABLE 3. Estimation of corrosion probability of buried rebars in concrete based on electrical resistivity [45, 46]

Electrical resistivity (kΩ-cm)	Estimation of corrosion probability
Less than 50	Very high
Between 50 and 100	High
Between 100 and 200	Low
More than 200	Very low

TABLE 4. Mixture design

Mix code	Cement	WPSA[a]	SN[b]	AlN[c]	Sand	Geavel	SP[d]
P0	550	0	0	0	705	1075	3.85
P0S	536.25	0	13.75	0	705	1075	3.98
P0AL	536.25	0	0	13.75	705	1075	4.10
P5	522.5	27.5	0	0	705	1075	4.85
P5S	508.75	27.5	13.75	0	705	1075	5.10
P5AL	508.75	27.5	0	13.75	705	1075	5.21
P10	495	55	0	0	705	1075	5.36
P10S	481.25	55	13.75	0	705	1075	5.46
P10AL	481.25	55	0	13.75	705	1075	589
P15	467.5	82.5	0	0	705	1075	6.21
P15S	453.75	82.5	13.75	0	705	1075	6.43
P15AL	453.75	82.5	0	13.75	705	1075	6.81

WPSA: Waste paper slag ash

SN: Silica nanoparticles

AlN: Aluminium oxide nanoparticles

SP: Superplasticizer

most important objectives of the present study. Concrete mixing using a mixer was initially done dry; The aggregates were mixed dry with cement and stirred in a mixer for 60 seconds. Then WPSA and nanoparticles were added to them, and the stirring operation was continued for another 45 seconds. Finally, water and superplasticizer were added to the mixture, and the mixing process was performed for another three minutes.

2. 4. Curing The samples' storage and curing conditions according to each series of specimens' condition were as follows: in the first 24 hours in the mold and then kept wet and transferred to the water tank, and stored in accordance with ASTM C192 [48].

3. RESULTS AND DISCUSSION

3. 1. Slump The workability depends on the slump and consistency of the concrete. The specimens' workability was evaluated using the slump test, and the results are presented in Figure 1. The slump of specimens containing 0, 5, 10, and 15% WPSA was 46, 51, 68, and 76 mm, respectively. The slump of specimens containing 0, 5, 10, and 15% WPSA in which 2.5% silica nanoparticles were used was 47, 54, 73, and 79 mm, respectively. The slump of specimens containing 0, 5, 10, and 15% WPSA in which 2.5% aluminum oxide nanoparticles were used was 56, 65, 78, and 81 mm, respectively. ACI 318-99 [47] proposed an allowable slump range for different RC building members (25 to 100 mm). All slump values obtained are in the mentioned range. Also, according to Figure 4, addition of 5 to 15% WPSA to the specimens with and without nanoparticles reduced the concrete slump from 11 to 76%, depending on the amount of nanoparticles compared to the control sample. The addition of nanoparticles and WPSA reduces slump and increases the possibility of locking. By adding superplasticizer to concrete specimens containing WPSA and nanoparticles, good workability can be achieved. By using appropriate amounts of WPSA, silica nanoparticles, and aluminum oxide nanoparticles along with suitable superplasticizer, operational mixing schemes used in projects can be achieved.

3. 2. Compressive Strength The mean compressive strength of the specimens at the ages of 7, 28, and 90 days is shown in Figure 5. The compressive strength of 7 days of specimens containing 0, 5, 10, and 15% WPSA were 38.9, 39.1, 40.3, and 37.1 MPa. The compressive strength of 7 days of specimens containing 0, 5, 10, and 15% WPSA in which 2.5% silica nanoparticles were used were 46.3, 47.6, 49.3, and 46.3 MPa, respectively. Also, the compressive strength of 7 days of specimens containing 0, 5, 10, and 15% WPSA in which 2.5% aluminum oxide nanoparticles were used were 44.6, 46.3, 48.7, and 45.2 MPa, respectively. According to the mentioned values, it is observed that adding WPSA to concrete samples without nanoparticles increased the 7-day compressive strength by about 1 to 4%, and addition of WPSA to concrete specimens containing nanoparticles increased the 7-day compressive strength about 14 to 27%.

Also, the 28-day compressive strength of P0, P5, P10, and P15 specimens were 57.1, 58.9, 60.3, and 55.6 MPa, respectively. By addition of 2.5% silica nanoparticles to the mentioned specimens caused their 28-day compressive strength to be 66.1, 67.2, 72.1, and 65.7 MPa, respectively. Also, the 28-day compressive strength of P0Al, P5Al, P10Al, and P15Al specimens were 65.3, 66.5, 70.2, and 65.5 MPa, respectively. Addition of 5 to 15% WPSA to specimens containing nanoparticles increased the 28-day compressive strength of the specimens by about 15 to 26%, depending on the amount of nanoparticles.

Using a combination of WPSA and nanoparticles was much better than using each one individually. For example, the 28-day compressive strength of a specimen using 10% WPSA and 2.5% silica nanoparticles was 26% higher than that of the control sample. The compressive

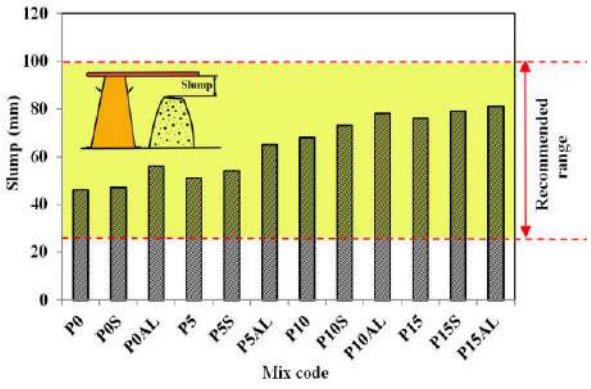


Figure 4. Slump value of concrete specimens containing different amounts of WPSA and nanoparticles

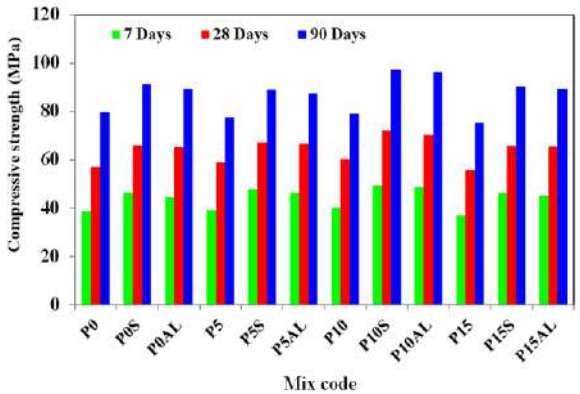


Figure 5. Compressive strength of specimens at 7, 28, and 90 days

strength of a specimen using 10% WPSA increased by only 4%, and the compressive strength of a specimen using only 2.5% silica nanoparticles was 16% higher than the control sample.

As the age of concrete increases, the compressive strength of the specimen increases. The 90-day compressive strength of concrete specimens containing nanoparticles and WPSA increased by about 10 to 22% compared to the control specimen, depending on the amount of nanoparticles and WPSA. Also, the highest 90-day compressive strength was obtained in the samples in which 10% WPSA and 2.5% silica nanoparticles were used, and its value was 97.3 MPa.

The use of silica nanoparticles and aluminum oxide nanoparticles in combination of WPSA has increased the compressive strength compared to the control sample. Excess silica is mixed with cement compounds to form C-S-H gels, and compressive strength was achieved by filling pores in concrete.

As the WPSA increased to more than 10%, the compressive strength decreased with a greater slope. The reason for this is that when the amount of WPSA in concrete exceeds a certain limit, WPSA is dehydrated due to the high intensity of the reaction. The use of superplasticizers up to a certain percentage of nanoparticles compensates for this deficiency, but due to the limited use of superplasticizers in higher percentages of WPSA, this problem persists, resulting in some porosity of the concrete, which results in decreases of the compressive strength.

Another point is that WPSA increases the compressive strength in a short time due to its specific surface area and high reaction intensity. WPSA reacts with calcium hydroxide formed due to cement curing and hydrates the production of calcium silica, the hard material that makes concrete strength. As shown in Figure 1, in the 7-day specimen, the upward trend and in the 28- and 90-day specimen, the growth has decreased. The reason for this behavior is probably that the 7-day samples did not suffer from water deficiency due to the intensity of the reaction at the time of the test. As a result, internal porosity has not yet formed in the concrete, but this has happened in the case of 28 and 90-days concrete.

3. 3. Splitting Tensile Strength The 28-day splitting tensile strength of specimens and the percentage of changes compared to the control specimen are shown in Figure 6. Single-use of WPSA in specimens does not play a significant role in splitting tensile strength of specimens and, in some cases, has even reduced the splitting tensile strength of specimens (P10 and P15). The combined use of 2.5% silica nanoparticles and 0, 5, 10, and 15% WPSA increased the splitting tensile strength by 14.8, 14.6, 15.6, and 12.8%, respectively. Also, the combined use of 2.5% aluminum oxide nanoparticles and 0, 5, 10, and 15% WPSA increased the splitting tensile strength by 7.7, 10.5, 9.2, and 6.9%, respectively. As with compressive strength results, the combined use of nanoparticles and WPSA has a more effective role in increasing splitting tensile strength than their single-use. The splitting tensile strength largely depends on the area of transition between the aggregate and the cement paste. By replacing part of the cement with WPSA, silica nanoparticles, and aluminum oxide nanoparticles, the transition zone between the aggregates is strengthened, and thus the tensile strength is increased.

Figure 7 shows the relationship between splitting tensile strength and compressive strength of specimens containing WPSA and nanoparticles at 28 days of age. The compressive strengths presented in this figure are the compressive strengths of cylindrical specimens. For this purpose, all 28-day compressive strengths of concrete cubic specimens were converted to cylindrical equivalent compressive strength using conversion coefficients. This diagram also presents the relationships between tensile strength and compressive strength proposed by ACI 318R-95 [49] and CEB-FIP [50].

The range of compressive and tensile strengths obtained is within the range of values of the mentioned

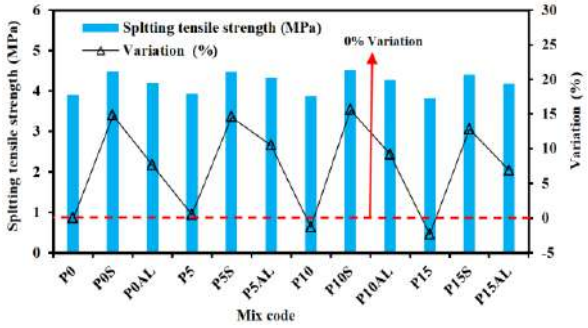


Figure 6. Splitting tensile strength and the percentage of changes compared to the control sample

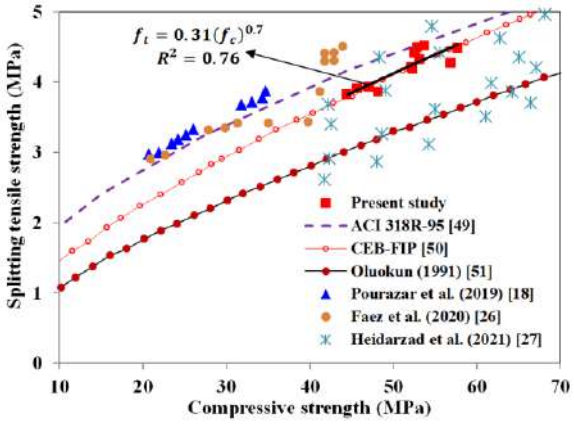


Figure 7. Relationship between splitting tensile strength and compressive strength and comparison with regulations and other studies

standards. Figure 7 shows the range of compressive and tensile strengths of some similar studies. Pourazar et al. [18] investigated the effect of WPSA and silica fume on concrete specimens. As can be seen, the range of compressive and tensile strengths obtained in the present study is in a higher range than concretes containing WPSA and silica fume. This result showed that the use of nanosilica in combination with WPSA has a much better performance compared to the use of silica fume. Nanoparticles have a more effective role in improving the mechanical properties of concrete due to higher filling. Also, the results reported by Faez et al. [26] and Heidarzad et al. [28] are shown in Figure 7, which used silica nanoparticles and aluminum nanoparticles to make concrete samples, respectively. The purpose of this chart is to ensure the accuracy of the results. The difference in the present study's values can be due to the water to cement ratio, quality of materials, type of additives, and the curing type.

3. 4. Flexural Strength The flexural strength of the specimens at the age of 28 days and their percentage change compared to the control specimen are shown in Figure 8. The use of 5 and 10% WPSA has increased the flexural strength by about 1.1 and 2%, respectively. But the use of 15% WPSA has reduced the flexural strength by about 1.5%. The use of nanoparticles in combination with WPSA has led to an increase in the specimens' flexural strength. The highest increase in flexural strength was obtained in the P10S specimen and the flexural strength of this specimen increased by 14.3% compared to the control sample. Comparing flexural strengths of specimens containing silica nanoparticles and aluminum oxide nanoparticles showed that the combined use of silica and WPSA nanoparticles has a more influential role in increasing the flexural strength of concrete specimens than specimens containing aluminum oxide and WPSA.

3. 5. Ultrasonic Pulse Velocity (UPV) Among the non-destructive methods available, the use of UPV, in addition to making it possible to measure hardness on the actual structure, can also reveal some defects of concrete [52]. UPV test can be used for determining the concrete uniformity, concrete strength, the modulus of elasticity, properties of concrete over time, the degree of hydration of cement, the durability of concrete, damaged layer action in concrete, and crack depth [53]. Figure 9 shows the UPV of the specimens in different states. This diagram also provides details on the classification of concrete types in terms of quality. This classification was proposed by Whitehurst [54], and concrete is divided into four categories in terms of quality: excellent, good, medium, and doubtful. The results showed that the specimens' UPV is in the range of 3.5 to 4.5 km/h and has good quality. Adding WPSA to the specimen increased the UPV. The highest increase of UPV was obtained in samples in which 10% WPSA was used in combination with 2.5% nanoparticles (silica and aluminum oxide).

3. 6. Water Absorption CEB divides concrete [55] into three categories based on initial water absorption: poor quality (5 to 6%), medium (between 3% and 5%), and good (less than 3%). According to the CEB classification, the control sample is in the middle category. Figure 10 shows the initial and final water absorption of the specimens in different states. Single-use of WPSA and nanoparticles and their combined use have reduced the water absorption percentage of the specimens. The initial water absorption percentage of concrete specimens containing WPSA is in the range of 2.25 to 3.67%. According to the mentioned classification, it can be stated that all the concrete specimens examined, in which WPSA, silica nanoparticles, and aluminum nanoparticles were used, are in the good quality category.

The combined use of nanoparticles and WPSA in all mixing designs have reduced the water absorption percentage of concrete. For example, the use of 10% WPSA and 2.5% silica nanoparticles reduced the initial and final water absorption of concrete by 28% and 29%, respectively. The combined use of nanoparticles and WPSA causes the concrete pores to fill more and decrease water absorption. Consumption of calcium

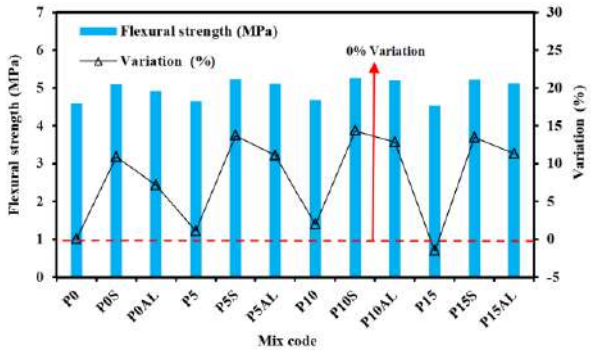


Figure 8. Flexural strength and percentage of changes compared to the control specimen

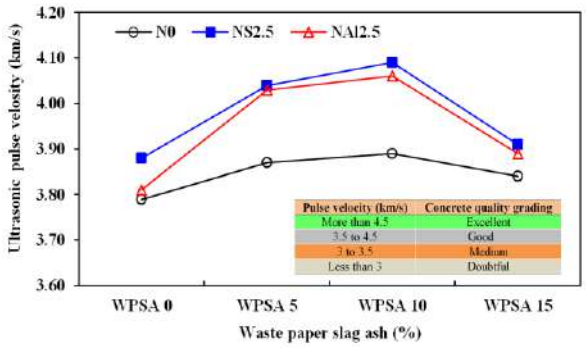


Figure 9. UPV of specimens

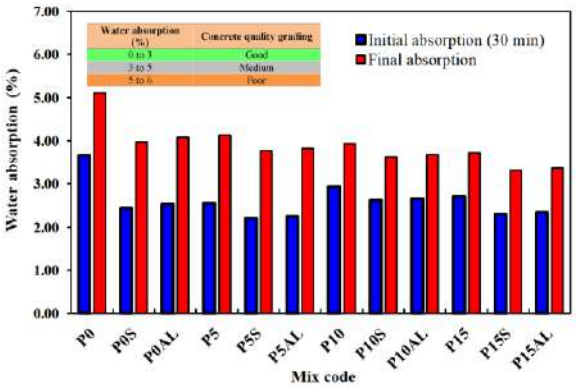


Figure 10. Percentage of initial and final water absorption of specimens in different states

hydroxide and production of more hydrated calcium silicate in the presence of nanoparticles and WPSA are other reasons for the decrease in concrete water absorption.

3. 7. Water Penetration Depth Figure 11 shows a graph of water penetration versus WPSA for specimens with and without nanoparticles. By replacing WPSA or silica nanoparticles and aluminum oxide nanoparticles, the water penetration depth index is improved. The P15S mixture, in which 15% WPSA and 2.5% silica nanoparticles were substituted for Portland cement, showed the lowest penetration depth, which is 3.76 mm. The use of WPSA alone in the specimens showed less effectiveness in improving the water penetration depth than the mixture containing WPSA and nanoparticles, which in this case can also be attributed to the lower pozzolanic activity of WPSA compared to the combination of WPSA and nanoparticles.

The use of 5, 10, and 15% WPSA reduced the water penetration depth by 16, 22, and 25%, respectively, compared to the control sample. Replacement of WPSA with cement has a greater contribution and effect on improving the absorption properties and permeability of concrete than the strength properties of concrete. The use of 5, 10, and 15% WPSA in specimens containing 2.5% silica nanoparticles reduced the water penetration depth by 22, 26, and 30%, respectively, compared to the control sample. Also, the use of 5, 10, and 15% WPSA in specimens containing 2.5% aluminum oxide nanoparticles reduced the water penetration depth by 21, 25, and 29%, respectively. The combined use of silica nanoparticles, aluminum oxide nanoparticles, and WPSA in high-strength concrete improved the microstructure and reduces the pores' size, leading to reduced permeability and increased resistance to aggressive materials as chlorine ions and the risk of carbonation.

3. 8. Electrical Resistivity Due to the fact that the corrosion process of steel rebars in concrete is an electrochemical process, so the electrical resistivity of concrete has a significant role in corrosion, and it can be expected that concretes with higher electrical resistivity showed better performance in reducing corrosion of steel rebars. Figure 12 shows the electrical resistivity of the specimens in different states. The use of WPSA has increased the electrical resistivity of concrete. With increasing WPSA value, the electrical resistivity of concrete specimens has increased more. P5, P10, P15 mixtures' electrical resistivity were 10, 19, and 29% higher than the control specimen (P0), respectively. Elkey and Sellevold [45] and Ong and Saraswathy [46] divided concrete specimens into four categories based on electrical resistivity and corrosion probability (Figure 12). Accordingly, specimens containing WPSA and nanoparticles are less likely to corrosion. The combined use of nanoparticles and WPSA has increased the electrical resistance by about 18 to 67 percent. A significant increase in electrical resistivity of specimens containing WPSA, silica nanoparticles, and aluminum oxide nanoparticles was due to pozzolanic reactions and reduced porosity of concrete pores. Examining these results, it is clear that the degree of permeability for all projects containing WPSA has been associated with a decrease.

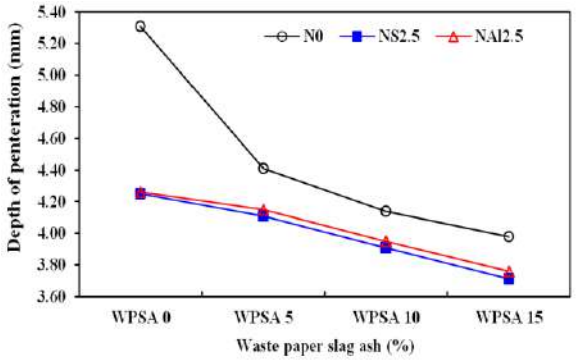


Figure 11. The water penetration depth of specimens in different conditions

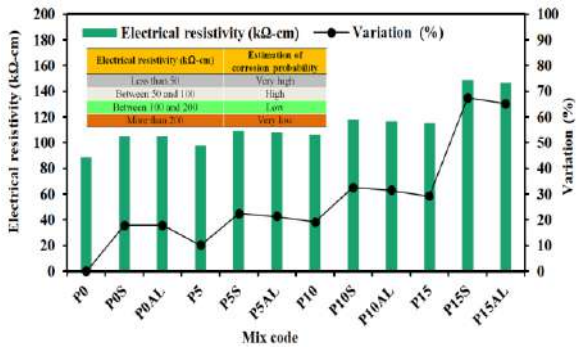


Figure 12. The electrical resistivity of specimens in different states

3. 9. Scanning Electron Microscope (SEM) Imaging

The results of experiments performed on concrete specimens containing nanoparticles and WPSA showed that 10% WPSA combined with 2.5 nanoparticles (silica and aluminum oxide) has an effective role in improving the mechanical properties and durability of concrete. Therefore, specimens P0, P10, P10S, and P10Al were selected to study the microstructure, and their SEM images were prepared. Figure 13 shows the SEM images of P0, P10, P10S, and P10Al. In the control sample (P0), there is no continuous continuity between the concrete components, and the components do not have an acceptable consistency. By creating C-H- S gel, nano-silica increases the compressive strength and adhesion of concrete components.

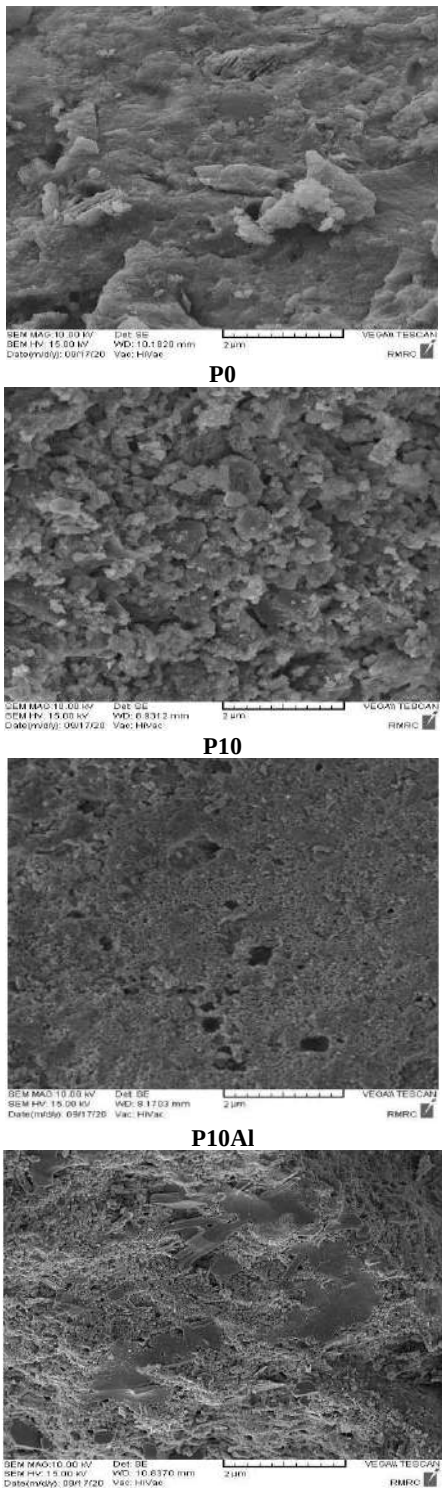
WPSA increases the durability of concrete by closing cavities. Comparison of SEM images of the control sample (without nanoparticles and WPSA) and the rest of the images containing nanoparticles and WPSA shows that the control sample has more component rupture and has more cavities. Nanoparticles reduce these cavities, resulting in increased compressive and tensile strengths and reduced water permeability.

In the microstructure of ordinary cement paste, it is observed that a large amount of needle-shaped hydrates surrounds the C-S-H gel. Comparison of this image and the image of cement paste made with 10% WPSA and 2.5% nanoparticles (silica and aluminum oxide) showed the better effect of this proposed compound on the microscopic structure of cement paste. However, in some parts of the figure, local growth of hydrated crystals was observed, which added to their non-uniformity.

Also, microscopic pores on a more or less uneven surface can be seen in these images. This image shows the effect of nanosilica on reducing the size of large crystals. It is also observed that WPSA and nanoparticles by filling the gel's pores reduce the size and number of microstructural cavities of the samples compared to samples without nanomaterials, resulting in greater coherence and integration for these specimens. As can be seen from the pictures, the results of the experiments are consistent with the photographs taken of the microstructure of the cement paste.

4. SUMMARY AND CONCLUSION

In this paper, the mechanical properties, durability, and microstructure of high-strength concrete containing WPSA and nanoparticles were investigated. The effect of WPSA on improving the mentioned characteristics was one of the main objectives of the present study. For this purpose, compressive strength, splitting tensile strength, flexural strength, UPV, water absorption, water penetration depth, and Electrical resistivity tests were performed. Also, the microstructure of concrete specimens containing WPSA and nanoparticles was evaluated using SEM images. This section summarizes the most important results:



P0

P10

P10Al

P10S

Figure 13. SEM images of specimens

- The combined use of WPSA and nanoparticles is better than the single-use of each. For example, the 28-day compressive strength of the specimen containing 10% WPSA and 2.5% silica nanoparticles was 26% higher than that of the control specimen. The compressive strength of the specimen containing 10% WPSA increased by only 4%, and the compressive strength of the specimen containing only 2.5% silica nanoparticles was 16% higher than that of the sample.
- The use of silica nanoparticles and aluminum oxide nanoparticles in combination with WPSA has increased the compressive strength compared to the control specimen. This is because of the excess silica mixes with the cement compounds to form the C-S-H gel. Increasing the strength is done by filling the pores of the concrete.
- WPSA increases compressive strength in a short time due to its specific surface area and high reaction intensity. WPSA reacts with calcium hydroxide formed due to cement curing and hydrates the production of calcium silica, which is the hard material that makes concrete strength.
- In 7-days specimens, an increase in compressive strength growth was observed, and, in 28 and 90-day specimens, a decrease in compressive strength growth was observed. The reason for this behavior is probably that the 7-day specimens did not suffer from water deficiency due to the intensity of the reaction at the time of the test. As a result, internal porosity has not yet formed in the concrete, but this has happened with 28 and 90-day-old concretes.
- Single-use of WPSA in specimens does not play a significant role in the splitting tensile strength of specimens and, in some cases, has even reduced the tensile strength of specimens. The combined use of nanoparticles and WPSA has a more effective role in increasing splitting tensile strength than their single-use. The amount of tensile strength of concrete depends to a large extent on the transition area between the aggregate and the cement paste. By replacing part of the cement with WPSA, silica nanoparticles, and aluminum oxide nanoparticles, the transition zone between the aggregates is strengthened, and thus the tensile strength has increased.
- The highest increase in flexural strength was obtained in the P10S specimen and the flexural strength of this specimen increased by 14.3% compared to the control specimen. Comparison of flexural strengths of specimen containing silica and aluminum oxide nanoparticles shows that the combined use of silica and WPSA nanoparticles has a more effective role in increasing the flexural strength of concrete specimens than specimen containing aluminum oxide and WPSA.
- The combined use of silica nanoparticles, aluminum oxide nanoparticles, and WPSA in high-strength concretes improves the microstructure and reduces porosity, leading to reduced permeability and increased resistance to aggressive materials such as chlorine ions and the risk of carbonation.
- The combined use of nanoparticles and WPSA has increased the electrical resistance by about 18 to 67 percent. A significant increase in electrical resistance of specimens containing WPSA, silica nanoparticles, and aluminum oxide nanoparticles was due to pozzolanic reactions and reduced porosity of concrete pores. Examining these results, it is clear that the degree of permeability for all projects containing WPSA has been associated with a decrease.
- WPSA and nanoparticles by filling the gel's pores reduce the size and number of microstructural cavities of the specimens compared to samples without nanomaterials, resulted in greater coherence and integration for these samples.

The paper industries use two common incineration and landfill methods to dispose of their waste, which is a matter of exorbitant costs. Hazardous compounds in this type of waste cause environmental pollution and endanger human health. According to the results, the use of WPSA in the production of concrete is more economical and environmental due to less use of cement. The use of WPSA from paper mills in the production of concrete leads to economic and ecological benefits. The present study results that were about the management of WPSA and their use in the concrete showed that these materials can be used as a substitute for a part of cement (up to 10%) and reduce the percentage of water absorption and strength of concrete. In this study, an attempt was made to overcome the strength reduction weakness by using silica nanoparticles and aluminum oxide nanoparticles. In these cases, higher percentages of waste paper waste can be used. In addition, the proposed concrete can be used to repair the piers of bridges and hydraulic structures.

5. REFERENCES

1. Azmi, AN., Fauzi, MA., Nor. MD., Ridzuan, AR., and Arshad, MF., "Production of controlled low strength material utilizing waste paper sludge ash and recycled aggregate concrete." In Matec web of conferences, EDP Sciences, Vol. 47, (2016). 01011.
2. Karada, AG, and Awchat, GD., "A review on partial replacement of fine aggregate by waste paper sludge in concrete." ***International Journal of Computer Science and Mobile Computing***, Vol. 6, No. 6, (2017), 369-376.
3. Ramirez, KG., Possan, E., Dezen, BG., and Colombo, M., "Potential uses of waste sludge in concrete production." ***Management of Environmental Quality: An International Journal***, Vol.28, No. 6, (2017), 821-838, DOI: org/10.1108/MEQ-09-2015-0178

4. Mavroulidou, M., and Awoliyi, S., "A study on the potential use of paper sludge ash in concrete with glass aggregate." ***Waste Management & Research***, Vol. 36, No. 11, (2018), 1061-1065, DOI: org/10.1177/0734242X18801196.
5. Patel, P., and Rajput, TS., "Practice of construction paper slush by restricted substitution of cement." ***Journal of Management and Architecture Research***. Vol. 1, No. 01, (2019), 1-3.
6. Vashistha, P., Kumar, V., Singh, SK., Dutt, D., Tomar, G., Yadav, P., "Valorization of paper mill lime sludge via application in building construction materials: A review." ***Construction and Building Materials***, Vol. 30, No. 211, (2019), 371-382, DOI: org/10.1016/j.conbuildmat.2019.03.085.
7. Babita, S., Saurabh, U., Abhishek, GK., Manoj, Y., Pranjal, B., Ravi, MK., and Pankaj, K., "Review paper on partial replacement of cement and aggregates with various industrial waste material and its effect on concrete properties". ***In Recycled Waste Materials***, Vol. 32, (2019), 111-117, DOI: org/10.1007/978-981-13-7017-5_13.
8. Fava, G., Ruello, ML., Corinaldesi, V., "Paper mill sludge ash as supplementary cementitious material." ***Journal of Materials in Civil Engineering***, Vol. 23, No. 6, (2010), 772-776, DOI: org/10.1061/(ASCE)MT.1943-5533.0000218.
9. Wong, HS, Barakat, R., Alhilali, A., Saleh, M., Cheeseman, CR., "Hydrophobic concrete using waste paper sludge ash." ***Cement and Concrete Research***, Vol. 70, (2015), 9-20, DOI: org/10.1016/j.cemconres.2015.01.005.
10. Dunster, A.M., "Paper sludge and paper sludge ash in Portland cement manufacture". ***MinRes Case Study, Building Research Establishment, Garston***, (2007).
11. Bai, J., Chaipanich, A., Kinuthia, JM., O'farrell, M., Sabir, BB., Wild, S., and Lewis, MH., "Compressive strength and hydration of wastepaper sludge ash–ground granulated blastfurnace slag blended pastes." ***Cement and Concrete Research***, Vol. 33, No. 8, (2003), 1189-1202, DOI: org/10.1016/S0008-8846(03)00042-5
12. Mozaffari, E., Kinuthia, JM., Bai, J., and Wild, S., "An investigation into the strength development of wastepaper sludge ash blended with ground granulated blastfurnace slag." ***Cement and Concrete Research***, Vol. 39, No. 10, (2009), 942-949, DOI: org/10.1016/j.cemconres.2009.07.001.
13. Pera, J., and Amrouz A. "Development of highly reactive metakaolin from paper sludge." ***Advanced Cement Based Materials***, Vol. 7, No. 2, (1998), 49-56, DOI: org/10.1016/S1065-7355(97)00016-3.
14. Sutcu, M., and Akkurt, S., "Utilization of recycled paper processing residues and clay of different sources for the production of porous anorthite ceramics." ***Journal of the European Ceramic Society***, Vol. 30, No. 8, (2010), 1785-1793, DOI: org/10.1016/j.jeurceramsoc.2010.01.038.
15. Ismail, M., Ismail, M. A., and Keok, L. S. "Fabrication of bricks from paper sludge and palm oil fuel ash". ***Concrete Research Letter*s**, Vol. 1, No. 2, (2010), 60-66.
16. Martínez, C., Cotes, T., and Corpas, FA., "Recovering wastes from the paper industry: Development of ceramic materials." ***Fuel Processing Technology***, Vol. 1, No. 103, (2012), 117-124, DOI: org/10.1016/j.fuproc.2011.10.017
17. Ahmad, S., Malik, MI., Wani, MB., Ahmad, R., "Study of concrete involving use of waste paper sludge ash as partial replacement of cement." ***IOSR Journal of Engineering***, Vol. 3, No. 11, (2013), 6-15, DOI: 10.9790/3021-031130615
18. Pourazar, J., Najarchi, M., Sanaati, B., Najafizadeh, MM., and Mirhosseini, SM., "The experimental assessment of the effect of paper waste ash and silica fume on improvement of concrete behavior." ***KSCE Journal of Civil Engineering***, Vol. 23, (2019), 4503-4515, DOI: org/10.1007/s12205-019-0678-x
19. Mavroulidou, M., and Shah, S., "Alkali-activated slag concrete with paper industry waste." ***Waste Management & Research***, Vol. 39, No. 3, (2021), 466-472, DOI: org/10.1177/0734242X20983890.
20. Arel, H. S., and Thomas, B. S. "The effects of nano-and micro-particle additives on the durability and mechanical properties of mortars exposed to internal and external sulfate attacks". ***Results in Physics***, Vol. 7, 843-851. (2017), 843-851, DOI: org/10.1016/j.rinp.2017.02.009.
21. Yang, H., Monasterio, M., Zheng, D., Cui, H., Tang, W., Bao, X., and Chen, X., "Effects of nano silica on the properties of cement-based materials: A comprehensive review." ***Construction and Building Materials***, Vol. 282, (2021), 122715, DOI: org/10.1016/j.conbuildmat.2021.122715.
22. Gopalakrishnan, R., and Jeyalakshmi, R., "Strength deterioration of nano-silica contained in ordinary Portland cement concretes in aggressive sulfate environments". The ***European Physical Journal Plus***, Vol. 133. No. 351, (2018), DOI: org/10.1140/epjp/i2018-12162-3.
23. Mirgozar Langaroudi, M. A., and Mohammadi, Y., "Effect of nano-clay on the freeze–thaw resistance of self-compacting concrete containing mineral admixtures". ***European Journal of Environmental and Civil Engineering***, (2019), 1-20. DOI: org/10.1080/19648189.2019.1665107.
24. Al-Tersawy, S. H., El-Sadany, R. A., and Sallam, H. E. M., "Long-term behavior of normal weight concrete containing hybrid nanoparticles subjected to gamma radiation." ***Archives of Civil and Mechanical Engineering***, Vol. 21, No. 1, (2021), 1-18, DOI: org/10.1007/s43452-020-00157-4.
25. Chu, S. H., Li, L. G., and Kwan, A. K. H., "Development of extrudable high strength fiber reinforced concrete incorporating nano calcium carbonate". ***Additive Manufacturing***, Vol. 37, (2021), 101617, DOI: org/10.1016/j.addma.2020.101617.
26. Faez, A., Sayari, a., and Manie, S., "Mechanical and rheological properties of self-compacting concrete containing Al_2O_3 nanoparticles and silica fume." ***Iranian Journal of Science and Technology, Transactions of Civil Engineering,*** Vol. 44, (2020), 217-227, DOI:10.1007/s40996-019-00339-y.
27. Ghanbari, M., Kohnehpooshi, O., and Tohidi, M., "Experimental Study of the Combined Use of Fiber and Nano Silica Particles on the Properties of Lightweight Self Compacting Concrete". ***International Journal of Engineering, Transactions B: Applications***, Vol. 33, No. 8, (2020), 1499-1511. DOI: 10.5829/ije.2020.33.08b.08.
28. Heidarzad Moghaddam, H., Maleki, A., and Lotfollahi-Yaghin, M. A. "Durability and Mechanical Properties of Self-compacting Concretes with Combined Use of Aluminium Oxide Nanoparticles and Glass Fiber". ***International Journal of Engineering, Transactions A: Basics***, Vol. 34, No. 1, (2021), 26-38, DOI: 10.5829/ije.2021.34.01a.04.
29. Mostafa, S. A., El-Deeb, M. M., Farghali, A. A., and Faried, A. S. "Evaluation of the nano silica and nano waste materials on the corrosion protection of high strength steel embedded in ultra-high performance concrete". ***Scientific Reports***, Vol. 11, No. 1, (2021), 1-16, DOI: org/10.1038/s41598-021-82322-0.
30. Fahmy, N. G. E., El-Mashery, R., Sadeek, R. A., and El-Hafaz, A. "Influence of Nano Particles in the Flexural Behavior of High-Strength Reinforced Concrete Beams." ***In Advanced Materials Research***, Vol. 1160, (2021), 25-43, DOI: org/10.4028/www.scientific.net/AMR.1160.25.
31. ASTM C330/ C330M-17a, Standard Specification for Lightweight Aggregates for Structural Concrete, ASTM International, West Conshohocken, PA, (2017).
32. ASTM C150/ C150M-20, Standard Specification for Portland Cement, ASTM International, West Conshohocken, PA, (2020), www.astm.org

33. Li, Z., Wang, H., He, S., Lu, Y., and Wang, M., "Investigations on the preparation andmechanical properties of the nano-alumina reinforced cement composite.", ***Materials Letters***, Vol. 60, No. 3, (2006), 356-359.
34. ASTM C494 / C494M-19, Standard Specification for Chemical Admixtures for Concrete, ASTM International, West Conshohocken, PA, (2019), www.astm.org
35. ASTM C1602/ C1602M-18, Standard Specification for Mixing Water Used in the Production of Hydraulic Cement Concrete, ASTM International, West Conshohocken, PA, (2018), www.astm.org
36. Shadmand, M., Hedayatnasab, A., and Kohnehpooshi, O. "Retrofitting of reinforced concrete beams with steel fiber reinforced composite jackets." ***International Journal of Engineering, Transactions B: Applications***, Vol. 33, No. 5, (2020), 770-783, DOI: 10.5829/ije.2020.33.05b.08.
37. ASTM C143 / C143M-20, Standard Test Method for Slump of Hydraulic-Cement Concrete, ASTM International, West Conshohocken, PA, (2020), www.astm.org
38. ASTM C39/C39M-18, Standard test method for compressive strength of cylindrical concrtete speciemens, ASTM International, West Conshohocken, PA, (2018).
39. ASTM C496/C496M-17, Standard test method for splitting tensile strength of cylindrical concrete specimens, ASTM International, West Conshohocken, PA, (2017).
40. ASTM C293 / C293M-16, Standard Test Method for Flexural Strength of Concrete (Using Simple Beam With Center-Point Loading), ASTM International, West Conshohocken, PA, (2016), www.astm.org
41. ASTM C597-16, Standard Test Method for Pulse Velocity Through Concrete, ASTM International, West Conshohocken, PA, 2016, www.astm.org
42. ASTM C642-13, Standard Test Method for Density, Absorption, and Voids in Hardened Concrete, ASTM International, West Conshohocken, PA, (2013), www.astm.org
43. DIN 1048-5., Test methods for concrete". Deutsches, (1991).
44. Shetty, MS. Concrete technology. S. Chand, Delhi, (2005).
45. W. Elkey and E. J. Sellevold, "Electrical Resistivity of Concrete", supplement 80, Norwegian Road Research Laboratory, Oslo, Norway, (1995).
46. Song, H.S., and Saraswathy, V., "Corrosion monitoring of reinforced concrete structures-a review." ***International journal of Electrochemical Science***, Vol. 2, (2007), 1–28.
47. ACI Committee 211, American Concrete Institute, Farmington Hills, Michigan, (1991).
48. ASTM C192 / C192M-18, Standard Practice for Making and Curing Concrete Test Specimens in the Laboratory, ASTM International, West Conshohocken, PA, 2018,
49. ACI 318-99, "Building Requirements for Structural Concrete and Commentary", American Concrete Institute, Farmington Hills, Michigan, 393, (1999).
50. CEB-FIP Model code. Comite Euro-International du Beton. London: Thomas Telford. Olukan, (1990).
51. Oluokun, F.A. "Prediction concrete tensile strength from its compressive strength: evaluation of existing relations for normal weight concrete". ***ACI Material Journal***, Vol. 88, No. 3, (1991), 302-309.
52. Grosse C.U., Reinhardt H.W., "New developments in quality control of concrete using ultrasound." Proceedings of the international symposium on NDT in civil engineering, Berlin (2003).
53. Krautkramer J., Krautkramer M., "Ultrasonic testing of materials", Berlin: Springer; (1990).
54. Whitehurst, E. A. "Evaluation of concrete properties from sonic tests". ***ACI Monograph***, No. 2, (1966), 94.
55. CEB-FIP. "Diagnosis and assessment of concrete" structures-state of art report, CEB Bulletin, 83. (1989).

Persian Abstract

چکیده

واحد های تولیدی صنایع کاغذسازی حجم قابل توجه‌ای از پسماند باقیمانده کاغذ را سوزانده و یا دفن می کنند. این موضوع از جنبه‌های زیست محیطی و اقتصادی مناسب نمی‌باشد. در مطالعه حاضر اثر استفاده ترکیبی از خاکستر لجن ضایعات کاغذ (WPSA) و نانو مواد بر خصوصیات مکانیکی و دوام بتن‌های با مقاومت بالا ارزیابی شده است. WPSA (۰، ۵، ۱۰ و ۱۵ درصد وزنی سیمان)، نانو ذرات سیلیس (۰ و ۲.۵ درصد وزنی سیمان) و نانودرات اکسید آلومینیوم (۰ و ۲.۵ درصد وزنی سیمان) متغیرهای مورد بررسی بودند. خصوصیات مکانیکی با آزمایش‌های مقاومت‌های فشاری، کششی شکافت، خمشی و سرعت امواج فراصوت ارزیابی شد. همچنین دوام بتن با انجام آزمایش‌هایی نظیر عمق نفوذ آب، جذب آب و مقاومت الکتریکی انجام گردید. ریزساختار نمونه‌ها با تهیه تصاویر میکروسکپی الکترونی تحلیل شد. اثر ترکیبی WPSA و نانوذرات بر بهبود خواص مکانیکی (مقاومت‌های فشاری، کششی و خمشی) و دوام بتن‌های با مقاومت بالا، بهتر از استفاده تکیِ از هر کدام از آنها می‌باشد. WPSA و نانو ذرات با کلسیم هیدروکسیدی که به دلیل هیدراتاسیون سیمان تشکیل میشود، واکنش می‌دهد و سیلیکا کلسیم هیدراته تولید میکند که همان ماده‌ی سخت عامل مقاومت بتن است. مصرف شدن هیدروکسید کلسیم و تولید سیلیکات کلسیم هیدراته‌ی بیشتر در حضور نانو ذرات و WPSA از جمله دلایل کاهش جذب آب، افزایش مقاومت الکتریکی و کاهش عمق نفوذ آب در نمونه‌های بتنی می‌باشد. با جایگزینی بخشی از سیمان با WPSA، نانوذرات سیلیس و نانوذرات اکسید آلومینیوم، ناحیه‌ی انتقال بین دانه‌های سنگی تقویت می‌شود و بدین ترتیب مقاومت‌های کششی و خمشی افزایش می‌یابد.

IJE Transactions A: Basics, Vol. 34, No. 7, (July 2021) 1667-1683

International Journal of Engineering

Journal Homepage: www.ije.ir

Concepts, Key Challenges and Open Problems of Federated Learning

Z. Iqbal[a,b], H. Y. Chan*[a]

[a] *School of Computer Sciences, Universiti Sains Malaysia, Pulau Pinang, Malaysia*
[b] *Department of Computer Science, University of Gujrat, Gujrat, Pakistan*

PAPER INFO

Paper history:
Received 21 January 2021
Received in revised form 28 April 2021
Accepted 21 May 2021

Keywords:
Federated Learning
On Device Learning
Decentralized Learning
Privacy Preserving Machine Learning

ABSTRACT

With the modern invention of high-quality sensors and smart chips with high computational power, smart devices like smartphones and smart wearable devices are becoming primary computing sources for routine life. These devices, collectively, might possess an enormous amount of valuable data but due to privacy concerns and privacy laws like General Data Protection Regulation (GDPR), this enormous amount of very valuable data is not available to train models for more accurate and efficient AI applications. Federated Learning (FL) has emerged as a very prominent collaborative learning technique to learn from such decentralized private data while reasonably satisfying the privacy constraints. To learn from such decentralized and massively distributed data, federated learning needs to overcome some unique challenges like system heterogeneity, statistical heterogeneity, communication, model heterogeneity, privacy, and security. In this article, to begin with, we explain some fundamentals of federated learning along with the definition and applications of FL. Subsequently, we further explain the unique challenges of FL while critically covering recently proposed approaches to handle them. Furthermore, this paper also discusses some relatively novel challenges for federated learning. To conclude, we discuss some future research directions in the domain of federated learning.

doi: *10.5829/ije.2021.34.07a.11*

1. INTRODUCTION

Recently, deep learning has gained an incredibly high peak of acceptance in artificial intelligence and machine learning research community. It has the ability to automatically extract and learn high-level complex features by the composition of low-level features. One of the most prominent features of deep learning, which typically makes it more attractive than traditional machine learning, is its remarkable ability to extract and sufficiently learn these complex features automatically. Where there is no need for hard-coded rules or domain expert knowledge or more intermediate steps to solve a problem. Deep learning has already outperformed the numerous traditional approaches in many fields including face detection, image recognition, speech recognition, health care, stock market prediction, and in many other fields [1-9]. The optimal performance of deep learning models greatly depends upon the availability of a significantly enormous amount of valuable data and the availability of high computational resources. So, to get a robust and efficient deep learning model, we typically need a large amount of valuable data and ample computational resources.

As smart devices (including smart mobile phones, tablets, and wearable devices) are being empowered with high computational resources including large memory storage and incredibly powerful sensors; people are rapidly switching their primary computing source from laptops and conventional desktop computers to these smart devices [10]. Particularly, the invention of Artificial Intelligence (AI) based smart-chips [11] have more significantly boosted this trend where companies' goal is to add the neural network power in smart devices. These devices are generating an enormous amount of valuable data including their location history, pictures, typing patterns, medical history, lifelogging data, etc. So, there is a lot of valuable real-world data, but in a decentralized fashion, which can be used to train deep learning models to get more accurate and intelligent applications.

*Corresponding Authors Institutional Email: hychan@usm.my (H.Y.Chan)

Please cite this article as: Z. Iqbal, H. Y. Chan, Concepts, Key Challenges and Open Problems of Federated Learning, International Journal of Engineering, Transactions A: Basics Vol. 34, No. 7, (2021) 1667-1683

Though these smart devices, collectively, possess a large amount of valuable data. However, usually, the nature of this data is highly sensitive. Therefore, due to different constraints including privacy concerns and privacy laws like GDPR [12], China's cybersecurity law [13], and California's privacy right act [14], it has become almost impossible for companies to collect, transfer, use or integrate users' data without their consent for any specific purpose.

Traditionally in distributed learning environment, to train a model, we typically accumulate all data at a central location, properly distribute it to separate parties for processing. But now, due to more privacy concerns of people and extremely strict privacy laws, it is almost impossible to collect updated real-time users' private data at the central location.

In such scenarios, it is intuitive to smartly leverage the private data of users stored locally and perform the necessary computation (model training) on these devices. Thus, ensuring the privacy guarantee of users' personal data and, on the other hand, also utilizing the computational resources of client devices. Different collaborative learning techniques [15-17] have been proposed to train deep learning models where different clients collaborate with each other to update their models by leveraging the learned knowledge of other clients rather than their private data. Specifically, a very promising decentralized learning technique called federated learning [16, 17] has been coined which has instantly attracted a large research community in Machine learning towards this research direction i.e. rather than transferring data to code (computing), we move the code to data.

FL has many advantages as compared to traditional distributed machine learning approaches [18-21] like *privacy*, where devices don't have to share their private data with other devices including a centralized server. *Low latency*, as devices would have updated model locally, so they do not need to wait for inferencing from cloud-server. *Huge computational resources*, as usually hundreds of devices, could participate in FL so a lot of computational resources would be available to train the model. Similarly, FL can help to more efficiently utilize the network bandwidth as, now, there is no need to transmit raw data to cloud-server rather just need to share the trained model parameters.

Though FL has emerged as a remarkably effective decentralized learning framework to leverage the massively distributed, highly unbalanced, and Non-independent and Identical Distribution (IID) private data of smart devices. Nevertheless, it comes with many (unique) challenges related to data, model architecture, communication, and privacy. Like, here, data is typically expected to be massively distributed, Non-IID, unbalanced, and inaccessible by other devices or centralized server due to privacy constraints. Similarly, Communication cost could be much higher as compared to computation cost and could experience challenges of limited and inconsistent bandwidth for various devices and of passive sampling. Furthermore, participating devices may naturally require specialized or more personalized models based on their specific requirements. Likewise, privacy is one of the primary foci of decentralized learning so local data of devices would be inaccessible to any other party. Key challenges of federated learning have been addressed thoroughly in section 2.

Extensive works [22-34] have been performed which more or less covered different aspects of FL effectively. However, most of them usually discussed FL in some particular context or discussed the core challenges of FL in a limited way like Yang et al. [22] put their major focus on different categories of FL based on a different distribution of data. Kulkarni et al. [23] put their primary focus on model personalization techniques for FL but do not discuss the other issues of FL. Xia [24] draws a comparison of FL with deep learning while putting the main focus on applying watermarking on deep neural networks in FL. Lyu et al. [25] put their main focus of discussion to the potential threats to FL. Similarly, Li et al. [26] provide a comprehensive survey on FL systems but they mostly discuss the design aspects of FL. Aledhari et al. [32] present a comprehensive survey of FL while focusing on protocols, applications, and use cases of FL in detail. Another comprehensive survey is performed by Li et al [33] but they do not adequately address the personalization issues where local clients may contain diverse model architectures. Similarly, these studies [25, 34] effectively cover the privacy and security aspects of FL but do not adequately cover the other challenges of FL. Furthermore, many authors [27-32] have explained FL in a particular context or some potential solutions in that particular context.

To put it concisely, there are many surveys on FL. However, most of them are tutorial-based or comprehensive in a particular context. Thus it stimulates us to perform a concise and comprehensive survey on FL adequately covering its key issues with possible solutions and future research directions.

The rest of the paper is divided into the following sections. Sections 1.1, 1.2, 1.3, and 1.4 present the basic flow, different frameworks, mathematical definition of FL, and applications of FL, respectively. In section 2, this paper comprehensively discusses the key challenges of FL including a critical overview of recent approaches to address those challenges. Section 3 discusses some open research areas and finally, the conclusion is presented in section 4.

1. 1. Basic Flow of Federated Learning Figure 1 illustrates the basic flow of FL. Here we typically assume that some clients want to collaborate for training a global model to perform some specific tasks. All participating devices collaborate with each other through a centralized server (aggregation server). In the first place, the centralized server forwards the copy of the global model to all active participants (active devices), then these devices train their copy of the global model on their

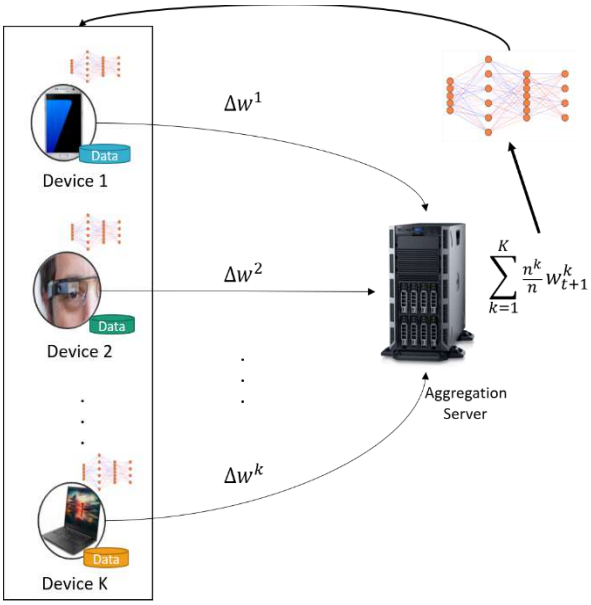


Figure 1. The basic architecture of Federated Learning

private data and send this updated model back to the server. After receiving all model updates from clients, a centralized server typically performs weighted aggregation on these local models' updates (parameters/gradients) to update the primary global model. Subsequently, the centralized server again sends this updated global model to all active clients, so active clients retrain this model. This process typically continues until the global model is converged.

1. 2. Different Frameworks To evaluate properly the performance of novel proposed solutions in federated learning, typically, a simulated environment is manually designed. For instance, datasets are manually split into different subsets in such a way these subsets mimic the behavior of Non-IID and unbalanced datasets. However, recently, different frameworks and benchmarks [35-44] have also been proposed for federated learning. Some of these frameworks also provide the federated datasets which fulfill the requirements of federated learning settings like distributed, unbalanced and Non-IID datasets. Similarly, some of these frameworks also provide the feature to compare different FL algorithms using different evaluation metrics. Though these datasets do not exactly mimic the real-world FL scenario still researchers can simulate their FL work using these datasets to mimic the behavior of near to real-world FL scenarios. A summary of these FL frameworks is given in Table 1.

1. 3. Definition of Federated Learning Suppose, K number of devices are participating in the federated learning process so dataset D is distributed among K devices as $D = \{D_1, D_2, D_3, .., D_k\}$ where each user i has a dataset $D_i (i \leq k)$, having n number of samples $\{(X_1^i, Y_1^i), (X_2^i, Y_2^i), (X_3^i, Y_3^i), \dots, (X_n^i, Y_n^i)\}$. Here each $X_i^k \in X_i$. Typically, the objective of federated learning is to minimize this objective function:

$$\min_w F(w) \quad where \quad F(w) = \sum_{k=1}^{K} \frac{n^k}{n} F_k(w) \tag{1}$$

TABLE 1. Some frameworks for federated learning Simulation

Approaches	Reference	Key Idea
TensorFlow Federated	[35]	An open-source framework that provides a platform for research experiments and large-scale simulation in FL. Additionally, it provides various federated datasets.
LEAF	[36]	Provide many datasets for benchmarking federated learning, MTL and Meta-Learning
PySyft	[37]	An open-source framework that combines FL and differential privacy, and integrates with deep learning frameworks like Keras, Tensorflow, or PyTorch to provide secure and private computations.
FATE	[38]	An open-source project which provides a secure computing framework
PaddleFL	[39]	Open-source framework based on PaddlePaddle which facilitates the researchers to compare different FL algorithms and deploy the FL system easily in large-scale distributed clusters.
NVIDIA Clara	[40]	A healthcare application framework but also provides full-stack GPU-accelerated libraries and SDKs to support FL.
OWKIN	[41]	Building a very large collaborative research network in health based on FL. They have also partially open-sourced their code for collaboration with other researchers and organizations.
Fedeval	[42]	It supports FedSGD and FedAvg algorithms. Use different evaluation features like accuracy, communication, time consumption, privacy, and robustness (ACTPR).
Fedml	[43]	Provide support for many machine-learning and FL algorithms. It supports mobile on-device training, distributed training, and standalone simulation.
OARF	[44]	Try to imitate real-world data distribution by collecting public datasets from distinct sources.

$where \quad F_k \ (w) = \frac{1}{n_k} \sum_{i \in |n_k|} f_i(w)$

Here n is the total number of samples in D while n_k is the total number of training samples in D_k on device $k \in K$. Here $f_i(w)$ represents the local objective function of each client k at a sample $i \in |n_k|$.

1. 4. Applications of Federated Learning Though the primary focus of federated learning is to learn from smart devices (cross-device) including smartphones, smart glasses, smartwatches, etc. where devices might have immensely valuable data but due to privacy concern. This enormous amount of valuable data is unavailable to train models for more accurate and efficient AI applications. But federated learning concept could be further extended to various organizations (cross-silo) like banks, hospitals which can collaborate with each other while preserving the confidentiality of users' data. For example, hospitals can securely collaborate with other hospitals or smart devices, containing health-related data, to train a shared global model for diagnosis or treatment of various medical disorders. Similarly, banks can also collaborate with each other, preserving the privacy of users' data, to train a shared global model for detecting the scam or fraudulent transactions.

There are many potential application areas [45-58] for federated learning including healthcare, sentiment analysis, recommendation systems, voice recognition, face detection, next-word prediction, predicting users' activities, autonomous vehicles, etc.

2. KEY CHALLENGES OF FEDERATED LEARNING

Federated learning has many unique challenges which typically make it different from traditional distribution optimization. These key challenges typically include model heterogeneity, statistical heterogeneity, communication (including system heterogeneity), and privacy. The following subsections concisely yet comprehensively explain these challenges with a critical review of recent approaches to handle these challenges.

2. 1. Statistical Heterogeneity In decentralized settings we typically assume, there is no centralized server to properly manage the distribution of data. Thus, it is very likely that clients would have highly unbalanced and Non-IID data as each device or user may have distinct preferences. For instance, let suppose there are two devices that want to collaborate in the federated learning scenario to train a unique global model (e.g., image classifier for fruit categories) while preserving the privacy of their data. Suppose the first device contains 50 samples of each of two classes (say apple and banana) while the second device contains 100 samples of each two classes (say orange and mango) so here these devices have unbalanced (varied number of samples for each class) and Non-IID (samples of distinct classes) data. Figure 2 presents an example of statistical heterogeneity.

FedAvg [17], a state of art algorithm based on SGD, shows that it can handle a certain amount of Non-IID data

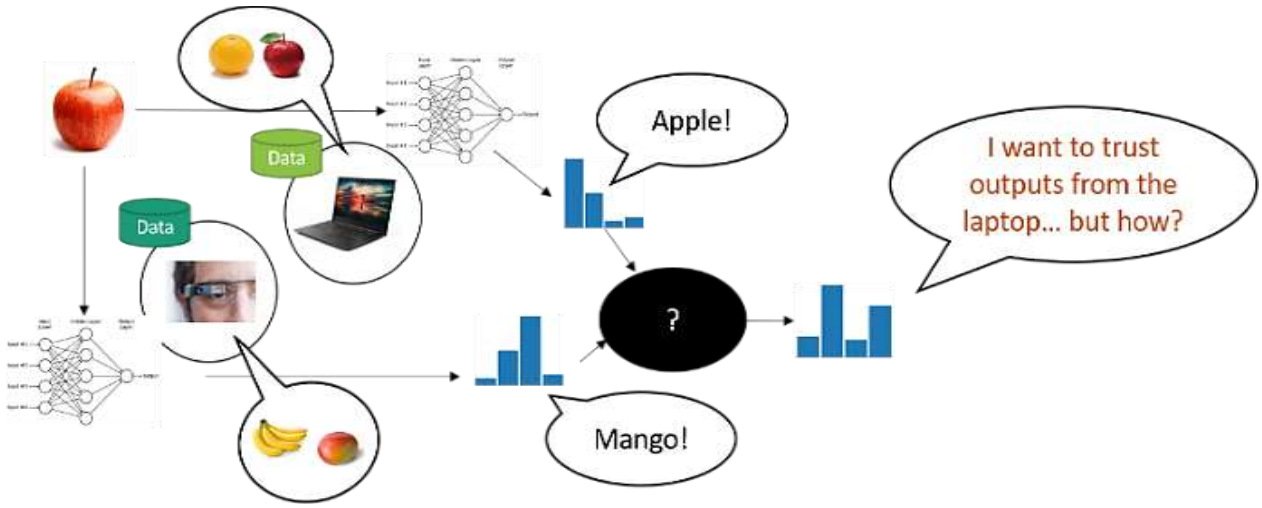


Figure 2. An example of Non-IID data. Suppose there are two devices, a laptop, and smart glasses, having their private data and a local model. The laptop has samples of orange and apple while smart glasses have samples of banana and mango. Here, they have Non-IID data distribution as they have samples of different classes. Now, suppose we are training a classifier to predict the fruit category. These local models would be well-trained to predict about classes for which they have training samples but not otherwise. For instance, if a sample of apple is provided to laptop-model then it would be reasonably confident that it is apple but if the same sample would be provided to the smart glasses model, then it would wrongly classify it like banana or mango. Now, the server needs to perform aggregation (as required in FL) but as you can see the server cannot take simple aggregation, rather it would like to give more preference to laptop output. But How? This is a simple example of a Non-IID problem in the context of FL.

TABLE 2. Some approaches addressing statistical heterogeneity

Approaches	Reference	Key Idea
Sharing some data	[59]	Using the conditional GAN to produce the missing label samples on devices by leveraging the private data of devices.
	[60]	Reducing the weight divergence b/w different distributions by leveraging the globally shared data
	[61]	Devices remove their distribution imbalance by data augmentation before local training.
Multitask Learning	[62]	Proposed MOCHA and show that MTL is a natural way to handle the statistical challenge.
	[63]	Employ the concept of Bayesian network and perform variational inference during learning.

but Smith et al. [61] have shown empirically that for high skewed Non-IID data, the performance of the convolutional neural network, trained using FedAvg can drop reasonably by 51% on CIFAR-10, 11% on MNIST and 55% for keyword spotting datasets. Therefore, FedAvg does not effectively handle the skewed Non-IID data which is natural and expected data distribution in the federated learning setting.

Some researchers proposed the idea to share some of the local data of devices or share some proxy data to handle the statistical heterogeneity [59, 60] to make the data distributions of devices as IID. Like, Zhao et al. [60] show that accuracy reduction, in the case of Non-IID data, could be attributed to *weight divergence* when two different training processes having the same weight initialization get different weights. They propose that if we can leverage the globally shared data (having uniform distribution over all classes) by distributing it to all clients, then it can reduce the weight divergence between distribution on the different devices. This weight divergence could be quantified using EMD (Earth Moving Distance) and in return, it would increase the accuracy of the model. Their approach does not look much practical as arranging and communicating uniform distribution of data over all classes could be challenging and can create overhead for communication.

Jeong et al. [59] proposed FAug (Federated Augmentation) that uses the concept of conditional Generative Adversarial Network (GAN) to produce the missing label samples on client devices by data augmentation. Where each client is required to identify and upload the missing target labels, in its distribution, to the server. Server oversamples these target labels to train the conditional GAN.

Finally, all devices download this trained GAN to produce missing target labels in their distribution. As target labels of each device may reveal some private and sensitive information with the server or with other devices (which have GAN, trained on all devices' data and that could be used to infer the other's target labels), therefore it requires all devices to additionally upload the redundant samples, other than target labels, on the server to handle the privacy issue at the cost of extra communication overhead. However, this method works with the assumption that client devices would agree to share their private data with the server. This seems an almost impractical solution and violates the key idea of FL i.e., privacy.

Some researchers [62-65] have shown that the natural way to address the statistical challenge (Non-IID) of data is Multitask Learning (MTL) where the goal is to learn from each node, having separate but related models, simultaneously. Here, each node represents a task that possesses their private data and the goal is to learn from these related but different tasks. Like Smith et al. [62] used the MTL in the federated learning setting. In MTL, an additional term is included in the loss function to model the relationship among tasks. They used the correlation matrix to measure the client similarity and trained separate but related models for each device (task) using a shared representation on the server. However, their method only works for convex optimization problems and is not scalable to a large population. Furthermore, Lim et al. [66] argue that this approach is not much suitable in federated learning scenarios when a specific task (model) doesn't possess its local data or may have very few training samples.

Corinzia et al. [63] employ the concept of a Bayesian network to connect all clients with the server and perform variational inference during learning. Their method can properly handle the non-convex problem, but it is much costly to scale it to a vast federated network as it refines the client models sequentially.

Duan et al. [61] revealed that model performance could also deteriorate due to global imbalance (when local distributions of data across all clients have class imbalance). It first removes the global imbalance by data augmentation where all devices first share their data distribution with the centralized server. Then, before performing local model training, each device first performs data augmentation on imbalance classes to make a balanced distribution. Subsequently, it employs the concept of mediators to combine training samples of relevant devices (selection is performed by calculating KL divergence between local and uniform distribution) based on their distributions to make it a uniform distribution. So, finally, this combined training (model) is shared with the global server for federated aggregation.

Recently, some researchers [67-71] have identified the limitations of the standard FedAvg algorithm, particularly, when clients have statistical heterogeneity. Lim et al. [66] questioned the performance of the standard FedAvg algorithm and suggested that standard aggregation is probably not the best aggregation way. By using the Mutual Information (MI) and different distance metrics, they demonstrate that with the increase in the number of iterations, correlation (MI) increases but in parallel, the distance of parameters also increases. Similarly, Xiao et al. [67] mention the three limitations of FedAvg i.e. 1) it cannot be applied on non-differentiable methods, 2) it usually requires many communications rounds, and 3) it is primarily designed for the cross-device setting. While addressing these limitations, they propose FedKT for cross-silo scenarios which can learn from both differentiable and non-differentiable models. Li et al. [68] explain that due to permutation invariance of NN, simple model parameter aggregation (FedAvg) may have a very negative impact so they propose PFNM, a probabilistic Federated Neural Matching algorithm that performs the matching among clients' NN neurons before averaging them. Yurochkin et al. [69] further extend this approach and propose a layer-wise matching approach (FedMA) and apply this approach to modern CNNs and LSTMs. Wang et al. [70] try to reduce Aggregation Error (AE) by constructing a definitely convex global posterior using a Gaussian product method to obtain the global expectation and co-variance by multiplying local posteriors. On the client-side, they proposed a new Federated Online Laplace Approximation (FOLA) method to obtain online local posterior probabilistic parameters which can directly be leveraged in the FL framework. Table 2 shows some recent approaches addressing statistical heterogeneity.

2. 2. Model Heterogeneity/Personalization As shown in Figure 3, typically, system heterogeneity is defined as where devices possess varied computational resources like different memory, processor, battery limit, active time, etc. More specifically, in model heterogeneity cases, due to varied computational resources and different business needs (trade-off between speed and accuracy), it is intuitive that devices may have the varying size of deep networks (different no. of layers) or may have completely different network architectures like some devices may be using CNN, some device may be using ResNet while some devices may opt for Inception.

Having the same model architecture would not only overburden the communication (already facing high communication challenges in federated learning) but would also increase the computation complexity for devices where low resourced devices may result in the form of stragglers or staled data. Probably, some devices might possess immensely valuable data but unable to train the same complex model. Therefore, it is intuitive that models should possess the proper number of nodes in their output layer to avoid unnecessary computation and communication overhead.

In this section, we focus on model heterogeneity, and regarding system (hardware) heterogeneity, the comprehensive discussion is performed in section 2.3. Table 3 presents some recent works to address model heterogeneity.

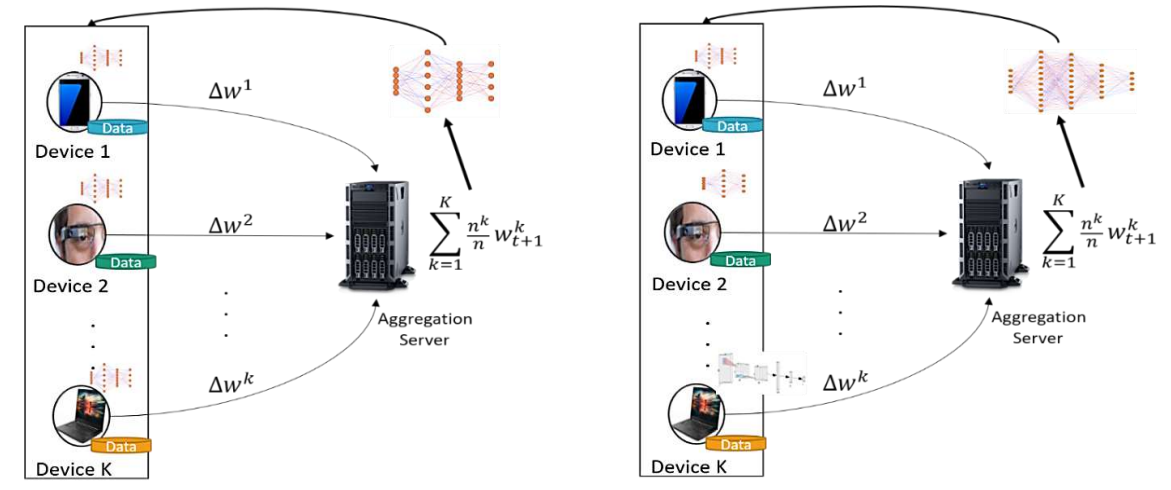


Figure 3. This illustrates the difference between (a) standard federated learning where all devices are required to have the same model architecture and (b) a More practical scenario of federated learning where different users might have different model architectures based on their computational resources and business needs

TABLE 3. Some approaches addressing model personalization/heterogenity

Approaches	Reference	Key Idea
Retraining global model	[72]	Retrain global model on client's private data.
Transfer Learning	[73-75]	Model personalization using transfer learning.
Meta Learning	[76-79]	Typically, the model is made adaptive by training it on multiple tasks.
Multitask Learning	[63]	Employs Bayesian Network and performs variational inference during learning.
	[62]	Extends MTL and uses a modified loss to represent relationships.
Knowledge Distillation	[80]	Allows the model heterogeneity.
	[81]	Address the model heterogeneity with Non-iid data.

We can divide the model heterogeneity into two general categories 1) where different models need to be personalized based on different geographical or personal preferences like in the case of next word predication, for a sentence "I love to visit ….", there would be customized predictions for different users living in separate geographical location or with distinct preferences. 2) where various models might have diverse architecture due to varied computational resources or different business needs. For instance, one might be using CNN with 5 layers, the second is using CNN with 10 layers, the other is using Random Forest, etc.

The primary goal of federated learning is to train a unique global model by leveraging the multiple clients' private data and their computational resources. In FL, it is assumed that all devices would be able to train the same copy of local solver (same model architecture) on their private data, and then these model's updates are sent to the aggregating server which further aggregates these updates.

Moreover, standard FL works with the assumption that all clients would possess reasonable computational and communicational resources to train the same model architecture. However, in the real scenario, this key assumption of FL does not seem logical as more complex deep learning models are being developed to achieve more accurate performance on real-world tasks. Furthermore, intuitively, all devices cannot be capable to train the same complex model. This incapability of models could be due to limited computational resources of devices or due to varying business needs (trade-off between accuracy and speed) or to reduce the computational and communication overhead.

Due, to such primary assumption of FL, most of the work has been performed with the same assumption of homogeneous models and most of the work has been performed for category 1 of model heterogeneity (model personalization based on personal preferences or distinct geographical locations) and very few works have addressed this potential problem of category 2 of model heterogeneity (clients contain diverse model architecture) and in a significantly limited fashion. To make a global model personalized, most personalization techniques suggest retraining the (collaboratively trained) global model on the users' local private data [72]. Some researchers have proposed model personalization approaches using transfer learning [73-75]. In transfer learning, usually, the last layers of a trained model are replaced with new layers to leverage the learned knowledge of the trained model on some new tasks. Some researchers suggest freezing the initial layers of a trained global model and retrained only the last few layers on local private data of individual clients.

Recently, some researchers [76-79] have also leveraged meta-learning to solve the personalization problem. Meta-learning is generally defined as "Learning to Learn" where a model is made adaptive by training it on multiple tasks in such a way that it can learn new tasks by providing very few examples of new tasks.

Multitask learning [82] has also been widely used by different researchers [62-65, 83, 84] to address the model personalization challenge. They leverage the distributed Multitask Learning to train separate but related models. They tried to train a personalized model for each distribution (as in FL, we assume the Non-IID distribution). Like, Smith et al. [62] extended the MTL in the federated learning setting. To model the relationship among tasks, they include an additional term in the loss function. They use the correlation matrix to measure the client similarity and train separate but related models for each device (task) using a shared representation on the server. However, their method only works for convex optimization problems and is not scalable to a massive population. Similarly, Corinzia et al. [63] employ the concept of a Bayesian network to connect all clients with the server and performs variational inference during learning. Their method can handle the non-convex problem, but it is much costly to scale it to a large federated network as it refines the client models sequentially. In addition to limitations of scalability and feasibility, these approaches do not address the other model heterogeneity scenarios like models having entirely diverse architectures and having different output layers.

Knowledge distillation [85] is a technique to distill the knowledge from a pre-trained cumbersome (teacher) model into a small (student) model so that the student model can also mimic the behavior of the teacher model. Li et al. [80] leverage the knowledge distillation to allow clients to use customized local models having a different number of layers. Therefore, it allows the clients to use diverse model architecture while collaborating in Federating Learning. But they have not evaluated their method with totally different model architectures.

Recently, Ma et al. [81] have proposed a very effective adaptive distillation approach where they address the model heterogeneity problem using Non-iid data. More specifically, they trained the local model using their own private labeled data whilst trained the global model using unlabeled Non-iid public data.

2. 3. Communication One of the primary challenges of federated learning is the Communication overhead of downloading the global model's parameters (or gradients) from the centralized server and then uploading the trained model's parameters (or gradients) back to the server in each round of communication. This communication overhead is proportional to the model size means for large-scale models; the number of model parameters could be in millions. More specifically, it becomes more challenging and extremely costly for devices having limited bandwidth and intermittent network connections. Because, intuitively, smart devices might possess good computational capabilities. However, these devices are likely to have different network bandwidth like some devices may support significantly efficient networks say 4G, 5G, or Wi-Fi, and some devices may only support significantly poor network connections. Similarly, some devices may bear good and stable connections whilst mostly mobile devices may bear unstable and intermittent limited connections in the real scenario. These network limitations including unpredictable network interruption could also result in the form of stragglers and passive sampling.

Researchers of naïve federated learning approach [16] have shown that the size of the update (parameters) would be independent of local training data of individual devices and the global model can be trained in few communication rounds. Therefore, as compared to sharing training datasets of all active devices with the centralized server, FL reduces the communication cost by order of magnitude. However, due to unpredictable network limitations of participating devices in FL, it is still an open challenge to reduce the communication cost during each round of communication with the centralized server. They have also proposed different methods like sketched updates to reduce the communication cost in the order of magnitude as compared to sharing data.

Usually, in centralized learning approach, there is less communication cost as compared to computation cost but, on the contrary, in decentralized learning approach, communication cost becomes more challenge as compared to computation cost because modern smart devices, generally, have more high processing power while having very less training data (as a fraction of total dataset) to be trained [86]. Therefore, we might, loosely, say that computation almost becomes free as compared to communication for many model types.

Intuitively, there are two straightforward approaches to reduce this communication cost; one is to add more parallelism by including more clients in each round of communication. The second approach is to add more complex computation on each device like perform many gradient steps on each client, instead of one, in each communication round. McMahan et al. [17] have empirically shown that adding more clients results in diminishing return after a particular limit but, in contrast, adding more complex computation produces more promising results.

Many works have been performed to reduce this communication cost of transferring the large weight matrices of deep networks, to handle unexpected interruption or dropout of participating devices, and synchronization latency caused by the computing power and network connectivity constraints. Asynchronous SGD [18, 87-93] also tries to handle this communication bottleneck. It accelerates the training process by updating the parameters immediately after a computing node has sent its gradients (asynchronous communication) instead of waiting for all computing nodes to send their gradients before updating parameters. Albeit it speeds up the training process and handles the stragglers' problem, but it comes with the staled gradient problem which could also affect the accuracy of the model. Some work has been performed to address the issues of active sampling where in each round of communication, clients fulfilling a particular criterion is selected rather than passive sampling where there are no criteria to select the effective clients like Nishio et al. [103] proposed a resource-based active device sampling technique for heterogeneous clients where the FL server first sends the resource request to maximum clients to get information about their available resources.

Subsequently, only those clients are selected who can complete the training process within a specified threshold. By selecting the maximum number of clients in each round, it assists the global model to attain high accuracy. But they do not address the data distribution issue like some devices, having high computational resources, might not have much data.

To reduce communication bandwidth, model compression schemes [95, 96] are also being used to reduce the model size, for communication, using different techniques. But most of the compression schemes work in the data center environment and system challenges of federated learning like participation of devices could be low, Non-IID data and local update schemes introduce new challenges for these techniques. Some researchers have proposed gradient quantization

methods [16, 98, 99], where gradients are quantized to low precision values to force the updating model to be sparse and low rank. Some of these approaches [98] could be difficult to extend to federated learning as errors accumulated locally may become staled if devices could not participate frequently in communication rounds. Some researchers have also proposed the gradient sparsification methods [100] where all gradients are not sent to the server rather only those gradients larger than a predefined specific threshold are sent to the server. As it could be much challenging to choose a correct threshold so some other techniques in gradient sparsification are also proposed like gradient dropping [101] to drop gradients after a particular absolute value or automatically tuning the compression rate based on the activity of local gradients [104]. Lin et al. [102] proposed a Deep Gradient Compression; a gradient compressing technique that employs various methods like momentum correction, local gradient clipping, momentum factor masking, and warm-up training to reduce the communication bandwidth (which could also result in the staleness problem) by two-order of magnitude without compromising the accuracy of the model. The author discussed the scenario of federated learning in his paper, so this technique could be extended in the federated learning setting.

Some researchers have leveraged distributed multi-task learning [62, 64, 83, 94, 105, 106] for communication efficient learning in the distributed environment but most of them do not mitigate the system challenges of the federated learning environment. Like Baytas et al. [94] allows for asynchronous updates to handle stragglers but doesn't address the fault-tolerance problem and it gives the convergence guarantee on the assumption of bounded delay which is not possible in federated learning setting where devices may experience significant delays or drop out completely due to sudden network interruption. Liu et al. [83], extend the distributed framework COCOA [107] to learn the relationship among distributed tasks along with the predictive models for each task but they do not explore the federated settings and make the assumption that data distribution is balanced means each device would perform a similar amount of work.

Smith et al. [62] have proposed an alternative of FedAvg (standard federated learning algorithm), called MOCHA for federated multitask learning setting. They more significantly extend the communication efficient algorithm COCOA to address the problems of fault tolerance and stragglers in the federated learning setting. But the effectiveness of their problem is limited to only convex problems.

Jeong et al. [59], have employed the concept of online distillation in federated learning. Online distillation is a simple model compression technique, called Codistillation [108] to reduce the communication overhead by many orders of magnitude. In this approach, communication overhead does not depend upon the model size or dataset size rather it depends upon the output dimension of the model. It shows that they have reduced the communication overhead around 26x as compared to the standard Federated Averaging [17] approach where parameters or gradients of models are shared with the parameter server. Similarly, Guha et al. [97] proposed one-shot federated learning. They additionally use the concept of ensemble learning and knowledge distillation to reduce the communication for the convex optimization problem. Table 4 presents some recent works to address communication challenge.

TABLE 4. Some approaches addressing communication challenges

Challenge	Approaches	Reference	Key Idea	Target
Fault tolerance, Stragglers	Asynchronous SGD	[18]	Perform the aggregation immediately after receiving data from any device	Synchronization latency due to low resourced devices, stragglers
	MTL approaches	[94]	Asynchronous distributed MTL	Stragglers
		[62]	Proposed MOCHA, generalizing the COCOA	Communication cost, stragglers, fault tolerance
		[64]	Extend the communication efficient algorithm COCOA.	Fault tolerance, stragglers
Communication overhead	Model compression schemes	[95, 96]	Model compression schemes	Reduce communication overhead
		[59]	Co-distillation, a model compression scheme	Reduce communication overhead
		[97]	Proposed one-shot federated learning using ensemble learning.	Reduce communication overhead
	Gradient compression schemes	[16, 98, 99]	Gradients are quantized to low precision values	Reduce communication overhead
		[100]	Gradient sparsification; only gradients larger than a particular threshold are sent to the server.	Reduce communication overhead
		[101]	Gradient dropping; drop gradients after specific threshold value.	Reduce communication overhead
		[102]	Employ various gradient compression schemes like momentum correction, local gradient clipping, momentum factor masking, and warm-up training.	Reduce communication overhead

2. 4. Privacy/Security It is the primary motivation behind federated learning, where we want to learn from user's data while securing the privacy of users' data. Typically, we can categorize privacy issues in three subproblems [15] as *privacy of input data*, the *privacy of the trained model*, and *privacy of the model's output*. Federated learning naturally ensures the privacy of input data as it only works on learned parameters from the private data rather than the original raw data. Though FL, try to secure the private data of clients by sharing only the trained parameters; however, recent studies [109, 110] have revealed that valuable information about client's training data could be reasonably inferred from its trained model update (learned parameters) with very high accuracy of up to 90%. It clearly explains that despite not utilizing private data in FL, still, private information of clients is vulnerable. Therefore, many solutions have been proposed to address various kinds of vulnerabilities.

Typically, we can categorize security issues into three general categories including data poisoning, model poisoning, and evasion attacks. In a *data poisoning attack*, some adversary clients may intentionally use the malicious data samples to mislead the global model by providing their local models trained on malicious data. In a *model poisoning attack*, adversaries' target is to misguide the machine learning models to produce malicious results [111]. Model poisoning attacks can further be divided into two categories. targeted and un-targeted adversarial attacks. *Targeted attacks*, also called backdoor attacks, usually do not compromise the overall accuracy of the global model rather just focus on some specific classes or examples. On the other hand, *un-targeted attacks* try to mislead the whole global model [112, 113]. In *evasion attacks*, also called inference-time attacks, adversaries puzzle the deployed machine learning model by providing such misleading and modified samples which seemingly look like original test samples [114].

To preserve the information from other clients, one of the most commonly applied privacy-preserving techniques is Differential Privacy (DP) [15, 115-117] due to its simple algorithm and relatively low communication overhead. In this technique, some noise is added in trained parameters of the model before uploading them to the aggregation server to make it impossible for the third party to distinguish the individuals. Abadi et al. [115] proposed a technique for deep learning algorithms where it includes a noise to trained parameters, before forwarding them to the server, using a Gaussian distribution. Similarly, Geyer et al. [116] further enhanced this method by introducing two steps; in the first step, the server selects some random clients to participate in a communication round, and then only those randomly selected clients would include noise in their trained parameters using Gaussian distribution before forwarding them to server. Hence, in this way, other participating clients would be unable to know which clients are participating in this round. so would not be able to infer the information from shared parameters of the global model. Differential privacy comes with the cost of a reduction in the accuracy. There exists a tradeoff between model accuracy and differential privacy because when we include more noise to ensure more privacy, it results in the reduction of model accuracy significantly. In the same way, we also need to consider the tradeoff between device performance and DP as system computing resources would be required for applying DP.

Another effective approach for privacy-preserving of distributed datasets is Secure Multiparty Computation (SMC) [118-123] where multiple parties collaboratively compute a function using their inputs without revealing their private inputs to other parties. This additionally requires extra computation and communication overhead. In addition, in this approach, typically, a minimum portion of users' data must be shared.

Similarly, fully homomorphic encryption and its variants [52, 123-127] are also being employed to improve the security of trained models. In this method, participating clients can only see the encrypted data and they need to perform some computation on this encrypted data. Results are sent to the owner and usually, only the owner has a private key to decrypt the data. Usually, homomorphic encryption can be divided into three categorized based on the number of operations allowed to perform on encrypted data. 1) Partially homomorphic encryption (PHE) 2) Somewhat homomorphic encryption (SWHE) and 3) Fully homomorphic encryption (FHE). Acar et al. [128], the authors have presented a comprehensive survey on homomorphic encryption schemes.

Bonawitz et al. [122] proposed a secure, failure-robust, and communication efficient protocol. The goal is to learn from a significant number of mobile devices by aggregating their contribution in a secured manner to prevent identifying the individual's contribution in collaborative learning. Though they proposed their work as a general secure communication protocol, however, they suggest that their method could be employed in federated learning settings where clients share their model rather than their private data, and these models need to be aggregated securely.

Another potential approach to address the privacy issue of FL is proposed by Mandal et al. [123] where authors guarantee the model and data privacy for user and server using homomorphic encryption. In addition, Xu et al. [129] proposed VerifyNet, which primarily addresses two problems 1) How to protect user's privacy during the training process and 2) How to trust the results from the server i.e. verification of aggregated results from the server. In their proposed approach, the server is supposed to provide the proof of correctness to all clients and there is almost no possibility of forging proof as the adversary needs to solve the adopted NP-hard problem to create a forging proof.

Chen et al. [52] proposed FedOpt to address the privacy and communication efficiency challenges of FL. They also design a Sparse Compression Algorithm (SCA) for communication efficiency and then integrate the additively homomorphic encryption with differential privacy to prevent data from being leaked. An overview of some recent approaches is given in Table 5.

3. FUTURE DIRECTIONS/OPEN CHALLENGES

In recent years, federated learning has emerged as an extremely promising domain to collaboratively learn from highly unbalanced, massively distributed Non-IID data. A lot of research work has been performed to address the numerous challenges of federated learning. However as critically discussed in section 2, there are many open research problems to be considered for more practical scenarios of federated learning. Some of them are discussed below.

3. 1. Supervised/Unsupervised Training Most of the work done in the field of FL is based on supervised learning where it is assumed that all devices have correct labels against their training samples. However, in a real scenario, some devices might contain unlabeled data or might bear incorrect labels against their data. Similarly, some devices might consist of missing classes in their training data. Therefore, it would be challenging for the global model to learn effectively or indicate similar confidence to all participating clients.

Recently, some researchers [130, 131] have leveraged semi-supervised and unsupervised learning in FL. In semi-supervised learning, it is supposed that very few label data are available, so a model is trained on both (available) labeled data and (public) unlabeled data. While in unsupervised learning, it is supposed that no label data is available to train the model, so the model is trained solely on unlabeled data. Recently, contrastive learning [132-135] has become a hot research direction to learn the representations of unlabeled data by training the model on the unlabeled data. Contrastive learning (also referred to as self-learning) is a pre-training process where the model tries to learn similar and distinct data samples in an unlabeled data distribution. Recently, Chen et al. [134] present an effective contrastive learning framework (simCLR) that outperforms the other state-of-the-art algorithms [132, 136, 137]. Very few works have been carried out to apply the FL with semi-supervised and unsupervised data. Therefore, it is still an open research area to be explored in FL settings.

3. 2. Synchronous/Asynchronous FL Typically, the most standard approach in current FL work is synchronous FL where the aggregation server waits for local updates from all participating devices. Then server performs aggregation on all these updates. However, it can result in a straggler effect where some devices might be significantly slower due to their local capacities like limited computational resources, limited power, or low bandwidth internet. Another potential direction is to use FL in the asynchronous mode [87-93] where the aggregation server would perform aggregation as it receives updates from participating devices. Though it could handle the straggler effects and could provide the flexibility to devices to join the FL process in halfway; however, it could result in the form of staled updates; where some devices with low capabilities might send the outdated outputs to the server which can affect the convergence of the global model. This asynchronous scheme has been applied successfully in centralized distributed computing where the bounded delay is expected. However, in the practical scenario of FL, it could be almost impossible to set a fixed bounded delay because, in FL settings, the delay could be up to a couple of hours or days.

Furthermore, as discussed above, that practical FL also experiences the challenges of system heterogeneity due to varied computational resources, memory, network infrastructure. Therefore, there are strong possibilities that some devices might drop during the FL process.

TABLE 5. Some approaches addressing privacy/security challenge

Approaches	Reference	Key Idea
Differential privacy	[115]	Include noise to trained parameters before aggregation.
	[116]	Two-fold protection, noise is added to random clients before aggregation.
	[52]	Integrate the additively homomorphic encryption with differential privacy.
Secure Multiparty Computation	[118-121]	Multiple parties compute collaboratively compute a function using their inputs.
	[122]	Proposed MOCHA and show that MTL is a natural way to handle the statistical challenge.
	[123]	Train their models based on an additive homomorphic encryption (HE) scheme and an aggregation protocol.
Homomorphic encryption	[52, 123-127]	Clients are required to perform some computations on encrypted data and only the owner can decrypt the data using the private key.

Therefore, FL systems should be capable to consider fault tolerance. Some researchers suggest permanently dropping such devices from the training process. However, it may result in the form of biased model training due to the loss of data samples possessing some specific characteristics. This drop-out (stateless) problem also led to other challenges of privacy and security where proposed solutions for privacy/security require stateful clients for validation and verification of trusted clients. Like, it could be more challenging for secure aggregation to work effectively if a significant number of clients suddenly leave the training process. Recently, some works have tried to address this issue like Wu et al. [89] proposed a semi-asynchronous learning client selection algorithm and a lag-tolerant mechanism to overcome different challenges like stragglers and model staleness in FL. Li et al. [93] introduce asynchronous updater to actively receive the unsynchronized local weights from stragglers. However, new algorithms and methods are required to address these challenges more effectively.

3. 3. Privacy/Security Privacy is the fundamental issue in FL and a lot of work has been done to address this issue but most of those approaches are not effectively applicable in real-world scenarios due to the limitation of computational and network resources including the availability of all devices in each round of communication (as discussed in section 2.3). For instance, secure aggregation is an effective approach to ensure privacy. Nevertheless, it has become a hurdle for other defense approaches to implement because the server can merely examine the aggregated outputs and usually unable to differentiate the outputs of clients.

As mentioned in section 2.4, there are many security challenges for FL including data poisoning, evasion attacks, and model poisoning [111-114]. Similarly, there are many other methods to target the FL system like some devices might work as free-rider [138]; where devices do not have relevant training samples but indeed want to get benefit from other devices. In such a way, it would equally compel other devices to provide more computational resources for the FL process. Contrary to centralized learning, it is challenging to figure out the trust mechanism for participating clients to identify whether clients are trustworthy or not. Though some approaches [139] have been proposed to address some of these issues, however, it usually requires a trade-off between performance and security of the FL system so still, there is a need for more robust and effective methods to address these challenges.

3. 4. Model Heterogeneity This paper also discusses some relatively new challenges for practical FL as discussed in section 2.2. Though model personalization is not a new research area, and some research works [62-65, 73-79, 83, 84] have been done in this area but very few works [80, 81] have been done to address the model heterogeneity challenge where devices might have entirely different model architectures based on their computational resources or business needs. This scenario becomes more challenging if we combine the missing classes problem with model heterogeneity.

3. 5. Different Federated Aggregation Algorithms As discussed in section 2.1, recently, some researchers have identified a relatively new challenge in FL where they demonstrate that the standard FL aggregation algorithm (FedAvg) is not the most effective method for clients' model aggregation particularly when clients have different data distributions. In the same way, FedAvg has many limitations including its applicability limitation to non-differentiable methods and its computation and communication overhead. Although many works have been, recently proposed to address the limitations of FedAvg including many alternatives of FedAvg like FedProx [28], FedMA [70], Scaffold [140]. Nevertheless, they still experience some limitations and can only be applied with some strong assumptions. Therefore, there is a need for some more robust approaches to address such challenges.

4. CONCLUSION

Federated learning has emerged as the de facto decentralized learning framework in privacy-preserving scenarios where the training data is not directly accessible. In federated learning, some devices train a statistical model on their private data and share only this trained model with centralized/aggregation server for collaborative learning. Though, federated learning is proven to be immensely beneficial in various domains including cross-device and cross-silo. But it equally has many challenges including system heterogeneity, statistical heterogeneity, communication, security, privacy, and model heterogeneity/personalization. This article briefly provides an overview of federated learning including its potential application areas. Subsequently, it comprehensively discusses its unique key challenges and a critical review of recent approaches to address these key challenges. Some relatively new challenges, like model heterogeneity and (global model) aggregation error, are also discussed in detail with potential approaches to address these issues. Furthermore, deriving from the key challenges' discussion, some open research areas are also discussed, in section 3, to be explored by the federated learning research community.

5. ACKNOWLEDGMENT

This research is sponsored by research grant (Malaysia Research University Network) 203.PKOMP.6777003.

6. REFERENCES

1. Taigman, Y., Yang, M., Ranzato, M.A. and Wolf, L., "Deepface: Closing the gap to human-level performance in face verification", in Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, IEEE, (2014), 1701-1708.
2. Simard, P.Y., Steinkraus, D. and Platt, J.C., "Best practices for convolutional neural networks applied to visual document analysis", in Proceedings of the International Conference on Document Analysis and Recognition, ICDAR, (2003), 958-963.
3. Salehi, S.M.M. and Pouyan, A.A., "Detecting overlapping communities in social networks using deep learning", ***International Journal of Engineering Transactions C: Aspects***, Vol. 33, No. 3, (2020), 366-376, DOI: 10.5829/ije.2020.33.03c.01.
4. Hannun, A., Case, C., Casper, J., Catanzaro, B., Diamos, G., Elsen, E., Prenger, R., Satheesh, S., Sengupta, S., Coates, A. and Ng, A.Y., *Deep speech: Scaling up end-to-end speech recognition*, in *ArXiv*. 2014.
5. Gheitasi, A., Farsi, H. and Mohamadzadeh, S., "Estimation of hand skeletal postures by using deep convolutional neural networks", ***International Journal of Engineering, Transactions A: Basics***, Vol. 33, No. 4, (2020), 552-559, DOI: 10.5829/ije.2020.33.04a.06.
6. Shaeiri, Z. and Kazemitabar, S.J., "Fast unsupervised automobile insurance fraud detection based on spectral ranking of anomalies", ***International Journal of Engineering, Transactions A: Basics***, Vol. 33, No. 7, (2020), 1240-1248, DOI: 10.5829/ije.2020.33.07a.10.
7. Savadi Hosseini, M. and Ghaderi, F., "A hybrid deep learning architecture using 3d cnns and grus for human action recognition", ***International Journal of Engineering, Transactions B: Applications***, Vol. 33, No. 6, (2020), 959-965, DOI: 10.5829/ije.2020.33.05b.29.
8. Sezavar, A., Farsi, H. and Mohamadzadeh, S., "A modified grasshopper optimization algorithm combined with convolutional neural network for content based image retrieval", ***International Journal of Engineering, Transactions A: Basics***, Vol. 32, No. 7, (2019), 924-930, DOI: 10.5829/ije.2019.32.07a.04.
9. Ghanbari Sorkhi, A., Hassanpour, H. and fateh, m., "Improvement of the r-fcn's deep network in object detection and annotation", ***Journal of Machine Vision and Image Processing***, Vol. 6, No. 2, (2020), 43-59.
10. Poushter, J. "Smartphone ownership and internet usage continues to climb in emerging economies." Pew research center 22, no. 1 (2016): 1-44.
11. Neuromation, *What's the deal with "ai chips" in the latest smartphones?* 2018.
12. Blackmer, W.S., *Eu general data protection regulation*. 2018.
13. KPMG, *Overview of china's cybersecurity law. Kpmg advisory (china) limited*. 2017.
14. *California privacy rights act | californians for consumer privacy*. 2020.
15. Shokri, R. and Shmatikov, V., "Privacy-preserving deep learning", in Proceedings of the ACM Conference on Computer and Communications Security, New York, USA, ACM Press, (2015), 1310-1321.
16. Konečný, J., McMahan, H.B., Yu, F.X., Richtárik, P., Suresh, A.T. and Bacon, D., *Federated learning: Strategies for improving communication efficiency*, in *ArXiv*. 2016.
17. Brendan McMahan, H., Moore, E., Ramage, D., Hampson, S. and Agüera y Arcas, B., "Communication-efficient learning of deep networks from decentralized data", in Proceedings of the 20th International Conference on Artificial Intelligence and Statistics, AISTATS 2017, (2017).
18. Dean, J., Corrado, G.S., Monga, R. and Chen, K., "Large scale distributed deep networks", in Neural Information Processing Systems'12, (2012), 1223-1231.
19. Shamir, O., Srebro, N. and Zhang, T., "Communication-efficient distributed optimization using an approximate newton-type method", in 31st International Conference on Machine Learning, ICML 2014, (2014), 2665-2681.
20. Reddi, S.J., Konečný, J., Richtárik, P., Póczós, B. and Smola, A., *Aide: Fast and communication efficient distributed optimization*, in *ArXiv*. 2016.
21. Ma, C., Konečný, J., Jaggi, M., Smith, V., Jordan, M.I., Richtárik, P. and Takáč, M., "Distributed optimization with arbitrary local solvers", ***Optimization Methods and Software***, Vol. 32, No. 4, (2017), 813-848, DOI: 10.1080/10556788.2016.1278445.
22. Yang, Q., Liu, Y., Chen, T. and Tong, Y., "Federated machine learning: Concept and applications", ***ACM Transactions on Intelligent Systems and Technology***, Vol. 10, No. 2, (2019), 1-19, DOI: 10.1145/3298981.
23. Kulkarni, V., Kulkarni, M. and Pant, A., "Survey of personalization techniques for federated learning", in Proceedings of the World Conference on Smart Trends in Systems, Security and Sustainability, WS4 2020, Institute of Electrical and Electronics Engineers Inc., (2020), 794-797.
24. Xia, Y., "Watermarking federated deep neural network models", (2020),
25. Lyu, L., Yu, H., Zhao, J. and Yang, Q., Threats to federated learning, in Lecture notes in computer science (including subseries lecture notes in artificial intelligence and lecture notes in bioinformatics). 2020.3-16.
26. Li, Q., Wen, Z., Wu, Z., Hu, S., Wang, N., Li, Y., Liu, X. and He, B., *A survey on federated learning systems: Vision, hype and reality for data privacy and protection*, in *ArXiv*. 2019.
27. Cui, L., Yang, S., Chen, F., Ming, Z., Lu, N. and Qin, J., "A survey on application of machine learning for internet of things", ***International Journal of Machine Learning and Cybernetics***, Vol. 9, No. 8, (2018), 1399-1417, DOI: 10.1007/s13042-018-0834-5.
28. Li, T., Sahu, A.K., Zaheer, M., Sanjabi, M., Talwalkar, A. and Smith, V., *Federated optimization in heterogeneous networks*, in *ArXiv*. 2018.1-28.
29. Niknam, S., Dhillon, H.S. and Reed, J.H., "Federated learning for wireless communications: Motivation, opportunities, and challenges", ***IEEE Communications Magazine***, Vol. 58, No. 6, (2020), 46-51, DOI: 10.1109/mcom.001.1900461.
30. Tran, N.H., Bao, W., Zomaya, A., Nguyen, M.N.H. and Hong, C.S., "Federated learning over wireless networks: Optimization model design and analysis", in IEEE INFOCOM 2019 - IEEE Conference on Computer Communications, Institute of Electrical and Electronics Engineers Inc., (2019), 1387-1395.
31. Kairouz, P., Brendan McMahan, H., Avent, B., Bellet, A., Bennis, M., Bhagoji, A.N., Bonawitz, K., Charles, Z., Cormode, G., Cummings, R., D'Oliveira, R.G.L., Rouayheb, S.E., Evans, D., Gardner, J., Garrett, Z., Gascón, A., Ghazi, B., Gibbons, P.B., Gruteser, M., Harchaoui, Z., He, C., He, L., Huo, Z., Hutchinson, B., Hsu, J., Jaggi, M., Javidi, T., Joshi, G., Khodak, M., Konečný, J., Korolova, A., Koushanfar, F., Koyejo, S., Lepoint, T., Liu, Y., Mittal, P., Mohri, M., Nock, R., Özgür, A., Pagh, R., Raykova, M., Qi, H., Ramage, D., Raskar, R., Song, D., Song, W., Stich, S.U., Sun, Z., Suresh, A.T., Tramèr, F., Vepakomma, P., Wang, J., Xiong, L., Xu, Z., Yang, Q., Yu, F.X., Yu, H. and Zhao, S., *Advances and open problems in federated learning*, in *ArXiv*. 2019.

32. Aledhari, M., Razzak, R., Parizi, R.M. and Saeed, F., "Federated learning: A survey on enabling technologies, protocols, and applications", ***IEEE Access***, Vol. 8, No., (2020), 140699-140725, DOI: 10.1109/access.2020.3013541.
33. Li, T., Sahu, A.K., Talwalkar, A. and Smith, V., "Federated learning: Challenges, methods, and future directions", ***IEEE Signal Processing Magazine***, Vol. 37, No. 3, (2020), 50-60, DOI: 10.1109/msp.2020.2975749.
34. Mothukuri, V., Parizi, R.M., Pouriyeh, S., Huang, Y., Dehghantanha, A. and Srivastava, G., "A survey on security and privacy of federated learning", ***Future Generation Computer Systems***, Vol. 115, No., (2021), 619-640, DOI: 10.1016/j.future.2020.10.007.
35. TensorFlow, *Tensorflow/federated: A framework for implementing federated learning*, in *GitHub*. 2020.
36. Caldas, S., Duddu, S.M.K., Wu, P., Li, T., Konečný, J., McMahan, H.B., Smith, V. and Talwalkar, A., *Leaf: A benchmark for federated settings*, in *ArXiv*. 2018.
37. Ryffel, T., Trask, A., Dahl, M., Wagner, B., Mancuso, J., Rueckert, D. and Passerat-Palmbach, J., *A generic framework for privacy preserving deep learning*, in *ArXiv*. 2018.
38. FedAI.org, *Fate-federated ai ecosystem*. 2019.
39. PaddleFL, *Paddlefl*. 2019.
40. Clara, N., *Nvidia clara | nvidia developer*. 2019.
41. OWKIN, *Federated learning - owkin*.
42. Chai, D., Chen, K., Wang, L. and Yang, Q., *Fedeval: A benchmark system with a comprehensive evaluation model for federated learning*, in *ArXiv*. 2020, arXiv.
43. He, C., Li, S., So, J., Zhang, M., Wang, H., Wang, X., Vepakomma, P., Singh, A., Qiu, H., Shen, L., Zhao, P., Kang, Y., Liu, Y., Raskar, R., Yang, Q., Annavaram, M. and Avestimehr, S., *Fedml: A research library and benchmark for federated machine learning*, in *ArXiv*. 2020.
44. Hu, S., Li, Y., Liu, X., Li, Q., Wu, Z. and He, B., *The oarf benchmark suite: Characterization and implications for federated learning systems*, in *ArXiv*. 2020.
45. Li, S., Cheng, Y., Liu, Y., Wang, W. and Chen, T., *Abnormal client behavior detection in federated learning*, in *ArXiv*. 2019.
46. Liu, Y.Y.Y., Huang, A., Luo, Y., Huang, H., Liu, Y.Y.Y., Chen, Y., Feng, L., Chen, T., Yu, H. and Yang, Q., *Fedvision: An online visual object detection platform powered by federated learning*, in *ArXiv*. 2020.
47. Huang, L., Shea, A.L., Qian, H., Masurkar, A., Deng, H. and Liu, D., "Patient clustering improves efficiency of federated machine learning to predict mortality and hospital stay time using distributed electronic medical records", ***Journal of biomedical informatics***, Vol. 99, (2019), 103291, DOI: 10.1016/j.jbi.2019.103291.
48. Ramaswamy, S., Mathews, R., Rao, K. and Beaufays, F., *Federated learning for emoji prediction in a mobile keyboard*, in *ArXiv*. 2019.
49. Jiang, J., Ji, S. and Long, G., "Decentralized knowledge acquisition for mobile internet applications", ***World Wide Web***, Vol. 23, No. 5, (2020), 2653-2669, DOI: 10.1007/s11280-019-00775-w.
50. Liu, Y., Liu, Y., Liu, Z., Liang, Y., Meng, C., Zhang, J. and Zheng, Y., "Federated forest", ***IEEE Transactions on Big Data***, Vol. 10.1109/tbdata.2020.2992755, (2020), 1-1, DOI: 10.1109/tbdata.2020.2992755.
51. Li, X., Gu, Y., Dvornek, N., Staib, L.H., Ventola, P. and Duncan, J.S., "Multi-site fmri analysis using privacy-preserving federated learning and domain adaptation: Abide results", ***Medical Image Analysis***, Vol. 65, (2020), 101765, DOI: 10.1016/j.media.2020.101765.
52. Chen, S., Xue, D., Chuai, G., Yang, Q. and Liu, Q., "Fl-qsar: A federated learning-based qsar prototype for collaborative drug discovery", ***Bioinformatics***, Vol. 36, No. 22-23, (2021), 5492-5498, DOI: 10.1093/bioinformatics/btaa1006.
53. Chen, D., Xie, L.J., Kim, B., Wang, L.C., Hong, C.S., Wang, L.C. and Han, Z., "Federated learning based mobile edge computing for augmented reality applications", in 2020 International Conference on Computing, Networking and Communications, ICNC 2020, (2020).
54. Hartmann, F., Suh, S., Komarzewski, A., Smith, T.D. and Segall, I., *Federated learning for ranking browser history suggestions*, in *ArXiv*. 2019.
55. Qi, T., Wu, F., Wu, C., Huang, Y. and Xie, X., *Fedrec: Privacy-preserving news recommendation with federated learning*, in *ArXiv*. 2020.
56. Asad, M., Moustafa, A. and Ito, T., "Fedopt: Towards communication efficiency and privacy preservation in federated learning", ***Applied Sciences***, Vol. 10, No. 8, (2020), 2864, DOI: 10.3390/app10082864.
57. Preuveneers, D., Rimmer, V., Tsingenopoulos, I., Spooren, J., Joosen, W. and Ilie-Zudor, E., "Chained anomaly detection models for federated learning: An intrusion detection case study", ***Applied Sciences***, Vol. 8, No. 12, (2018), 2663, DOI: 10.3390/app8122663.
58. Brisimi, T.S., Chen, R., Mela, T., Olshevsky, A., Paschalidis, I.C. and Shi, W., "Federated learning of predictive models from federated electronic health records", ***International Journal of Medical Informatics***, Vol. 112, (2018), 59-67, DOI: 10.1016/j.ijmedinf.2018.01.007.
59. Jeong, E., Oh, S., Kim, H., Park, J., Bennis, M. and Kim, S.-L., "Communication-efficient on-device machine learning: Federated distillation and augmentation under non-iid private data", in Neural Information Processing Systems, (2018).
60. Zhao, Y., Li, M., Lai, L., Suda, N., Civin, D. and Chandra, V., *Federated learning with non-iid data*, in *ArXiv*. 2018.
61. Duan, M., Liu, D., Chen, X., Tan, Y., Ren, J., Qiao, L. and Liang, L., "Astraea: Self-balancing federated learning for improving classification accuracy of mobile deep learning applications", in Proceedings - 2019 IEEE International Conference on Computer Design, ICCD 2019, (2019), 246-254.
62. Smith, V., Chiang, C.-k., Sanjabi, M. and Talwalkar, A., "Federated multi-task learning", in Neural Information Processing Systems, (2017), 4427-4437.
63. Corinzia, L. and Buhmann, J.M., "Variational federated multi-task learning", in NeurIPS 2019 Workshop on Federated Learning for Data Privacy and Confidentiality, (2019).
64. Wang, J., Kolar, M. and Srebro, N., "Distributed multi-task learning", in Proceedings of the 19th International Conference on Artificial Intelligence and Statistics, AISTATS 2016, (2016), 751-760.
65. Sattler, F., Müller, K.-R. and Samek, W., *Clustered federated learning: Model-agnostic distributed multi-task optimization under privacy constraints*, in *ArXiv*. 2019.1-12.
66. Lim, W.Y.B., Luong, N.C., Hoang, D.T., Jiao, Y., Liang, Y.-C., Yang, Q., Niyato, D. and Miao, C., "Federated learning in mobile edge networks: A comprehensive survey", ***IEEE Communications Surveys & Tutorials***, Vol. 22, No. 3, (2020), 2031-2063, DOI: 10.1109/comst.2020.2986024.
67. Xiao, P., Cheng, S., Stankovic, V. and Vukobratovic, D., "Averaging is probably not the optimum way of aggregating parameters in federated learning", ***Entropy (Basel)***, Vol. 22, No. 3, (2020), 314, DOI: 10.3390/e22030314.
68. Li, Q., He, B. and Song, D., *Model-agnostic round-optimal federated learning via knowledge transfer*, in *ArXiv*. 2020.

69. Yurochkin, M., Agarwal, M., Ghosh, S., Greenewald, K., Hoang, T.N. and Khazaeni, Y., "Bayesian nonparametric federated learning of neural networks", in 36th International Conference on Machine Learning, ICML 2019, International Machine Learning Society (IMLS), (2019), 12583-12597.
70. Wang, H., Yurochkin, M., Sun, Y., Khazaeni, Y. and Papailiopoulos, D., *Federated learning with matched averaging*, in *ArXiv*. 2020.
71. Liu, L. and Zheng, F., *A bayesian federated learning framework with multivariate gaussian product*, in *ArXiv*. 2021.
72. Sim, K.C., Zadrazil, P. and Beaufays, F., "An investigation into on-device personalization of end-to-end automatic speech recognition models", in Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH, ISCA, (2019), 774-778.
73. Wang, K., Mathews, R., Kiddon, C., Eichner, H., Beaufays, F. and Ramage, D., *Federated evaluation of on-device personalization*, in *ArXiv*. 2019.
74. Schneider, J. and Vlachos, M., *Personalization of deep learning*, in *ArXiv*. 2019.
75. Mansour, Y., Mohri, M., Ro, J. and Suresh, A.T., *Three approaches for personalization with applications to federated learning*, in *ArXiv*. 2020.
76. Finn, C., Abbeel, P. and Levine, S., "Model-agnostic meta-learning for fast adaptation of deep networks", in 34th International Conference on Machine Learning, ICML 2017, (2017), 1856-1868.
77. Jiang, Y., Konečný, J., Rush, K. and Kannan, S., *Improving federated learning personalization via model agnostic meta learning*, in *ArXiv*. 2019.
78. Khodak, M., Balcan, M.-F. and Talwalkar, A., *Adaptive gradient-based meta-learning methods*, in *ArXiv*. 2019.
79. Fallah, A., Mokhtari, A. and Ozdaglar, A., *Personalized federated learning: A meta-learning approach*, in *ArXiv*. 2020.
80. Li, D. and Wang, J., "Fedmd: Heterogenous federated learning via model distillation", in NeurIPS 2019 Workshop on Federated Learning for Data Privacy and Confidentiality, (2019).
81. Ma, J., Yonetani, R. and Iqbal, Z., " Adaptive Distillation for Decentralized Learning from Heterogeneous Clients,", 2020 25th International Conference on Pattern Recognition (ICPR), 2021, pp. 7486-7492, doi: 10.1109/ICPR48806.2021.9412356.
82. Caruana, R., "Multitask learning, Learning to learn, ed. S. Thrun and L. Pratt, Boston, MA, Springer US, 10.1007/978-1-4615-5529-2_5, (1998), 95-133.
83. Liu, S., Pan, S.J. and Ho, Q., "Distributed multi-task relationship learning", in Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, (2017), 937-946.
84. Ruder, S., *An overview of multi-task learning in deep neural networks*, in *ArXiv*. 2017.
85. Hinton, G., Vinyals, O. and Dean, J., *Distilling the knowledge in a neural network*, in *ArXiv*. 2015.
86. Konečný, J., McMahan, H.B., Ramage, D. and Richtárik, P., *Federated optimization: Distributed machine learning for on-device intelligence*, in *ArXiv*. 2016.
87. Sprague, M.R., Jalalirad, A., Scavuzzo, M., Capota, C., Neun, M., Do, L. and Kopp, M., "Asynchronous federated learning for geospatial applications", in Communications in Computer and Information Science, Springer Verlag, (2019), 21-28.
88. Stripelis, D. and Ambite, J.L., *Semi-synchronous federated learning*, in *ArXiv*. 2021.
89. Wu, W., He, L., Lin, W., Mao, R. and Jarvis, S., *Safa: A semi-asynchronous protocol for fast federated learning with low overhead*, in *ArXiv*. 2019.
90. Liu, Y., Yuan, X., Zhao, R., Zheng, Y. and Zheng, Y., *Rc-ssfl: Towards robust and communication-efficient semi-supervised federated learning system; rc-ssfl: Towards robust and communication-efficient semi-supervised federated learning system*, in *ArXiv*. 2018.
91. Chen, T., Jin, X., Sun, Y. and Yin, W., *Vafl: A method of vertical asynchronous federated learning*, in *ArXiv*. 2020.
92. van Dijk, M., Nguyen, N.V., Nguyen, T.N., Nguyen, L.M., Tran-Dinh, Q. and Nguyen, P.H., *Asynchronous federated learning with reduced number of rounds and with differential privacy from less aggregated gaussian noise*, in *ArXiv*. 2020.
93. Li, X., Qu, Z., Tang, B. and Lu, Z., *Stragglers are not disaster: A hybrid federated learning algorithm with delayed gradients*, in *ArXiv*. 2021.
94. Baytas, I.M., Yan, M., Jain, A.K. and Zhou, J., "Asynchronous multi-task learning", in Proceedings - IEEE International Conference on Data Mining, ICDM, (2017), 11-20.
95. Zhang, H., Li, J., Kara, K., Alistarh, D., Liu, J. and Zhang, C., "Zipml: Training linear models with end-to-end low precision, and a little bit of deep learning", in 34th International Conference on Machine Learning, ICML 2017, (2017), 6132-6140.
96. Wang, H., Sievert, S., Charles, Z., Liu, S., Wright, S. and Papailiopoulos, D., "Atomo: Communication-efficient learning via atomic sparsification", in Advances in Neural Information Processing Systems, (2018), 9850-9861.
97. Guha, N., Talwalkar, A. and Smith, V., *One-shot federated learning*, in *ArXiv*. 2019.
98. Seide, F., Fu, H., Droppo, J., Li, G. and Yu, D., "1-bit stochastic gradient descent and its application to data-parallel distributed training of speech dnns", in Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH, (2014), 1058-1062.
99. Wen, W., Xu, C., Yan, F., Wu, C., Wang, Y., Chen, Y. and Li, H., "Terngrad: Ternary gradients to reduce communication in distributed deep learning", in Advances in Neural Information Processing Systems, (2017), 1510-1520.
100. Strom, N., "Scalable distributed dnn training using commodity gpu cloud computing", in Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH, (2015), 1488-1492.
101. Aji, A.F. and Heafield, K., "Sparse communication for distributed gradient descent", in EMNLP 2017 - Conference on Empirical Methods in Natural Language Processing, Proceedings, Stroudsburg, PA, USA, Association for Computational Linguistics, (2017), 440-445.
102. Lin, Y., Han, S., Mao, H., Wang, Y. and Dally, W.J., "Deep gradient compression: Reducing the communication bandwidth for distributed training", in ICLR 2017, (2017).
103. Nishio, T. and Yonetani, R., "Client selection for federated learning with heterogeneous resources in mobile edge", in IEEE International Conference on Communications, (2019), 1-7.
104. Chen, C.Y., Choi, J., Brand, D., Agrawal, A., Zhang, W. and Gopalakrishnan, K., "Adacomp: Adaptive residual gradient compression for data-parallel distributed training", in 32nd AAAI Conference on Artificial Intelligence, AAAI 2018, (2018), 2827-2835.
105. Ahmed, A., Das, A. and Smola, A.J., "Scalable hierarchical multitask learning algorithms for conversion optimization in display advertising", in WSDM 2014 - Proceedings of the 7th ACM International Conference on Web Search and Data Mining, New York, USA, ACM Press, (2014), 153-162.
106. Mateos-Núñez, D., Cortés, J., Mateos-Núñez, D. and Cortes, J., "Distributed optimization for multi-task learning via nuclear-norm approximation∗∗the authors are with the department of mechanical and aerospace engineering, university of california,

san diego, USA", ***IFAC-PapersOnLine***, Vol. 48, No. 22, (2015), 64-69, DOI: 10.1016/j.ifacol.2015.10.308.

107. Smith, V., Forte, S., Ma, C., Takác, M., Jordan, M.I. and Jaggi, M., "Cocoa: A general framework for communication-efficient distributed optimization", ***Journal of Machine Learning Research***, Vol. 18, (2018), DOI: 10.3929/ethz-b-000282738.
108. Anil, R., Pereyra, G., Passos, A., Ormandi, R., Dahl, G.E. and Hinton, G.E., "Large scale distributed neural network training through online distillation", in 6th International Conference on Learning Representations, ICLR 2018 - Conference Track Proceedings, (2018), 1-12.
109. Tramèr, F., Zhang, F., Juels, A., Reiter, M.K. and Ristenpart, T., "Stealing machine learning models via prediction apis", in Proceedings of the 25th USENIX Security Symposium, USENIX Association, (2016), 601-618.
110. Melis, L., Song, C., De Cristofaro, E. and Shmatikov, V., "Exploiting unintended feature leakage in collaborative learning", in Proceedings - IEEE Symposium on Security and Privacy, (2019), 691-706.
111. Wei, W., Liu, L., Loper, M., Truex, S., Yu, L., Gursoy, M.E. and Wu, Y., *Adversarial examples in deep learning: Characterization and divergence*, in *ArXiv*. 2018.
112. Bagdasaryan, E., Veit, A., Hua, Y., Estrin, D. and Shmatikov, V., *How to backdoor federated learning*, in *ArXiv*. 2018.
113. Wang, H., Sreenivasan, K., Rajput, S., Vishwakarma, H., Agarwal, S., Sohn, J.-y., Lee, K. and Papailiopoulos, D., *Attack of the tails: Yes, you really can backdoor federated learning*, in *ArXiv*. 2020.
114. Biggio, B., Corona, I., Maiorca, D., Nelson, B., Šrndić, N., Laskov, P., Giacinto, G. and Roli, F., "Evasion attacks against machine learning at test time", in Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), (2013), 387-402.
115. Abadi, M., McMahan, H.B., Chu, A., Mironov, I., Zhang, L., Goodfellow, I. and Talwar, K., "Deep learning with differential privacy", in Proceedings of the ACM Conference on Computer and Communications Security, (2016), 308-318.
116. Geyer, R.C., Klein, T. and Nabi, M., *Differentially private federated learning: A client level perspective*, in *ArXiv*. 2017.
117. Hitaj, B., Ateniese, G. and Perez-Cruz, F., "Deep models under the gan: Information leakage from collaborative deep learning", in Proceedings of the ACM Conference on Computer and Communications Security, New York, NY, USA, Association for Computing Machinery, (2017), 603-618.
118. Bogdanov, D., Laur, S. and Willemson, J., "Sharemind: A framework for fast privacy-preserving computations", in Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), Berlin, Heidelberg, Springer Berlin Heidelberg, (2008), 192-206.
119. Araki, T., Furukawa, J., Lindell, Y., Nof, A. and Ohara, K., "High-throughput semi-honest secure three-party computation with an honest majority", in Proceedings of the ACM Conference on Computer and Communications Security, New York, NY, USA, Association for Computing Machinery, (2016), 805-817.
120. Mohassel, P. and Zhang, Y., "Secureml: A system for scalable privacy-preserving machine learning", in Proceedings - IEEE Symposium on Security and Privacy, (2017), 19-38.
121. Mohassel, P. and Rindal, P., "Aby3: A mixed protocol framework for machine learning", in Proceedings of the ACM Conference on Computer and Communications Security, New York, NY, USA, Association for Computing Machinery, (2018), 35-52.
122. Bonawitz, K., Ivanov, V., Kreuter, B., Marcedone, A., McMahan, H.B., Patel, S., Ramage, D., Segal, A. and Seth, K., "Practical secure aggregation for privacy-preserving machine learning", in Proceedings of the ACM Conference on Computer and Communications Security, New York, USA, ACM Press, (2017), 1175-1191.
123. Mandal, K. and Gong, G., "Privfl: Practical privacy-preserving federated regressions on high-dimensional data over mobile networks", in Proceedings of the ACM Conference on Computer and Communications Security, (2019).
124. Gentry, C., "Fully homomorphic encryption using ideal lattices", in Proceedings of the Annual ACM Symposium on Theory of Computing, New York, New York, USA, ACM Press, (2009), 169-178.
125. Coron, J.S., Lepoint, T. and Tibouchi, M., "Scale-invariant fully homomorphic encryption over the integers", in Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), Springer Verlag, (2014), 311-328.
126. Brakerski, Z., Gentry, C. and Vaikuntanathan, V., "(leveled) fully homomorphic encryption without bootstrapping", in ITCS 2012 - Innovations in Theoretical Computer Science Conference, New York, New York, USA, ACM Press, (2012), 309-325.
127. Bourse, F., Minelli, M., Minihold, M. and Paillier, P., "Fast homomorphic evaluation of deep discretized neural networks", in Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), Springer Verlag, (2018), 483-512.
128. Acar, A., Aksu, H., Uluagac, A.S. and Conti, M., "A survey on homomorphic encryption schemes", ***ACM Computing Surveys***, Vol. 51, No. 4, (2018), 1-35, DOI: 10.1145/3214303.
129. Xu, G., Li, H., Liu, S., Yang, K. and Lin, X., "Verifynet: Secure and verifiable federated learning", ***IEEE Transactions on Information Forensics and Security***, Vol. 15, No., (2020), 911-926, DOI: 10.1109/tifs.2019.2929409.
130. Itahara, S., Nishio, T., Koda, Y., Morikura, M. and Yamamoto, K., *Distillation-based semi-supervised federated learning for communication-efficient collaborative training with non-iid private data*, in *ArXiv*. 2020.
131. Van Berlo, B., Saeed, A. and Ozcelebi, T., "Towards federated unsupervised representation learning", in EdgeSys 2020 - Proceedings of the 3rd ACM International Workshop on Edge Systems, Analytics and Networking, Part of EuroSys 2020, New York, NY, USA, Association for Computing Machinery, (2020), 31-36.
132. Tian, Y., Krishnan, D. and Isola, P., "Contrastive multiview coding", in Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), Springer Science and Business Media Deutschland GmbH, (2020), 776-794.
133. Hénaff, O.J., Srinivas, A., De Fauw, J., Razavi, A., Doersch, C., Eslami, S.M.A. and Van Den Oord, A., *Data-efficient image recognition with contrastive predictive coding*, in *ArXiv*. 2020.
134. Chen, T., Kornblith, S., Norouzi, M. and Hinton, G., *A simple framework for contrastive learning of visual representations*, in *ArXiv*. 2020.
135. Li, Q. and Song, D., *Model-contrastive federated learning*, in *ArXiv*. 2021.
136. Misra, I. and van der Maaten, L., "Self-supervised learning of pretext-invariant representations", in Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, IEEE Computer Society, (2020), 6706-6716.
137. He, K., Fan, H., Wu, Y., Xie, S. and Girshick, R., "Momentum contrast for unsupervised visual representation learning", in Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, IEEE Computer Society, (2020), 9726-9735.

138. Lin, J., Du, M. and Liu, J., *Free-riders in federated learning: Attacks and defenses*, in *ArXiv*. 2019.

139. Zong, B., Song, Q., Min, M.R., Cheng, W., Lumezanu, C., Cho, D.-k. and Chen, H., "Deep autoencoding gaussian mixture model for unsupervised anomaly detection", in 6th International Conference on Learning Representations, ICLR 2018 - Conference Track Proceedings, (2018).

140. Karimireddy, S.P., Kale, S., Mohri, M., Reddi, S.J., Stich, S.U. and Suresh, A.T., *Scaffold: Stochastic controlled averaging for on-device federated learning*, in *ArXiv*. 2019.

Persian Abstract

چکیده

با اختراع مدرن حسگرهای با کیفیت بالا و تراشه های هوشمند با توان محاسباتی بالا ، دستگاه های هوشمند مانند تلفن های هوشمند و دستگاه های پوشیدنی هوشمند در حال تبدیل شدن به منابع محاسباتی اصلی برای زندگی روزمره هستند. این دستگاه ها ، در مجموع ، ممکن است دارای مقدار زیادی داده ارزشمند باشند ، اما به دلیل نگرانی درباره حریم خصوصی و قوانین حریم خصوصی مانند GDPRمقررات عمومی حفاظت از داده ها، این مقدار عظیم داده های بسیار ارزشمند برای آموزش مدل ها برای کاربردهای دقیق تر و کارآمد هوش مصنوعی در دسترس نیست. . آموزش فدراسیون (FL)به عنوان یک روش یادگیری مشارکتی بسیار برجسته برای یادگیری از چنین داده های خصوصی غیرمتمرکز ظهور کرده و در عین حال محدودیت های حریم خصوصی را نیز برآورده می کند. برای یادگیری از چنین داده های غیرمتمرکز و توزیع شده گسترده ، یادگیری فدرال باید بر برخی چالش های منحصر به فرد مانند ناهمگنی سیستم ، ناهمگنی آماری ، ارتباطات ، ناهمگنی مدل ، حریم خصوصی و امنیت غلبه کند. در این مقاله ، برای شروع ، برخی از اصول یادگیری فدراسیون به همراه تعریف و کاربردهای FL را توضیح می دهیم. پس از آن ، ما چالش های منحصر به فرد FL را بیشتر توضیح می دهیم در حالی که رویکردهای اخیراً پیشنهادی برای کنترل آنها را به طور انتقادی پوشش می دهیم. علاوه بر این ، این مقاله همچنین برخی از چالش های نسبتاً جدید برای یادگیری فدراسیون را مورد بحث قرار می دهد. برای نتیجه گیری ، ما در مورد برخی از مسیرهای تحقیقاتی آینده در حوزه یادگیری فدراسیون بحث می کنیم.

IJE TRANSACTIONS A: Basics Vol. 34, No. 7, (July 2021) 1684-1693

International Journal of Engineering

Journal Homepage: www.ije.ir

Deep Convolutional Neural Network for Finger-knuckle-print Recognition

A. Zohrevand*[a], Z. Imani[a], M. Ezoji[b]

[a] *Computer Engineering Department, Kosar University of Bojnord, Bojnord, North Khorasan, Iran*
[b] *Department of Electronics, Faculty of Electrical and Computer Engineering Babol Noshirvani University of Technology, Babol, Mazandaran, Iran*

PAPER INFO

Paper history:
Received 17 April 2021
Received in revised form 04 May 2021
Accepted 08 June 2021

Keywords:
Human Biometric
Hand-based Biometric
Finger Knuckle Print
Transfer Learning
Convolutional Neural Network

ABSTRACT

Finger-Knuckle-Print (FKP) is an accurate and reliable biometric in compare to other hand-based biometrics like fingerprint because of the finger's dorsal region is not exposed to surfaces. In this paper, a simple end-to-end method based on Convolutional Neural Network (CNN) is proposed for FKP recognition. The proposed model is composed only of three convolutional layers and two fully connected layers. The number of trainable parameters hereby has significantly reduced. Additionally, a straightforward method is utilized for data augmentation in this paper. The performance of the proposed network is evaluated on Poly-U FKP dataset based on 10-fold cross-validation. The best recognition accuracy, mean accuracy and standard deviation are 99.83%, 99.18%, and 0.76, respectively. Experimental results show that the proposed method outperforms the state-of-the-arts in terms of recognition accuracy and the number of trainable parameters. Also, in compare to four fine-tuned CNN models including AlexNet, VGG16, ResNet34, and GoogleNet, the proposed simple method achieved higher performance in terms of recognition accuracy and the numbers of trainable parameters and training time.

doi: *10.5829/ije.2021.34.07a.12*

1. INTRODUCTION[1]

Using a reliable recognition method is a critical challenge in both academic and industrial contexts [1]. Biometrics can be defined as unique behavioral or physical characteristics of humans. Given the convenience and accuracy of biometric-based methods, they have huge potentials in diverse applications such as e-marketing, security, access control, e-banking, etc. [1]. In the past decades, many studies have investigated the advantages of some biometric characteristics, including face [2], iris [3], fingerprint [4], palm-print [5], hand geometry [6], finger-vein [7], and Finger Knuckle Print (FKP) [8]. Compared to different types of biometrics, hand-based biometrics has received significant attention in recent years [1].

Recent studies show that skin wrinkles on the outer area of finger knuckle have a unique pattern. Due to the unique features of FKP, it can be used as a discriminative biometric method [9]. FKP has several advantages over other hand-based biometrics. For example, people tend to grasp stuff by the inner side of their hands, so the FKP surface is not damaged or abraded. Since the data collection process in the FKP is contactless, it is usually more popular among users [10]. In this article, attempts have been made to design a simple method based on Convolutional Neural Network (CNN) for FKP recognition.

To the best of our knowledge, despite the excellent performance of CNN in different computer vision applications, the use of CNN models for the recognition of FKP image has received scant scholarly attention. Therefore, this study uses a novel CNN model to recognize FKP image. The main points of this work are as follows:

1. To the best of our knowledge, this is the first work that focuses on the effectiveness of CNN model for FKP recognition.

* Corresponding Author Institutional Email: a.zohrevand@kub.ac.ir
(A. Zohrevand)

Please cite this article as: A. Zohrevand, Z. Imani, M. Ezoji, Deep Convolutional Neural Network for Finger-Knuckle-Print Recognition, International Journal of Engineering, Transactions A: Basics Vol. 34, No. 7, (2021) 1684-1693

2. A simple end-to-end learning method is proposed without the need of handcrafted feature extraction for recognizing FKP images.
3. A straightforward data augmentation approach is utilized to collect appropriate data for training proposed CNN model.
4. As an investigation, four well-known CNN architectures including VGG16 [11], Google-Net [12], and ResNet34 [13], and AlexNet [14] are fine-tuned for FKP recognition.

The rest of this article is formed as follows. Section 2 reviews the major works in FKP recognition. Section 3 describes the proposed methodology. The details of experimental results are presented in section 4. Discussion and comparison are explained in section 5. Finally, conclusions and future works are explained in section 6.

2. LITERATURE REVIEW

Due to the importance of FKP in user recognition, many researchers have conducted extensive studies on the FKP recognition. To the best of our knowledge, all state-of-the-art in FKP recognition can be assigned to two main categories: classic and deep-based approaches. As shown in Figure 1, studies in both categories comprise almost two parts feature extraction and classification. As the features are extracted manually in the classic works, deep-based methods extract appropriate features automatically. In the following, the most seminal works in both categories are reviewed.

Woodard and Flynn [15], one of the first researchers to investigate the surface of finger knuckle as a biometric identifier, designed a dataset contain 3D finger back knuckle surface with the Minolta 900/910 sensor. Then, they used a curvature-based shape indicator for extracting desired features. Kumar and Ravikanth [10] presented a person identification system based on a 2D finger-back surface. They applied several approaches like PCA, LDA, and ICA to extract feature. Kumar and Zhou [16] applied the Robust Line Orientation Code to the FKP images for extracting the local orientation information as feature vectors. Zhang et al. [17] applied the Gabor filter for extracting the magnitude and orientation information of FKP images. Zhang et al. [18] have applied the Fourier Transform (FT) to the FKP images, and the FT coefficient was considered as a feature representation. Morales et al. [19] utilized the Gabor filter for enhancing the FKP lines, adopting SIFT descriptor for the feature extraction. Zhu [20] utilized SURF descriptor to extract features and then matching. Badrinath et al. [21] combined SURF and SIFT to enhance the FKP texture images, and FKP recognition. The reflectance and illumination were extracted from each FKP image [22]. Then they used serial feature fusion to create a huge vector of feature for each individual. The Gabor filter was utilized for extracting desired features [8]. They also used Hamming distance along with Support Vector Machine (SVM) for reducing False Acceptance Rate (FAR). Vidhyapriya and Lovelyn Rose [23] used the Gabor filter and Expectation-Maximization (EM) to extract texture patterns. They also utilized SIFT descriptor to create a feature vector. Heidari and Chalechale [24] have extracted the features by a pattern histogram based on entropy and some statistical texture features. Then, the genetic algorithm was used to extract the optimized sub-features. Attia et al. [25] used Multilayer Deep Rule Based (DRB) method for FKP recognition. First, two types of features including Binarized Statistical Image feature and Gabor Filter bank are extracted from input images. These features are then fed to fuzzy rule based DRB classifier for person authentication. Chlaoua et al. [26] used PCA-Net beside SVM for FKP recognition. They applied PCA to extract two feature banks and SVM for classification. Even though these works were impressive, but the features are extracted manually. Unlike classical works, features are extracted automatically in deep-based methods. CNN is the most significant models in deep-based works. Thus, in the second category, the CNN models were used for FKP recognition.

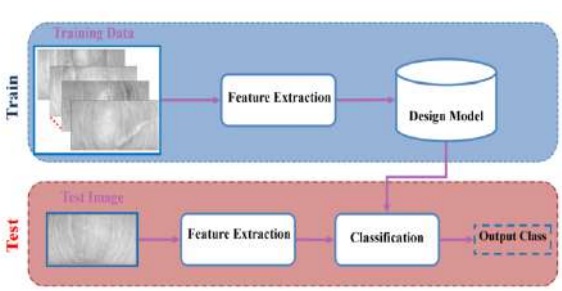


Figure 1. General diagram for the FKP recognition

LeCun and Bengio [27] were the first to propose CNN architecture. CNN models have been the subject of considerable attention in most computer vision applications [28-31]. These families of neural networks, which combine feature extraction and classification roles, are intended to recognize images based on their scale, shift, and distortions. Generally, CNNs contain input and output layers as well as several sequentially-connected convolutional layers that are followed by a fully connected layer(s). In each convolutional layer, inputs of the previous layer are convolved with trainable filters. Relu used as an usual activation function in CNN layers [32]. To diminish data size and reduce the over-fitting phenomena [33], the pooling operation was performed for the output of the current layer. Irrespective of the great performance of CNN models in diverse computer visions, the utilization of CNN models for the recognition of FKP images has been largely overlooked.

As far as we are concerned there are studies [34, 35] that utilized CNN models for recognizing FKP images. In this paper, we address the FKP recognition issue and attempt to develop a new design for the recognition of the FKP image.

3. PROPOSED METHODOLOGY

Despite the excellent performance of CNN in different applications in computer vision community, the use of CNN models to recognize FKP image has been a low research priority. Therefore, this paper aims to propose FKP recognition by CNN models. In the following subsections, Transfer Learning (TL) and the proposed CNN architecture are reviewed for the FKP recognition.

3. 1. Transfer Learning Providing adequate labeled training data with a distribution similar to the test data is the best scenario for machine learning [36]. However, collecting sufficient training samples is time-consuming, expensive, or even impossible in some cases [36]. TL is a research problem in machine learning that retains knowledge obtained from the solution of a problem (source domain) to be applied to different but relatively similar problems (target domain). TL has been the most popular approach in CNN models in recent years [37, 38]. In fact, few people train an entire CNN models from scratch (with random initialization), since it is usually uncommon to have a database of adequate size. Instead, it is customary to pre-train a CNN on a large dataset, e.g. Image-Net [14], and then utilize the CNN either as a feature extractor or an initialization for the desired task. Fine-tuning is the most common approach in the TL scenarios. In this approach, some earlier layers are frozen (due to over-fitting concerns), and the network is re-trained by further layers. This is inspired by the observation that the CNN's earlier features include more typical features (e.g. edge detectors), which are common features in numerous tasks; however, further layers become gradually more specific to the details of the classes contained in the original dataset.

3. 2. Proposed CNN Model For Recognizing FKP Images As shown in Figure 2, the source and target domains are entirely different in FKP recognition. Therefore, TL-based methods may not be the best choice for FKP recognition. In this article, we presented a CNN model for FKP recognition. The proposed CNN architecture is depicted in Figure 3. As can be seen, the proposed architecture consists of one input layer, data augmentation block, three convolutional layers with two corresponding max-pooling (L_1 and L_2) and one average-pooling layers (L_3) for feature extraction, two fully-connected layers (L_4, and L_5) for classification, and finally one output layer. According to Figure 3, feature extraction and classification were conducted automatically. In the training phase, first each input image is augmented and then the CNN's trainable parameters are updated with respect to its label by specific learning algorithm. Then, in a network with adjusted parameters a feature is extracted from the input image in the test phase and in the forward path the label is predicted after classification. In the test phase, there is no data augmentation for the input image. The main goal of this model design is to achieve best performance with the adequate number of layers and weights.

4. EXPERIMENTAL RESULTS

The aim goal of this article is to propose a simple CNN architecture for recognizing FKP images. All experiments were utilized by a machine with these features: Intel® core i3 - 6300 CPU @3:70GHz, 48GB RAM, and NVidia® 1060Ti 6GB GPU. The experiments were conducted using the PyTorch® framework installed on Microsoft® Windows10.

4. 1. Detail of Simulation There are several important parameters to design the proposed CNN model such as: the kernel size, number of filters, strides, and etc. that are summarized in Table 1. Also, the proposed CNN architecture requires some hyper-parameters such as: batch sizes, initial learning rate, regularization, and number of epochs that are shown in Table 2. It should be noted all parameter values in Table 1 and also the hyper-parameter values in Table 2 have been set experimentally. Finally The back-propagation algorithm beside the Adam[39] optimizer is utilized for training the proposed CNN architecture.

4. 2. Database For evaluating the proposed personal recognition system, the Poly-U FKP database [17] was utilized. This database was collected from 165 participants. In the data collection process, 12 samples were taken from each finger knuckle in two distinct

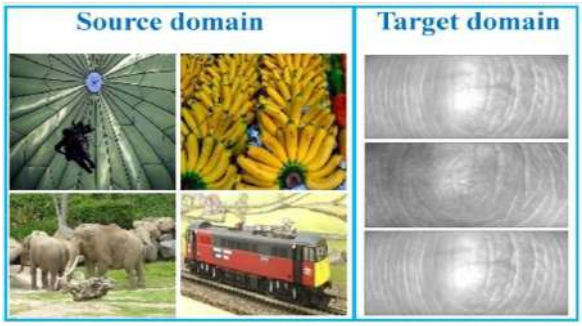


Figure 2. Some image samples in the source and target domains. As can be seen, the source and target domains are different in terms of texture, semantics, etc.

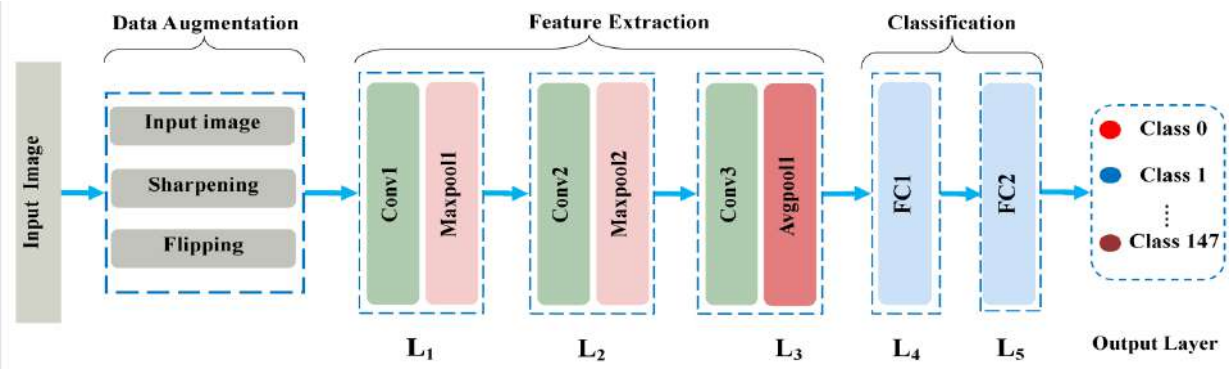


Figure 3. Proposed CNN model for recognizing FKP images. This architecture contains five layers (L_1, L_2, L_3, L_4, and L_5) for extracting feature and classification

TABLE 1. The details of the proposed CNN architecture depicted in Figure 3

Layer Name	Layer Type	No. of Filters	Kernel Size	Stride	Input Features	Output Features	No. of Parameters
L_1	Convolution (Conv1)	16	(3×3)	(1×1)	(1,220,110)	(16,218,108)	448
	Batch Normalization	---	---	---	(16,218,108)	(16,218,108)	32
	RELU	---	---	---	(16,218,108)	(16,218,108)	0
	Max Pooling (Maxpool1)	---	(4×4)	(4×4)	(16,218,108)	(16,54,27)	0
L_2	Convolution (Conv2)	32	(3×3)	(1×1)	(16,54,27)	(32,52,25)	4,640
	Batch Normalization	---	---	---	(32,52,25)	(32,52,25)	64
	RELU	---	---	---	(32,52,25)	(32,52,25)	0
	Max Pooling (Maxpool2)	---	(2×2)	(2×2)	(32,52,25)	(32,26,12)	0
L_3	Convolution (Conv3)	64	(3×3)	(1×1)	(32,26,12)	(64,24,10)	18,496
	Batch Normalization	---	---	---	(64,24,10)	(64,24,10)	128
	RELU	---	---	---	(64,24,10)	(64,24,10)	0
	Average Pooling (Avgpool1)	---	(2×2)	(2×2)	(64,24,10)	(64,2,2)	0
L_4	Fully Connected (FC1)	---	---	---	256	256	65,792
	RELU	---	---	---	256	256	0
L_5	Fully Connected (FC2)	---	---	---	256	148	38,036
Totally = 127,636							

TABLE 2. Hyper-parameter setting of the proposed CNN model

Hyper-parameter	value
Batch size	8
Number of epochs	100
Initial learning rate	0.001
L2regularization	0.001

sessions (six samples per session). In each session, the images of a participant's Right Index (RI), Right Middle (RM), Left Middle (LM), and Index (LI), fingers were collected, respectively. A total of 7,104 samples from 592 distinct fingers were public available from 148 participants which used in this work. After data collection, the region of interest was extracted from each finger in the preprocessing. Finally, 7,104 samples (size: 220×110) were obtained. Figure 4 shows some samples from the Poly-U FKP database.

The existence of many trainable parameters is one of challenges in CNN models [36]. Thus, sufficient data is required to avoid over-fitting in training CNN models. In many real-life applications, appropriate data is not available. There are several approaches for data augmentation [40]. In this paper, two data augmentation

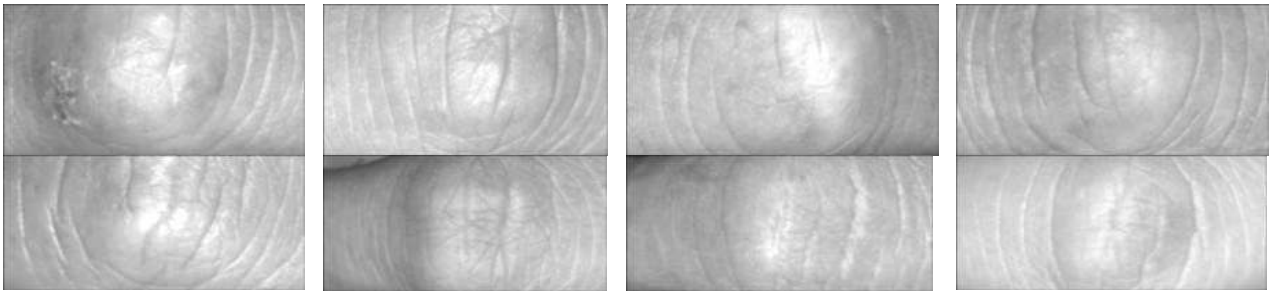


Figure 4. Some samples of Poly-U FKP database. Upper row, from left to right, LI, LM, RI, and RM fingers of the same person, respectively. Lower row, from left to right, four different LI fingers of different persons

methods have been employed. First, as described by Bloice et al. [41], we attempted to augment data by different zooming scales, different rotation angles and elastic distortions. In this augmenting process, 50% of data are chosen for testing in each class and remaining 50% is initially augmented by Bloice et al. [41] and then saved for the training set. Moreover, the data in each class will quadruple after the augmentation. The second approach includes simple transformations such as sharpening and flipping on the input finger image. As shown Figure 5, the Gaussian low-pass filter with four different parameter values was used for image sharpening. Image flipping was performed by flipping original the image horizontally and vertically. All of these transformations were conducted in MATLAB® 2020 after which each image sample was augmented into six new image samples. The K-fold cross validation is one of standard methods for training models, especially when limited label data are available. Thus, for first time, this paper uses 10-fold cross validation for training CNN model. In the second augmentation process, the principal 10% of data was chosen for the testing set, and remaining data (=90% of all data) was first augmented with the proposed augmentation method and then saved for the training set.

4. 3. Experiment and Evaluation In the first, the four well-known CNN models, AlexNet, VGG16, ResNet34, and GoogleNet, which trained on Image-Net dataset, were fine-tuned for FKP recognition. To do so, the three last layers of the VGG16 and AlexNet and the last layer of two other CNNs were re-trained on the augmented FKP dataset. The recognition accuracy of the training set and the training loss of these CNN models are shown in Figure 6. As can be seen, VGG16 and AlexNet outperformed other CNN models.

As noted in the TL section, when the target and source domains have different distributions, the TL approach will be ineffective. Thus, we seek to train the proposed CNN architecture on dataset that is augmented by two approaches separately. The results of the experiment are presented in Table 3. The first augmentation method employs a number of transformations like elastic distortion, which may deform the structure of FKP images. In contrast, the second augmentation method effectively retains the structure of images. Hence, as

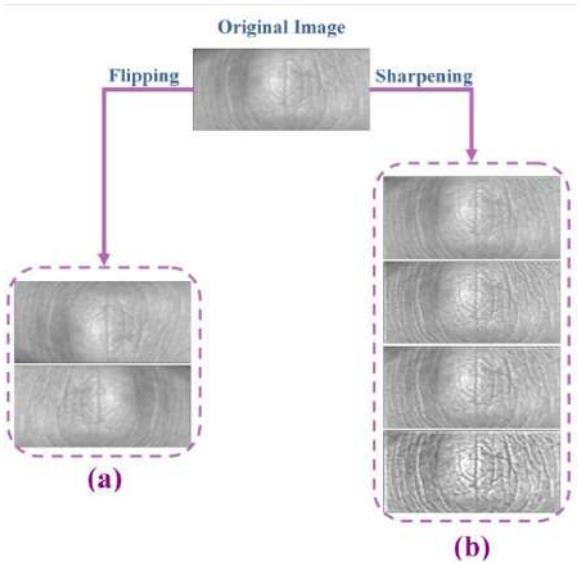


Figure 5. Data augmentation includes: a) flipping, b) sharpening

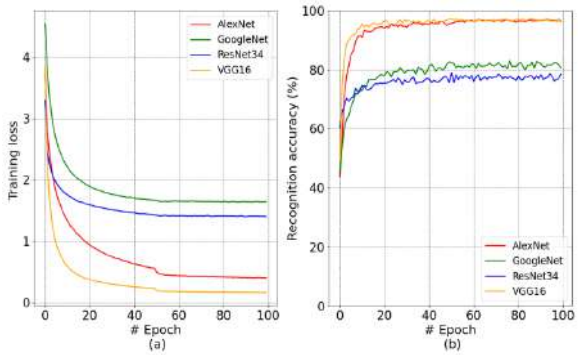


Figure 6. a: Training loss, b: recognition accuracy of FKP recognition by fine-tuning the four conventional CNN models

shown in Table 3, compared to the above two augmentation methods, the proposed CNN model with the second augmentation has higher recognition accuracy. Also, Table 3 compares the efficiency of the proposed approach with four conventional CNN architectures in terms of the number of trainable, training time, and recognition accuracy. As shown in this table, the proposed CNN not only has fewer parameters but also demonstrates a higher performance. Finally, the experimental result of proposed method with second augmentation in 10-fold cross validation shown in Figure 7. Receiver Operating Characteristic (ROC) curves is great tools for evaluating the performance of a biometric system. This diagram can be plotted by drawing True Positive Rate (TPR) versus False Positive Rate (FPR) for all thresholds [24]. Figure 8 depicted the ROC of the presented approach compared to the four CNN models. As can be seen, this diagram confirms the effectiveness of the presented CNN model with second augmentation compare to four fine-tuned CNN models.

TABLE 3. Comparing the performance of four CNN models with the proposed

Method	#Trainable parameters	Recognition Accuracy (%)	Training Time(GPU)
ResNet34	75,924	79.02	167m
VGG-16	120,152,212	97.29	546m
GoggleNet	151,700	83.08	124m
AlexNet	55,140,500	97.12	200m
Proposed CNN + first Augmentation method [41]	**127,636**	**96.50**	**286m**
Proposed CNN + second Augmentation method	**127,636**	**99.83†/98.18***	**157m**

*: The average recognition accuracy in 10-fold cross-validation

†: The best recognition accuracy in 10-fold cross-validation

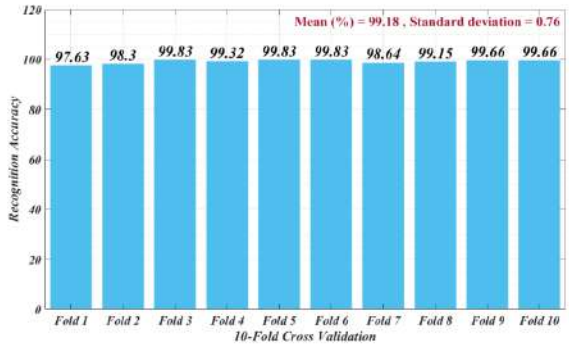


Figure 7. The recognition accuracy of the proposed CNN architecture in 10-fold cross validation

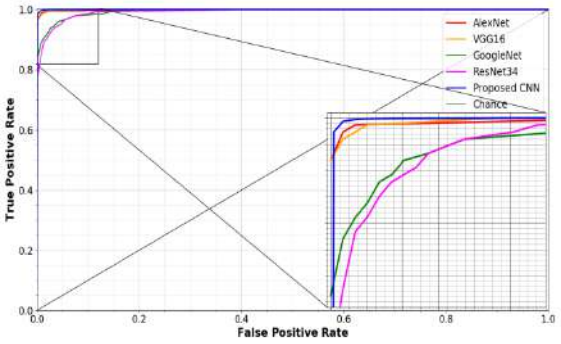


Figure 8. The ROC of the proposed CNN architecture and four conventional CNN models

Based on two augmentation methods, Table 4 depicted the performance of proposed model for different fingers. As can be see, the proposed CNN architecture with the second augmentation method has higher recognition accuracy. In the last step of this experiment, the performance of the presented approach with the second augmentation was assessed for different fingers in 10-fold cross validation, as depicted in Figure 9.

TABLE 4. The performance of the presented approach with respect to various fingers in two different augmentation process

Method	Recognition accuracy of each finger (%)			
	LI	LM	RI	RM
Proposed CNN + first Augmentation method [41]	95.4	95.6	94.4	96.6
Proposed CNN+ second Augmentation method	99.6†/99.3*	100.0†/99.8*	100.0†/99.5*	100.0†/99.5*

*: The average recognition accuracy in 10-fold cross-validation

† : The best recognition accuracy in 10-fold cross-validation

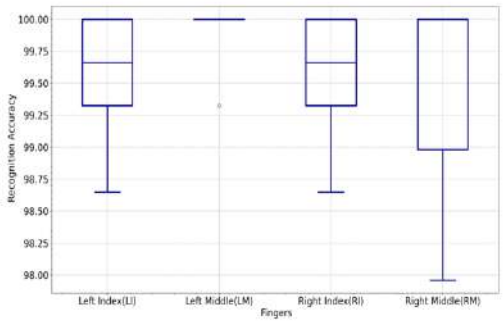


Figure 9. The performance of the proposed approach (with second augmentation) with respect to different fingers (Right Index (RI), Right Middle (RM), Left Middle (LM), and Index (LI)) in 10-fold cross validation

5. DISCUSSION AND COMPARISON

This section draws a comparison between our study and the most recent state-of-the-art works. The literature review suggests that the research on FKP recognition can be split into two groups, classic and deep-based methods. Unlike all classic methods in which appropriate features are extracted manually, in this paper, the feature extraction and classification are conducted automatically. As depicted in Table 5, the proposed model outperforms the classic works in term of recognition accuracy.

One of the essential challenges in the CNN model is the number of trainable parameters, which makes these models inapplicable for low-computing devices such as mobile phones. This paper, for the first time, investigated the effectiveness of CNN model for FKP recognition and strived to propose a simple end-to-end architecture for recognizing FKP images. In addition, in this paper, a straightforward method was utilized for data augmentation. In this regard, Table 6 compares the proposed model with the latest single finger-based works.

As it is evident, our proposed method is superior to other classic works. As shown in Table 5, compared to other works in the deep-based methods the proposed CNN structure not only demonstrates competitive performance in term of recognition accuracy but also the trainable parameters reduced significantly. In fact, compared to the deep-based methods, our proposed method offers several advantages such as fewer trainable parameters, simplicity, and efficient data augmentation.

6. CONCLUSION AND FUTURE WORKS

Given that behaviors and attributes are unique features, human biometrics (e.g. palm-print, finger vein, hand geometry, fingerprint, and Finger Knuckle Print), can be used to improve personal validation. In this paper, attempts were made to recognize FKP using a new CNN model. In the experiments, first four CNN models, including AlexNet, VGG16, ResNet34, and GoogleNet, were fine-tuned for FKP recognition. Then, a new CNN model was developed in the second set of experiments.

TABLE 5. Comparing the performance of the presented approach (with second augmentation) with the state-of-the-art methods

Category	Ref.	Method	#Trainable Parameters	Database: Poly-U FKP		Recognition Accuracy (%)
				(Train%,Test%)	# of Class	
Classic	[42]	Log-Gabor + Gray Level Intensity	----	(50, 50)	165	96.56
	[43]	DFB + LDA + Norm	----	(50, 50)	165	99.29
	[44]	Differential box-counting + fractal dimension	----	N/A	165	97.29
	[45]	Gabor filter + fractal dimension	----	(50, 50)	165	93.73
	[8]	Gabor features + Support Vector Machine	----	N/A	165	89.33
	[23]	Gabor features + Expectation Maximization + SIFT	----	N/A	165	98.00
	[24]	Entropy pattern histogram + texture features	----	(60,40)	165	94.91
	[25]	Multilayer Deep Rule Based (DRB)	----	N/A	165	99.65
	[26]	PCANet + Support Vector Machine	----	N/A	165	100
Deep	[34]	Siamese convolutional neural network model	253,000	(90, 10)	150	99.28
	----	**Proposed CNN + first Augmentation method**[45]	**127,636**	**(50,50)**	**145**	**96.50**
	----	**Proposed CNN + second Augmentation method**	**127,636**	**10-fold**	**145**	**99.83†/99.18***

*: The average recognition accuracy in 10-fold cross-validation

†: The best recognition accuracy in 10-fold cross-validation

TABLE 6. Comparing the performance of the presented approach (with second augmentation) with respect to different fingers (RI: Right Index, and RM: Right Middle, LI: Left Index, LM: Left Middle)

Category	Ref.	#Trainable Parameters	Database: Poly-U FKP		Recognition accuracy of each finger (%)			
			(Train%, Test%)	# of Class	LI	LM	RI	RM
Classic	[42]	----	(50, 50)	165	89.9	88.5	88.4	89.4
	[43]	----	(50, 50)	165	90.3	88.6	89.7	89.7
	[22]	----	N/A	165	91.0	94.8	91.4	91.8
	[8]	----	N/A	165	97.4	97.5	94.2	99.1
	[25]	----	N/A	165	93.5	94.3	93.9	94.1
	[35]	6,985,812	(50, 50)	165	99.1	98.90	99.4	98.3
Deep	**Proposed CNN + first Augmentation method [41]**	**127, 636**	**(50,50)**	**145**	**95.4**	**95.6**	**94.4**	**96.6**
	Proposed CNN + second Augmentation method	**127, 636**	**10-fold**	**145**	**99.6†/99.3***	**100.0†/99.8***	**100.0†/99.5***	**100.0†/99.5***

*: The average recognition accuracy in 10-fold cross-validation

†: The best recognition accuracy in 10-fold cross-validation

The open-source Poly-U FKP dataset was used for training and testing. Due to limited data, straightforward methods were adopted for data augmentation, which compared to other data augmentation, can effectively keep the structure of FKP images. In the TL approach, VGG16 surpasses the other three CNN models. In the subsequent experiment, the proposed CNN model outperformed four conventional CNN in terms of recognition accuracy, quantity of trainable parameters, and training time. In the last part of the experiment, the proposed CNN was compared with the latest works, with the results indicating its higher performance in term of recognition accuracy. In a future study, we intend to expand TL learning in two parallel domains. For this purpose, we will consider the fingerprint dataset as the source domain and the FKP dataset as the target domain. The proposed CNN will be trained on the fingerprint dataset and the obtained knowledge will be used for FKP.

7. REFERENCES

1. Dargan, S. and Kumar, M., "A comprehensive survey on the biometric recognition systems based on physiological and behavioral modalities", ***Expert Systems with Applications***, Vol. 143, (2020). DOI: https://doi.org/10.1016/j.eswa.2019.113114.
2. Guo, G. and Zhang, N., "A survey on deep learning based face recognition", ***Computer Vision and Image Understanding***, Vol. 189, (2019), 102805. DOI: https://doi.org/10.1016/j.cviu.2019.102805.
3. Proença, H. and Neves, J.C., "Segmentation-less and non-holistic deep-learning frameworks for iris recognition", in 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). (2019), 2296-2305.DOI: 10.1109/CVPRW.2019.00283.
4. Win, K.N., Li, K., Chen, J., Viger, P.F. and Li, K., "Fingerprint classification and identification algorithms for criminal investigation: A survey", ***Future Generation Computer Systems***, Vol. 110, (2020), 758-771. DOI: https://doi.org/10.1016/j.future.2019.10.019.
5. Zhong, D., Du, X. and Zhong, K., "Decade progress of palmprint recognition: A brief survey", ***Neurocomputing***, Vol. 328, (2019), 16-28. DOI: https://doi.org/10.1016/j.neucom.2018.03.081.
6. Klonowski, M., Plata, M. and Syga, P., "User authorization based on hand geometry without special equipment", ***Pattern Recognition***, Vol. 73, (2018), 189-201. DOI: https://doi.org/10.1016/j.patcog.2017.08.017.
7. Liu, H., Yang, G., Yang, L. and Yin, Y., "Learning personalized binary codes for finger vein recognition", ***Neurocomputing***, Vol. 365, (2019), 62-70. DOI: https://doi.org/10.1016/j.neucom.2019.07.057.
8. Muthukumar, A. and Kavipriya, A., "A biometric system based on gabor feature extraction with svm classifier for finger-knuckle-print", ***Pattern Recognition Letters***, Vol. 125, (2019), 150-156. DOI: https://doi.org/10.1016/j.patrec.2019.04.007.
9. Zhang, L., Zhang, L., Zhang, D. and Zhu, H., "Ensemble of local and global information for finger–knuckle-print recognition", ***Pattern Recognition***, Vol. 44, No. 9, (2011), 1990-1998.DOI: https://doi.org/10.1016/j.patcog.2010.06.007.
10. Kumar, A. and Ravikanth, C., "Personal authentication using finger knuckle surface", ***IEEE Transactions on Information Forensics and Security***, Vol. 4, No. 1, (2009), 98-110. DOI: 10.1109/TIFS.2008.2011089.
11. Simonyan, K. and Zisserman, A., "Very deep convolutional networks for large-scale image recognition", ***arXiv preprint arXiv:1409.1556***, (2014).
12. Szegedy, C., Wei, L., Yangqing, J., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V. and Rabinovich, A., "Going deeper with convolutions", in 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). (2015), 1-9.DOI: 10.1109/CVPR.2015.7298594.
13. He, K., Zhang, X., Ren, S. and Sun, J., "Deep residual learning for image recognition", in 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). (2016), 770-778.DOI: 10.1109/CVPR.2016.90.
14. Krizhevsky, A., Sutskever, I. and Hinton, G.E., "Imagenet classification with deep convolutional neural networks", ***Commun. ACM***, Vol. 60, No. 6, (2017), 84-90. DOI: 10.1145/3065386.

15. Woodard, D.L. and Flynn, P.J., "Finger surface as a biometric identifier", ***Computer Vision and Image Understanding***, Vol. 100, No. 3, (2005), 357-384. DOI: https://doi.org/10.1016/j.cviu.2005.06.003.
16. Kumar, A. and Zhou, Y., "Personal identification using finger knuckle orientation features", ***Electronics Letters***, Vol. 45, No. 20, (2009), 1023-1025.
17. Zhang, L., Zhang, L., Zhang, D. and Zhu, H., "Online finger-knuckle-print verification for personal authentication", ***Pattern Recognition***, Vol. 43, No. 7, (2010), 2560-2571. DOI: https://doi.org/10.1016/j.patcog.2010.01.020.
18. Zhang, L., Zhang, L. and Zhang, D., "Finger-knuckle-print verification based on band-limited phase-only correlation", in Computer Analysis of Images and Patterns, Berlin, Heidelberg, Springer Berlin Heidelberg. (2009), 141-148.DOI: https://doi.org/10.1007/978-3-642-03767-2_17.
19. Morales, A., Travieso, C., Ferrer, M. and Alonso, J., "Improved finger-knuckle-print authentication based on orientation enhancement", ***Electronics Letters***, Vol. 47, No. 6, (2011), 380-381.
20. Zhu, L., "Finger knuckle print recognition based on surf algorithm", in 2011 Eighth International Conference on Fuzzy Systems and Knowledge Discovery (FSKD). Vol. 3, (2011), 1879-1883.DOI: 10.1109/FSKD.2011.6019781.
21. Badrinath, G.S., Nigam, A. and Gupta, P., "An efficient finger-knuckle-print based recognition system fusing sift and surf matching scores, Berlin, Heidelberg, Springer Berlin Heidelberg. Vol., (2011), 374-387. DOI: 10.1007/978-3-642-25243-3_30.
22. Chaa, M., Boukezzoula, N.-E. and Meraoumia, A., "Features-level fusion of reflectance and illumination images in finger-knuckle-print identification system", ***International Journal on Artificial Intelligence Tools***, Vol. 27, No. 03, (2018), 1850007.DOI: 10.1142/s0218213018500070.
23. Vidhyapriya, R. and Lovelyn Rose, S., "Personal authentication mechanism based on finger knuckle print", ***Journal of Medical Systems***, Vol. 43, No. 8, (2019), 232.DOI: 10.1007/s10916-019-1332-3.
24. Heidari, H. and Chalechale, A., "A new biometric identity recognition system based on a combination of superior features in finger knuckle print images", ***Turkish Journal of Electrical Engineering & Computer Sciences***, Vol. 28, No. 1, (2020), 238-252.DOI: 10.3906/elk-1906-12.
25. Attia, A., Akhtar, Z., Chalabi, N.E., Maza, S. and Chahir, Y., "Deep rule-based classifier for finger knuckle pattern recognition system", ***Evolving Systems***, (2020). DOI: 10.1007/s12530-020-09359-w.
26. Chlaoua, R., Meraoumia, A., Aiadi, K.E. and Korichi, M., "Deep learning for finger-knuckle-print identification system based on pcanet and svm classifier", ***Evolving Systems***, Vol. 10, No. 2, (2019), 261-272. DOI: 10.1007/s12530-018-9227-y.
27. LeCun, Y. and Bengio, Y., "Convolutional networks for images, speech, and time series", ***The Handbook of Brain Theory and Neural Networks***, Vol. 3361, No. 10, (1995), 1995.
28. Feizi, A., "Convolutional gating network for object tracking", ***International Journal of Engineering, Transactions A: Basics***, Vol. 32, No. 7, (2019), 931-939. DOI: 10.5829/ije.2019.32.07a.05.
29. Gholamalinejad, H. and Khosravi, H., "Irvd: A large-scale dataset for classification of iranian vehicles in urban streets", ***Journal of AI and Data Mining***, Vol. 9, No. 1, (2021), 1-9. DOI: 10.22044/jadm.2020.8438.1982.
30. Khatami, A., Babaie, M., Tizhoosh, H.R., Nazari, A., Khosravi, A. and Nahavandi, S., "A radon-based convolutional neural network for medical image retrieval", ***International Journal of Engineering, Transactions C: Aspects***, Vol. 31, No. 6, (2018), 910-915. DOI: 10.5829/ije.2018.31.06c.07.
31. Zohrevand, A., Sattari, M., Sadri, J., Imani, Z., Suen, C.Y. and Djeddi, C., "Comparison of persian handwritten digit recognition in three color modalities using deep neural networks, Cham, Springer International Publishing. (2020), 125-136. DOI: 10.1007/978-3-030-59830-3_11.
32. Eckle, K. and Schmidt-Hieber, J., "A comparison of deep networks with relu activation function and linear spline-type methods", ***Neural Networks***, Vol. 110, (2019), 232-242. DOI: https://doi.org/10.1016/j.neunet.2018.11.005.
33. Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I. and Salakhutdinov, R., "Dropout: A simple way to prevent neural networks from overfitting", ***The Journal of Machine Learning Research***, Vol. 15, No. 1, (2014), 1929-1958.
34. Joshi, J.C., Nangia, S.A., Tiwari, K. and Gupta, K.K., "Finger knuckleprint based personal authentication using siamese network", in 2019 6th International Conference on Signal Processing and Integrated Networks (SPIN). (2019), 282-286. DOI: 10.1109/SPIN.2019.8711663.
35. Zhai, Y., Cao, H., Cao, L., Ma, H., Gan, J., Zeng, J., Piuri, V., Scotti, F., Deng, W., Zhi, Y. and Wang, J., "A novel finger-knuckle-print recognition based on batch-normalized cnn, Cham, Springer International Publishing. (2018), 11-21. DOI: 10.1007/978-3-319-97909-0_2.
36. Zhuang, F., Qi, Z., Duan, K., Xi, D., Zhu, Y., Zhu, H., Xiong, H. and He, Q., "A comprehensive survey on transfer learning", ***Proceedings of the IEEE***, Vol. 109, No. 1, (2021), 43-76. DOI: 10.1109/JPROC.2020.3004555.
37. Liu, S., Tian, G. and Xu, Y., "A novel scene classification model combining resnet based transfer learning and data augmentation with a filter", ***Neurocomputing***, Vol. 338, (2019), 191-206. DOI: https://doi.org/10.1016/j.neucom.2019.01.090.
38. Mersa, O., Etaati, F., Masoudnia, S. and Araabi, B.N., "Learning representations from persian handwriting for offline signature verification, a deep transfer learning approach", in 2019 4th International Conference on Pattern Recognition and Image Analysis (IPRIA). (2019), 268-273. DOI: 10.1109/PRIA.2019.8785979.
39. Kingma, D.P. and Ba, J., "Adam: A method for stochastic optimization", ***arXiv preprint arXiv:1412.6980***, Vol., No., (2014).
40. Shorten, C. and Khoshgoftaar, T.M., "A survey on image data augmentation for deep learning", ***Journal of Big Data***, Vol. 6, No. 1, (2019), 60. DOI: 10.1186/s40537-019-0197-0.
41. Bloice, M.D., Stocker, C. and Holzinger, A., "Augmentor: An image augmentation library for machine learning", ***Journal of Open Source Software***, Vol. 2, No. 19, (2017), 432. DOI: 10.21105/joss.00432.
42. Shariatmadar, Z.S. and Faez, K., "An efficient method for finger-knuckle-print recognition by using the information fusion at different levels", in 2011 International Conference on Hand-Based Biometrics. (2011), 1-6. DOI: 10.1109/ICHB.2011.6094325.
43. Zeinali, B., Ayatollahi, A. and Kakooei, M., "A novel method of applying directional filter bank (dfb) for finger-knuckle-print (fkp) recognition", in 2014 22nd Iranian Conference on Electrical Engineering (ICEE). (2014), 500-504. DOI: 10.1109/IranianCEE.2014.6999594.
44. Nunsong, W. and Woraratpanya, K., "Modified differential box-counting method using weighted triangle-box partition", in 2015 7th International Conference on Information Technology and Electrical Engineering (ICITEE). (2015), 221-226. DOI: 10.1109/ICITEED.2015.7408945.
45. Nunsong, W. and Woraratpanya, K., "An improved finger-knuckle-print recognition using fractal dimension based on gabor wavelet", in 2016 13th International Joint Conference on Computer Science and Software Engineering (JCSSE). (2016), 1-5. DOI: 10.1109/JCSSE.2016.7748904.

Persian Abstract

چکیده

به دلیل اینکه ناحیه پشتی انگشت در معرض تماس با سطوح قرار نمی‌گیرد، در مقایسه با سایر بایومتریک ها مبتنی بردست، Finger-Knuckle-Print (FKP) یک بایومتریک دقیق و قابل اعتماد می‌باشد. در این مقاله یک روش ساده بر اساس شبکه های عصبی عمیق برای تشخیص FKP پیشنهاد شده است. روش پیشنهادی تنها از سه لایه کانوولوشنی و دو لایه کاملا متصل تشکیل شده است. به همین دلیل تعداد وزن های قابل آموزش به شدت کاسته شده است. همچنین یک روش سرراست برای افزودن داده در این مقاله پیشنهاد شده است. بهترین دقت، دقت متوسط و انحراف معیار روش پیشنهادی روی پایگاه داده Poly-U FKP و براساس10-fold cross-validation به ترتیب ۹۹.۸۳٪ ، ۹۹۱۸٪ و ۰.۷۶ می باشد. نتایج پیاده سازی نشان می دهد که روش پیشنهادی از لحاظ دقت تشخیص و تعداد پارامترهای قابل آموزش، از روش های موجود بهتر عمل می‌کند. همچنین در مقایسه با شبکه های عصبی مانند AlexNet، VGG16، ResNet34 و GoogleNet که برای تشخیص FKP اصلاحا fine-tune شده‌اند روش ساده پیشنهادی از نظر دقت تشخیص و تعداد پارامترهای قابل آموزش و زمان آموزش عملکرد بالاتری را به ارائه می‌دهد.

IJE TRANSACTIONS A: Basics Vol. 34, No. 7, (July 2021) 1694-1702

International Journal of Engineering

Journal Homepage: www.ije.ir

A Disturbance Observer Based Fuzzy Feedforward Proportional Integral Load Frequency Control of Microgrids

S. Asgari*[a], M. B. Menhaj[b], A. Abolfazl Suratgar[a], M. G. Kazemi[c]

[a] *Distributed Intelligent Optimization Research Laboratory, Department of Electrical Engineering, Amirkabir University of Technology, Tehran, Iran*

[b] *Department of Electrical Engineering, Amirkabir University of Technology, Tehran, Iran*

[c] *Department of Electrical and Computer Engineering, Islamic Azad University of Gachsaran, Gachsaran, Iran*

PAPER INFO

Paper history:
Received 29 November 2020
Received in revised form 16 March 2021
Accepted 17 April 2021

Keywords:
Microgrid
Load Frequency Control
Fuzzy Feedforward PI Controller
Disturbance Observer

ABSTRACT

In this paper, the load frequency control (LFC) problem of microgrids in islanded operation mode is tackled using Fuzzy Feedforward PI (FFPI) controller. To this end, a feedforward loop is considered in the control structure of the microgrid in addition to the classical feedback controller wherein a proportional integrator controller is used. The disturbance signal, which can be load variations or renewable energy resources uncertainties, is estimated using a disturbance observer. The understudy microgrid includes wind turbine and solar cells as renewable sources, a diesel generator and loads. A fuzzy controller is also used for pitch angle control of the wind turbine, which may smooth out the generated power and improve the frequency control of microgrid. To show the capability of the proposed strategy, two different scenarios are considered and the obtained simulation results easily approve the efficiency of the proposed structure for LFC of microgrid in islanded operation mode.

doi: *10.5829/ije.2021.34.07a.13*

1. INTRODUCTION

Electrical energy transmission cost in power systems leads to pay much attention to new technologies such as microgrid, which results in some other benefits such as more reliability and power quality improvement. On the other hand, fossil fuel sources are substantially reduced in these years, which causes to much attention to renewable energy sources such as wind energy and solar cells. These renewable sources have unpredictable nature due to change in wind speed or solar irradiance. The variations in produced power or load in a power system leads to frequency deviations, which is an unfavorable phenomenon and must be compensated by the controller as fast as possible.

Microgrids operation mode is defined as the grid-connected and islanded one. In the islanded operation of a microgrid, a local controller is needed to control the frequency deviations due to change in produced power or load. In this regard, various control strategies and controllers are studied in the literature.

Fuzzy logic sets were the beginning of a new worldview into mathematics and a novel step into real world problems solution in a realistic manner. The fuzzy logic provides the more practical and understandable concept for understudy process with linguistic variables as if-then rules, which further provides more efficient control of the system. A fuzzy controller includes fuzzifier, inference mechanism, knowledge base and de-fuzzifier [1]. Fuzzy logic finds its path into different applications such as modeling, control [2], and fault diagnosis [3].

Diverse control strategies towards frequency, voltage, current and power control of microgrids have been considered. Conventional feedback structure, computational intelligence based controllers, decentralized and hierarchical control strategies [4, 5], and cascade control structure [6] are used for the

*Corresponding Author Institutional Email: asgari.shadi@aut.ac.ir (S. Asgari)

Please cite this article as: S. Asgari, M. B. Menhaj, A. Abolfazl Suratgar, M. G. Kazemi, A Disturbance Observer Based Fuzzy Feedforward Proportional Integral Load Frequency Control of Microgrids, International Journal of Engineering, Transactions A: Basics Vol. 34, No. 7, (2021) 1694-1702

microgrids. Due to importance of frequency control of the microgrids in islanded operation, various control strategies have been proposed in the literatures. In order to deal with the control challenges of islanded microgrids, an efficient control strategy is required. Toward this end, a Central Management Agent which maintains the stability of the microgrid by controlling an Energy Storage System and a Central Synchronous Generator is presented by Farzinfar et al. [7].

Adaptive controller [8], sliding mode controllers [9], fractional order controllers [10], model predictive controllers (MPC) [11], and multi agent approach [12] are studied for LFC of microgrids. Robust control strategy also noticed for LFC of microgrids in order to make the control system robust to model uncertainties and disturbances [13,14]. Quantitative feedback theory is another robust control method, which consists of a feedback compensator and a pre-filter, and is designed by shaping the frequency responses of the system to satisfy some design constraints [15]. A nonlinear controller for a distribution static compensator of a microgrid based on partial feedback linearization theory and PID controller is proposed by Ara et al. [16], which uses a combination of a fuzzy system and Galaxy-based Search Algorithm to optimize the parameters of the PID controllers. Optimal PI-settings for first-order with delay systems to achieve a specific level of robustness was presented by Grimholt and Skogestad [17] and the results were compared with the simple SIMC-rule. Some surveys on the modeling and control of microgrids, and their control strategies have been published recently [18,19].

On the other hand, computational intelligence based methods have received much attention due to their benefits in modeling and control design of microgrids. Boutabba et al. [20] presented a modeling and implementation of new control schemes for an isolated photovoltaic using a fuzzy logic controller. The proposed fuzzy logic controller provides the appropriate duty cycle to the DC-DC converter for the PV system to achieve Maximum Power Point Tracking in PV. Mahmoud et al. [21] have achieved the LFC of microgrid using neural networks controller. Different metaheuristic methods such as GA, PSO and HSA are used to optimize the controller for LFC of microgrids [22-24]. a fractional order fuzzy PID controller is designed for LFC of a microgrid in isolated mode [25]. LFC of isolated microgrids in a ship power system is considered by Khooban et al. [26], which is based on a new optimal fractional order fuzzy PD+I controller. An adaptive multi-objective fractional-order fuzzy PID controller for frequency control of an islanded microgrid including Electric Vehicles is presented by Khooban et al. [27].

The feedback control strategy for frequency deviation control of microgrid is based upon the effects of disturbance, which leads to frequency deviation. The feedforward controller acts in a different manner and change the input of the system according to the disturbance signal or its estimation. In fact, prior to have the effects of disturbances on the system outputs, the controller takes the necessary action by properly setting the system input. In this paper, based on a disturbance observer the disturbance signal, which is defined as the power variations due to load or renewable sources, is estimated in the islanded microgrid. The fuzzy logic controller is used in the feedforward controller of the microgrid wherein the disturbance and its variations are used to change the parameters of the controller.

Succinctly, the salient contribution of this paper is stated as follows:

- Presenting a new control structure including Fuzzy Feedforward PI controller for LFC of an islanded microgrid using disturbance observer.

Moreover, a fuzzy logic controller is used to pitch angle control of wind turbine for more smoothing out the power of wind turbine, and consequently possessing a better frequency control of the microgrid.

The rest of the paper is organized as follows. In section 2, the structure of the microgrid and mathematical model for different parts of the microgrid are presented. The proposed LFC strategy for the islanded operation of microgrid is presented in section 3. Section 4 gives the simulation results of the proposed method, which is followed by a conclusion in section 5.

2. PROBLEM FORMULATION and PRELIMINARIES

Due to various sources of energy, different structures of the microgrid may be presented. The considered microgrid in this paper is shown in Figure 1, which includes both renewable energy resources and diesel generator as the deterministic source of energy.

In Figure 1, the generated power of wind turbine, solar cells and diesel generator are represented as P_W, P_S and P_{DG}. The power of the load is also shown as P_L. The considered mathematical models for the above mentioned parts are given as follows.

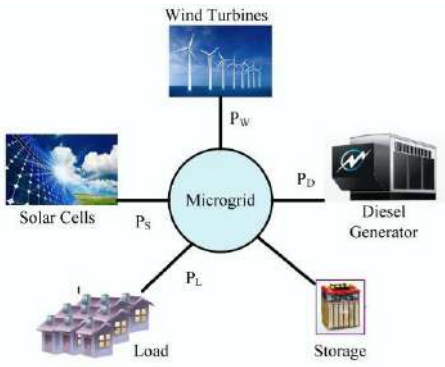


Figure 1. Structure of the microgrid

2. 1. Wind Turbine The mathematical model of a wind turbine is generally involved the mathematical model of wind, generator, aerodynamics and mechanical drive train. Wind model is defined by the wind speed, which is given in Equation (1):

$$v = v_m + v_t(t) \tag{1}$$

v_m is the constant part of wind model with slow variation rate, which shows the mean value of wind speed. v_t is the turbulent part of wind speed with fast variation rate. More details about the turbulent part may be found in literature [28]. The kinetic energy of wind as stated in Equation (2) defines the aerodynamic part of the wind turbine.

$$P = \frac{1}{2}\dot{m}v^2 = \frac{1}{2}\rho\pi R^2 v^3 \tag{2}$$

in which ρ is defined as air density and R is the radius of the rotor. The part of the kinetic energy that converts to mechanical energy is defined by power coefficient (C_p) as (3) as given by Burkart et al. [29].

$$P_r = C_p.P \tag{3}$$

Equation (4) represents C_p, which is dependent on the tip speed ratio (λ) and pitch angle (β).

$$C_p = 0.5176(\frac{116}{\lambda_i} - 0.4\beta - 5)e^{\frac{-21}{\lambda_i}} + 0.0068\lambda \tag{4}$$

where

$$\frac{1}{\lambda_i} = \frac{1}{\lambda+0.08\beta} - \frac{0.035}{\beta^3+1} \tag{5}$$

The maximum value of C_p is given as 0.59 based on Albert Betz's law. This value is 0.49 for the understudy wind turbine $(V47/660kW)$, operating in the Manjil site [30,31].

The generator of the understudy wind turbine is an induction one. Although the induction generator has a nonlinear mathematical model, the steady state mathematical model due to the much faster dynamics of the generator in comparison to the mechanical parts is assumed here [32].

$$T_g(\omega_i) = \frac{\rho R_2 L_m^2 V_e^2 \omega_i}{[R_1R_2 - \omega_a\omega_i(L_1L_m - L_m^2)]^2 + (\omega_a R_2 L_1 + \omega_i R_1 L_2)^2} \tag{6}$$

The root mean square value of the supply voltage, frequency of supply voltage, rotor frequency, number of pole pairs, mutual inductance, resistance of stator and rotor, stator inductance and rotor inductance are represented by V_e, ω_a, ω_i, p, L_m, R_1, R_2, L_1 and L_2, respectively [32].

The considered one-mass model of the mechanical drive train is given in Equation (7) [33]:

$$J\frac{d\omega_{rot}}{dt} = T_{rot} - nT_g, \omega_{rot} = \frac{\omega_g}{n} \tag{7}$$

where J is the inertia, ω_{rot} is the speed of rotor, ω_g the speed of generator, n is the gearbox ratio, T_{rot} is rotor torque and T_g is generator torque.

2. 2. PV Cells Model Solar cells convert sun irradiance into electrical energy. The considered microgrid includes PV cells as a renewable source of energy, which its power is dependent on the environmental temperature as well. Generally, the mathematical model of PV cells is obtained using the equivalent electrical circuit, which is assumed as a nonlinear current source [34]. Equation (8) is achieved based on KCL law for the equivalent one-diode model of PV cell [34].

$$I = I_{ph} - I_{Rp} - I_d \tag{8}$$

Assuming N_s cells in series and N_p cells in parallel, the relation between I and V for a PV array is given by following equations.

$$I = N_p I_{Ph} - N_p I_0[e^{\frac{q[\frac{V}{N_s} + \frac{IR_s}{N_p}]}{nkT}} - 1] - \frac{[\frac{N_p V}{N_s} + IR_s]}{R_{sh}} \tag{9}$$

$$I_{Ph} = [I_{scr} + K_i(T - T_r)]\frac{S}{100} \tag{10}$$

where I_{ph}, I_0 and I stand for insulation current, reverse saturation current and cell current, respectively. The series and parallel resistance, thermal voltage, irradiance, the temperature of PV array, reference temperature and the short circuit current are depicted as R_s, R_p, S, T, T_r and I_{scr}, respectively

2. 3. Diesel Generator A controllable and deterministic source of power is required for better control of a microgrid and to compensate the unpredictable conditions of available renewable energies in the microgrid. To this end, a diesel generator is considered here. The mathematical model of the power system is as follows [9]:

$$\Delta\dot{f}(t) = -\frac{1}{T_p}\Delta f(t) + \frac{K_p}{T_p}\Delta P_d(t) - \frac{K_p}{T_p}\Delta P_L(t) \tag{11}$$

$$\Delta\dot{P}_d(t) = -\frac{1}{T_t}\Delta P_d(t) + \frac{1}{T_t}\Delta X_g(t) \tag{12}$$

$$\Delta\dot{X}_g(t) = -\frac{1}{RT_g}\Delta f(t) - \frac{1}{T_g}\Delta X_g(t) - \frac{1}{T_g}\Delta E(t) + \frac{1}{T_g}u(t) \tag{13}$$

$$\Delta\dot{E} = K_e\Delta f(t) \tag{14}$$

where frequency deviation $\Delta f(t)$, diesel generator power changes $\Delta P_d(t)$, load deviation $\Delta X_g(t)$ and incremental change in governor position ΔE are considered as the states of the system. The definitions and values of other parameters are tabulated in Table 1.

3. CONTROLLER DESIGN

The proposed structure for LFC of the microgrid is depicted in Figure 2.

TABLE 1. Parameters of power system and their values [9]

Parameter	Definition	Value
T_p	Time constant of plant	20
K_p	Plant gain	120
T_t	Time constant of turbine	0.3
T_g	Governor time constant	0.08
R	Speed regulation	2.4
K_e	Integral gain	0.1

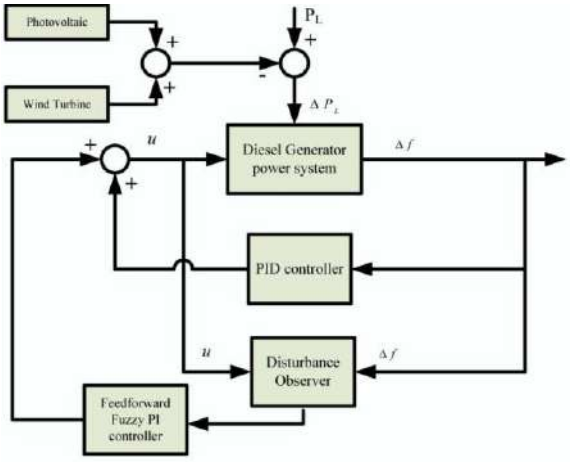


Figure 2. Structure of the proposed fuzzy feedforward control

This structure includes a FFPI controller and a disturbance observer in addition to the conventional feedback controller. The disturbance observer is used to estimate the disturbance from the input data and frequency deviation at each moment. In the proposed control strategy, the disturbance is defined as the deviation in produced power and load. The disturbance observer, fuzzy feedforward controller and fuzzy pitch angle controller are given in the following subsections.

3. 1. Disturbance Observer Unlike the feedback control strategy that acts based upon the effects of disturbances on the outputs, the feedforward control acts based on the known disturbances prior to affect on the system. In other words, we need the value of disturbance to use the feedforward controller. As the considered disturbance in the microgrid is unknown, a disturbance observer is utilized to estimate the value of the disturbance. The estimated disturbance is given to the fuzzy feedforward controller, which produces an appropriate control signal.

The disturbance observer is a modified version of the given one by wang et al. [9]. As discussed by Wang et al. [9], by augmentation the disturbance as one state of the system, the augmented state space representation form may be given as follows:

$$\begin{bmatrix}\dot{x}\\ \dot{W}\end{bmatrix}=\begin{bmatrix}A & I\\ 0 & 0\end{bmatrix}\begin{bmatrix}x\\ W\end{bmatrix}+\begin{bmatrix}B\\ 0\end{bmatrix}u+\begin{bmatrix}0\\ \dot{W}\end{bmatrix} \quad (15)$$

The observer for the augmented system, which simultaneously estimates the states and disturbances, has the following form:

$$\begin{bmatrix}\dot{\hat{x}}\\ \dot{\hat{W}}\end{bmatrix}=\begin{bmatrix}A & I\\ 0 & 0\end{bmatrix}\begin{bmatrix}\hat{x}\\ \hat{W}\end{bmatrix}+\begin{bmatrix}B\\ 0\end{bmatrix}u-\begin{bmatrix}L_1\\ L_2\end{bmatrix}(\hat{y}-y) \quad (16)$$

$$\hat{y}=C\hat{x} \quad (17)$$

where $\hat{x}$ and $\hat{W}$ are the state and output estimation of the observer, respectively. The two-part gain of the observer is considered as L_1 and L_2. Equation (17) may be written as Equation (18):

$$\begin{bmatrix}\dot{\hat{x}}\\ \dot{\hat{W}}\end{bmatrix}=\begin{bmatrix}A & I\\ 0 & 0\end{bmatrix}\begin{bmatrix}\hat{x}\\ \hat{W}\end{bmatrix}+\begin{bmatrix}B\\ 0\end{bmatrix}u-\begin{bmatrix}L_1.C\\ L_2.C\end{bmatrix}\tilde{x} \quad (18)$$

where $\tilde{x}$ is defined as the state estimation error of the observer as Equation (19).

$$\tilde{x}=\hat{x}-x \quad (19)$$

The state estimation error of the observer can be calculated as follows:

$$\begin{bmatrix}\dot{\tilde{x}}\\ \dot{\tilde{W}}\end{bmatrix}=\hat{A}\begin{bmatrix}\tilde{x}\\ \tilde{W}\end{bmatrix}-\begin{bmatrix}0\\ \dot{W}\end{bmatrix} \quad (20)$$

where

$$\tilde{W}=\hat{W}-W \quad (21)$$

$$\hat{A}=\begin{bmatrix}A-L_1C & I\\ -L_2C & 0\end{bmatrix} \quad (22)$$

The lemma by Wang et al. [9] is used to calculate the gain of the disturbance observer.

Lemma [9]: For the given L_1 and L_2 as Equations (23) and (24), the state estimation error of the disturbance observer Equation (18) is asymptotically stable.

$$L_1=A+2\Lambda \quad (23)$$

$$L_2=\Lambda^2 I \quad (24)$$

wherein

$$\Lambda=diag\{\lambda_1,\dots,\lambda_n\},\lambda_i>0,i=1,2,\dots,n \quad (25)$$

Using the above-mentioned lemma, the values for L_1 and L_2 are obtained as follows for $\Lambda=\{50,50,100,100\}$.

$$L_1=\begin{bmatrix}87.5 & 0 & -5.2 & 5.2\\ 0 & 1.65 & 87.5 & 10\end{bmatrix}^T,$$

$$L_2=1e3\begin{bmatrix}2.5 & 0 & 0 & 0\\ 0 & 0 & 5 & 5\end{bmatrix}^T$$

3. 2. Pitch Angle Fuzzy Controller In order to obtain a smooth power from the wind turbine in its third

operation mode in the microgrid, a fuzzy logic controller is used in this paper. The structure of the fuzzy pitch angle controller is shown in Figure 3.

The input of the controller is the variation in rotational speed and the output is considered as the pitch angle variation. The input and output membership functions are given in Figures 4 and 5, respectively.

Fuzzy rules of the controller are given in Table 2. The Mamdani inference and center of gravity de-fuzzifier method is used in the fuzzy controller.

3. 3. Feedforward Fuzzy Controller Feedforward control is used to reduce the effect of disturbances in the system, while the main functions of feedback control are stability and the reference signal tracking. Having a fixed feedforward controller may lead to unfavorable results in some conditions due to the nonlinear dynamics of the system and their limitations. In this paper, a fixed structure feedforward controller is regarded in which the parameters are updated based upon a fuzzy rule. The fuzzy rule structure for updating of the feedforward controller parameters is shown in Figure 6, which leads to better performance of LFC.

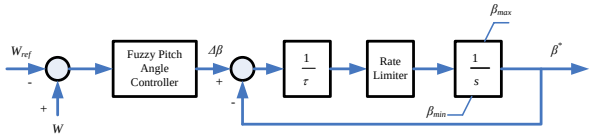


Figure 3. Structure of the fuzzy controller

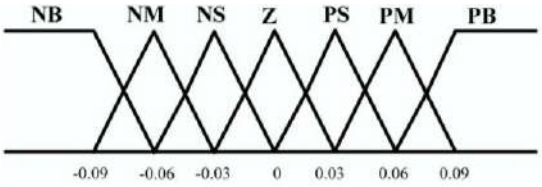


Figure 4. Input membership function for pitch angle control

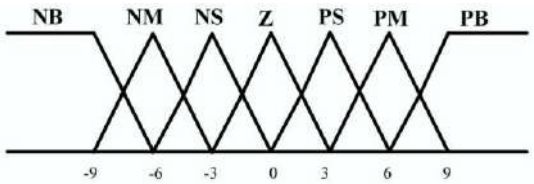


Figure 5. Output membership function for pitch angle control

TABLE 2. Fuzzy rules for pitch angle control.

ΔW	NB	NM	NS	Z	PS	PM	PB
$\Delta\beta$	NB	NM	NS	Z	PS	PM	PB

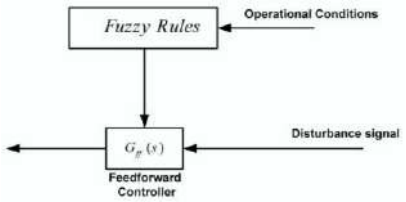


Figure 6. Fuzzy rules for feedforward controller design

A Proportional-Integral (PI) controller is used in the feedforward controller wherein their parameters including K_p and K_I are tuned based on the fuzzy logic. The normalized fuzzy rules for these two parameters are given in Tables 3 and 4.

The membership functions of the input and output of the fuzzy system are depicted in Figures 7 and 8, respectively.

TABLE 3. Fuzzy rules for K_p of the FFPI controller

Disturbance		NB	NM	Z	PM	PB
Disturbance Variation	NB	S	S	S	S	S
	NM	M	M	S	S	S
	Z	S	S	S	S	S
	PM	M	M	M	B	B
	PB	B	B	B	B	B

TABLE 4. Fuzzy rules for K_I of the FFPI controller

Disturbance		NB	NS	Z	PS	PB
Disturbance Variation	NB	B	B	M	S	Z
	NM	B	B	M	Z	S
	Z	M	M	Z	M	M
	PM	M	M	M	B	B
	PB	Z	S	M	B	B

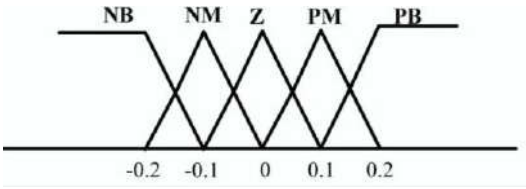


Figure 7. Input membership function of FFPI controller

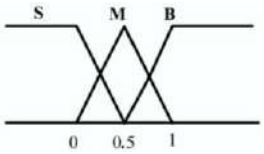


Figure 8. Output membership function of FFPI controller

The Mamdani inference mechanism and center of gravity de-fuzzifier method is used in the feedforward fuzzy controller as well.

4. SIMULATION RESULTS

The efficacy of our proposed feedforward fuzzy controller for LFC purpose is shown through some simulation studies for the considered microgrid. In the following, two scenarios including high integrations of wind power, solar power and load disturbance case and sudden load change case are considered. These two scenarios are studied by Khooban et al. [8].

4. 1. Scenario 1: High Integrations of Wind Power, Solar Power and Load Disturbance In this scenario, the performance of the proposed controller is verified in the presence of high integration of wind, solar and load power. The load model in this scenario is assumed as follows as shown in Figure 9. The wind speed and solar irradiation are also depicted in Figures 10 and 11, respectively.

The power variation of the wind turbine with the fuzzy pitch angle controller is shown in Figure 12.

As it can be seen from the figure, the fuzzy pitch angle controller leads to more smooth power in the wind turbine. Therefore, a better LFC may be achieved by the main controller of the microgrid. The obtained power from PV array considering given solar radiation is shown in Figure 13.

The estimation of disturbance signal and its real values is shown in Figure 14.

The disturbance observer efficiently estimates the disturbance signal in the microgrid. This estimated signal is given to the fuzzy feedforward controller, which further changes the diesel generator power to alleviate the unfavorable frequency deviation. The frequency deviation of the proposed FFPI control strategy compared to the conventional PID controller is given in Figure 15.

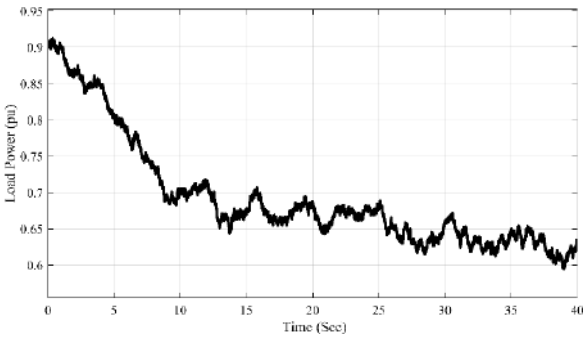


Figure 9. Load model in the microgrid

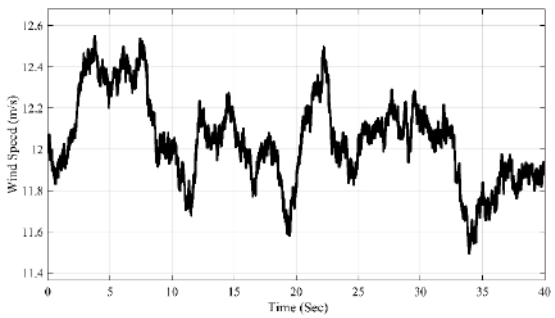


Figure 10. Wind speed

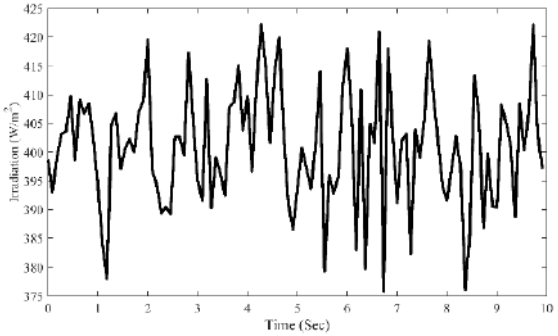


Figure 11. Solar irradiation

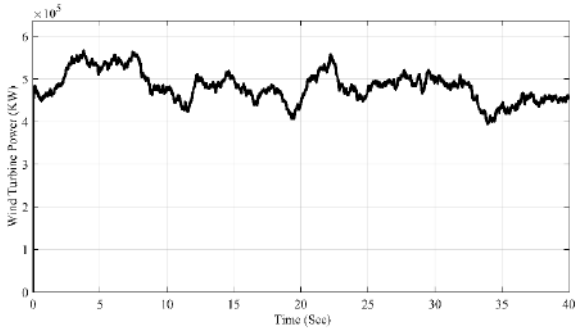


Figure 12. WT power with fuzzy pitch angle controller

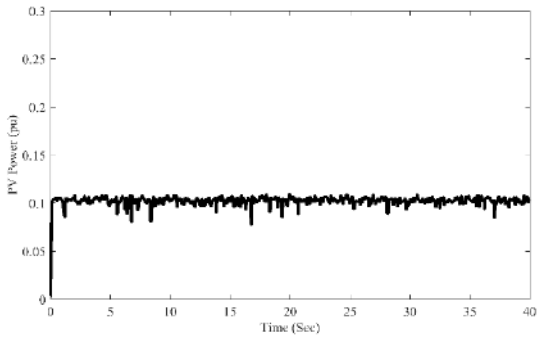


Figure 13. PV array power

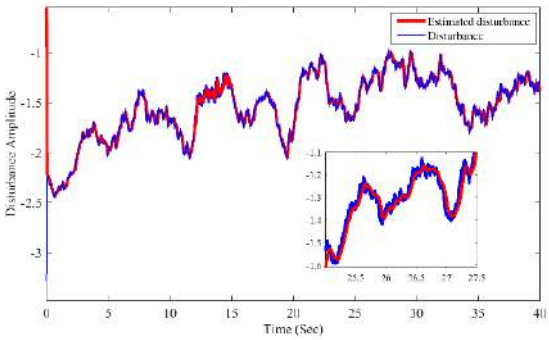


Figure 14. Estimated disturbance signal and its real value

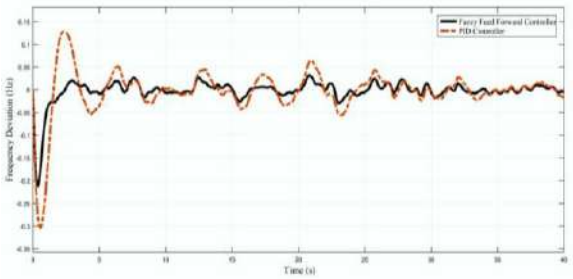


Figure 15. Frequency deviation for the first scenario

The proposed fuzzy feedforward controller and fuzzy pitch angle controller lead to great efficiency in the frequency deviation of the microgrid. For better comparison between the FFPI controller and the conventional PID controller, some criteria of the frequency deviation are given in Table 5, which are nominated as IAFD (Integral Absolute Frequency Deviation), ITAFD (Integral Time Absolute Frequency Deviation) and ITSFD (Integral Time Sqaure Frequency Deviation). The conventional PID control strategy is tunned using modified Harmony Search (HS) algorithm as presented by Shivaie et al. [35]. In the FFPI controller case, we simultaneously have both conventional PID feedback controller and feedforward fuzzy PI controller, while the second case is the PID feedback controller without feedforward controller.

4. 2. Scenario 2: Sudden Load Change In this case, the performance of the controller is studied with series step change in the load profile. The wind and solar power fluctuations are not considered in the simulation. In other words, it is assumed that the wind and PV powers are equal to the average wind power and irradiation power in the considered period, respectively. The wind turbine power and PV power are assumed as 0.6 and 0.1 p.u., respectively. The disturbance signal and its estimation are given in Figure 16, which shows the great performance of the disturbance observer.

The frequency deviation of the proposed FFPI control strategy compared to conventional PID controller for the second scenario is given in Figure 17.

As it can be seen from the figure, the proposed method shows much better performance compared with conventional PID controller. The frequency deviation criteria of microgrid for the second scenario are also given in Table 6.

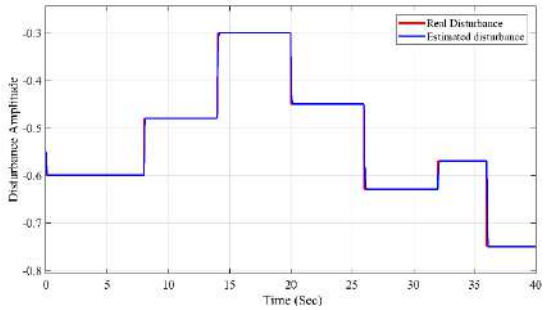


Figure 16. Estimation of disturbance in the second scenario

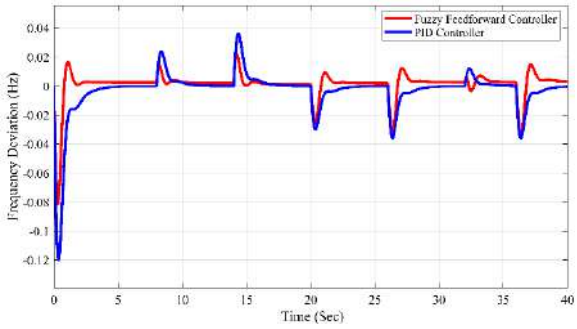


Figure 17. Frequency deviation for the second scenario

TABLE 5. Frequency deviation criteria of the first scenario for different control strategies of microgrid

Criterion vs. control method	FFPI controller	PID controller
IAFD	0.4851	1.103
ITAFD	6.302	12.72
ITSFD	0.0910	0.4182
Maximum deviation (Absolute value)	0.211	0.309

TABLE 6. Frequency deviation criteria of the first scenario for different control strategies of microgrid

Criterion vs. control method	FFPI controller	PID controller
IAFD	0.2102	0.2535
ITAFD	3.902	4.077
ITSFD	0.046	0.068
Maximum deviation (Absolute Value)	0.081	0.1198
Settling time (s) (first peak)	2.44	4.99

As it can be seen from the values of the tables, the proposed fuzzy feedforward controller with the fuzzy pitch angle controller has a much better performance compared to the conventional PID controller.

5. CONCLUSION

A novel control strategy including fuzzy logic feedforward controller based on disturbance observer and wind turbine fuzzy logic pitch angle controller for load frequency control of microgrid in its islanded operation mode has been presented. The difference between the renewable sources and load has been assumed as the disturbance in the system. The disturbance has been represented as disturbance term in the microgrid model, which has been estimated by the modified disturbance observer. The understudy microgrid contains load, wind turbine, solar cells and diesel generator. The considered fuzzy pitch angle control in wind turbine leads to more smooth power of wind turbine. Two diverse scenarios including high integration of different power sources and sudden load change have been studied to show the efficacy of both the observer and FFPI controller. The proposed FFPI control strategy has been shown to have a great impact on the performance on the frequency deviation control of the microgrid. Simulation results and numerical values of different criteria of frequency deviations for the proposed controller compared to those of the conventional form easily confirm the outperformance of the method presented in the paper.

The design and performance of the proposed control strategy for converter-based or over modulation cases can be considered as future work

6. REFERENCES

1. Yager, R. R., and Zadeh, L. A. “An introduction to fuzzy logic applications in intelligent systems.” ***Springer Science & Business Media***. Vol. 165, (2012).
2. Bakhoda, O. Z., Menhaj, M. B. and Gharehpetian, G. B. “Fuzzy logic controller vs. PI controller for MPPT of three-phase grid-connected PV system considering different irradiation conditions.” ***Journal of Intelligent & Fuzzy Systems***, Vol. 30, No. 3, (2016), 1353-1366. DOI: 10.3233/IFS-152049.
3. Yan, H., Xu, Y., Cai, F., Zhang, H., Zhao, W. and Gerada, C. “PWM-VSI Fault Diagnosis for PMSM Drive Based on Fuzzy Logic Approach.” ***IEEE Transactions on Power Electronics***. Vol. 34, No. 1, (2018), 759-768. DOI: 10.1109/TPEL.2018.2814615.
4. Zhang, B., Dou, C., Yue, D., Zhang, Z., and Zhang, T. “A cyber-physical cooperative hierarchical control strategy for islanded microgrid facing with random communication failure.” ***IEEE Systems Journal***, Vol. 14, No. 2, (2020), 2849-2860. DOI: 10.1109/JSYST.2019.2956774.
5. Sedghi L, Emam M, Fakharian A and Savaghebi M. “Decentralized control of an islanded microgrid based on offline model reference adaptive control.” ***Journal of Renewable and Sustainable Energy***. Vol. 10, No. 6, (2018), 065301. DOI: 10.1063/1.5046803.
6. Gholami, S., Mohammad A., and Sajeeb S. “Control strategy for dispatchable distributed energy resources in islanded microgrids.” ***IEEE Transactions on Power Systems***, Vol. 33. No. 1, (2018), 141-152. DOI: 10.1109/TPWRS.2017.2692801.
7. Farzinfar, M., Nair, N. K. C., and Bahadornejad, M. “A New Adaptive Load-Shedding and Restoration Strategy for Autonomous Operation of Microgrids: A Real-Time Study.” ***International Journal of Engineering, Transactions A: Basics***, Vol. 33, No. 1 (2020), 82-91. doi:10.5829/ije.2020.33.01a.10.
8. Khooban, M. H., Niknam, T., Blaabjerg, F., Davari, P. and Dragicevic, T. “A robust adaptive load frequency control for micro-grids.” ***ISA Transactions***, Vol. 65, (2016), 220-229. DOI: 10.1016/j.isatra.2016.07.002.
9. Wang, C., Mi, Y., Fu, Y. and Wang, P. “Frequency control of an isolated micro-grid using double sliding mode controllers and disturbance observer.” ***IEEE Transactions on Smart Grid***, Vol. 9, No. 2, (2016), 923-930. DOI: 10.1109/TSG.2016.2571439.
10. Pandey, S. K., Mohanty, S. R., and Kishor, N. “A literature survey on load–frequency control for conventional and distribution generation power systems.” ***Renewable and Sustainable Energy Reviews***, Vol. 25, (2013), 318-334. DOI: 10.1016/j.rser.2013.04.029.
11. Pahasa, J. and Ngamroo, I. “Coordinated control of wind turbine blade pitch angle and PHEVs using MPCs for load frequency control of microgrid.” ***IEEE Systems Journal***, Vol. 10, No. 1, (2016), 97-105. DOI: 10.1109/JSYST.2014.2313810.
12. Yu, D., Zhu, H., Han, W., and Holburn, D. “Dynamic multi agent-based management and load frequency control of PV/Fuel cell/wind turbine/CHP in autonomous microgrid system.” ***Energy***, Vol. 173, (2019), 554-568. DOI: 10.1016/j.energy.2019.02.094.
13. Nandar, C. S. A. “Robust PI control of smart controllable load for frequency stabilization of microgrid power system.*”* ***Renewable Energy***, Vol. 56, (2013), 16-23. doi:10.1016/j.renene.2012.10.032.
14. Bevrani, H., Feizi, M. R. and Ataee, S. “Robust Frequency Control in an Islanded Microgrid: H∞ and μ-Synthesis Approaches.” ***IEEE Transactions on Smart Grid***. Vol. 7, No. 2, (2016) 706-717. DOI: 10.1109/TSG.2015.2446984.
15. Naghash, R., Alavi, S. M. M., and Afjei, S. E. "Robust Control of Wireless Power Transfer despite Load and Data Communications Uncertainties." ***IEEE Journal of Emerging and Selected Topics in Power Electronics***, (2020) DOI: 10.1109/JESTPE.2020.3033001.
16. Ara, A. L., Tolabi, H. B. and Hosseini, R. “Dynamic modeling and controller design of distribution static compensator in a microgrid based on combination of fuzzy set and galaxy-based search algorithm.” ***International Journal of Engineering-Transactions A: Basics***, Vol. 29, No. 10, (2016). 1392-1400. DOI: 10.5829/idosi.ije.2016.29.10a.10.
17. Grimholt, C., Skogestad, S., “Optimal PI-Control and Verification of the SIMC Tuning Rule.” ***IFAC Proceedings Volumes***, Vol. 45, No. 3, (2012), 11-22. DOI: 10.3182/20120328-3-IT-3014.00003.
18. Rajesh, K. S., Dash, S. S. and Rajagopal, R. “Load Frequency Control of Microgrid: A Technical Review.” In Green Buildings and Sustainable Engineering. 115-138, Springer, Singapore (2020). DOI: 10.1007/978-981-15-1063-2_9.
19. Abhishek, A., Ranjan, A., Devassy, S., Verma, B. K., Ram, S. K., and Dhakar, A. K. “Review of hierarchical control strategies for DC microgrid.” ***IET Renewable Power Generation***, Vol. 14, No. 10, (2020), 1631-1640. DOI: 10.1049/iet-rpg.2019.1136.

20. Boutabba, T., Drid, S., Chrifi-Alaoui, L., and Benbouzid, M. E. "A new implementation of maximum power point tracking based on fuzzy logic algorithm for solar photovoltaic system." ***International Journal of Engineering, Transactions A: Basics***, Vol. 31, No. 4, (2018), 580-587. DOI: 10.5829/ije.2018.31.04a.09.
21. Mahmoud, M. S., Hussain, S. A. and Abido, M. A. "Modeling and control of microgrid: An overview." ***Journal of the Franklin Institute***, Vol. 351, No. 5, (2014), 2822-2859. doi:10.1016/j.jfranklin.2014.01.016.
22. Khalghani, M. R., Khooban, M. H., Mahboubi-Moghaddam, E., Vafamand, N. and Goodarzi, M. "A self-tuning load frequency control strategy for microgrids: Human brain emotional learning." ***International Journal of Electrical Power & Energy Systems***, Vol. 75, (2016), 311-319. DOI: 10.1016/j.ijepes.2015.08.026.
23. Esmaeili, M., Shayeghi, H., Mohammad Nejad, H. and Younesi, A. "Reinforcement learning based PID controller design for LFC in a microgrid." ***COMPEL-The International Journal for Computation and Mathematics in Electrical and Electronic Engineering***, Vol. 36, No. 4, (2017), 1287-1297. DOI: 10.1108/COMPEL-09-2016-0408.
24. Ghafouri, A., Milimonfared, J. and Gharehpetian, G. B. "Fuzzy-Adaptive Frequency Control of Power System Including Microgrids, Wind Farms, and Conventional Power Plants." ***IEEE Systems Journal***, Vol. 12, No. 3, (2017), 2772-2781. DOI: 10.1109/JSYST.2017.2715819.
25. Kayalvizhi, S. and Kumar, D. V.: Load Frequency Control of an Isolated Micro Grid Using Fuzzy Adaptive Model Predictive Control. IEEE Access. 5, 16241-16251 (2017). DOI: 10.1109/ACCESS.2017.2735545.
26. Khooban, M. H., Dragicevic, T., Blaabjerg, F. and Delimar, M. "Shipboard Microgrids: A Novel Approach to Load Frequency Control." ***IEEE Transactions on Sustainable Energy***. Vol. 9, No. 2, (2018), 843-852. DOI: 10.1109/TSTE.2017.2763605.
27. Khooban, M. H., Niknam, T., Shasadeghi, M., Dragicevic, T. and Blaabjerg, F. "Load Frequency Control in Microgrids Based on a Stochastic Non-Integer Controller." ***IEEE Transactions on Sustainable Energy***. Vol. 9, No. 2, (2018), 853-861. DOI: 10.1109/TSTE.2017.2763607.
28. Kassem, A. M. "Modeling and control design of a stand-alone wind energy conversion system based on functional model predictive control." ***Energy Systems***. Vol. 3, No. 3, (2012), 303-323. DOI: 10.1007/s12667-012-0051-3.
29. Burkart, R., Margellos, K. and Lygeros, J. "Nonlinear control of wind turbines: An approach based on switched linear systems and feedback linearization." In Decision and Control and European Control Conference (CDC-ECC), 2011 50th IEEE Conference on, 5485-5490 (2011). DOI: 10.1109/CDC.2011.6160709.
30. Asgari, Sh, and Yazdizadeh A. "Robust model-based fault diagnosis of mechanical drive train in V47/660 kW wind turbine." ***Energy Systems***, Vol. 9, No. 4, (2018), 921-952. DOI: 10.1007/s12667-017-0231-2.
31. Asgari, Sh, Yazdizadeh, A. and Kazemi M. G. "Robust Model-Based Fault Detection and Isolation for V47/660kW Wind Turbine." ***AUT Journal of Modeling and Simulation***. Vol. 45, No. 1, (2015), 55-66. DOI: 10.22060/MISCJ.2015.489.
32. Alonge, F., D'Ippolito, F., Cangemi, T., Magazzu, A. and Maniscalchi, M. "A model-based control strategy for wind turbines with asynchronous generator." In Clean Electrical Power, ICCEP'07. International Conference on, 506-513 (2007). DOI: 10.1109/ICCEP.2007.384262.
33. Ugalde-Loo, C. E., Ekanayake, J. B. and Jenkins, N. "State-space modeling of wind turbine generators for power system studies." ***IEEE Transactions on Industry Applications***, Vol. 49, No. 1, (2013), 223-232. DOI: 10.1109/TIA.2012.2228836.
34. Lopes, J. P., Moreira, C. L. and Madureira, A. G. "Defining control strategies for microgrids islanded operation." ***IEEE Transactions on Power Systems***. Vol. 21, No. 2, (2006), 916-924. DOI: 10.1109/TPWRS.2006.873018.
35. Shivaie, M., Kazemi, M. G., and Ameli, M. T. "A modified harmony search algorithm for solving load-frequency control of non-linear interconnected hydrothermal power systems." ***Sustainable Energy Technologies and Assessments***, Vol. 10, (2015), 53-62. DOI: 10.1016/j.seta.2015.02.001.

Persian Abstract

چکیده

در این مقاله به مساله کنترل بار – فرکانس ریزشبکه در مود جزیره ای با استفاده از کنترل کننده پیش سوی فازی پرداخته شده است. به این منظور علاوه بر کنترل کننده فیدبک معمول از یک کنترل کننده پیش سو در ساختار کنترلی ریزشبکه بهره برده شده است. سیگنال اغتشاش که می تواند تغییرات بار یا عدم قطعیت ناشی از منابع تولید پراکنده باشد توسط یک مشاهده گر اغتشاش تخمین زده می شود. ریزشبکه مورد مطالعه شامل توربین بادی، سلول خورشیدی، دیزل ژنراتور و بار است. همچنین یک کنترل کننده فازی برای کنترل زاویه پیچ توربین در نظر گرفته شده است که از تغییرات زیاد توان تولیدی توسط توربین جلوگیری کرده و کنترل فرکانس ریزشبکه را بهبود می دهد. نتایج شبیه سازی بخوبی نشان دهنده کارایی روش پیشنهادی برای کنترل بار – فرکانس ریزشبکه در مود جزیره ای است.

International Journal of Engineering

Journal Homepage: www.ije.ir

Operating Wheelchair using Flex Sensor

D. Pant*, N. Singh, P. Gupta

Department of Electronics and Communication Engineering,Indira Gandhi Delhi Technical University for Women, Delhi, India

P A P E R I N F O

Paper history:
Received 27 January 2021
Received in revised form 26 April 2021
Accepted 29 May 2021

Keywords:
Flex Sensor
Arduino
DC Motors
Wheelchair
Nodemcu
ThingSpeak

A B S T R A C T

Disabled people find their movements extremely difficult with their current gadgets. Although there are numerous devices and instruments accessible to empower their mobility, they require fine and precise control, which may be impractical in instances of higher disability. The haptic innovations are very beneficial, but demands a user-friendly environment too. The number of strategies, designs and frameworks in this field are more in number, which depends on Joystick. But this can prove to be uneconomical. Therefore, the aim of this paper is to control a wheelchair through a utility that is budget-sensitive and practical to use. A flex sensor-based wheelchair is presented in this paper, which is user friendly within a low budget. The proposed system is also accompanied by a movement monitoring system that can help the user's kin to observe his/her movement. The aim is to provide such features which is affordable to the society and can be used with ease.

***doi**: 10.5829/ije.2021.34.07a.14*

NOMENCLATURE

R_{flex}	Flex Resistance	V_{In}	Input Voltage
V_{CC}	Voltage common collector	R_{const}	Constant resistance
V_0	Output Voltage		

1. INTRODUCTION

The wheelchair is one of the most utilized equipment to improve movement and enhance personal satisfaction for individuals, who experience issues in walking. This may be due to spinal line wounds bringing about paraplegia or quadriplegia, cerebrovascular accident, or a stroke, leading to impaired movement. Thus, people may suffer from various kinds of disabilities and this percentage cannot be neglected. A wheelchair provides comfort, improves mobility and provides a solution for an impaired person to partake in social activities, access organizations, human administrations and reliance on himself improving self-confidence. A properly assisted wheelchair benefits the physical well-being and personal satisfaction of the clients. Regardless of quick, innovative, and logical advancement in gadgets for people with disabilities, there has been practically little headway in the wheelchair plan in the course of the most recent years but still, efforts are being made [1]. In 1932, Harry Jennings, an engineer, fabricated the principal collapsing, cylindrical steel wheelchair. Then came the first motorized wheelchair, which was operated by turning the wheels of the seat physically. In the event where a patient could not do this, someone else would need to push the wheelchair and the patient from behind. A motorized or force/power wheelchair is one where a little engine drives the wheels to spin. Thereafter, the electric wheelchair was designed by the Canadian innovator, George Klein and his group of associates. These wheelchairs got progressively significant as assistive innovation and the number of clients soared. For instance, improving the suspension performances of an electric wheelchair [2]. Along these lines, a programmed wheelchair [3] was created to take care of the issue for route and safe development in the ideal course. Diverse information techniques can be utilized to perform this task. This human-machine interface includes eye-gaze following technique, voice recognition, joystick-controlled, and so on. But utilization of these techniques can prove to be a tough task for people with a physical disability as they may find it tough to operate it around.

*Corresponding author email: divya27pant@gmail.com (D. Pant)

Please cite this article as: D. Pant, N. Singh, P. Gupta, Operating Wheelchair using Flex Sensor, International Journal of Engineering, Transactions A: Basics Vol. 34, No. 7, (2021) 1703-1709

Some shortcomings for instance, in an eye following wheelchair [4], the user is not allowed to envision the adjoining environmental factors when the framework is dynamic, and a voice acknowledgment-based wheelchair cannot work appropriately in background noise conditions. Furthermore, a traditional joystick controller requires much strength which is more than the strength of seriously crippled individuals. Even if this situation is ignored, it is as yet a very troublesome assignment for physically incapacitated individuals to get authority over the proper utilization of a joystick. This implies that although these methods are available to the disabled, however, satisfaction and command control of the wheelchair is impaired in certain circumstances. Motion capture Systems using flex sensors in robotics has also helped in human locomotion [5]. This paper throws light on the usage of a Flex sensor [6] in a wheelchair to make the functional capacity as near normal as possible. Here flex sensors will be the spine of the wheelchair. A flex sensor or bending sensor can be put to use for navigating the wheelchair in numerous directions [7]. The proposed wheelchair objective is to provide the largest possible satisfaction to the individuals who are crippled/amputees. This motion-controlled wheelchair is partitioned into two segments, (i) Transmitter: Flex sensor and (ii) Receiver: Wheelchair. Furthermore, the data now gathered from this setup can be used to keep a tab on the position and the motion of the wheelchair. This is accomplished by uploading the generated information/data to ThingSpeak via NodeMCU. Section 1 explains the introduction and background. Section 2 emphasizes the technical stack of the proposed model. Section 3 exhibits the execution (equipment and programming) of the venture and section 4 explains the planning of our proposed model. Section 6 concludes the paper.

2. TECHNICAL STACK

The proposed mechanism will consist of a flex sensor, a Microcontroller and DC Motors. A flex sensor controls the motors, which will in turn control the wheels of the wheelchair. The aim is to blend hardware as well as a software interface to achieve this. ThingSpeak acts as a cloud to store the transmitted data via Hypertext Transfer Protocol, with the help of RESTful API. For constructing a monitoring system of the wheelchair, the data received from the flex sensor is used to analyze the movement of the wheelchair remotely. To access the data, ThingSpeak is chosen which eases the process of accessing the data remotely.

2. 1. Flex Sensor A Flex sensor is a thin, lightweight, and flexible strip where the resistance is proportional to the amount it is bent. It can operate from 0° to 180°. A flex sensor produces an analog signal. The carbon surface is engineered on a strip with the working voltage varying from 0 V to 5V and working temperature ranges from -45ºC to +80ºC. The scope of curve opposition is from 45 K to 125 K with a resilience obstruction of ±30%. The flex sensor works whenever the plastic strip is bent. Table 1 shows the value of the flex sensor at 0° and 90°. The resistance increases/escalates as the strip is bend from 0° to 90°.

This adjustment in obstruction relies on surface linearity, which implies that the resistance of this sensor would be diverse at different points. A Proper Filter can significantly improve the smoothness of fingers which will ease the movement of the wheelchair [8].

2. 2. Arduino Arduino is an open-source platform that can be utilized to develop certain objects by interacting with an assortment of physical entities. Arduino is chosen as a microcontroller for the proposed mechanism because it is highly compatible with a large no of sensors, has 13 GPIO pins and 5 analog pins making it expandable with many sensor modules.

2. 3. Nodemcu NodeMCU is chosen for the proposed mechanism as a result of its modest cost and the degree of network support. It can also be modified using Arduino IDE so it can run the libraries utilized by Arduino [9]. This tool is particularly helpful for IoT applications, because of its compact nature and inherent WiFi support. The connections of a flex sensor to a NodeMcu are akin to that of Arduino and flex sensor.

2. 4. L293D Motor Driver IC L293D is a (motor driver) IC that lets the DC motor drive on both routes. This 16-pin IC can manage a hard and fast of two DC motors concurrently in any course. 36V can be the maximum voltage supply. L293D consists of 4 inputs; 2,7 for the left motor and pin 15,10 for the right one. The benefit of utilizing the motor driver is that it can turn around the polarity of applied voltage without readjusting the circuit and can provide us current in both directions up to 600mA.

3. PROPOSED IMPLEMENTATION

3. 1. Control System To command the wheelchair to move in different directions just by bending the fingers, a control system is proposed. Detailed implementation of this process is discussed in further sections.

TABLE 1. Values of resistance at 0°and 90°

Angle Values	Resistance (ohm)
0°	40910.20
90°	93697.67

3. 1. 1. Interfacing the Flex Sensor with Arduino

To interface Arduino with flex sensor a voltage divider circuit is implemented to analyze the change in resistance of the flex sensor generated by bending of fingers, which produces different values of angles. The Flex sensor is interfaced with Arduino as shown in the pin configuration of the Flex sensor in Figure 1. One of the pins is connected to Arduino's power supply pin, VCC. The next pin is connected to Arduino's GPIO pin i.e the analog pin. Since there are five analog pins, it can be connected to any one of them because it generates an output in analog. As the aim is to create a voltage divider circuit, a resistor measuring 33k ohms is connected at A0 with a flex sensor.

Analog values received from the flex sensor via pin A0 are the output voltages obtained but these values need to be converted into discrete volts and then the flex resistance will be obtained by the formula as expressed in Equation (1):

$$V_0 = V_{in} * \left(\frac{R_{const}}{R_{const}+R_{flex}}\right) \quad (1)$$

The above equation can also be rewritten as shown in Equation (2):

$$R_{flex} = \left(\frac{V_{in}}{V_0} - 1\right) * R_{const} \quad (2)$$

The obtained resistance is then fed to an inbuilt function called map(). This function helps in converting the obtained resistance value to the exact angle which actuates the motors further. Table 2 shows the values of angles for different values of resistance.

3. 1. 2. Interfacing Arduino with Motor Driver

As mentioned under the technology stack, the L293D IC can control 2 motors simultaneously. Thus, its objective is to actuate the motors according to the command sent by Arduino to the motor driver.

For the hardware implementation of the motor driver (L293D) with Arduino, the output pins for each motor present in the motor driver are connected to Arduino GPIO pins. As shown in Figure 2, the ground and VCC of the motor driver are connected to Arduino's ground and VCC pin, respectively [10] To actuate the motors, an effective power source is required as their expected power input is much more than the maximum power supplied by Arduino. Hence the motor driver will be connected to an external power supply as well.

For the software implementation of the motor driver in Arduino IDE, the connected GPIO pins are set in output mode to actuate the motors. Now the decision-making part is analyzed. First, three functions in Arduino IDE are formed that are the stop ( ), forward ( ) and backward( ). The whole actuation process depends on the values of the angle generated by the map ( ) function. For the decision-making part, the flow of the program is controlled by using the if-else statement. For instance, if the angle is less than -10°, the backward () function is called and the motors start moving in the reverse direction hence making the wheelchair move backwards.

TABLE 2. Values of resistance and angle

Resistance (ohm)	Angle
74389	57°
88128	81°
107739.7	114°
108342.9	115°
150011.2	185°
134178.5	158°
44463.06	7°
40719.46	1°
40397.46	0°
36744.19	-5°
32286.95	-13°
29393.96	-17°

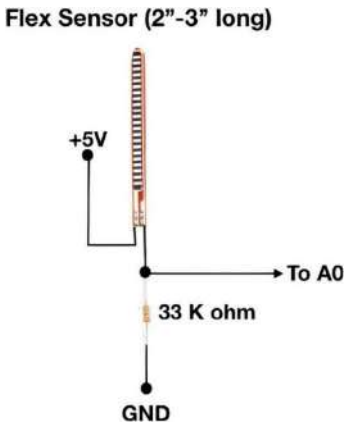


Figure 1. Pinout configuration of flex sensor

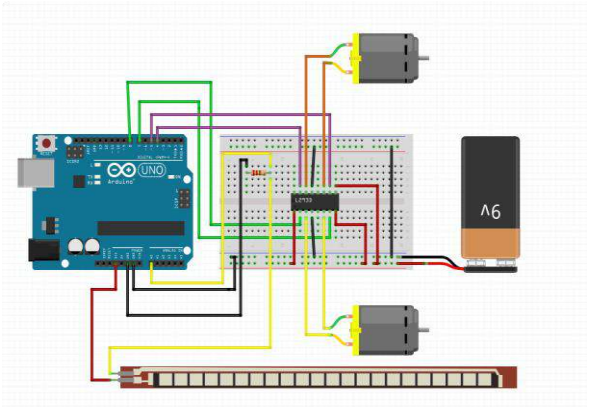


Figure 2. L293D Motor Driver IC with DC Motors and Arduino

3. 1. 3 Interfacing Motors with Motor Driver To drive the motors according to the bending of the finger, the motor driver will receive a request from Arduino to perform a particular function i.e. forward, backward, stop, etcetera. As L293D is an H-bridge, so motors can rotate clockwise and anti-clockwise according to the discrete signal sent by Arduino to the input pin of L293D. After the input is received, L293D will generate output according to the input signals. For instance, consider the case of a stop function. When both values are sent high or low, then the circuit remains open which stops the motors. Similarly, outputs of different functions can be generated.

3. 1. 4. Compilation of Control System of Wheelchair All the implementations covered above are compiled together in this section. Figure 3 is a flowchart depicting the flow of the process. The flex sensor is connected to the Arduino, along with the voltage divider circuit. Then the motor driver, L293D, is connected to the Arduino with a nine-volt battery. After these connections are established, a pair of motors are connected to the motor driver. This completes the hardware compilation and it is ready for the software simulation. The code written in Arduino IDE is uploaded and is executed by the sequence of commands written in embedded C. The compilation starts with the setup in Arduino IDE. All the GPIO pins are set to INPUT/OUTPUT mode. The value of the constant resistor (33k ohms) and the resistance of the flex sensor at 0° and 90° are set which is measured by the multimeter. Finally, the board rate is set at 9600 to print the data on Serial Monitor for the Arduino board. After execution of the loop, the first value generated from the flex sensor is retrieved which is then converted to a discrete value as flex voltage. The resistance of the flex sensor when the finger is bent is calculated by Equation (2). This value is then sent to the map function which will generate the angle created by bending the flex sensor. After the angle is generated, it is sent to perform a logical operation, which compares the generated value of the angle with predefined ranges, to execute the particular function. If a proper match is found, then that function is executed, which will send the request to the motor driver to drive the motors. Now the compiler will again return to the start of the loop and repeat the procedure i.e a value is received and then it will execute another function. By this procedure, the aim to control the wheelchair by bending the fingers is achieved successfully.

3. 2. Movement Monitoring System To monitor the movement of the person sitting in the wheelchair, a monitoring system is framed. The concept of the Internet of Things is implemented to achieve this. The monitoring system aims to send real-time data to the guardian of the person sitting in the wheelchair. This data will signify whether the person is moving or not. It will reveal if it is moving forward, backward, left or right. So, with the help of this data, the guardian will be able to keep track of the person.

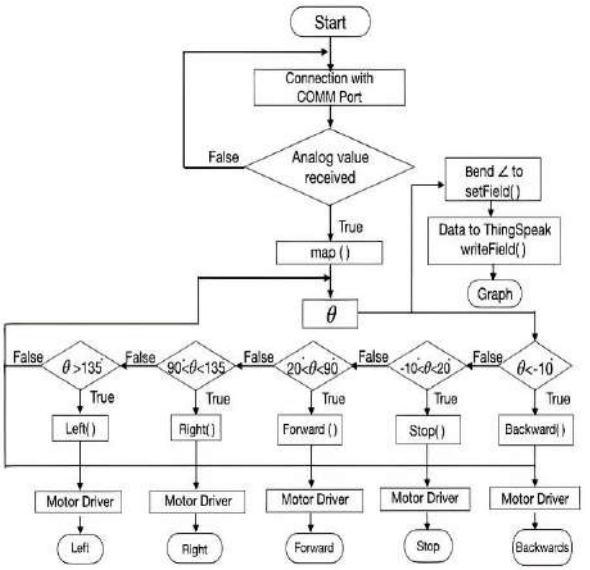


Figure 3. Flowchart

3. 2. 1. Implementation of Monitoring System To implement the movement monitoring system, the NodeMCU board is interfaced with the flex sensor. The NodeMCU receives data from the flex sensor and sends the data to the ThingSpeak server. Once the data is posted, a graph is plotted. With the help of the graph, the guardian can track the movement of the patient in real time. Detailed implementation of this process is discussed in further sections.

3. 2. 2. Hardware Implementation of NodeMCU with Flex Sensor The power pin of the flex sensor will be connected to NodeMCU's power supply pin, VCC (3.3 volts). The other pin will be connected to NodeMCU's GPIO pin i.e the analog pin. It is connected to A0 since it has only one analog pin. This connection is made because the flex sensor generates output in analog. Both the Arduino and the NodeMCU will be connected with the same voltage divider circuit as mentioned before.

3. 2. 3. Software Implementation of NodeMCU with Flex Sensor The data generated from the flex sensor is required to be sent to the cloud with the help of the map () function to reveal the movement of the wheelchair. For instance, a wheelchair moves backward when the angle is less than -10°.This data from the flex sensor is received by NodeMCU. The aim is to post the data to the ThingSpeak website. This is achieved by using

the RESTfull API and HTTP. To write data to ThingSpeak, the write API Key is included in the code. As NodeMCU acts as a client and ThingSpeak acts as a server, the client will send a request to write the data by using the API Key. Flex sensor data needs to be sent to the selected field. The channel id needs to be specified in the code as well so that the data is sent to the right channel id created. To accomplish all this, the wifi module is connected to the internet network to send a request to access the ThingSpeak website and post data into it. This is performed by the function "WiFiClient". Then the compiler will check the status of the wifi whether it is connected or not and will repeat till the ESP8266 wifi module gets connected to the network. Furthermore, the ThingSpeak.begin( ) function will help in initializing ThingSpeak libraries. The main function of posting the data starts when setField( ) is called, where field specified is set as a parameter of this function along with data to be posted. Finally, the writeField( ) function along with the channel number and the write API Key calls the setField function to post the data to ThingSpeak.

3. 2. 4. Final Compilation of Monitoring System

After Node MCU is connected with the Flex sensor and the power supply, the data of the Flex sensor is ready to be posted to the ThingSpeak cloud. The graph obtained will help in monitoring the movement of the wheelchair. When the data is around the x-axis, it will signify that the wheelchair is not moving. When the data is approximately 10 units below the x-axis, this will imply that the wheelchair is moving in a backward direction. If the data is around 20 units above the x-axis, this will mean that the wheelchair is moving in a forward direction and whenever the data is about 90 units above the x-axis, this will signify that the wheelchair is moving in the right direction. Finally, if the data is around 135 units above the x-axis, this will denote that the wheelchair is moving in the left direction. The data is posted in the intervals of approximately 17 seconds which is revealed in Table 3 and simultaneously the graph is plotted as shown in Figure 4.

TABLE 3. Angles uploaded to thingspeak

Time (s)	Angle Values
0	-17°
17	62°
34	-19°
51	62°
68	92°
85	78°
102	74°

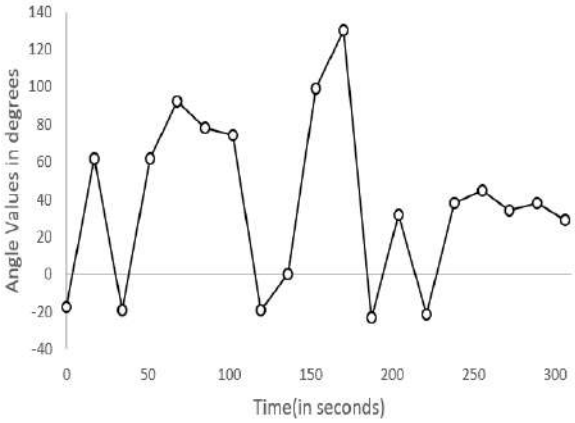


Figure 4. Graph for Thingspeak

4. DESIGNING OF THE WHEELCHAIR

To demonstrate the working of the wheelchair a prototype is assembled as shown in Figure 5. The prototype comprises of a flat chassis with the DC motors attached to it. Furthermore, wheels are appended to the motors. For the assembly of the prototype, a breadboard, Arduino, and NodeMCU are placed at the top of the chassis but for the actual wheelchair, all the components will be present inside a box named processing module and that will be placed below it for a steady flow. Only flex sensor is present at the grip of the user who is sitting on the wheelchair. The systems that are controlling and monitoring the framework will simultaneously update the data. Figure 6 shows the actual block diagram for the same.

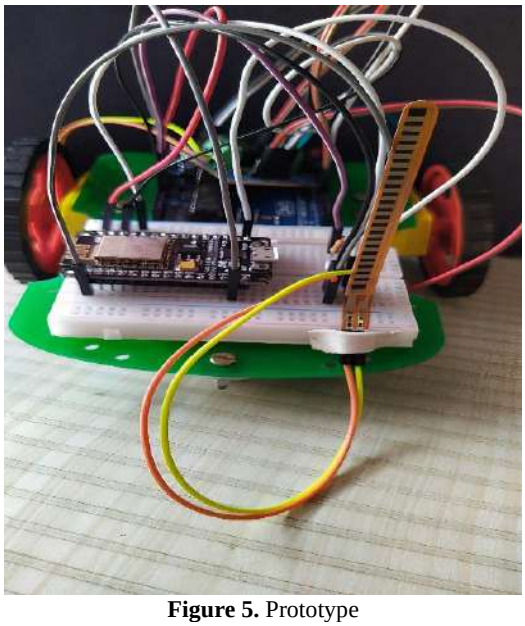


Figure 5. Prototype

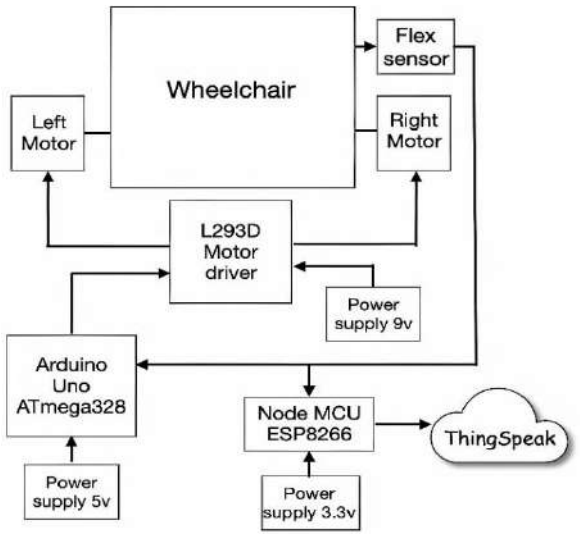


Figure 6. Block Diagram for Wheelchair Using Flex Sensor

5. COMPARISON WITH EXISTING MODELS

The point of our exploration is to structure and improvise the existing scenario for individuals with higher in-adaptability by the means of the flex sensor. The existing wheelchairs are mostly based on a joystick that controls the direction as per the user's wishes. But this framework is susceptible to internal faults and hence requires fault diagnosis and fault-tolerant control [11]. Other works include a camera-based and voice-based controlled wheelchair. The restriction with camera-based structure is that they require the client to keep up the head in a fixed position and a voice-based framework does not fare well in noisy conditions [12].

This proposed framework intends to improve the lives of physically disabled people at an affordable range. The most critical piece of this wheelchair is that it can be easily controlled by the movement of the user/patient fingers. This won't just assist the patient, but will also benefit the specialist and their family members who are in contact with the patient through the monitoring system. Hence, this system will provide better execution than the previously mentioned ongoing applications. Moreover, the direction can be very easily controlled with lesser strength as compared to existing wheelchair systems.

6. CONCLUSION AND FUTURE SCOPES

In this paper, a method has been proposed to interface the flex sensor and the processing module with the existing wheelchair. The control of the wheelchair has been achieved by bending the flex sensor and actuating the motors. Furthermore, a monitoring system has been successfully implemented by sending the data generated from the flex sensor to the cloud which would help the guardian to track the movement of the person sitting in the wheelchair. The future scopes may include implementation of a health monitoring system that can provide numerous provisions such as blood pressure, temperature, heart rate monitoring and many more. To extend the movement monitoring system, GPS can be installed to give real time location of the user.

7. REFERENCES

1. Leaman, J. and La, H. M., "A comprehensive review of smart wheelchairs: past, present, and future." ***IEEE Transactions on Human-Machine Systems***, Vol. 47, No. 4, (2017), 486-499.DOI: 10.1109/THMS.2017.2706727
2. Saghafi, A., Bagheri Hosseinabadi M. and Khomarian, N., "Control of Electric Wheelchair Suspension System based on Biodynamic Response of Seated Human Body." ***International Journal of Engineering, Transactions B: Applications,*** Vol. 33, No. 8, (2020), 1629-1636. DOI: 10.5829/ije.2020.33.08b.21
3. Simpson, R., LoPresti, E., Hayashi, S., Nourbakhsh, I. and Miller, D., "The smart wheelchair component system." ***Journal of Rehabilitation Research & Development***, Vol. 41, No. 3B, (2004), 429- 442. DOI: 10.1682/jrrd.2003.03.0032
4. Shahzad, K. and Bilal Khan, M., "Control of a robotic wheel-chair prototype for people with walking disabilities." ***International Journal of Engineering, Transactions B: Applications,*** Vol. 31, No. 5, (2018), 693-698. DOI: 10.5829/ije.2018.31.05b.02
5. Pastor, S.S., Rivera, C.T., Avilés, O.F. and Mauledoux, M.F., "A Real-time Motion Tracking Wireless System for Upper Limb Exosuit Based on Inertial Measurement Units and Flex Sensors." ***International Journal of Engineering, Transactions C: Aspects,*** Vol. 32, No. 6, (2019), 820-827. DOI: 10.5829/ije.2019.32.06c.04
6. Saggio, G., Riillo, F., Sbernini, L. and Quitadamo, L.R., "Resistive flex sensors: a survey." ***Smart Materials and Structures,*** Vol. 25, No. 1, (2015): 013001. DOI: 10.1088/0964-1726/25/1/013001
7. Wallam, F. and Asif, M.," Dynamic finger movement tracking and voice commands based smart wheelchair." ***International Journal of Computer and Electrical Engineering,*** Vol. 3, No. 4, (2011), 497-502. DOI: 10.7763/ijcee.2011.v3.368
8. Eu, K.S., Yong, S.L., Yip, M.W., Lee, Y.K., Ko, Y.H. and Yap, K.M., 2014. "Fingers Bending Motion Controlled Electrical Wheelchair by using Flexible Bending Sensors with Kalman filter Algorithm." In 3rd International Conference on Convergence and its Application, Vol. 7, No. 13, (2014), 637-647. DOI: 10.12988/ces.2014.4670
9. Amri, Y. and Setiawan, M.A., "Improving Smart Home Concept with the Internet of Things Concept Using RaspberryPi and NodeMCU." In IOP Conference Series: Materials Science and Engineering, Vol. 325, No. 1, (2018), 012021. DOI: 10.1088/1757-899X/325/1/012021
10. Shetti, P.R. and Mangave, A.G., "DC motor speed control with feedback monitor based on C# application." ***International Journal of Research in Engineering and Technology,*** Vol. 3, No. 3, (2014) 398-401. DOI: 10.15623/ijret.2014.0303073
11. Hashimoto, M., Nakamura, Y., Oba, F. and Takahashi, K., "Fault Diagnosis and Fault-Tolerant Control of a Joystick Controlled Wheelchair." ***IFAC Proceedings Volumes,*** Vol. 39, No. 16, (2006), 211-216. DOI: 10.3182/20060912-3-de-2911.00039
12. Nishimori, M., Saitoh, T. and Konishi, R., "Voice controlled intelligent wheelchair." In SICE Annual Conference 2007, IEEE, (2007), 336-340. DOI: 10.1109/sice.2007.4421003

Persian Abstract

چکیده

افراد معلول با وسایل فعلی خود حرکات خود را بسیار دشوار می دانند. اگرچه دستگاه ها و ابزارهای بی شماری برای توانمندسازی تحرک آنها در دسترس است ، اما آنها به کنترل دقیق نیاز دارند ، که ممکن است در موارد معلولیت بالاتر غیر عملی باشد. نوآوری های لپ تاپ بسیار سودمند است ، اما یک محیط کاربر پسند را نیز می طلبد. تعداد استراتژی ها ، طرح ها و چارچوب ها در این زمینه تعداد بیشتری دارد که به جوی استیک بستگی دارد. اما این می تواند غیر اقتصادی باشد. بنابراین ، هدف این مقاله کنترل ویلچر از طریق ابزاری است که از نظر بودجه حساس و کاربردی باشد. صندلی چرخدار مبتنی بر سنسور فلکس در این مقاله ارائه شده است که با بودجه کم کاربر پسند است. سیستم پیشنهادی همچنین با یک سیستم نظارت بر حرکت همراه است که می تواند به خویشاوندان کاربر در مشاهده حرکت او کمک کند. هدف تهیه چنین ویژگی هایی است که برای جامعه مقرون به صرفه باشد و بتوان از آن به راحتی استفاده کرد.

IJE TRANSACTIONS A: Basics Vol. 34, No. 7, (July 2021) 1710-1717

International Journal of Engineering

Journal Homepage: www.ije.ir

Using Universal Nand-nor-inverter Gate to Design D-latch and D Flip-flop in Quantum-dot Cellular Automata Nanotechnology

H. Alamdar[a], G. Ardeshir[a], M. Gholami*[b]

[a] *Faculty of Electrical and Computer Engineering, Babol Noshirvani University of Technology, Babol, Iran*
[b] *Department of Electrical Engineering, Faculty of Engineering and Technology, University of Mazandaran, Babolsar, Iran*

PAPER INFO

Paper history:
Received 19 Feruary 2020
Received in revised form 06 May 2021
Accepted 26 May 2021

Keywords:
Quantum-dot Cellular Automata
Flip-flop
Latch
Phase-frequency Detector

ABSTRACT

The process of reducing dimensions in CMOS technology and also making digital devices more portable, faces serious challenges such as increasing frequency and reducing power consumption. For this reason, scientists are looking for a solution such as replacing CMOS technology with other technologies including Quantum-dot Cellular Automata (QCA) technology and many researches have designed digital circuits by using QCA technology. Flip-flops are one of the main blocks in most digital circuits. In this paper, a D-type flip-flop (D-FF) is presented in QCA technology that a majority gate has been used in its feedback path to reset. The D-FF is designed by the proposed D Latch which is based on Nand-Nor-Inverter (NNI) and a new inverter gate that the proposed D latch has 24 cells and 0.5 clock cycle latency and 0.02 μm^2 area. The new inverter gate of the D-FF has output signal with high polarization level and lower area than previous inverters and the NNI gate of the D-FF is a universal gate. One of the applications of D-FFs with reset pin is the use in Phase-frequency detector (PFD). In the proposed scheme, a reset feature has been added to D-FF since the PFD structure can be designed. All of the proposed schemes are evaluated by the QCADesigner software and energy consumption simulations are estimated using QCAPro software for all proposed circuits.

***doi**: 10.5829/ije.2021.34.07a.15*

1. INTRODUCTION

CMOS technology is faced with serious challenges such as short-channel effects, cut-off leakage, and high power consumption. Hence, the researchers have proposed alternatives to replace CMOS technology. One of these technologies is QCA technology [1]. The circuits designed in QCA technology use a square cell with size of a molecule or an atom. These QCA circuits can operate at around the terahertz frequencies because of not containing any output capacitors. Since there is no current in QCA circuits, their power consumption is less than CMOS circuits [2, 3].

Today, many digital systems include storage elements, sequential logic circuits and combinational logic circuits. The storage elements have capable of storing binary information. Fip-flop as a storage element has ability of storing one bit. The flip-flops are based on latches which operate with signal levels [4]. Flip-flops are used in phase/frequency detector (PFD) of a Phase-Locked Loop (PLL) and a Delay-Locked Loop (DLL). For an example, the convensional PFD is designed by using D flip-flops [5-8].

Different D latches and D-FFs have been designed in QCA technology [9-16]. In QCA technology, digital circuits such as latches are compared in terms of area, latency and number of cells. Some of these circuits are suitable and capable of using in large circuits, but some other are not. By latency we mean number of clocks needed for generating the valid output signal.

Dehkordi, and Sadeghi [9], have designed D latch with 43 cells and a 1.25 clock cycle latency. This D latch design has a NOT gate with larger number of cells and area than the convetional NOT gate. Also, this D latch has high latency compared to current work. Hashemi and Navi [10] have proposed the D latch which has large number of cells, occupying massive area and high latency compared to current works. Also, a NOT gate uses as in

*Corresponding Author Email: m.gholami@umz.ac.ir (Mohammad Gholami)

Please cite this article as: H. Alamdar, G. Ardeshir, M. Gholami, Using Universal Nand-nor-inverter Gate to Design D-latch and D Flip-flop in Quantum-dot Cellular Automata Nanotechnology, International Journal of Engineering, Transactions A: Basics Vol. 34, No. 7, (2021) 1710-1717

the previous work [9]. Abutaleb [11], has proposed a design for D latch with 28 cells and 0.02 μm^2 area and 0.5 clock cycle latency, but it does not follow the basic design rules of QCA well. Sasamal et al. [12] have designed D latch with rotated majority gate, however, this design is less stable than other designs existing in literature. A D latch is designed with a 2-input multiplexer gate in literature [13] that reduces cell number and stability compared to the previous design but increases the latency. Zoka and Gholami [14] have designed D-FF with reset pin that has a large number of cells and high latency, and also in order to become resetting, the input associated with resetting must be constantly active leading to high energy dissipation. D-FF with reset terminal in literature [15], has reduced latency while number of cells and area are large, however, the D-FF uses coplanar crossovers. Binaei and Gholami [16] have designed D-FF with reset pin has only reduced number of cells when comparing to our work.

In this work, first, a universal gate, called NNI gate, is used in D-latch and D flip-flop. Second, a method for creating D flip-flop with reset pin is introduced in which the D flip-flop resets by interrupting the feedback path. Third, a new PFD is designed in QCA Technology. The NNI gates used in the D latch hass low latency. The D flip-flop designed based on the D-latch is sensitive with rising edges of the clock.

The next section describes the QCA structure. In section 3, the proposed structures for D latch, D-FFs and PFD will be explained. Section 4 is about simulation and the results.

2. THE BASIC PRINCIPLES OF QCA

In QCA technology, logic circuits are designed based on a square cell. This cell has 4 quantum holes or dots in its four corners where two of the dots are charged with two electrons. The electrons have moving ability among these dots by tunneling between them. The logical state is determined by the arrangement of electrons of within this cell. Due to repulsive force of the same charges, these electrons have two stable situations that the two stable situations are considered as ZERO and ONE logical in digital circuits. There are two forms of arrangement for these cells, 45° and 90° [17, 18].

According to coulomb repulsion force law, the electrons within dots of adjacent cells act on each other. So, in this way the data is transferred from one cell to its adjacent cell. Thus a wire can be implemented by putting a row of cells together. Because of the arrangement of cells in the two forms of 90° and 45°, there are two models of wire alignment [13]. To generate different signals in wire crossings, there are two crossover options in QCA technology: coplanar crossings [19] and multilayer crossover [20].

Two basic gates known in QCA technology are majority gate and inverter. By using the colony force and mutual interaction between cells, these basic gates can be made in QCA technology.

To have a synchronous circuits in QCA technology, in addition to control data flow in the wire to the same as that in a pipeline, a clock zone method is used. To operate correctly, Functional QCA circuits need clock [13, 21]. The cells also operate on clock. Clocking has two roles in QCA. It controls data flow and serves as power supply [3]. Although mapping conventional logic functions to majority logic can be done easily, the clock scheme in QCA makes directly conversion of a CMOS architecture into QCA counterpart difficult. The main unit of such designs is majority gate while this gate isn't a universal gate and cannot realize the logical NOT operation. The designers must consider a separate costly QCA cell arrangements for realization of the logical NOT. By combining these functions with inverters, these logical components are used to implement other logical functions. Therefore, a universal gate structure called the Nand-Nor-Inverter (NNI) has been introduced by Sen and Sikdar [22] which can be used as a logical element for QCA-based designs. This gate implements the function of $F = NNI(A, B, C) = maj(A', B, C') = A'B + BC' + C'A'$.

3. THE PROPOSED STRUCTURE

In this section, a D latch and a D-FF are proposed by using the NNI gate [22]. Then, a PFD is designed using the proposed D-FF along with adding a reset pin to demonstrate applicability and usability in larger circuits.

The block diagram of proposed D latch and the design of its structure in QCA technology are shown in Figures 1 and 2, respectively. The proposed D latch has 24 cells and 0.5 clock cycle latency and $0.02\ \mu m^2$ area. As shown in Figure 1, a 2:1 multiplexer is designed by using basic gates of QCA technology, and it is converted to D latch by applying feedback path which is shown with dashed line in Figure 1. This proposed D latch is designed with NNI gate [22] and the new inverter gate introduced by Zahmatkesh et al. [23].

Then, the proposed D latch is converted to a D-FF by adding a level-to-edge converter on the clock input. The D-FF is shown in Figure 3. The level-to-edge converter is a logic circuit that converts level to falling, rising or dual edge. The level-to-rising edge converter used in the proposed D-FF is shown by dashed rectangle in Figure 3. The NNI gate is used in this converter to reduce number of cell, area, and also the complexity of the circuit. Function of rising edge converter is $CLK_{new}(t) = CLK(t).\overline{CLK(t-1)}$ [13]. The proposed D-FF of Figure 3 has 43 cells and 1 clock cycle latency and $0.04\ \mu m^2$ area.

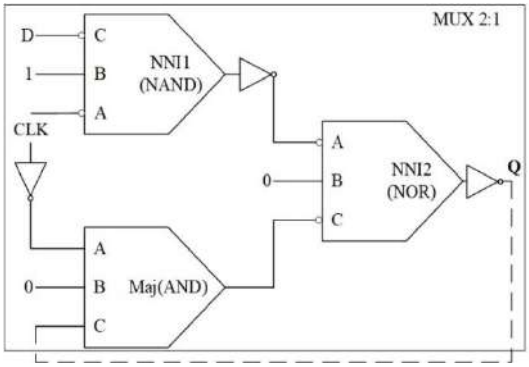


Figure 1. block diagram of proposed D latch

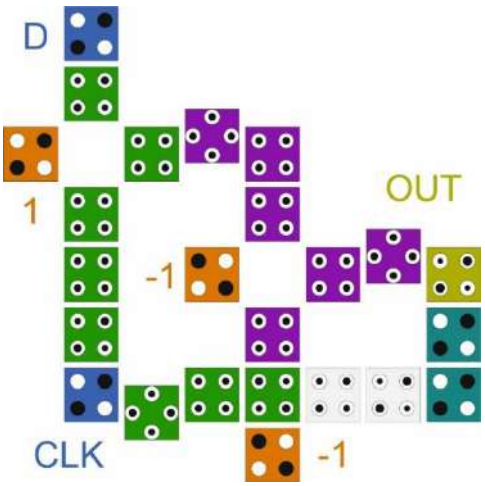


Figure 2. proposed D latch in QCA technology

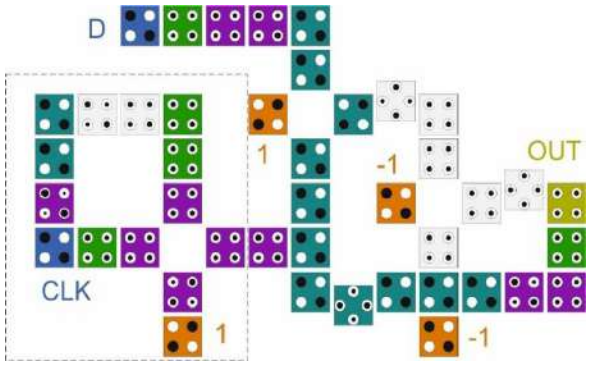


Figure 3. Proposed D-FF in QCA technology

As shown in Figure 4, the proposed D-FF (Figure 3) can be easily converted to a D-FF with reset pin by inserting a 3-input majority gate (Maj2) and changing this majority gate to 2-input AND gate. To reset the D-FF, a logical ONE is applied to one of the inputs of this AND gate (Maj2) while reset pin is set to logic ZERO (because an inverter is used on reset path). Due to this work, applying any value to another inputs of Maj2, same value is resulted in output of Maj2. If reset pin is set to logical ONE, logical ZERO is applied to one of the Maj2 and output of Maj2 becomes logical ZERO and this shows how resetting of the proposed D-FF is reset with the reset pin. The block diagram for the proposed D-FF with reset pin and its circuit in QCA technology are shown in Figures 4 and 5, respectively.

The proposed D-FF with reset pin consists of 54 cells, 1.25 clock cycle latency and 0.047 μm^2 area. Finally, the proposed PFD structure can be designed by using two the proposed D-FFs with reset pin and an AND gate. If the PFD circuit is designed based on conventional block diagram [5], the outputs of PFD will have no correct values due to the pipeline feature of QCA circuits. Therefore, the PFD circuit in QCA technology is designed based on block diagram shown in Figure 6 in which a subtraction operation is performed at the end of the proposed PFD circuit illustrated in this figure by dashed rectangle. The proposed PFD in QCA technology is depicted in Figure 7. As it can be seen in this figure, in the proposed PFD, the outputs of two D-FFs with reset pin are connected to the AND gate. Theoutput of the AND gate is connected to the reset pins of the D-FFs. As both D-FFs become high output simultaneously, the reset operation is completed by the AND gate, this implies that the PFD has calculated phase and frequency difference. According to Figure 6, the subtraction operation is as follows. Output of D-FF1 with reverse of output of D-FF2 are connected to the inputs of AND gate (1). The

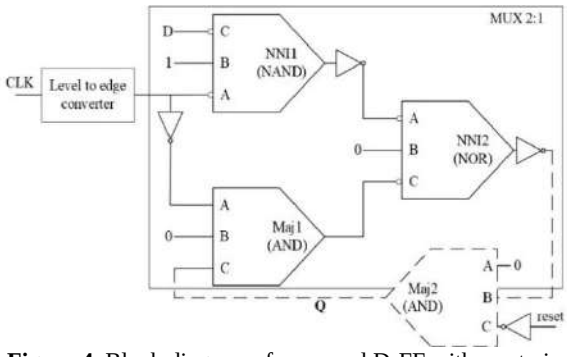


Figure 4. Block diagram of proposed D-FF with reset pin

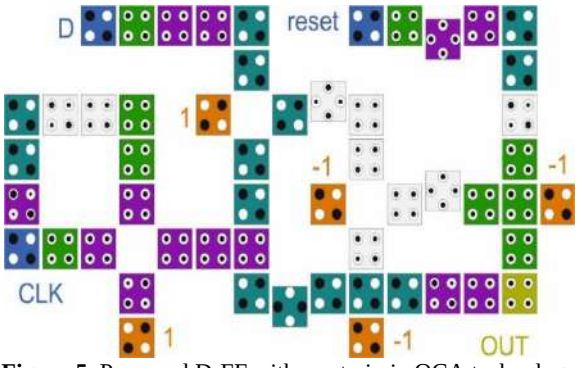


Figure 5. Proposed D-FF with reset pin in QCA technology

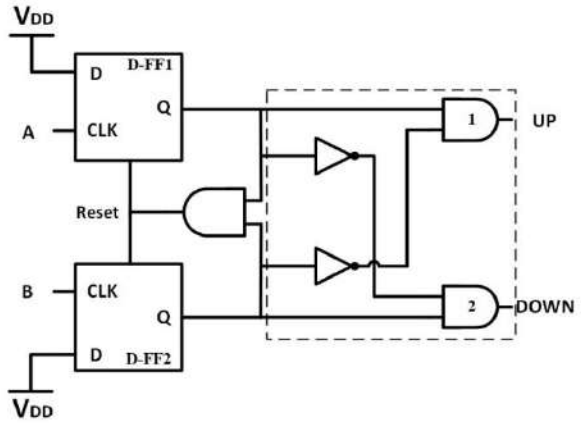


Figure 6. Block diagram of proposed PFD

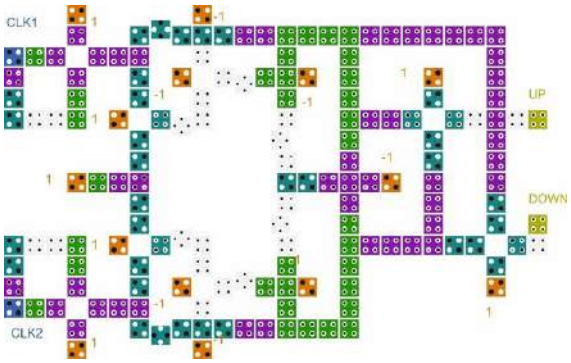


Figure 7. Proposed PFD circuit in QCA technology

output of AND gate (1) is correct UP output of the PFD. Then, the output of D-FF2 with reverse of output of D-FF1 are connected to the inputs of another AND gate (2), and the output of this AND gate (2) is DOWN output of the PFD. By doing this, latency does not affect outputs of the PFD. In this design the AND gates are designed based on NNI gate. This helps omitting inverter gate on the paths that where inputs are required to be reversed. So, this acting reduces the complexity and area of the circuit. The proposed PFD consists of 161 cells, 2 clock cycle latency and 0.175 μm^2 area.

4. SIMULATIONS AND RESULTS

In this section, the results of simulations and the energy dissipation for the proposed circuits are expressed. The simulations of the circuits' schemes has been done by QCADesigner software version 2.0.3 [24]. For better performance of QCA circuit analysis, all proposed QCA schemes were investigated using both coherence vector and bistable approximation simulation engines with default parameters.

First, the simulation results for the proposed D latch of Figure 2 can be seen in Figure 8 that the output depends on logical ONE level of the clock input. When the input of the clock becomes logical ONE level, the D input is sampled. This means that it responds to logical ONE level of the clock pulse. On the other hand, when the level becomes logical ZERO, the proposed D latch is stored the last value of its output. All of these actions are marked with arrows. The state of a latch or flip-flop is switched when a change in the control input occurs. This momentary change is called a *trigger*, and the transition caused by it is called flip-flop triggering. This means that clock input of the latch is sensitive to level, but clock input of the flip-flop is sensitive to edge and latch is converted to flip-flop by placing a level-to-edge converter at the clock input of latch. This flip-flop can be sensitive to falling, rising or dual edge. The simulation results for proposed D-FF with reset pin is illustrated in Figure 9. As shown in Figure 9 is marked with arrows, this D flip-flop is sensitive to rising edge. At any

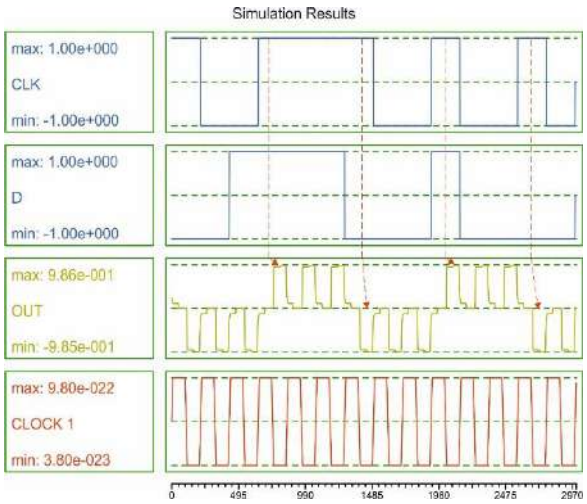


Figure 8. Simulation results of proposed D-latch

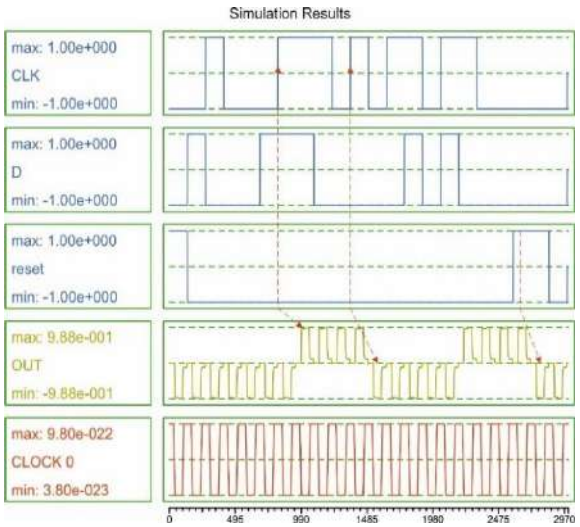


Figure 9. Simulation results of proposed D-FF with reset pin

moments that the clock input sense rising edges, value of D input is transferred to the output. When the reset input is set to logical ONE, output of D-FF with reset pin reaches logical ZERO. Tables 1 and 2 show the comparison of proposed D latch and D-FF with related designs. As it can be seen in these the proposed D-FF has the best delay performance while it contains few number of cells, and also energy dissipation of the proposed D-FF has improved.

Finally, the simulation results of the proposed PFD are shown in Figures 10, 11 and 12. The PFD can detect phase and frequency differences of its input signals. For this purpose, simulations of Figures 10 and 11 is performed, two signals are shown with the same frequency and different phases in Figure 10 and the proposed PFD detects phase difference of these two signals. Two signals with different frequencies are applied to inputs of the proposed PFD and the results are shown in Figure 11. In Figure 12, two different inputs are applied to inputs of proposed PFD, in order to test its detecting ability of phase differences for both inputs. Table 3 is listed comparison of the proposed PFD with other related works. As shown in Table 3, structure of the proposed PFD has improved in terms of cells, area, latency and energy dissipations.

TABLE 1. Comparison of the proposed D latch with other related works

Structure	Cells count	Area (μm^2)	Latency	Average switching energy dissipation (meV) $0.5E_k$	Average leakage energy dissipation (meV) $0.5E_k$
D latch in [9]	43	0.04	1.25	-	-
D latch in [10]	48	0.05	1	-	-
D latch in [11]	28	0.02	0.5	31.50	8.49
D latch in [12]	23	0.02	0.5	-	-
D latch in [13]	19	0.02	0.75	0.00467	0.03033
Proposed D latch in figure 2	24	0.02	0.5	0.01689	0.01041

TABLE 2. Comparison of the proposed D-FF with reset pin with other related works

Structure	Cells count	Area (μm^2)	Latency	Average switching energy dissipation (meV), $0.5E_k$	Average leakage energy dissipation (meV), $0.5E_k$
D-FF in [14]	82	0.11	2	-	-
D-FF in [15]	95	0.11	1	-	-
asynchronous D-FF in [16]	73	0.11	3	0.05643	0.02574
synchronous D-FF in [16]	73	0.1	2.5	0.04319	0.02656
Proposed D-FF in fig. 5	54	0.047	1.25	0.04414	0.02205

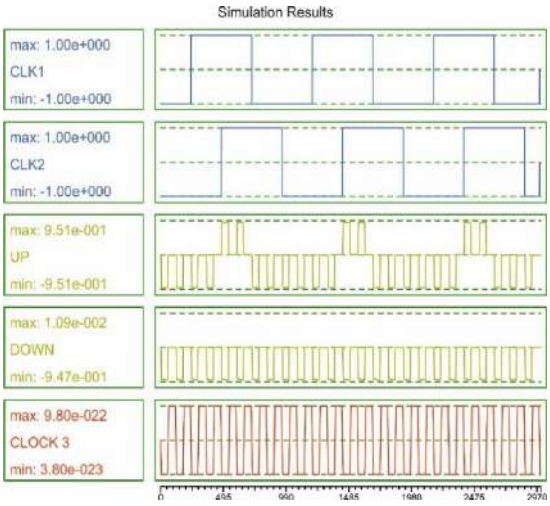


Figure 10. Detect different phases

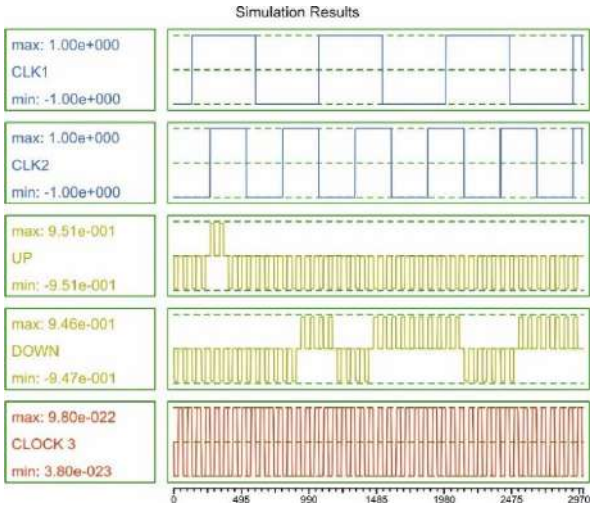


Figure 11. Detect different frequencies

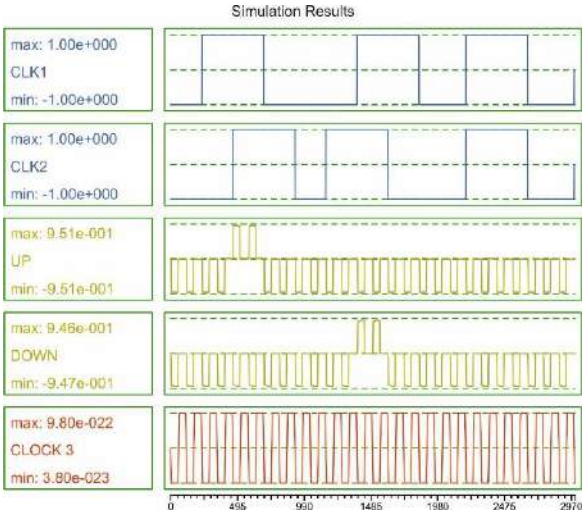


Figure 12. Correct examination two input of proposed PFD

Power consumption analysis of proposed designs is done by QCAPro software that the software uses a fast approximation method to estimate the most erroneous cells in the QCA circuit design [25]. Power dissipation maps for the proposed circuits of Figures 2, 5, and 7 with $0.5E_k$ are illustrated in Figures 13 to 15, respectively. QCA logic gates are thought to be ideal, so there is no charge transfer between cells and current does not flow. Also, the electric charges do not leave any cell, no current is released. These gates cause energy dissipation and we compare the energy consumption of conventional QCA logic gates in electrostatic and thermodynamic approaches. The results show that by increasing the number of inputs, the geometry concentration and the unbalanced numbers of "0" and "1" output modes in the gate truth table will add to the energy dissipation of a QCA gate. We believe that electron transfer between points is a wasteful phenomenon, though the net current is zero. Figures 13, 14 and 15 show that there are dissipations [29]. Darker points define points with more energy dissipation in these figures. The vectors of Figures 8 to 10 are employed for power dissipation maps in Figures 13 to 15, respectively. In Table 4, average for switching energy dissipation and leakage energy dissipation have been listed for the proposed structures and for the power dissipation maps in Figures 13 to 15.

TABLE 3. Comparison of the proposed PFD with other related works

Structure	Cells count	Area (μm^2)	Latency	Average switching energy dissipation (meV), $0.5E_k$	Average leakage energy dissipation (meV), $0.5E_k$
PFD in [26]	199	0.22	2	0.09642	0.06169
PFD in [27]	170	0.26	2.75	0.08667	0.06530
PFD in [28]	159	0.2	2.25	0.03771	0.05370
PFD in [29]	141	0.17	2	0.08679	0.05061
Proposed PFD in figure 7	161	0.175	2	0.05555	0.06202

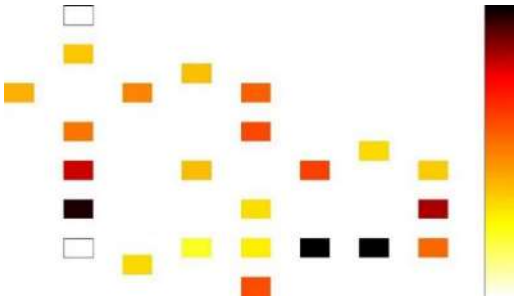


Figure 13. Power dissipation maps for proposed D latch with $0.5E_k$

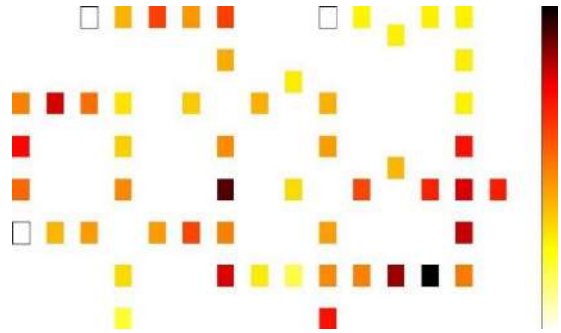


Figure 14. Power dissipation maps for proposed D-FF with reset pin with $0.5E_k$

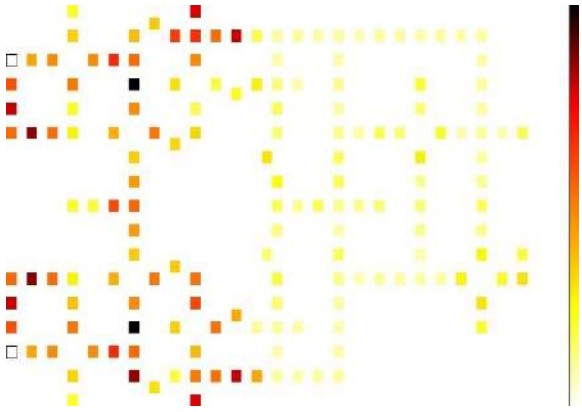


Figure 15. Power dissipation maps for proposed PFD with $0.5E_k$

TABLE 4. Power analysis results

	Average switching energy dissipation (meV)			Average leakage energy dissipation (meV)		
	$0.5E_k$	$1E_k$	$1.5E_k$	$0.5E_k$	$1E_k$	$1.5E_k$
Proposed D latch in figure 2	0.01689	0.01358	0.01101	0.01041	0.02637	0.04328
Proposed D-FF with reset pin in figure 5	0.04414	0.03510	0.02819	0.02205	0.05909	0.09869
Proposed PFD in figure 7	0.05555	0.04347	0.03464	0.06202	0.17174	0.29340

5. CONCLUSION

In this paper a novel D-latch has been designed based on QCA nanotechnology. The proposed D-latch uses NNI gate. The proposed D-latch has only 24 QCA cells, 0.02μm^2 and delay of 0.5 cycle of QCA clock. These parameters indicate that the proposed D-latch has improved compared to previous work, and this D-latch improves the design of larger circuits. Then, the proposed D-latch is used to have a new D-FF in QCA nanotechnology. Reset ability is added to proposed D-FF since this pin is needed in many applications of D-FFs such as counter, shift register and PFD. To show the correct performance of the proposed D-FF with reset ability, it is used in PFD structure. Also, PFD can be used in PLL and DLL. power simulations for all of the proposed designs have been reported.

6. REFERENCES

1. Lent, C.S., Tougaw, P.D., Porod, W. and Bernstein, G.H., "Quantum cellular automata", ***Nanotechnology***, Vol. 4, No. 1, (1993), 49.
2. Seminario, J.M., Derosa, P.A., Cordova, L.E. and Bozard, B.H., "A molecular device operating at terahertz frequencies: Theoretical simulations", ***IEEE Transactions on Nanotechnology***, Vol. 3, No. 1, (2004), 215-218.
3. Timler, J. and Lent, C.S., "Power gain and dissipation in quantum-dot cellular automata", ***Journal of Applied Physics***, Vol. 91, No. 2, (2002), 823-831.
4. Mano, M.M. and Ciletti, M., "Digital design: With an introduction to the verilog hdl, Pearson, (2013).
5. Razavi, B., "Design of analog cmos integrated circuits. 2nd edition, New York, NY, USA: McGraw-Hill, (2017).
6. Gholami, M. and Ardeshir, G., "Jitter of delay-locked loops due to pfd", ***IEEE Transactions on Very Large Scale Integration Systems***, Vol. 22, No. 10, (2014), 2176-2180.
7. Rahimpour, H., Gholami, M., Miar-Naimi, H. and Ardeshir, G., "All digital fast lock dll-based frequency multiplier", ***Analog Integrated Circuits and Signal Processing***, Vol. 78, No. 3, (2014), 819-826.
8. Gholami, M., "Phase detector with minimal blind zone and reset time for gsamples/s dlls", ***Circuits, Systems, and Signal Processing***, Vol. 36, No. 9, (2017), 3549-3563.
9. Dehkordi, M.A. and Sadeghi, M., "A new approach to design d-ff in qca technology", in Proceedings of 2012 2nd International Conference on Computer Science and Network Technology, IEEE. (2012), 2245-2248.
10. Hashemi, S. and Navi, K., "New robust qca d flip flop and memory structures", ***Microelectronics Journal***, Vol. 43, No. 12, (2012), 929-940.
11. Abutaleb, M., "Robust and efficient quantum-dot cellular automata synchronous counters", ***Microelectronics Journal***, Vol. 61, (2017), 6-14.
12. Sasamal, T.N., Singh, A.K. and Ghanekar, U., Design of qca-based d flip flop and memory cell using rotated majority gate, in Smart innovations in communication and computational sciences. 2019, Springer.233-247.
13. Roshan, M.G. and Gholami, M., "Novel d latches and d flip-flops with set and reset ability in qca nanotechnology using minimum cells and area", ***International Journal of Theoretical Physics***, Vol. 57, No. 10, (2018), 3223-3241.
14. Zoka, S. and Gholami, M., "Two novel flip flop with level triggered reset in quantum dot cellular automata", ***International Journal of Engineering***, ***Transactions C: Aspects,*** Vol. 31, No. 3, (2018), 415-421.
15. Zoka, S. and Gholami, M., "A novel rising edge triggered resettable d flip-flop using five input majority gate", ***Microprocessors and Microsystems***, Vol. 61, (2018), 327-335.
16. Binaei, R. and Gholami, M., "Design of novel d flip-flops with set and reset abilities in quantum-dot cellular automata nanotechnology", ***Computers & Electrical Engineering***, Vol. 74, (2019), 259-272.
17. Snider, G.L., Orlov, A.O., Kummamuru, R.K., Ramasubramaniam, R., Amlani, I., Bernstein, G.H., Lent, C.S., Merz, J.L. and Wolfgang, P., "Quantum-dot cellular automata: Introduction and experimental overview", in Proceedings of the 2001 1st IEEE Conference on Nanotechnology. IEEE-NANO 2001 (Cat. No. 01EX516), IEEE. (2001), 465-470.
18. Xiao, L.-r., Chen, X.-x. and Ying, S.-y., "Design of dual-edge triggered flip-flops based on quantum-dot cellular automata", ***Journal of Zhejiang University SCIENCE C***, Vol. 13, No. 5, (2012), 385-392.

19. Tougaw, P.D. and Lent, C.S., "Logical devices implemented using quantum cellular automata", ***Journal of Applied Physics***, Vol. 75, No. 3, (1994), 1818-1825.
20. Gin, A., Tougaw, P.D. and Williams, S., "An alternative geometry for quantum-dot cellular automata", ***journal of Applied Physics***, Vol. 85, No. 12, (1999), 8281-8286.
21. Amirzadeh, Z. and Gholami, M., "Counters designs with minimum number of cells and area in the quantum-dot cellular automata technology", ***International Journal of Theoretical Physics***, Vol. 58, No. 6, (2019), 1758-1775.
22. Sen, B. and Sikdar, B., "Characterization of universal nand-nor-inverter qca gate", in Proceedings of 11th IEEE VLSI Design and Test Symposium. (2007), 433-442.
23. Zahmatkesh, M., Tabrizchi, S., Mohammadyan, S., Navi, K. and Bagherzadeh, N., "Robust coplanar full adder based on novel inverter in quantum cellular automata", ***International Journal of Theoretical Physics***, Vol. 58, No. 2, (2019), 639-655.
24. Walus, K., Dysart, T.J., Jullien, G.A. and Budiman, R.A., "Qcadesigner: A rapid design and simulation tool for quantum-dot cellular automata", ***IEEE Transactions on Nanotechnology***, Vol. 3, No. 1, (2004), 26-31.
25. Srivastava, S., Asthana, A., Bhanja, S. and Sarkar, S., "Qcapro-an error-power estimation tool for qca circuit design", in 2011 IEEE international symposium of circuits and systems (ISCAS), IEEE. (2011), 2377-2380.
26. Alamdar, H., Ardeshir, G. and Gholami, M., "Phase-frequency detector in qca nanotechnology using novel flip-flop with reset terminal", ***International Nano Letters***, Vol. 10, (2020), 111-118.
27. Binaei, R. and Gholami, M., "Using d flip-flop with reset terminal to design pfd in qca nanotechnology", ***International Journal of Electronics***, Vol. 107, No. 12, (2020), 1940-1962.
28. Gholamnia Roshan, M., Gholami, M. and Mahdiyan Jouibari, S., "Low power and low latency phase-frequency detector in quantum-dot cellular automata nanotechnology", ***International Journal of Nanoscience and Nanotechnology***, Vol. 16, No. 3, (2020), 145-152.
29. Alamdar, H., Ardeshir, G. and Gholami, M., "Novel quantum-dot cellular automata implementation of flip-flop and phase-frequency detector based on nand-nor-inverter gates", ***International Journal of Circuit Theory and Applications***, Vol. 49, No. 1, (2021), 196-212.

Persian Abstract

چکیده

روند کاهش ابعاد در فناوری CMOS و همچنین قابل حمل تر شدن دستگاه های دیجیتال، با چالش‌های جدی از جمله افزایش فرکانس و کاهش توان مصرفی روبرو است. به همین دلیل دانشمندان به دنبال راه‌حلی مانند جایگزینی فناوری CMOS با فناوری دیگری از جمله فناوری آتاماتای سلولی دات–کوانتومی (QCA) است و بساری از محققان مدارهای دیجیتالی با استفاده از فناوری QCA طراحی کرده‌اند. فلیپ‌فلاپ‌ها یکی از بلوک‌های اصلی در اکثر مدارهای دیجیتالی هستند. دراین مقاله، یک فلیپ‌فلاپ نوع D در تکنولوژی QCA ارائه می‌شود که از یک گیت اکثریت در مسیر فیدبک آن جهت بازنشانی استفاده شده است. فلیپ‌فلاپ D توسط لچ D پیشنهادی مبتنی‌بر گیت Nand-Nor-Inverter (NNI) و گیت وارونگر جدید طراحی شده که لچ D پیشنهادی ۲۴ سلول و ۰/۵ دوره چرخش کلاک تاخیر و مساحت $0.02\ \mu m^2$ دارد. گیت وارونگر فلیپ‌فلاپ D سیگنال خروجی با سطح قطبیت زیاد و مساحت کمتر نسبت به وارونگرهای گذشته دارد و گیت NNI فلیپ‌فلاپ D گیتی یونیورسال می‌باشد. یکی از کاربردهای فلیپ‌فلاپ D استفاده در آشکارساز فاز–فرکانس (PFD) است. در طرح پیشنهادی ویژگی بازنشانی به فلیپ‌فلاپ D اضافه شده تا ساختار PFD را بتوان طراحی کرد. تمام طرح‌های پیشنهادی توسط نرم‌افزار QCADesigner و شبیه‌سازی انرژی مصرفی برای همه مدارات پیشنهادی با استفاده از نرم‌افزار QCAPro سنجیده شده‌اند.

International Journal of Engineering

Journal Homepage: www.ije.ir

A Compact Model of Gate Capacitance in Ballistic Gate-all-around Carbon Nanotube Field Effect Transistors

A. Dixit, N. Gupta*

Nanomaterial Device Laboratory, Department of Electrical and Electronics Engineering,Birla Institute of Technology and Science, Pilani, Rajasthan, India

P A P E R I N F O

Paper history:
Received 24 February 2021
Received in revised 06 May 2021
Accepted 09 May 2021

Keywords:
Carbon Nanotube Field Effect Transistor
Quantum Capacitance
Gate Capacitance
Gate-all-around Structure

A B S T R A C T

This paper presents a one-dimensional analytical model for calculating gate capacitance in Gate-All-Around Carbon Nanotube Field Effect Transistor (GAA-CNFET) using electrostatic approach. The proposed model is inspired by the fact that quantum capacitance appears for the Carbon Nanotube (CNT) which has a low density of states. The gate capacitance is a series combination of dielectric capacitance and quantum capacitance. The model so obtained depends on the density of states (DOS), surface potential of CNT, gate voltage and diameter of CNT. The quantum capacitance obtained using developed analytical model is 2.84 pF/cm for (19, 0) CNT, which is very close to the reported value 2.54 pF/cm. While, the gate capacitance comes out to be 24.3×10^{-2} pF/cm. Further, the effects of dielectric thickness and diameter of CNT on the gate capacitance are also analysed. It was found that as we reduce the thickness of dielectric layer, the gate capacitance increases very marginally which provides better gate control upon the channel. The close match between the calculated and simulated results confirms the validity of the proposed model.

doi: *10.5829/ije.2021.34.07a.16*

NOMENCLATURE

C	*gate capacitance*	**Greek Symbols**	
t_{ins}	thickness of insulator	ε_r	relative permittivity of gate dielectrical material
C_{cen}	centroid capacitance	ε_0	the permittivity of free space
C_q	quantum capacitance	ψ_s	surface potential (eV)
Q_{CNT}	linear charge density	V_{CNT}	surface voltage (V)
E_g	energy band gap	**Subscripts**	
U	fermi energy	s	source
$D_P(E)$	density of state at p^{th} subband	d	drain
		g	gate

1. INTRODUCTION

Due to scaling, the Complementary Metal-Oxide Semiconductor (CMOS) technology is facing lot of challenges which motivated researchers to look into different materials and structures for fabricating FET. Actually, the scaling of CMOS devices mainly degrades the performance of device due to increase in leakage current. This leakage current increases [1] further due to various short channel effects [2] including source/drain charge sharing, drain-induced barrier lowering (DIBL), subsurface punch through and velocity saturation. Due to limitations of MOSFET, conventional nano-scale SOI MOSFET which uses novel configuration gives better electrical performance but the complexity of these devices pushes the need to find out an alternative of FET devices [3]. In order to reduce these effects and enhance the device performance, carbon allotropes *i.e.* Carbon

*Corresponding Author Institutional Email: ngupta@pilani.bits-pilani.ac.in (N. Gupta)

Please cite this article as: A. Dixit, N. Gupta, A Compact Model of Gate Capacitance in Ballistic Gate-all-around Carbon Nanotube Field Effect Transistors, International Journal of Engineering, Transactions A: Basics Vol. 34, No. 7, (2021) 1718-1724

Nanotubes (CNTs) [4] are very promising material, due to their small size and remarkable electromechanical and thermal properties. Carbon nanotube field effect transistors (CNFETs) are better alternative to replace the Si-MOSFETs [5] due to better control of short channel effects, less propagation delay, high on-off current ratio and less power consumption [6-11] in nanometer (nm) dimension for the use in future computing system.

In CNFET devices, CNTs are channels between source and drain and the channel current through CNT is controlled by gate terminal. The CNFET has two geometrical structures: planar structure and Gate-All-Around (GAA) structure. The planar device performance is affected by fringing effect whereas the GAA structure is free from such effect due to its geometry hence, GAA structure is expected to be ideal geometry which maximizes the electrostatic gate control in FETs [12, 13].

Apart from various advantages, the major concern with CNFET is the effect of quantum capacitance on the performance of device. This is due to the fact that when the device has nanometer dimension, the quantum capacitance is comparable with electrostatic dielectric capacitance. In Si-MOSFET, it has been observed that quantum capacitance increases with increase in gate voltage [14]. The quantum capacitance is introduced in device when density of state (DOS) is very low or the energy level separation between states is very high [15]. So, it is important to investigate quantum capacitance of CNFET.

For CNFET, gate capacitance is a function of quantum capacitance and dielectric capacitance. The quantum capacitance for semiconducting single wall CNT (SWNT) (16, 8) has been already measured experimentally [16]. In past, various analytical models for gate capacitance have been proposed. Most of the proposed models are for planar structure only. Ahmed et al. [17] developed an analytical model for aligned CNT based FET with screening effects. This conformal mapping-based model is limited to the planar structure of CNFET. Singh [18] proposed an analytical expression for quantum capacitance. This analytical expression is based on the total charge density and analyzed that leakage current reduces for lower quantum capacitance but this expression is validated only for planar CNFET. Deyasi and Sarkar [19] developed a model for GAA-CNFET using Non-Equilibrium Green's Function (NEGF) method. They considered armchair (metallic) nanotube for CNFET but CNFET has semiconductor channel between source and drain. The novelty of the work lies in the choice the device structure with single CNT as channel which was further considered for modeling quantum capacitance for GAA-CNFET.

In this work, Gate-All-Around (GAA) structure is considered with single wall CNT as a channel. It has been assumed that conduction occurs in first sub-band and that there are no fringing effect and screening effect, due to single wall CNT between source and drain and the structure is free from sharp edges.

This paper is organized in four sections. Section 2 describes the theory and model of quantum capacitance for channel (CNT). Section 3 explains the result and discussion obtained from analytical model and simulations and finally the conclusion of this study is given in section 4.

2. THEORY AND MODEL

Figure 1 shows the schematic view of GAA- CNFET considered for gate capacitance modeling. The gate capacitance (C_{total}) in CNFET is a series combination of dielectric capacitance (C_{ins}) and inversion layer capacitance (C_{inv}) which is given as follows:

$$C_{total} = \frac{C_{ins} \times C_{inv}}{C_{ins} + C_{inv}} \quad (1)$$

where C_{ins} for gate-all-around structure is given by:

$$C_{ins} = \frac{2\pi\varepsilon_0\varepsilon_r}{ln\left(\frac{r_{CNT} + t_{ins}}{r_{CNT}}\right)} \quad (2)$$

where ε_0 is the permittivity of free space, ε_r is the relative permittivity of gate dielectric material, r_{CNT} is the radius of CNT and t_{ins} is the dielectric thickness.

Further, the inversion layer capacitance (C_{inv}) is a series combination of centroid capacitance (C_{cen}) and quantum capacitance (C_q) [15]:

$$C_{inv} = \frac{C_{cen} \times C_q}{C_{cen} + C_q} \quad (3)$$

C_{cen} can be ignored because we assume all changes to be located at the same position in the CNT layer. So:

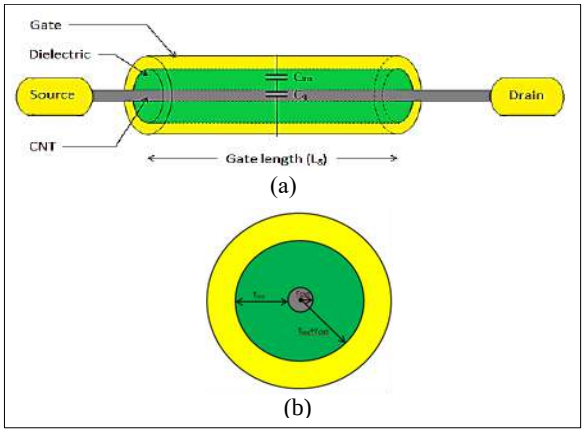


Figure 1. Schematic view of CNFET for modelling gate capacitance (a) front view (b) side view

$$C_{inv}=C_q \tag{3-a}$$

Equation (1) and (3-a), shows that the gate capacitance is a series combination of C_{ins} and C_q, which is given as:

$$C_{total}=\frac{C_{ins}\times C_q}{C_{ins}+C_q} \tag{4}$$

Quantum capacitance can be calculated by measuring the change in the linear charge density with surface potential [20].

$$C_q=\frac{d(\text{linear charge density})}{d(\text{surface potential})}=\frac{d(Q_{CNT})}{d(\psi_S)} \tag{5}$$

$$\psi_S = qV_{CNT} \tag{6}$$

so, Equation (5) can be rewritten as:

$$C_q=\frac{1}{q}\times\frac{d(Q_{CNT})}{d(V_{CNT})} \tag{7}$$

Linear charge density is the sum of charge density at source and drain terminals. The charge densities are different because V_{gs} and V_{ds} are independent and different. The total charge density (Q_{CNT}) on CNT surface is given as;

$$Q_{CNT} = Q_s + Q_d \tag{8}$$

where Q_s and Q_d are the charge densities at the source end and drain end of the channel, respectively.

Q_{CNT} is a function of Fermi-Dirac distribution and density of state (DOS). Hence, the total Q_{CNT} can be written as [21];

$$Q_{CNT}=\int_{E_C}^{\infty}\left[\left\{\frac{D(E)}{2}F(E-U_S)\right\}+\left\{\frac{D(E)}{2}F(E-U_d)\right\}\right]dE \tag{9}$$

D(E) is density of state which is a function of E, U_s and U_d are the fermi energy of source and drain, respectively.

There are n subbands, so the total charge for p^{th} subband is given as:

$$Q_{PCNT}=\int_{E_{CP}}^{\infty}\left[\begin{array}{l}\left\{\frac{D_P(E)}{2}F(E-E_{CP}-U_S)\right\}+\\ \left\{\frac{D_P(E)}{2}F(E-E_{CP}-U_d)\right\}\end{array}\right]dE \tag{10}$$

where E_{CP} is the conduction band minimum of p^{th} subband. The conduction band minima is the difference of p^{th} band energy (Δ_p) and surface potential (ψ_s) *i.e*:

$$E_{CP} = \Delta_P - qV_{CNT} \tag{11}$$

For one-dimensional structure, the general equation for density of state is given by [22]:

$$D(E)=\frac{D_0 E}{\sqrt{E^2-\left(\frac{E_g}{2}\right)^2}} \tag{12}$$

where

$$D_0=\frac{8}{3\pi V_\pi a_{c\text{-}c}}=2\times10^9\, m^{-1}eV^{-1}$$

$a_{c\text{-}c}$= Carbon-carbon length = 1.42 Å

$V\pi$= carbon-carbon bonding energy=3.0 eV

The DOS of pth subband is given by:

$$D_P(E)=\frac{D_0(E+\Delta_P)}{\sqrt{(E+\Delta_P)^2-(\Delta_P)^2}} \tag{13}$$

Charge density at source is given by:

$$Q_S=\int_{E_{CP}}^{\infty}\frac{D_P(E)}{2}F(E-E_{CP}-U_s) \tag{14}$$

By using Equations (11) and (13), charge density of source becomes:

$$Q_S=\int_{E_{CP}}^{\infty}\left[\frac{D_0(E+\Delta_P)}{2\sqrt{(E+\Delta_P)^2-(\Delta_P)^2}}\times\frac{1}{1+exp\left(\frac{E-\Delta_P+qV_{CNT}-U_s}{kT}\right)}\right]dE \tag{15}$$

Similarly charge density at drain can be written as:

$$Q_d=\int_{E_{CP}}^{\infty}\frac{D_0(E+\Delta_P)}{2\sqrt{(E+\Delta_P)^2-(\Delta_P)^2}}\times\frac{1}{1+exp\left(\frac{E-\Delta_P+qV_{CNT}-U_d}{kT}\right)}dE \tag{16}$$

Equations (8), (15) and (16), give the total charge density:

$$Q_{CNT}=\int_{E_{CP}}^{\infty}\left[\frac{D_0(E+\Delta_P)}{2\sqrt{(E+\Delta_P)^2-(\Delta_P)^2}}\times\left\{\frac{1}{1+e^{\left(\frac{E-\Delta_P+qV_{CNT}-U_s}{kT}\right)}}+\frac{1}{1+e^{\left(\frac{E-\Delta_P+qV_{CNT}-U_d}{kT}\right)}}\right\}\right]dE \tag{17}$$

where; $U_s = qV_{gs}$; $U_d = qV_{ds}$

For SWCNT, as we considered 1st subband, so $\Delta_p=\Delta_1$ which is equal to the corresponding energy level minima (E) (*i.e.* Δ_1=E). So, from Equation (17):

$$Q_{CNT}=\frac{D_0}{2}\times\frac{2}{\sqrt{3}}\times\left[\frac{1}{1+e^{\left(\frac{V_{CNT}-V_{gs}}{(V_T)}\right)}}+\frac{1}{1+e^{\left(\frac{V_{CNT}-V_{ds}}{(V_T)}\right)}}\right] \tag{18}$$

In CNFET, two types of electric field exist: horizontal and vertical. Horizontal electric field is due to V_{ds}, which is responsible for the flow of current in the channel while the vertical electric field is due to V_{gs}, which is responsible to generate the charge carriers in channel. In this study for the calculation of total charge carriers on CNT, it was assumed that $V_{ds}=0$ and $V_{gs}=V_g$. So, Equation (18), can be written as follows:

$$Q_{CNT} = \frac{D_0}{\sqrt{3}} \times \left[\frac{1}{1+exp\left(\frac{V_{CNT}-V_g}{V_T}\right)} + \frac{1}{1+exp\left(\frac{V_{CNT}}{V_T}\right)} \right] \quad (19)$$

where, $(V_g - V_{CNT})$ can be calculated by [23]:

$$V_{g,CNT} = V_g - V_{CNT} = \begin{cases} 0 & for\ V_g < \Delta_1 \\ \alpha(V_g - \Delta_1) & for\ V_g > \Delta_1 \end{cases} \quad (20)$$

where α is slope of curve (V_g versus ψ_s), which is a function of device parameter, *d* (diameter of CNT).

From Equations (5) and (19), the quantum capacitance for SWCNT with first subband is as follows:

$$C_q = \frac{Q_{CNT}}{qV_{CNT}} = \frac{D_0}{\sqrt{3}\cdot qV_{CNT}} \times \left[\frac{1}{1+exp\left(\frac{V_{CNT}-V_g}{V_T}\right)} + \frac{1}{1+exp\left(\frac{V_{CNT}}{V_T}\right)} \right] \quad (21)$$

Quantum capacitance also depends on the chirality and diameter of CNT. As we increase the diameter of CNT, the energy bandgap and first sub-band minima also reduces and V_{CNT} increases.

3. RESULTS AND DISCUSSION

Quantum capacitance (Cq) is calculated using Equation (19) and further analyzed. In this study, CNT (19, 0) has been assumed for which the diameter is calculated using QuantumWise Atomistix Tool Kit (ATK) (QuantumATK O-2018.06) which comes out to be 1.49nm as illustrated in Figure 2. The thickness of dielectric layer and the channel length are considered as 8nm and 30nm, respectively which is feasible dimension for a practical GAA-CNFET [24]. The parameters considered for calculation used in Equation (19), are shown in Table 1.

In our previous work, it was demonstrated that La_2O_3 is the best gate dielectric material followed by HfO_2 and ZrO_2 [25, 26]. So, we used La_2O_3 (ε_r =30) as a dielectric layer in CNFET. The surface potential of CNT is calculated for different values of gate voltage (V_g). The value corresponding to 1.49nm diameter CNT for V_g and V_{CNT} are listed in Table 2 and $V_{g,CNT}$ calculated using Equation (20). For first sub-band energy minima of CNT, the value is calculated using simulation which comes out to be 0.281eV. Figure 3 shows the plot of quantum

Figure 2. Illustration of diameter calculation for (19, 0) CNT simulated using QuantumWise Atomistix Tool Kit (ATK)

TABLE 1. Parameters for quantum capacitance

S. No.	Parameters	Values
1.	CNT	Zigzag (19,0)
2.	Diameter of CNT	1.49nm
3.	Dielectric material	La_2O_3 (ε_r=30)
4.	Channel length	30nm
5.	Source to Drain Voltage (V_{ds})	0V

TABLE 2. Surface potential of CNT for different value of V_g [27]

V_g (volt)	V_{CNT} (volt)	$V_{g,CNT}$ (volt)
0	0	0
0.05	0.5	0
0.1	0.1	0
0.15	0.15	0
0.2	0.2	0
0.25	0.25	0
0.3	0.3	0
0.35	0.3276	0
0.4	0.3490	0.06305
0.45	0.3629	0.0871
0.5	0.3727	0.1261
0.55	0.3801	0.1699
0.6	0.3861	0.18915
0.65	0.3910	0.2590
0.7	0.3952	0.2522

capacitance (C_q) as a function of gate voltage. It is observed that initially C_q increases sharply with increase in V_g upto 0.4 Volt thereafter the slope of curve decreases

and peak is observed at V_g= 0.5V and after that it gradually decreases with increase in V_g. The peak value of C_q obtained at V_g of 0.5V is 2.84pF/cm. The value obtained from this analytical work is compared with published results, which is listed in Table 3.

The gate capacitance of CNFET is calculated using Equation (1) for which C_{ins} is calculated with the help of Equation (2). By this analysis, the gate capacitance of CNFET comes out to be 24.3×10^{-2} pF/cm. In order to validate finding of this work, analytical work is compared with simulations [29]. Figure 4 shows the simulated results of gate capacitance versus gate voltage, which gives the C_{total} as 22.5×10^{-2} pF/cm at V_g of 0.5V.

Figure 5 shows the variation of gate capacitance with gate voltage for different dielectric layer thickness (t_{ox}= 5nm, 7nm, 8nm and 9nm). This graph shows that as the dielectric thickness increases, gate capacitance decreases very marginally. This is inconsistent with the work presented by Deyasi and Sarkar [19], in which they have observed that quantum capacitance is not much affected by the thickness of dielectric layer. Table 4 shows the comparative analysis of analytical and simulated gate capacitance for the different thickness of dielectric layer. The close match between the two conform the validity of our proposed analytical model for gate capacitance of CNFET.

Figure 6 shows the variation of gate capacitance with gate voltage for different diameter of CNTs

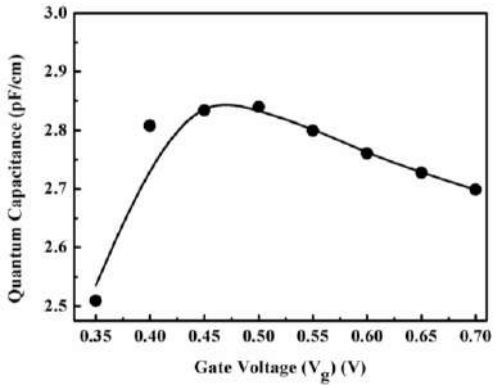


Figure 3. Variation of quantum capacitance versus gate voltage for (19, 0) CNT

TABLE 3. Comparative analysis of quantum capacitance at V_g = 0.5V

Research Group	Quantum capacitance (C_q) (pF/cm)
Mozahid and Ali [28]	3
Deyasi and Sarkar [19]	2.54
Proposed work	2.84

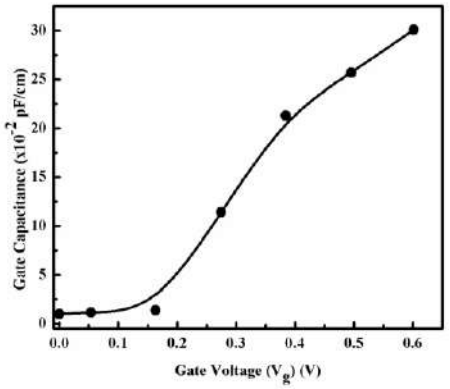


Figure 4. Simulation plot of gate capacitance versus gate voltage for CNFET

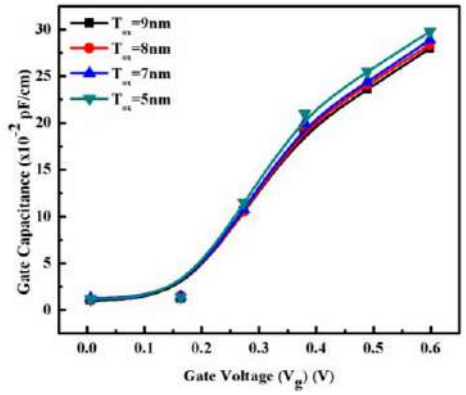


Figure 5. Simulation plot of gate capacitance versus gate voltage for CNFET with different gate dielectric thickness

TABLE 4. Comparison between analytical and simulated results for different thickness of dielectric layer

Thickness of dielectric layer (t_{ox})	Gate Capacitance ($\times10^{-14}$ F/cm)	
	Analytical Work	Simulated Work
5nm	25.44	24.06
7nm	25.06	23.05
8nm	24.91	22.66
9nm	24.78	22.23

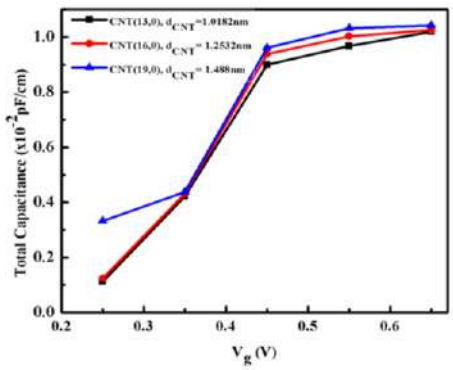


Figure 6. Simulation plot of gate capacitance versus gate voltage for CNFET with different diameter of CNT

($d_{(13,0)}$=1.0182nm, $d_{(16,0)}$=1.2532nm and $d_{(19,0)}$=1.488nm). The total gate capacitance for CNT (19,0) is higher than CNT (13,0) and CNT (16,0) at V_{gs} of 0.25V but it is increases marginally after V_{gs} of 0.35V.

4. CONCLUSION

A new analytical model of gate capacitance for ballistic GAA-CNFET with SWCNT has been proposed and successfully investigated. To verify this model we used zigzag semiconductive CNT (19, 0) with the channel length of 30 nm. For La_2O_3 (ε_r=30) gate dielectric material, we obtained the quantum capacitance of 2.84 pF/cm and gate capacitance of 22.5×10^{-2} pF/cm at gate voltage of 0.5V. In order to verify the validity of proposed model, the results obtained were compared with experimental work available in literature. The close match between the calculated and experimental results confirms the validity of the proposed model. Further, the effect of gate dielectric thickness (5nm, 7nm, 8nm and 9nm) and diameter of CNT (1.0182nm, 1.2532nm and 1.488nm), on gate capacitance is also studied. It has been analysed that gate capacitance marginally decreases with increase in the thickness of gate dielectric layer, while increases with the rise in the diameter of CNT. Moreover, this model can be very well used to interpret the response of the CNFET towards variation of other dependent physical parameters. The results were compared with simulation results, and close match between the two, shows the validity of the proposed model for gate capacitance and also shows that the outcome of the performance of the device would remain consistent.

5. ACKNOWLEDGMENTS

Authors acknowledge the financial support of Defence Research and Development Organisation (DRDO), Govt. of India [ERIP/ER/DGMED&OS/990416502/ M/01/1657] and Nanomaterial device laboratory, BITS Pilani for carrying research out work reported this paper.

6. REFERENCES

1. Bala S. and Khosla M., "Comparative Study and Analysis of CNTFET and Tunnel CNTFET", ***Journal of Nanoelectronics and Optoelectronics***, Vol. 13, (2018), 324-330. DOI: 10.1166/jno.2018.2234.
2. Anvarifard M. K., Ramezani Z. and Amiri I. S., " Proposal of an Embedded Nanogap Biosensor by a Graphene Nanoribbon Field-Effect Transistor for Biological Samples Detection", ***Physica Status Solidi (a***), Vol. 217, (2019), 1900879 (1-7). DOI: 10.1002/pssa.201900879.
3. Anvarifard, M. K., "Modeling a Double-Halo-Doping Carbon Nanotube FET in DC and AC Operations***", ECS Journal of Solid State Science and Technology***, Vol. 7, (2018), M209-M216. DOI: 10.1149/2.0191812jss.
4. Bousari N. B. and Anvarifard M. K., "A Theoretical Study on Charge Transfer of Twisted T-Graphene Nanoribbon Surface", ***ECS Journal of Solid State Science and Technology*** 9, (2020), 021001.
5. Khadem Hosseini V., Dideban D., Ahmadi M. T., and Ismail R., " An analytical approach to model capacitance and resistance of capped carbon nanotube single electron transistor", ***AEU-International Journal of Electronics and Communications***, Vol. 90, (2018), 97–102. DOI: 10.1016/j.aeue.2018.04.015.
6. Prakash P., Mohana Sundaram K. and Anto Bennet M., "A review on carbon nanotube field effect transistors (CNTFETs) for ultra-low power applications", ***Renewable and Sustainable Energy Reviews***, Vol. 89, (2018), 194–203. DOI: 10.1016/j.rser.2018.03.021.
7. Sinha S. K. and Chaudhury S., "Comparative study of leakage power in CNTFET over MOSFET device", ***Journal of Semiconductors***, Vol. 35, (2014), 114002 (1-6). DOI: 10.1088/1674-4926/35/11/114002.
8. Moaiyeri M H, Rahi A, Sharifi F and Navi K, "Design and evaluation of energy-efficient carbon nanotube FET-based quaternary minimum and maximum circuits", ***Journal of Applied Research and Technology***, Vol. 15, (2017), 233-241 DOI: 10.1016/j.jart.2016.12.006.
9. Shirazi S. G. and Mirzakuchaki S., "High on/off current ratio in ballistic CNTFETs based on tuning the gate insulator parameters for different ambient temperatures", ***Applied Physics A***, Vo. 113, (2013), 447-457. DOI: 10.1007/s00339-012-7543-9.
10. Murthy G. R., Singh A. K., Hossen J. and Velrajkumar P., "Performance analysis of electrical characteristics for Short Channel Effects (SCE) in Carbon Nanotube Field Effect Transistor (CNTFET) Devices", ***Journal of Engineering and Applied Sciences***, Vol. 12, (2017), 5116-5120. DOI: 10.3923/jeasci.2017.5116.5120
11. Khaleqi, M., Mir, A., Mirzakuchaki, S., Farmani, A., "Design and performance analysis of wrap-gate CNTFET-based ring oscillators for IoT applications", ***Integration***, Vol. 70, 116-125. DOI: 10.1016/j.vlsi.2019.10.005.
12. Jena B., Pradhan K. P., S. Dash, Mishra G. P., Sahu P. K. and Mohapatra S. K., "Performance analysis of undoped cylindrical gate all around (GAA) MOSFET at subthreshold regime", ***Adv. Nat. Sci. Nanosci. Nanotechnology***, Vol. 6, (2015), 035010-4. DOI: 10.1088/2043-6262/6/3/035010.
13. Chaudhury S. and Sinha S. K., "Carbon Nanotube and Nanowires for Future Semiconductor Devices Applications", ***Nanoelectronics***, (2019), 375-398. DOI: 10.1016/B978-0-12-813353-8.00014-2.
14. Shailendra S. R. and Ramakrishnan V. N., "Analysis of quantum capacitance on different dielectrics and its dependence on threshold voltage of CNTFET**",** International Conference on Nextgen Electronic Technologies: Silicon to Software (ICNETS2) Chennai India, (2017), 213-217. DOI: 10.1109/ICNETS2.2017.8067933.
15. Wong H-S. P. and Akinwande D., "Carbon Nanotube and Graphene Device Physics". Cambridge University Press New York, 2011.
16. Dai J., Li J., Zeng H. and Cui X., "Observation of Quantum Capacitance of individual single walled carbon nanotubes", ***Applied Physics Letters***, Vol. 94, (2008), 1-13. DOI: 10.1063/1.3093443.
17. Ahmed Z., Zhang L., and Chan M., "Gate Capacitance Model for Aligned Carbon Nanotube FETs with Arbitrary CNT Spacing",

IEEE Transactions on Electron Devices, Vol. 62, (2015), 4327-4332. DOI: doi:10.1109/ted.2015.2484384.

18. Singh A. K., "Analytical analysis of quantum capacitance in nano-scale single-wall carbon nanotube field effect transistor (CNTFET)", ***International Journal of Nanoelectronics and Materials***, Vol. 11, (2018), 249-262
19. Deyasi A. and Sarkar A., "Analytical computation of electrical parameters in GAAQWT and CNTFET with identical configuration using NEGF method", ***International Journal of Electronics***, Vol. 105, (2018), 1-16. DOI: 10.1080/00207217.2018.1494339.
20. Kordrostami Z., Sheikhi M. H. and Zarifkar A., "Influence of Channel and Underlap Engineering on the High-Frequency and Switching Performance of CNTFETs", ***IEEE Transactions on Nanotechnology***, Vol. 11, (2012), 526-533. DOI: 10.1109/tnano.2011.2181998.
21. Djamil R., Salima B. and Kheireddine L., "Performance Enhancement of CNTEFT with High-K Dielectric", ***Advanced Material Research***, Vol. 685, (2013), 340-344. DOI: 10.4028/www.scientific.net/amr.685.340.
22. Lundstrom M. S. and Guo J., Nanoscale Transistors: Device Physics, Modeling and Simulation, Springer Berlin, 2006.
23. Bala S. and Khosla M., "Electrostatically doped tunnel CNTFET model for low-power VLSI circuit design", ***Journal of Computational Electronics***, Vol. 17, (2018), 1528-1535. DOI: 10.1007/s10825-018-1240-7.
24. Franklin A. D., Koswatta S. O., Farmer D, Tulevski G S, Smith J T, Miyazoe H and Haensch W, "Scalable and fully self-aligned n-type carbon nanotube transistors with gate-all-around", ***International Electron Devices Meeting,*** (2012). DOI: 10.1109/iedm.2012.6478979.
25. Dixit A. and Gupta N., "Analysis of Different Gate Dielectric Materials in Carbon Nanotube Field Effect Transistor (CNFET) using Optimization Technique", IEEE Electron Device Kolkata Conference 2018 IEEE EDKCON), Kolkata, (2018). DOI: 10.1109/edkcon.2018.8770503.
26. Dixit A. and Gupta N., "Investigation into gate dielectric material using different optimization techniques in carbon nanotube field effect transistors," ***Journal of Micromechanics and Microengineering***, Vol. 29, 2019, 094002, DOI: 10.1088/1361-6439/ab2a61.
27. Singh A., Khosla M. and Raj B., "Compact model for ballistic single wall CNTFET under quantum capacitance limit", ***Journal of Semiconductors***, Vol. 37, (2016), 104001 (1-8). DOI: 10.1088/1674-4926/37/10/104001
28. Mozahid F. and Ali M. T., "Simulations of Enhanced CNTFET with HfO2 Gate Dielectric", ***International Journal of Scientific and Research Publications***, Vol. 5, (2015), 1-6.
29. Marani R. and Perri A. G., "CNTFET-Based Design of Current Mirror in Comparison with MOS Technology", ***ECS Journal of Solid State Science and Technology***, Vol. 6, (2017), M60-M68. DOI: 10.1149/2.0261705jss

Persian Abstract

چکیده

در این مقاله یک مدل تحلیلی یک بعدی برای محاسبه ظرفیت گیت در ترانزیستور اثر میدان نانولوله کربن گیت (GAA-CNFET) با استفاده از روش الکترواستاتیک ارائه می شود. مدل پیشنهادی از این واقعیت الهام گرفته شده است که ظرفیت کوانتومی برای نانولوله کربنی (CNT)ظاهر می شود که دارای تراکم کم حالت است. ظرفیت گیت ترکیبی از ظرفیت دی الکتریک و ظرفیت کوانتومی است. مدل بدست آمده به چگالی حالتها (DOS) ، پتانسیل سطح CNT ، ولتاژ گیت و قطر CNT بستگی دارد. ظرفیت کوانتومی به دست آمده با استفاده از مدل تحلیلی توسعه یافته ۲.۸۴ pF / cm برای (۰/۰۱۹) CNT است که بسیار نزدیک به مقدار گزارش شده ۲.۵۴ pF / cm است. در حالی که ، ظرفیت گیت ۲۴.۳ × ۱۰-۲ pF / cm بیرون می آید. علاوه بر این ، اثرات ضخامت و قطر دی الکتریک در ظرفیت گیت نیز مورد تجزیه و تحلیل قرار می گیرد. مشخص شد که هرچه ضخامت لایه دی الکتریک را کاهش می دهیم ، ظرفیت گیت بسیار کم افزایش می یابد که کنترل بهتر گیت روی کانال را فراهم می کند. تطابق نزدیک بین نتایج محاسبه شده و شبیه سازی شده ، اعتبار مدل پیشنهادی را تأیید می کند.

IJE TRANSACTIONS A: Basics Vol. 34, No. 7, (July 2021) 1725-1735

International Journal of Engineering

Journal Homepage: www.ije.ir

A New Consensus-based Distributed Adaptive Control for Islanded Microgrids

M. Keshavarz[a], A. Doroudi*[a], M. H. Kazemi[a], N. Mahdian Dehkordi[b]

[a] *Electrical Engineering Department, Shahed University, Tehran, Iran*
[b] *Control Department of Electrical Engineering Faculty, Shahid Rajaee Teacher Training University, Tehran, Iran*

PAPER INFO

Paper history:
Received 15 March 2021
Received in revised form 21 April 2021
Accepted 04 June 2021

Keywords:
Adaptive Control
Distributed Control
Consensus Protocol
Disturbances
Microgrids
Secondary Control

ABSTRACT

This paper proposes a novel distributed adaptive secondary controller for microgrids (MGs) in islanded operation. To enhance the dynamic behaviour of a microgrid considering uncertainties and disturbances, the proposed controller uses a consensus-based adaptive control structure. A novel consensus protocol is proposed to restore the frequency and voltage (f &V) of a microgrid to their rated values. A Lyapunov function is presented to assure the asymptotic stability of the controller and the ultimate boundedness of the global neighborhood consensus error. The nonlinear nature of MGs has been also considered in the algorithm. Unlike other methods in this field that require complete information of distributed generators (DGs), the proposed controller requires only power droop coefficients and is independent of DGs parameters. Different simulations are conducted in MATLAB/SimPower Toolbox on a typical microgrid and under various disturbances to judge the performance of the adaptive controller. The simulation results show the effect of the proposed controller on increasing the resilience of an MG.

***doi**: 10.5829/ije.2021.34.07a.17*

NOMENCLATURE

$Z_{\omega i}$	Avirtual signal	f_i , k_i , g_i , h_i	Nonlinear Functions
D_i	An Uncertainty	a, b	Nonnegative Real Number
V_{ref}	DG Output Voltage	i^{th}	Receiving Node
DG	Distributed Generator	j^{th}	Sending Node
DAFC	Distributed Adaptive Frequency Controller	L	Laplacian Matrix
DSC	Distributed Secondary Control	$u_i = [\tau_i \quad \varpi_i]$	Output Of Secondary Controller
DVAC	Distributed Adaptive Voltage Controller	G	Pinning Matrix
D_{Pi} , D_{Qi}	Droop Coefficients	p, q	Positive Numbers
$\xi_{\omega i}$	Frequency Consensus Error	P_i , Q_i	Output Real and Reactive Powers
$f\&V$	Frequency and Voltage	ξ_{Pi}	Real Power Neighbors Tracking Error
ξ_{vi}	local Neighborhood Consensus Error	SCL	Secondary Control Layer
MG	Microgrid	τ_i , ϖ_i	Set Points
λ_0	Minimum Eigenvalue	β	Steady State Error
σ_{min}	Minimum Singular Value	VSC	Voltage Source Converter
$Z_{\omega i}$	Avirtual signal	f_i , k_i , g_i , h_i	Nonlinear Functions
D_i	An Uncertainty	i^{th}	Receiving Node
V_{ref}	DG Output Voltage		

*Corresponding Author Institutional Email: doroudi@shahed.ac.ir
(A. Doroudi)

Please cite this article as: M. Keshavarz, A. Doroudi, M. Hossein Kazemi, N. Mahdian Dehkordi, A New Consensus-based Distributed Adaptive Control for Islanded Microgrids, International Journal of Engineering, Transactions A: Basics Vol. 34, No. 7, (2021) 1725-1735

1. INTRODUCTION

Currently, renewable energy sources have a significant share of modern power networks [1]. MG is a cluster of small power networks made of several renewable energy sources (photovoltaics, wind turbines, micro-turbines, etc.), energy storage units and, electrical loads [1-2]. MGs can operate in both grid-connected and islanded operating modes. In grid-connected mode, the main grid directly control the frequency and voltage (f & V) of MG, whereas, in islanded mode, all DGs are responsible to maintain these quantities within pre-specified limits [2, 3]. Proper control of microgrid is a prerequisite for stable and economically efficient operation [4,5]. To maximize the use of resources and improve the effectiveness of DGs, a modern control approach is required to increase the microgrid reliability and to provide global stability for the system [6, 7].

In general, the MGs control ststem is carried out at three levels of hierarchy including primary, secondary, and tertiary control levels [7]. The primary control level is related to the local control loops along with droop control of the DGs without requirements for the communication infrastructure. It stabilizes the microgrid f & V and shares the active and reactive power between individual DG units using local current, voltage, and, droop power control loops [8, 9]. The main drawback of this control scheme is the steady-state f & V deviations from their reference values. The secondary control layer (SCL) is utilized to overcome this drawback and compensate for f & V deviations result from the droop control method. The SCL can be implemented in three manners: centralized [10-12], decentralized [13, 14], and distributed [15-21]. The centralized control method employs a central control module and a communication grid between each DG and the central controller. Therefore, it demands extensive communication system to collect and process the massive information of all DGs. As a result, this control method suffers from the drawbacks of the presence of a single point of failure and low reliability [10]. The decentralized control strategy is not usually employed in the secondary control due to the lack of sufficient information to coordinate and harmonize all available MGs resources [13, 14]. To overcome the centralized and decentralized controls problems, inspired by the idea of multi-agent systems, distributed control methods using sparse communication networks have been recently presented [15-21]. In the distributed control scheme, each agent only communicates with the neighboring agents employing a sparse communication grid. The control method reduces the communication grid complexity and is needless of a central controller, and consequently enhance the overall microgrid reliability [22-26]. The third and highest level of control system structure is the tertiary control and essentially deals with economic dispatch and optimal power flow [27].

So far, different types of distributed secondary control (DSC) algorithms have been developed [22-25]. Thanks to the graph theory, consensus protocol is the base of the great number of these algorithms [20,21]. In multi-agent systems, reaching an agreement on a certain amount that depends on the status of all agents is called consensus. When each DG is considered as an agent, returning f & V to the reference values (synchronization process) can be a matter of consensus. Distributed synchronization process necessitates that each agent (i.e., each DG) exchange information with other agents by a consensus protocol. A DSC scheme for islanded microgrids is presented by Dehkordi et al. [25]. Two individual f & V controllers are separately designed and examined. A distributed finite-time approach is first proposed to ensure the finite time restoration of voltage magnitude. The frequency restoration algorithm is then presented so that correct real power sharing is achieved. A two-layer DSC prorocol is proposed by Bidram et al. [28],. Voltage source inverters (VSIs) are utilized in the first layer for maintaining f & V of the microgrid. The second layer which consists of current source inverters (CSIs) shares the reactive and active powers among DGs in a appropriate manner. The effects of delays and noises in communication channels among DGs have been discussed by Shahab et al. [29] thorough stochastic and/or distributed time-delay control methods. Moreover, detail discussion about communication delays' effect on the control of an islanded MG has been investigated by a small-signal model in literature [10,18]. A distributed voltage averaging method was proposed by Amoateng et al. [22] to provide a simple compromise between the incompatible objectives of bus voltages regulating and good reactive power sharing.

A novel DSC approach is proposed by Bidram et al. [30] . The proposed approach is fully distributed and its structure needs a sparse communication grid. To transform the secondary voltage control to a linear synchronization problem, input–output feedback linearization technique was employed. Anyway, the control method is so complex and heavily depends on DGs parameters and network dynamics. Multi-functional DSC structure was proposed by Li et al. [31] with individual frequency, voltage and active power regulator modules. The method, however, needs normalized power information of all DGs.

Microgrid along with its control system is an unknown nonlinear and time-variable system and always prone to uncertainties which are caused by internal disturbances such as parametric variations and unmodeled dynamics or external disturbances such as load changes and DG tripping events [29-32]. Currently to face these challenges, the development of model-free distributed adaptive controllers has become the focus of researchers. Bidram and Davoudi [33] proposed an

adaptive secondary voltage control scheme for microgrids in islanded mode. The paper has used the input output feedback linearization technique like [29] to achieve a linear dynamic system. The fault tolerant consensus-based control of multi agent systems, when the faults and disturbances coexist, have recently been presented in the literature to control multi-agent systems [34]. Anyway, the control method is very complex and heavily depends on DGs parameters and network dynamics. A new cooperative adaptive distributed consensus protocol was presented in the presence of unmodated dynamics by Amoateng et al. [22]. To achieve this goal, a smart secondary adaptive control protocol is proposed which uses two neural networks based identifier and controllers. High computational complexity, knowing DGs information for training neural networks and, need for a specially designed switching scheme between the two controllers are the three disadvantages of this method. Dehkordi et al. [27] proposed a fully DSC to restore f & V, irrespective of parametric uncertainties. However, the paper still needs DGs parameters.

In this paper, a new adaptive fully DSC is proposed. The MG system is considered as an unknown nonlinear dynamic system. A Novel control scheme as a consensus problem is introduced and a Lyapunov proof is presented to assure the asymptotic stability of the control system. The upper bound of the consensus error is also explicitly derived. The proposed controller covers the uncertain dynamics and nonlinear nature of MGs and requires only active and reactive droop coefficients. The followings are the main novelty of the paper:

a. The proposed adaptive protocol is fully distributed and ensures MG asymptotic stability.
b. The goal of any controller design is to maintain system performance despite of inaccuracies and model changes. The proposed adaptive protocol is less dependent on system parameters and DGs than existing distributed adaptive control methods.
c. The proposed secondary controller is adaptive and robust and it can easily respond to severe uncertainties such as communication link failures and DG outages.

The reminder of this paper is structured as follows: section 2 presents the modeling approach. The proposed Algorithm is presented in section 3. In section 4, simulation results are discussed and finally, conclusions are summarized in sections 5.

2. MODELING APPROACH

2. 1. Preliminaries of Graph Theory A rooted-out branching (directed tree) topology is commonly assigned as a sparse communication network for microgrid systems. In this topology (Figure 1), DGs are considered as the nodes of the communication digraph and the edges denote the communication links. Each edge establishes a connection between a pair of nodes, and each node receives information, only from one node (except the root node). In fact, the information exchange among DGs is done on a directed graph $G = (\upsilon, \varepsilon, A)$ that consists of N nodes called $\upsilon = (1,2,\ldots,N)$, a set of links$\varepsilon$, and its associated adjacent matrix $A(N \times N)$. a_{ij} is the weight of edge (v_j, v_i). For $a_{ij}=1$, i^{th} node receives information from j^{th} node, otherwise $a_{ij}=0$. The Laplacian matrix L is assigned as $L_{ii}=\sum_{j=1}^{N}\alpha_{ij}$ and $L_{ij}=-a_{ij}$. The eigenvalues of L has one zero entry $(\lambda_1=0)$ with all other have positive real parts, i.e., $0 \le \lambda_2 \le \lambda_3 \le \ldots \le \lambda_N$ [25].

It is noteworthy that in the proposed control strategy, any type of directed graph (digraph) can be employed and the only necessary condition for the selected digraph is to have a spanning tree. If there is a directed path between the root node and any other node in a digraph, the digraph will actually have a spanning tree.

2. 2. Modeling of MGs Figure 2 shows an islanded MG depicting its power network, communication grid, and control layers. The power network interconnects the primary DC source to the voltage source converter (VSC) including power, voltage, and current control loops. VSC is connected to the network thorough an LCL filter. The primary control loops regulate the desired outputs of the inverter bridge. D–q reference frame is utilized to formulate the nonlinear dynamics of the system [30].

Frequency and voltage values must be readjusted after any disruption. The local droop method is adopted to balance the generation and consumption of active and reactive power. A relationship between the active power and frequency and the voltage amplitude and reactive power is assigned by the power controller block as follows:

$$\omega_i = \varpi_i - D_{Pi}P_i \quad (1)$$

$$v_{odi} = \Im_i - D_{Qi}Q_i \ , v_{oqi} = 0 \quad (2)$$

where P_i and Q_i are the measured real and reactive power at DG output, respectively. $\Im_i$ and ϖ_i are the set points and D_{Pi} and D_{Qi} are the droop coefficients.

A microgrid has a nonlinear nature and resembles a multiagent system. In distributed control structure, each

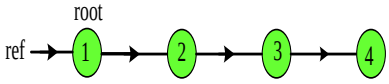


Figure 1. Topology of the communication graph

DG communicates with its neighbor units through a communication grid. The nonlinear dynamics of the i^{th} DG can be written as Equation (3) in state equations form:

$$\begin{cases} \dot{x}_i = f_i(x_i) + k_i(x_i)D_i + g_i(x_i)\,u_i \\ y_i = h_i(x_i) \qquad\qquad\qquad\qquad \text{i=0,...,N} \end{cases} \tag{3}$$

Details of DG and its components equations are given in literature [33]. D_i is considered as an uncertainty. $u_i = [\tau_i \quad \varpi_i]$ is the output of secondary controllers. They are assigned such that f & V of all DGs regulated to the reference values. f_i, k_i, g_i and h_i are nonlinear functions.

So far, two types of methods have been proposed to design an adaptive secondary control for a microgrid modeled with nonlinear Equation (3). In the first type, neural networks are used instead of nonlinear model. In order to train these networks, the microgrid information must be available. In the second type, input–output feedback linearization technique is utilized to transform the nonlinear dynamics into the linear form. This technique needs $f_i(x_i)$, $g_i(x_i)$ and their derivative to x_i for the design and implementation of the distributed secondary controller. Hence, they have high dependence on the system and DGs parameters.

In this paper, the modelling of the MG system is performed in the form of an unknown nonlinear dynamic system. However, having active and reactive droop coefficients of DGs is sufficient to ensure the synchronization process. The controller protocol only requires the measured output signals of the DGs and there is no need to modify the controller parameters if there exists inaccuracies or changes in the power network parameters. The set point for primary control is provided by secondary controller. The secondary control also restores the outputs of individual DG units to a reference value by the communication network that shares information among DGs.

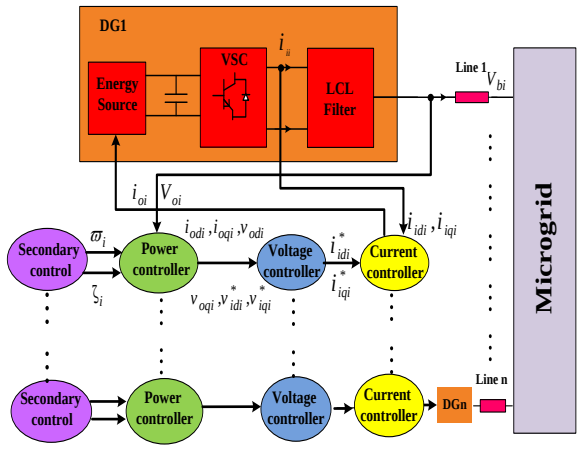


Figure 2. Schematic diagram of a MG including inverter-based DGs with their primary and secondary controllers

3. PROPOSED ALGORITHM

3. 1. Distributed Adaptive Voltage Controller (DAVC) This section aims to develop a novel consensus based DAVC for microgrids in islanded mode. DAVC is responsible for elimination of the voltage deviations caused by primary droop mechanism. The compact form of secondary voltage control equations can be written as Equation (2). Differentiating these equation yields:

$$\dot{v}_{odi} = \dot{\mathfrak{T}}_i - D_{Qi}\dot{Q}_i \equiv Z_{vi} \tag{4}$$

DAVC should guarantee that the DG output voltage magnitudes all synchronises with V_{ref}. The proposed adaptive control law is as follows:

$$\begin{aligned} Z_{vi} &= -(d_{vi} + \rho_{vi})\xi_{vi} \\ \dot{d}_{vi} &= \xi_{vi}^T\xi_{vi} \;\;,\; \rho_{vi} = \xi_{vi}^T\xi_{vi} \end{aligned} \tag{5}$$

where ξ_{vi} is the local neighborhood consensus error:

$$\xi_{vi} = \sum_{j=1}^{N} \tilde{L}_{ij}(v_{odi} - v_{odj}) + G_{ij}(v_{odi} - v_{ref}) = \sum_{j=1}^{N} \tilde{L}_{ij}(v_{odi} - v_{ref}) \tag{6}$$

G is a diagonal matrix and called the pinning matrix. The pinning gain (the diagonal element of the pinning matrix) is nonzero only for the nodes that are directly connected to the root (leader) node. The nodes for which the pinning gain is nonzero are referred to as the pinned or controlled nodes.

In matrix form, Equation (6) may be given as:

$$\xi_v = (L + G)\varepsilon_v = \tilde{L}\varepsilon_v \tag{7}$$

It is easy to prove that [35]:

$$\|\varepsilon_v\| \le \frac{\|\xi_v\|}{\sigma_{\min}(\tilde{L})} \tag{8}$$

$\sigma_{\min}(\tilde{L})$ depicts the minimum singular value of matrix $\tilde{L}$. In multi-agent systems the main problem is the definition of a consensus protocol. The protocol describes the rules by witch each agent interacts with its neighboring agents so that all agents can reach the desired state.

Figure 3 shows a rooted-tree graph with 6 nodes used in this paper. Each node (except the root node) takes information from its neighbor and sends it to the next node. The numbers shown on the links are the contribution of the state of each node to the next node.

As the root node is the only pinned node in Figure 3, all elements of G are zero except the first element which is one. Therefore, $\tilde{L}$ is as follows:

$$\tilde{L} = \begin{bmatrix} 1 & 0 & 0 & 0 & \cdots & 0 & 0 \\ -1 & 1 & 0 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & -1 & 1 & 0 \\ 0 & 0 & 0 & \cdots & 0 & -1 & 1 \end{bmatrix} \quad (9)$$

From Equation (6), we have:

$$\dot{\xi}_{vi} = \tilde{L}\dot{\varepsilon}_{vi} = \tilde{L}\dot{v}_{odi} = \tilde{L}Z_{vi} \quad (10)$$

Let the Lyapunov function candidate is [36]:

$$V = \sum_{i=1}^{N} \frac{1}{2} m_i (2d_{vi} + \rho_{vi})\rho_{vi} + \sum_{i=1}^{N} \frac{1}{2} m_i (d_{vi} - \beta)^2 \quad (11)$$

where $m_i \geq 0$ are the diagonal elements of a diagonal matrix M. The derivative of Lyapunov function is:

$$\dot{V} = \sum_{i=1}^{N} (m_i \dot{d}_{vi} \rho_{vi} + m_i \dot{\rho}_{vi} d_{vi} + m_i \dot{\rho}_{vi} \rho_{vi}) + \sum_{i=1}^{N} m_i (d_{vi} - \beta)\dot{d}_{vi} = \sum_{i=1}^{N} m_i (d_{vi} + \rho_{vi})\dot{\rho}_{vi} + m_i (d_{vi} + \rho_{vi} - \beta)\dot{d}_{vi} \quad (12)$$

The matrix form of Equation (12) can be written as:

$$\dot{V} = (D + \rho)M\dot{\rho} + (D + \rho - \beta)M\dot{D} \quad (13)$$

By substituting Equation (5) into Equation (13), we have:

$$\dot{V} = 2(D + \rho)M\xi_v^T \dot{\xi}_v + (D + \rho - \beta)M\xi_v^T \xi_v = 2\xi_v^T (D + \rho)M\dot{\xi}_v + \xi_v^T (D + \rho - \beta)M\xi_v \quad (14)$$

Placing $\dot{\xi}$ from Equation (10) into Equation (14) yields: $\dot{V} = 2\xi_v^T[-(D + \rho)^2 M\tilde{L}]\xi_v + \xi_v^T[(D + \rho - \beta)M]\xi_v$ (15) $M\tilde{L}$ can be written as $\frac{1}{2}\left(M\tilde{L} + \tilde{L}^T M\right)$[37, 38]. Then:

$$\dot{V} = \xi_v^T[-(D + \rho)^2 (M\tilde{L} + \tilde{L}^T M) + (D + \rho - \beta)M]\xi_v \quad \dot{V} \leq \xi_v^T[-\lambda_0 (D + \rho)^2 + (D + \rho)M - \beta M]\xi_v \quad (16)$$

where λ_0 is the minimum eigenvalue of $\left(M\tilde{L} + \tilde{L}^T M\right)$.

Lemma 1 [39]: If a and b are nonnegative real number and p and q are positive numbers such that $\frac{1}{P} + \frac{1}{q} = 1$, then $ab \leq \frac{a^P}{p} + \frac{b^q}{q}$.

By applying Lemma 1 and assumed that $a = \sqrt{\lambda_0}(D + \rho),\ b = \sqrt{\beta M}\ , p = q = 2$, Equation (16) can be written as:

$$\dot{V} \leq \xi_v^T[-2\sqrt{\lambda_0 \beta M}(D + \rho) + (D + \rho)M]\xi_v \quad \dot{V} \leq \xi_v^T[(D + \rho)(M - 2\sqrt{\lambda_0 \beta M})]\xi_v \quad (17)$$

$$\beta \geq \frac{\max\{m_1, \ldots, m_N\}}{\lambda_0}$$

Hence, if then $\dot{V} \leq 0$ and the proof is completed. Indeed, according to Equation (10), β represents steady state error and qualifies the ultimate boundedness of consensus error.

Figure 4 shows the DVAC block diagram. The controller output τ_i could be written as:

$$\tau_i = \int (Z_{vi} + D_{Qi}\dot{Q}_i)dt \quad (18)$$

3. 2. Distributed Adaptive Frequency Controller (DAFC) DFAC should assign ϖ_i in Equation (1) for synchronization of all DGs frequency. In fact, this controller simulates the governor and its set point mechanism from a synchronous generator. By differentiating of Equation (1), we have:

$$\dot{\omega}_i = \ddot{\varpi}_i - D_{Pi}\dot{P}_i \equiv Z_{\omega i} \quad (19)$$

where $Z_{\omega i}$ is a virtual signal. As the frequency is a global quantity among a MG, the DAFC can be proposed in such a way that in addition to frequency synchronization, the sharing of DGs output real powers are divided according to their nominal (rated) powers. It means that:

$$\frac{P_j}{P_i} = \frac{D_{Pi}}{D_{Pj}}, \quad \forall i,j \in N \quad (20)$$

Similar to the previous section, the control law is introduced as follows:

$$Z_{\omega i} = -(d_{\omega i} + \rho_{\omega i})\xi_{\omega i} \quad (21)$$

where $\xi_{\omega i}$ is the frequency consensus error:

$$\xi_{\omega i} = \sum_{j=1}^{N} L_{ij}(\omega_i - \omega_j) + G_{ij}(\omega_i - \omega_{ref}) \quad (22)$$

ϖ_i is written as:

$$\varpi_i = \int (Z_{\omega i} + D_{Pi}\dot{P}_i)dt \quad (23)$$

For appropriate active power sharing, the following signal Z_{Pi} is introduced:

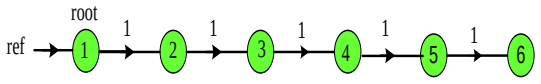


Figure 3. Topology of the communication grid with six nodes

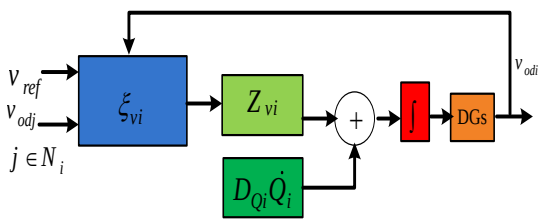


Figure 4. Schematic of the proposed DAVC

$$Z_{Pi} = -(d_{Pi} + \rho_{Pi})\xi_{Pi} \quad (24)$$

where ξ_{Pi} is the real power neighbors tracking error:

$$\xi_{Pi} = \sum_{j=1}^{N} a_{ij}(D_{Pi}P_i - D_{Pj}P_j) \quad (25)$$

Now, Equations (21) and (24) can be combined to design the ϖ_i as the controller output.

$$\varpi_i = \int (Z_{\omega i} + Z_{Pi})dt \quad (26)$$

Figure 5 shows the DAFC block diagram.

4. SIMULATION RESULTS

An typical MG is used for demonstration the effectiveness of the proposed control strategy. For this purpose, A 50 Hz, 380 V, islanded microgrid with six DGs, five loads and several lines is considered (Figure 6). The communication topology is depicted in Figure 3. DG1 (the reference DG) is the only agent which accesses to the f & V reference values. The specifications of DGs, RL loads, transmission lines, and the control system are summarized in Table 1. MATLAB/SimPower software environment is used to test all the simulation scenarios.

It should be note that although the proposed distributed control method is analyzed by a 6-node microgrid, the design procedure is modular and scalable and can be implemented in larger microgrid with more DGs.

The results of the proposed control strategy are also compared with the data reported in literature [28], which is one of the main and last activities accomplished in this field. By doing this, better capabilities of the proposed method are represented in compare to the conventional distributed methods. Four different cases, with different degree of uncertainty and disturbance levels are used for evaluation the behavior of the proposed controllers.

4. 1. Case 1: Load Changes In this section, the behavior of the proposed method is investigated in response to the load change and compared with the method of Bidram et al. [28]. The following is the simulation scenario:

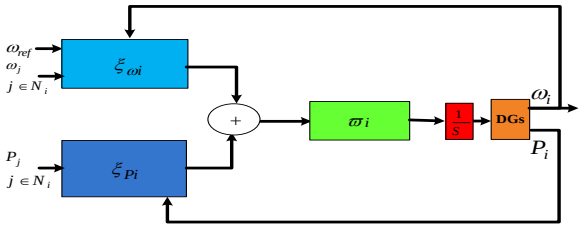


Figure 5. Schematic of the proposed DAFC

TABLE 1. Parameters of the test MG

		DG #1&2&5		DG #3&4&6
DGs	DP	1.06×10^{-4}	DP	1.06×10^{-4}
	DQ	5.02×10^{-3}	DQ	5.02×10^{-3}
	Z_c	0.015+j0.65Ω	Z_c	0.03+j0.65Ω
	L_{f1},L_{f2}	1.35, 0.27mH	L_{f1},L_{f2}	1.35, 0.27mH
	R_{f1},R_{f2}	0.1, 0.05Ω	R_{f1},R_{f2}	0.1, 0.05Ω
	C_f	47μF	C_f	47μF
	K_{PV}	0.1	K_{PV}	0.05
	K_{IV}	420	K_{IV}	390
	K_{PC}	15	K_{PC}	10.5
	K_{IC}	20000	K_{IC}	16000
Lines	Z_{line1}	0.12+j0.1Ω	Z_{line5}	0.175+j0.58Ω
	Z_{line2}	0.175+j0.58Ω	Z_{line6}	0.12+j0.1Ω
	Z_{line3}	0.12+j0.1Ω	Z_{line7}	0.175+j0.58Ω
	Z_{line4}	0.12+j0.1Ω	Z_{line8}	0.175+j0.58Ω
RL Loads	Load #1	P=13 kW, Q=7.5 kVar	Load #3	P=7 kW, Q=7 kVAR
	Load #2	P=13 kW, Q=7.5 kVar	Load #4	P=6 kW, Q=6 kVAR
	Load #5	P=14 kW, Q=6 kVar		

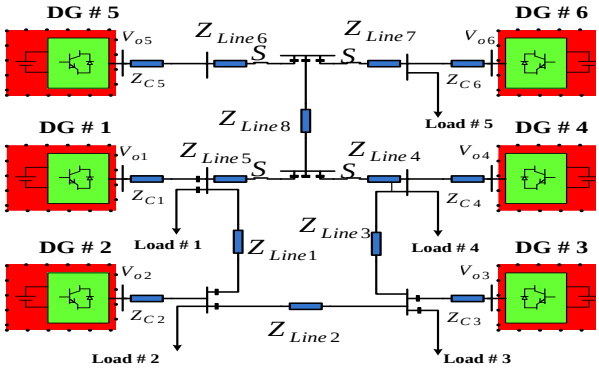


Figure 6. Islanded test MG

1) t = 0 s is the simulation starting time. The secondary controllers are off and only the primary control is working.

2) At t = 0.5 s, the secondary controllers are applied. It should be noted that in practical applications, the secondary voltage control should be applied immediately after the disturbance occurs. However, in this paper, the secondary controller is intentionally delayed by 0.5 s. to highlight its effectiveness.

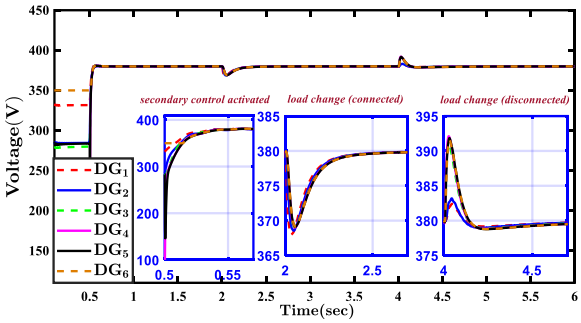
3) At t = 2 s, a load with P= 6 kW and Q = 6 kVar is connected to the bus1 (parallel to the existing load).

4) At t = 4 s, loads 3 and 4 are disconnected from the MG.

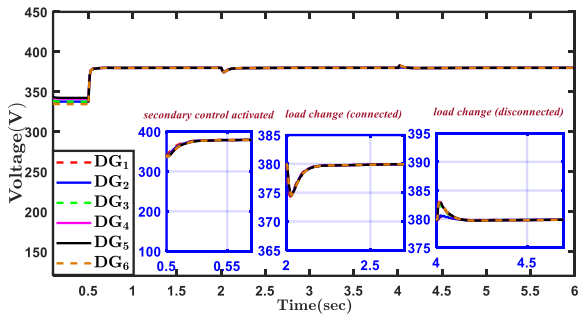
Figures 7a and (8a) show the voltage (frequency) of DGs deduced by the method of Bidram et al. [28] and Figures 7b and (8b) represent the voltage (frequency) of DGs obtained by the proposed method. As it can be seen in these figures, at the beginning and once the primary

control is applied, DGs operating f & V all go to a common values that deviate from the nominal values. After the secondary controllers are applied at t = 0.5 s, a short transient occurs and both f & V quickly return to their nominal values.Then, load changes occur at t=2 s and t=4 s. Transient periods are zoomed to show details that are not shown in the original figures. The zoomed regions show that both overshoot and settling time resulting from the proposed method are less than those parameters obtained from the method [28].

Figures 9a and 9b represent the output real power ratio of the six DGs using Bidram et al. [28] method and the proposed strategy, respectively. As shown, the proposed strategy presents appropriate real power sharing among all DGs.

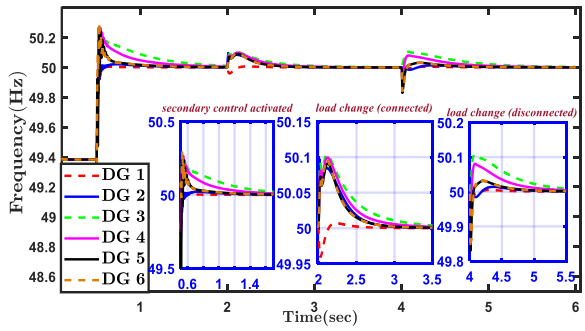


(a) method of Bidram et al. [28]

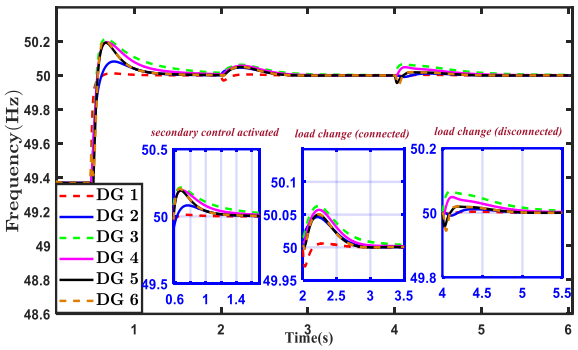


(b) the proposed method

Figure 7. DGs output voltage magnitudes

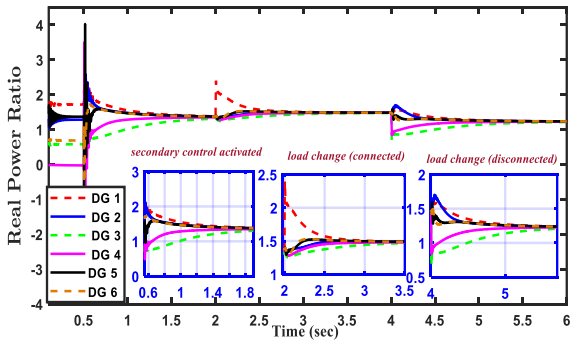


(a) method of Bidram et al. [28]

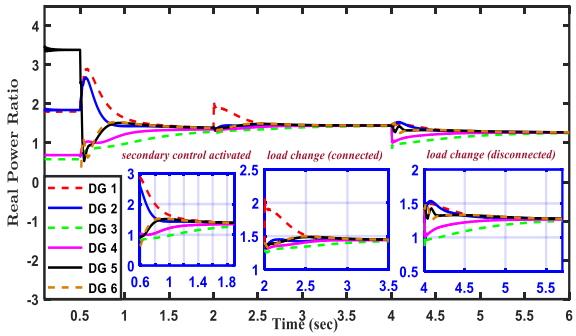


(b) the proposed method

Figure 8. DGs output frequency



(a) method of Bidram et al. [28]



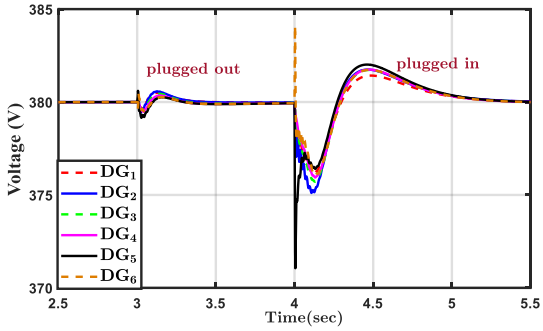
(b) the proposed method

Figure 9. DGs output real power ratio

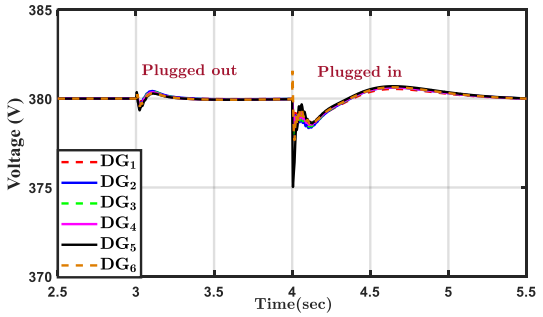
4. 2. Case 2: Plug and Play Capability In this case, a higher disturbance level is assumed and the plug-and-play capability of the proposed controllers is investigated. For this purpose, at t = 3 s, DG 6 is unplugged from the MG and is plugged back in at t = 4 s. Although this DG is instantly turned off, the power measurements exponential decay to zero because of the existing low-pass filters. The control parameters are the same as in case 1. The results are displayed in Figures 10(a), 11(a) and 12(a) for method of Bidram et al. [28] and in Figures 10(b), 11(b) and 12(b) for the proposed method. As it is seen, the proposed controller responds well to the outage and reconnection of the DG unit, which

can be considered as a large disturbance. Also, the proposed controller maintains accurate proportional power sharing and frequency and voltage regulation before, during, and after the plug-and-play event with much less transients than those of Bidram et al. [28]. After unplugging of DG6 at t=3 s, the other units share the excess power among themselves in proportion to their power ratings. It should be noted that although DG 6 is disconnected from Bus 6, the bus voltage is still available.

4. 3. Case 3: Failure of Communication Links In this case, resiliency to a single link failure is investigated. It is assumed that at t = 3 s, the comunication link between DG1 and DG2 is deliberately disconnected and reconnected after 1 m.s.

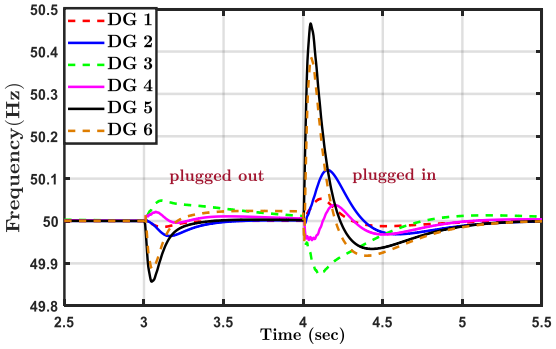


(a) method of Bidram et al. [28]

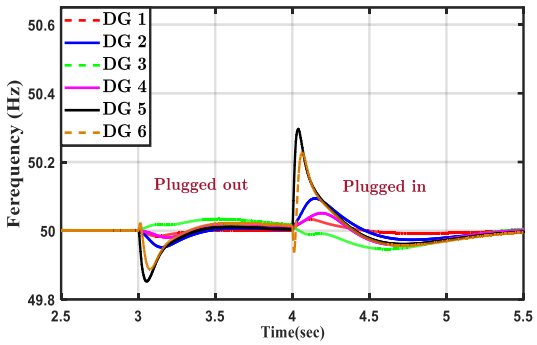


(b) the proposed method

Figure10. DGs output voltage magnitudes

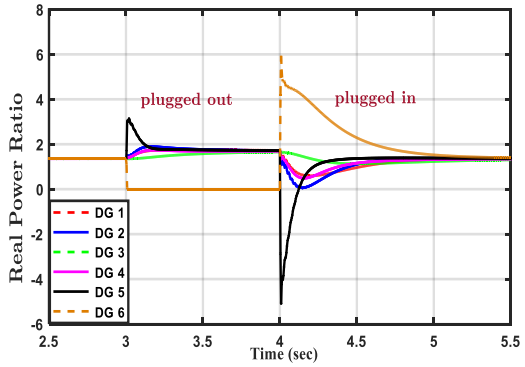


(a) method of Bidram et al. [28]

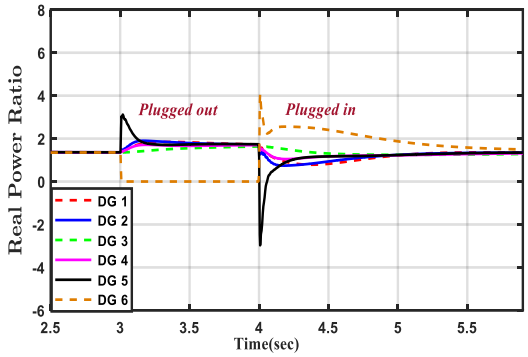


(b) the proposed method

Figure 11. DGs output frequency



(a) method of Bidram et al. [28]



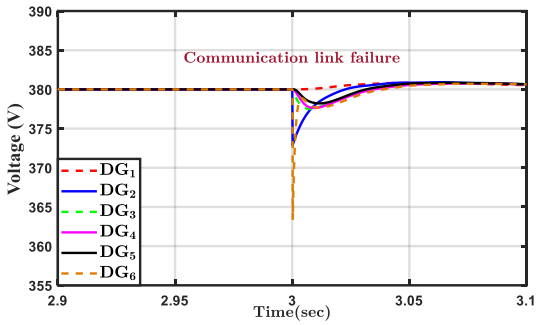
(b) the proposed method

Figure 12. DGs output real power ratio

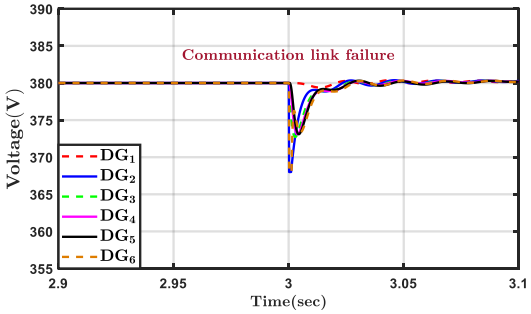
Immediately after reconnection of this link, the link between DG 5 and DG6 is disconnected for 1 m.s and then reconnected. The results are given in Figures 13(a) and 14(a) for the method of Bidram et al. [28] and 13(b) and 14(b) for the proposed method. The results depicts that under such conditions, the islanded MG is still stable and f & V restores to their nominal values, but the proposed method has much better transient response. For example, for the proposed method, voltage magnitude achieves a steady state after 10 m.s. while for the method of Bidram et al. [28], this time is 25 msec. Moreover, the

rate of frequency change obtained from [28] is 2.2 Hz/m. s while in the proposed method this parameter is equal 0.3 Hz/m.s As the figures shown, the worst behavior belongs to DG6, because this DG is the farthest one from the leader node. Remarkably, DGs and their control mechanism typically have much slower dynamics than communication systems, which commonly use low-delay, wide-bandwidth communication protocols. This is especially true for distributed control methods where each controller only communicates with its neighbors. Because of that, we do not discuss about delays in this paper.

4. 4. Case 4: DG Parameters Change In this case, the performance of the proposed ADVC is verified subsequent to the changes in network parameters. It is assumed that at t = 1 s, the filters capacitors are randomly changed from 49 to 51 μF. Figure 15 shows the results of the output voltages of DGs over time under this condition. As the figure shows, the DGs voltages remain in the permissible range and the system tolerates parameter changes. In fact, these changes are the events that occur in real situations due to the tolerances, temperature, aging and etc. The result shows that performance of the adaptive voltage controller does not deteriorate as a result of the changes in filter capacitors.

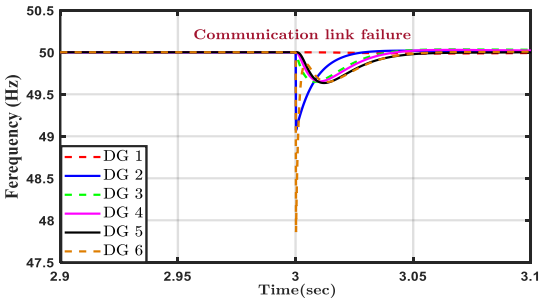


(a) method of Bidram et al. [28]

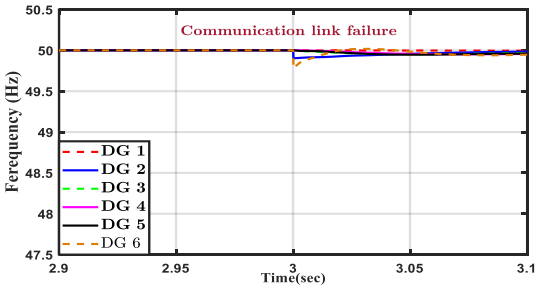


(b) the proposed method

Figure 13. DGs output voltage magnitudes in case 3



(a) method of Bidram et al. [28]



(b) the proposed method

Figure 14. DGs output frequency in case 3

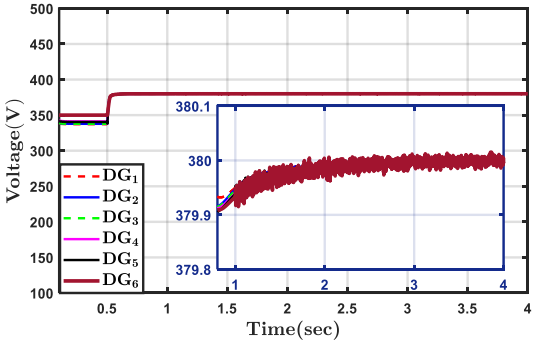


Figure 15. DGs output voltage magnitudes in case 4- The proposed protocol

5. CONCLUSION AND FUTURE WORKS

Based on an adaptive fully distributed algorithm, two secondary controllers were presented in this work to improve the behavior of MGs during islanded operation. The nonlinear nature of MGs has been covered in the algorithm. In spite of other methods that require complete DGs information, in this method only the power droop coefficient is needed. A consensus-based strategy was developed to returns and to synchronize f & V of MG to their referenced values after any disruption. The stabilization of the synchronization problem was proved using a rigorous Lyapunov analysis. The effectiveness of the proposed method was examined using a typical six-DGs network. The obtaining results were compared to method introduced by Bidram and his coworkers to validate the method. The results show the efficacy of the proposed secondary controllers which are less sensitive

to intense disturbances and presents low overshoot and settling time during load changes, plug and play and communication links failure. Indeed, the proposed controllers enhance the resiliency of MGs. The proposed distributed adaptive control is very simple in terms of structure and unlike centralization methods, does not require a large and complex communication network. Also, it can be easily implemented and used in software.

Limitations and future applications: Communication is indispensable to access neighbor data and, thus, to the operation of the distributed control system. Accordingly, channel non-idealities such as communication transmission/propagation delay/noises and limited bandwidth, may compromise the overall system performance. A few aspects listed below attract the authors' interest to do research on them in future works:

- A detailed discussion about the voltage drop as a result of a communication link failure.
- The proposed secondary control scheme can be extended to address the communication network related issues, such as data loss, packet jamming, link failure, cyber attack, etc.

6. REFERENCES

1. Keshavarz, M., Doroudi, A., Kazemi, M. H., and Dehkordi, N, M., "A Novel Hybrid Droop-Isochronous Control Strategy for Microgrid Management", ***Iranian Journal of Electrical and Electronic Engineering***, Vol. 17, No. 2, (2021), 1798-1798, Doi: 10.22068/IJEEE.17.2.1798.
2. Bidram, A., Poudel, B., Damodaran, L., Fierro, R., and Guerrero, J. M., "Resilient and cybersecure distributed control of inverter-based islanded microgrids", ***IEEE Transactions on Industrial Informatics***, Vol. 16, No. 6, (2020) 3881-3894, Doi: 10.1109/TII.2019.2941748.
3. Lou, G., Gu, W., Sheng, W., Song, X., and Gao, F., "Distributed model predictive secondary voltage control of islanded microgrids with feedback linearization", ***IEEE Access***, Vol. 6, (2018), 50169-50178, Doi: 10.1109/ACCESS.2018.2869280.
4. Dahiya, R., "Stability analysis of islanded DC microgrid for the proposed distributed control strategy with constant power loads", ***Computers & Electrical Engineering***, Vol. 70, (2018), 151-162, Doi: 10.1016/j.compeleceng.2018.02.020**.**
5. Lavanya, V., & Senthil Kumar, N. "Seamless Transition in Grid-connected Microgrid System using Proportional Resonant Controller", ***International Journal of Engineering, Transactions A: Basics***, Vol. 33, No. 10, (2020), 1951-1958. Doi: 10.5829/IJE.2020.33.10A.13.
6. Farzinfar, M., Nair, N. K. C., & Bahadornejad, M. "A New Adaptive Load-Shedding and Restoration Strategy for Autonomous Operation of Microgrids: A Real-Time Study", ***International Journal of Engineering***, ***Transactions A: Basics***, Vol. 33, No. 1, (2020), 82-91. Doi: 10.5829/IJE.2020.33.01A.10.
7. Han, Y., Zhang, K., Li, H., Coelho, E. A. A., and Guerrero, J. M. "MAS-based distributed coordinated control and optimization in microgrid and microgrid clusters: A comprehensive overview", ***IEEE Transactions on Power Electronics***, Vol. 33, No. 8, (2017), 6488-6508, Doi: 10.1109/TPEL.2017.2761438.
8. Ko, B. S., Lee, G. Y., Choi, K. Y., Kim, R. Y., Kim, S., Cho, J., and Kim, S. I. , "Flexible Control Structure for Enhancement of Scalability in DC Microgrids", ***IEEE Systems Journal***, Vol. 14, No. 3, (2020), 4591-4601, Doi: 10.1109/JSYST.2019.2963707.
9. Sharma, R., and Sathans, S., "Virtual impedance based phase locked loop for control of parallel inverters connected to islanded microgrid", ***Computers & Electrical Engineering***, Vol. 73, (2019), 58-70, Doi: 10.1016/j.compeleceng.2018.11.005.
10. Muhammad, M., Chul-Hwan, K., Muhammad, S., "Robust Centralized Control for DC Islanded Microgrid Considering Communication Network Delay", ***IEEE Access***, Vol. 8, (2020), 77765-77778, Doi: 10.1109/ACCESS.2020.2989777.
11. Qian, T., Liu, Y., Zhang, W., Tang, W., Shahidehpour, M., "Event-triggered updating method in centralized and distributed secondary controls for islanded microgrid restoration", ***IEEE Transactions on Smart Grid***, Vol. 11, No. 2, (2019), 1387-1395, Dio.10.1109/TSG.2019.2937366.
12. Amiri, H., Markadeh, G, A., Dehkordi, N, M., and Blaabjerg, F., "decentalized robust backstepping voltage control of photovoltaic systems for DC islanded microgrid based on disturbance observer method", ***ISA Transactions***, Vol .101, (2020), 471-481, Doi: 10.1016/j.isatra.2020.02.006.
13. Arani, A. A. K., Gharehpetian, G. B., Abedi, M., "Decentralized primary and secondary control strategies for islanded microgrids considering energy storage systems characteristics", ***IET Generation, Transmission & Distribution***, Vol. 13, No. 14, (2019), 2986-2992, 2019, Doi: 10.1049/iet-gtd.2019.0362.
14. Dehkordi, N. M., Moussavi, S. Z. "Distributed resilient adaptive control of islanded microgrids under sensor/actuator faults", ***IEEE Transactions on Smart Grid***, Vol. 11, No. 3, (2019), 2699-2708, Doi: 10.1109/TSG.2019.2960205.
15. Peng, J., Fan, B., and Liu, W., "Voltage-Based Distributed Optimal Control for Generation Cost Minimization and Bounded Bus Voltage Regulation in DC Microgrids", ***IEEE Transactions on Smart Grid***, Vol. 12, No. 1, (2020), 106-116, Doi: 10.1109/TSG.2020.3013303.
16. Espina, E., Llanos, J., Burgos-Mellado, C., Cardenas-Dobson, R., Martinez-Gomez, M., and Saez, D., "Distributed Control Strategies for Microgrids: An Overview", ***IEEE Access***. Vol. 8, (2020), 193412-193448, Doi: 10.1109/ACCESS.2020.3032378.
17. Li, Z., Zang, C., Zeng, P., Yu, H., and Li, S., "Fully distributed hierarchical control of parallel grid-supporting inverters in islanded AC microgrids", ***IEEE Transactions on Industrial Informatics***, Vol. 14, No. 2, (2018), 679-690, Doi: 10.1109/TII.2017.2749424.
18. Zhang, Z., Dou, C., Yue, D., Zhang, B., Xu, S., Hayat, T., and Alsaedi, A., "An event-triggered secondary control strategy with network delay in islanded microgrids", ***IEEE Systems Journal***, Vol.13, No. 2, (2018), 1851-1860.
19. Zuo, S., Altun, T., Lewis, F. L., and Davoudi, A. "Distributed resilient secondary control of DC microgrids against unbounded attacks", ***IEEE Transactions on Smart Grid***, Vol. 11, No. 5, (2020), 3850-3859, Doi: 10.1109/TSG.2020.2992118.
20. Mathew, P., Madichetty, S., Mishra, S., "A multilevel distributed hybrid control scheme for islanded DC microgrids", ***IEEE Systems Journal***, Vol. 13, No. 4, (2019), 4200-4207, Doi: 10.1109/JSYST.2019.2896927.
21. Li, Q., Peng, C., Wang, M., Chen, M., Guerrero, J. M., and Abbott, D., "Distributed secondary control and management of islanded microgrids via dynamic weights", ***IEEE Transactions on Smart Grid***, Vol. 10, No. 2, (2018), 2196-2207. Doi: 10.1109/TSG.2018.2791398.
22. Amoateng, D. O., Al Hosani, M., Elmoursi,M. S., Turitsyn, K., and Kirtley, J. L., "Adaptive voltage and frequency control of islanded multi-microgrids", ***IEEE Transactions on Power Systems***, Vol. 33, No. 4, (2017), 4454-4465, Doi: 10.1109/TPWRS.2017.2780986.

23. Wu, X., Shen, C., and Iravani, R., "A distributed, cooperative frequency and voltage control for microgrids", ***IEEE Transactions on Smart Grid***, Vol. 9, No. 4, (2018), 2764-2776, Doi: 10.1109/TSG.2016.2619486.
24. Li, Z., Zang, C., Zeng, P., Yu, H., and Li, S. "Fully distributed hierarchical control of parallel grid-supporting inverters in islanded AC microgrids", ***IEEE Transactions on Industrial Informatics***, Vol. 14, No. 2, 679-690, Doi: 10.1109/TII.2017.2749424.
25. Dehkordi, N. M., Sadati, N., and Hamzeh, M., "Distributed robust finite-time secondary voltage and frequency control of islanded microgrids", ***IEEE Transactions on Power Systems***, Vol. 32, No. 5, (2017), 3648-3659, Doi: 10.1109/TPWRS.2016.2634085.
26. Jin, Z., Meng, L., Guerrero, J. M., and Han, R., "Hierarchical control design for a shipboard power system with DC distribution and energy storage aboard future more-electric ships", ***IEEE Transactions on Industrial Informatics***, Vol. 14, No. 2, (2017), 703-714, Doi: 10.1109/TII.2017.2772343.
27. Dehkordi, N.M., Sadati, N., and Hamzeh, M., "Fully distributed cooperative secondary frequency and voltage control of islanded microgrids", ***IEEE Transactions on Energy Conversion***, Vol.32, No. 2, (2016), 675-685, Doi: 10.1109/TEC.2016.2638858,
28. Bidram, A., Davoudi, A., and Lewis, F. L., "A multiobjective distributed control framework for islanded AC microgrids", ***IEEE Transactions on Industrial Informatics***, Vol. 10, No. 3, (2014), 1785-1798, Doi: 10.1109/TII.2014.2326917.
29. Shahab, M. A., Mozafari, B., Soleymani, S., Dehkordi, N. M., Mohammadnezhad Shourkaei, H., and Guerrero, J. M., "Stochastic Consensus-Based Control of μGs With Communication Delays and Noises", ***IEEE Transactions on Power Systems***, Vol. 34, No. 5, (2019), 3573-3581, Doi: 10.1109/TPWRS.2019.2905433.
30. Bidram, A., Davoudi, A., Lewis, F. L., and Guerrero, J. M., "Distributed cooperative secondary control of microgrids using feedback linearization", ***IEEE Transactions on Power Systems***, Vol. 28, No. 3, (2013), 3462-3470, Doi: 10.1109/TPWRS.2013.2247071.
31. Olivares, D. E, Mehrizi-Sani, A., Etemadi, A. H., Cañizares, C. A., Iravani, R., Kazerani, M., Hajimiragha, A. H., Gomis-Bellmunt, O., Saeedifard, M., and Palma-Behnke, R., "Trends in microgrid control", ***IEEE Transactions on Smart Grid***, Vol. 5, No. 4, (2014), 1905-1919, Doi: 10.1109/TSG.2013.2295514.
32. Dehkordi, N. M., Baghaee,H. R., Sadati, N., and Guerrero, J. M., "Distributed noise-resilient secondary voltage and frequency control for islanded microgrids", ***IEEE Transactions on Smart Grid***, Vol. 10, No. 4, (2018). 3780-3790, Doi: 10.1109/TSG.2018.2834951.
33. Bidram, A., Davoudi, A., Lewis, F. L., and Ge, S. S., "Distributed adaptive voltage control of inverter-based microgrids", ***IEEE Transactions on Energy Conversion***, Vol. 29, No. 4, (2014). 862-872, Doi: 10.1109/TEC.2014.2359934.
34. Afshari, A., Karrari, M., Baghaee, H. R., and Gharehpetian, G. B, "Distributed fault-tolerant voltage/frequency synchronization in autonomous AC microgrids", ***IEEE Transactions on Power Systems***, Vol. 35. No. 5, (2020), 3774-3789, Doi: 10.1109/TPWRS.2020.2975115.
35. Qu, Z., "Cooperative control of dynamical systems: applications to autonomous vehicles", Springer Science & Business Media, (2009).
36. Lv, Y., Li, Z., and Duan, Z., "Distributed adaptive consensus protocols for multiple Lur'e systems over directed graphs", ***IET Control Theory & Applications***, Vol. 10, No. 4, (2016), 443-450, Doi: 10.1049/iet-cta.2015.0682.
37. Li, Z., Wen, G., Duan, Z., and Ren, W., "Designing fully distributed consensus protocols for linear multi-agent systems with directed graphs", ***IEEE Transactions on Automatic Control***, Vol. 60, No. 4, (2014), 1152-57, Doi; 10.1109/TAC.2014.2350391.
38. Zhang, H., Lewis, F. L., and Qu, Z., "Lyapunov, adaptive, and optimal design techniques for cooperative systems on directed communication graphs", ***IEEE Transactions on Industrial Electronics***, Vol. 59, No. 7, (2011), 3026-6041, Doi: 10.1109/TIE.2011.2160140.
39. Bernstein, D. S., "Matrix mathematics: theory, facts, and formulas ", Princeton university press, (2009).

Persian Abstract

چکیده

این مقاله یک کنترل‌کننده ثانویه تطبیقی توزیع‌شده جدید را برای میکروگریدها (MGs) در حالت عملکرد جزیره‌ای پیشنهاد می‌کند. کنترل‌کننده پیشنهادی برای بهبود رفتار دینامیکی یک میکروگرید تحت اختلالات، از ساختار کنترل تطبیقی مبتنی بر اجماع استفاده می‌کند. یک پروتکل اجماع جدید برای بازگرداندن فرکانس و ولتاژ (f & V) میکروگرید به مقادیر نامی آن‌ها و یک تابع لیاپانوف برای اطمینان از پایداری مجانبی سیستم و حد نهایی خطای همسایگی ارائه شده است. ماهیت غیرخطی میکروگرید نیز در الگوریتم پوشش داده شده است. کنترل‌کننده پیشنهادی بر خلاف روش‌های دیگر که به اطلاعات کامل ژنراتورهای توزیع‌شده (DG) نیاز دارند، فقط به اطلاعات ضرایب افت توان نیاز دارد و مستقل از پارامترهای ژنراتورهای توزیع‌شده است. شبیه‌سازی‌های مختلفی برای بررسی عملکرد کنترل‌کننده تطبیقی، در جعبه ابزار MATLAB/SimPower روی یک میکروگرید نمونه و تحت اختلالات مختلف انجام شده است. نتایج شبیه‌سازی، کارایی کنترلر پیشنهادی را در افزایش تاب آوری میکروگریدها نشان می‌دهد.

International Journal of Engineering

Journal Homepage: www.ije.ir

Prediction of Drug-target Protein Interaction Based on the Minimization of Weighted Nuclear Norm and Similarity Graph between Drugs and Target Proteins

A. Ghanbari Sorkhi*[a], S. M. R. Hashemi[b], H. Yarmohammadi[c], M. Iranpour Mobarakeh[d]

[a] *Faculty of Electrical and Computer Engineering, University of Science and Technology of Mazandaran, Behshahr, Iran*
[b] *Young Researchers and Elite Clu, Qazvin Branch, Islamic Azad University, Qazvin, Iran*
[c] *Faculty of Computer Engineering, Shahrood university of technology, Shahrood, Iran*
[d] *Computer engineering and It department, Payam Noor University, Tehran, Iran*

PAPER INFO

Paper history:
Received 11 February 2021
Received in revised form 1 March 2021
Accepted 13 March 2021

Keywords:
Drug-target Interactions
Drug Discovery Process
Computational Prediction
Weighted Nuclear Norm Minimization
Similarity Graph
Low-rank Matrix

ABSTRACT

Identification of drug-target protein interaction plays an important role in the drug discovery process. Given the fact that prediction experiments are time-consuming, tedious, and very costly, the computational prediction could be a proper solution for decreasing search space for evaluation of the interaction between drug and target. In this paper, a novel approach based on the known drug-target interactions based on similarity graphs is proposed. It was shown that use of this method was a low-ranking issue and WNNM (weighted nuclear norm minimization) method was applied to detect the drug-target interactions. In the proposed method, the interaction between the drug and the target is encoded by graphs. Also known drug-target interaction, drug-drug similarity, target-target and combination of similarities were used as input. The proposed method was performed on four benchmark datasets, including enzymes (Es), ion channels (IC), G protein-coupled receptors (GPCRs), and nuclear receptors (NRs) based on the AUC and AUPR criteria. Finally, the results showed the improved performance of the proposed method.

doi: *10.5829/ije.2021.34.07a.18*

1. INTRODUCTION

The evaluation of the drug-target interactions (DTIs) has attracted the attention of researchers in the field of pharmaceutical science, recently [1]. Accordingly, extensive efforts have been dedicated to the assessment of drug repositioning to discover the interaction between new targets and the existing drugs. In fact, DTI is defined as detection factor between the target and drug interaction that leads to changes in the drug's behavior/use. On the other hand, the identification of these interactions will minimize the adverse side effects of drugs [2]. Wet-lab experiments to recognize these potential interactions are cost and time consuming. Therefore, computational prediction (CP) methods have been used in recent years [3]. In general, CP methods can be divided into three categories of ligand-based [4], docking approaches [5], and chemogenomic approaches [6].

However, many chemogenomic approaches have been attracted attention of many researchers lately. These methods can be extensively used on accessible biological data [7]. In fact, these methods use data that includes process information simultaneously to predict. Here, information about processes means the diagram of the chemical structure and genomic sequence for drugs and targets. This general technique is divided into two categories of feature-based and similarity-based methods. Supervised machine learning methods are exploited in the feature-based technique. In fact, the methods include feature vectors of sets of drug-target pairs along with class labels that show the presence of interaction (positive instances) and absence of interaction (negative instances) [8].

*Corresponding author email: ali.ghanbari@mazust.ac.ir (A. Ghanbari Sorkhi)

Please cite this article as: A. Ghanbari Sorkhi, S. M. R. Hashemi, H. Yarmohammadi, M. Iranpour Mobarakeh, Prediction of Drug-target Protein Interaction Based on the Minimization of Weighted Nuclear Norm and Similarity Graph between Drugs and Target Proteins, International Journal of Engineering, Transactions A: Basics Vol. 34, No. 7, (2021) 1736-1742

In similarity-based methods, two similarity matrices related to drugs and similarity targets along with the interaction matrix which represents the interaction between drug pairs and targets are used, respectively [9].

These similarities usually arise through the chemical structures for the drug as well as through the protein sequence alignment for the target. Similarity-based methods have many positive features [10].

Unlike feature-based methods, similarity-based methods do not require a feature extraction or feature selection, which is a difficult and complex process.

Computational similarity criteria have recently been developed and widely used, the similarity of the chemical structure of drugs as well as the similarity of genomic sequences of targets are examples of this.

Due to the direct relationship between similarity-based approaches and kernel methods, similarity-based methods have better performance in prediction.

Similarity matrices show relationships between drugs and genes through chemical space and genomic space, respectively.

These features represent the superiority of similarity-based approaches over other approaches.

In the present research, we used a method based on Low-Rank Matrix Approximation (LRMA) according to weighted nuclear norm minimization (WNNM). In addition, the graph of drug-drug similarity and target-target similarity, and drug-target interaction was used to improve the performance of the proposed method. The details of the proposed method and the steps of the algorithm are described in the following sections.

2. Proposed Method

The drug-target interaction was shown with x-matrix, where the rows represent the drugs, and the columns represent the target. The matrix value is indicative of drug-target interactions. Since all interactions are not known, they are a relative matrix of observations which are expressed as follows:

$$Y = R.X \tag{1}$$

In Equation (1), R is a subsampling operator. In this binary matrix, the value of 1 is indicative of known interaction and the value of zero shows unknown interaction or absence of interaction. A sampled DTI relative matrix is available. The goal of this equation is to estimate the x-matrix from known Ys and Rs. X is a low-rank matrix that needs to be retrieved. To this end, Equation (2) was applied:

$$\min_{X} \quad rank(X) \quad such\ that\ Y = R.X \tag{2}$$

There are numerous methods for minimization of rank in various fields of vision and machine learning, which have attracted the attention of many researchers in this field. One of the most important methods is nuclear norm minimization (NNM), which can guarantee the matrix rank exactly under some limited and theoretical conditions. Nonetheless, the NNM method is unable to make an exact approximation of the matrix rank for various real applications since it often tends to minimize the grade components too much. These methods are used to reconstruct the data by applying additional rank constraints to the estimated matrix. Given the fact that the direct minimization of the rank is an NP-hard problem, it is difficult to solve. In general, WNNM is used to minimize the matrix's rank. The nuclear norm of x-matrix, shown by $\|X\|_*$, is the sum of its singular values. For instance, in $\|X\|_* = \sum_i \sigma_i$ is the singular value of the x-matrix. The goal of NNM is retrieving the low-rank x-matrix from its degraded observation Y matrix by minimizing the $\|X\|_*$.

Recently, NNM-based methods are used in various areas, including removing noise from video, background extraction, and subspace clustering. Nonetheless, nuclear norm is often accepted as convex substitution of matrix rank. Although it has a theoretical guarantee, singular value thresholding (SVT) model reduces degree variables too much for NNM since it treats components of different degrees equally, and therefore, cannot estimate the matrix rank accurately. Numerous methods are proposed to improve the NNM performance. For intrinsic reconstruction, by solving an NNM problem, low-grade noise input can most likely be solved. In this method, nuclear norm proximal (NNP) can be defined as follows:

$$\hat{X} = prox_{\lambda\|.\|_*}(Y) = argmin_X \|Y - R.X\|_F^2 + \lambda\|X\|_* \tag{3}$$

Equation (3) can be solved by applying a norm threshold action on singular values of the observation matrix in the form of Equation (4) [11, 12]:

$$\hat{X} = US_{\frac{\lambda}{2}}(\Sigma)V^T \tag{4}$$

where $Y = U\Sigma V^T$ is a SVD of Y, and $S_{\frac{\lambda}{2}}(\Sigma)$ is the norm threshold in the Σ convex matrix with the $\frac{\lambda}{2}$ parameter. For each convex component, Σ_{ii} exists in Σ. The norm threshold function can be defined in the form of Equation (5):

$$S_{\frac{\lambda}{2}}(\Sigma)_{ii} = max\left(\Sigma_{ii} - \frac{\lambda}{2}, 0\right) \tag{5}$$

While solving the equation is simple, the NNM has some limitations. The nuclear norm treats all singular values equally and ignores previous knowledge that often exists for matrix values. For instance, larger singular values of the data matrix are usually more important than smaller values in the most vision applications since they show the main components of the data. Different weights must be visually assigned to

different individual values so that the NNM flexibility is commensurate with the real scenarios. To correct the NNM's weakness, recent advancements have shown that the minimization of weighted nuclear norm can achieve a better matrix rank approximation, compared to NNM, which innovatively equates inverse weight with singular values. Researchers have proposed the WNNM method to improve NNM flexibility. The weighted imbalanced norm of the matrix is defined in the form of Equation (6):

$$\|X\|_{w,*} = \sum_i |w_i \sigma_i(X)| \quad (6)$$

where, $\sigma_1(X) \geq \sigma_2(X) \geq \cdots \geq \sigma_n(X)$, $w = [w_1, w_2, \ldots, w_n]$ and $w_i \geq 0$, the non-negative weight is allocated to $\sigma_i(X)$. The weight factor increases the ability to show the main nuclear norm. Logical weights determined based on prior knowledge and understanding of the problem use the model of the corresponding nuclear norm minimization of WNNM to make a better estimation of latent data from corrupted input. In this research, the WNNM method was applied for the DTI problem. As mentioned, the present research exploited the adjacent matrix, which shows the drug-target interaction matrix.

To analyze the WNNP problem, a lemma is presented [13] which following special Lemma 1 is derived from this lemma [14]:

Lemma 1. For any $m \times n$ matrices A and B, $tr(A^T B) \leq \sum_i \sigma_i(A)\sigma_i(B)$, where $\sigma_1(A) \geq \sigma_2(A) \ldots \geq 0$ and $\sigma_1(B) \geq \sigma_2(B) \ldots \geq 0$ are the descending singular values of A and B, respectively. Equality occurs if it is only possible to find units U and V which concurrently singular value analyze A and B because
$A = U\Sigma_A V^T$, and $B = U\Sigma_B V^T$,

where the ordered eigenvalue matrices are showed by Σ_A and Σ_B with singular value $\sigma(A)$ and $\sigma(B)$ along the diagonal with the same order, respectively.

The following main theorem is concluded based on the result of Lemma 1 [15].

Theorem 1 Given $Y \in \Re^{m \times n}$, without loss of generality, it is assumed that $m \geq n$, and let $Y = U\Sigma V^T$ be the SVD of Y, where $\Sigma = \begin{pmatrix} diag(\sigma_1, \sigma_2, \ldots, \sigma_n) \\ 0 \end{pmatrix} \in \Re^{m \times n}$. $\hat{X} = U\hat{D}V^T$ is expressed as the universal optimal WNNP problem in (3), where $D = \begin{pmatrix} diag(d_1, d_2, \ldots, d_n) \\ 0 \end{pmatrix}$ is a diagonal non-negative matrix and the solution $(d_1, d_2, \ldots, d_n)$ is for the following convex optimization problem:

$$\min_{d_1, d_2, \ldots, d_n} \sum_{i=1}^{n} (\sigma_i - d_i)^2 + w_i d_i \quad s.t. \; d_1 \geq d_2 \geq \cdots \geq d_n \quad (7)$$

According to theorem 1, the WNNP problem is a new quadratic optimization problem with linear constraints whose global optimization is easily calculated by off-the-shelf convex optimization solvers. Therefore, for the non-convex WNNP problem, a global solution can be obtained through (7). The next results show that when the weights are arranged in non-descending order, the global solution (7) can be obtained in closed-form [15].

Result 1 If $\sigma_1 \geq \sigma_2 \geq \cdots \geq 0$ and the weights convince $0 \leq w_1 \leq w_2 \leq \cdots \leq w_n$, then the global optimization of (7) is $d = \max\left(\sigma - \frac{w_1}{2}, 0\right)$

The conclusion in result 1 is very useful considering that the singular values of a matrix are arranged in non-ascending order and the larger singular values usually correlate with the subspaces of the most important components of the data matrix.

Larger singular values have shrinked less to preserve original and valid information of the underneath data. Therefore, through result 1, there is an optimal closed-form solution to the WNNP problem using the weighted singular value soft-thresholding operation [15]:

$$prox_{\lambda\|.\|_*}(Y) = US_{\frac{W}{2}}(\Sigma)V^T \quad (3)$$

where $Y = U\Sigma V^T$ is the SVD of Y, and $S_{\frac{w}{2}}(\Sigma)$ is the generalized soft-thresholding operator with weight vector w

$$S_{\frac{w}{2}}(\Sigma)_{ii} = \max\left(\Sigma_{ii} - \frac{w_i}{2}, 0\right) \quad (3)$$

Also the above WNNP solver exactly decadents to the NNP solver for the traditional NNM problem when all the weights w_i are set the same.

In this matrix, the value is 1 in case of the presence of a known interaction between the drug (d_t) and target (t_j); otherwise, the value is zero. In the present study, the drug similarity matrix (S_d) and target similarity matrix (S_s) were applied in addition to the interaction matrix .

In addition, we applied the SIMCOMP similarity method [16] based on the number of common substructures in chemical structure. In fact, S_d shows the similarity of the chemical structure of drug pairs. Also S_t shows the degree of similarity between the two proteins, estimated according to the genome sequence similarity based on the amino acid sequence of target protein.

Notably, the normalized Smith-Waterman method [17] was applied for estimating this case.

in addition to the application of the introduced similarity matrix there are four other similarity matrices, including cosine (S_{cos}), correlation (S_{cor}), hamming (S_{ham}), and jaccard (S_{jac}) which were used for DTI prediction [5].

The current research also exploited five similarity matrices estimated by the drug-target interaction matrix. In fact, similarity matrices are used for DTI, as shown in Equation (7):

$$\min_{Z} \quad \|Y - W(Z))\|_F^2 + \lambda\|Z\|_* \\ +\alpha_1 Tr\left(Z^T \sum_{i=1}^{nsim} L_d^i X\right) + \alpha_2 Tr\left(Z^T \sum_{i=1}^{nsim} L_t^i X^T\right) \tag{7}$$

In Equation (7), $\alpha_1 > 0$ and $\alpha_2 > 0$ are balancing parameters, Tr(.) is the operator of the matrix's transposes, nsim shows the number of similarity matrixes. Here, five similarity matrixes were considered. Moreover, L_d and L_t are Laplacian graph for S_d and S_t, estimated in the form of $L_d = D_d - S_d$ and $L_t = D_t - S_t$, respectively.

In this regard, D_d and D_t are degree matrices for drugs and targets, computed by $D_d^{ii} = \sum_j S_d^{ij}$ and $D_t^{ii} = \sum_j S_t^{ij}$.

In this section, the WNNM method shown in algorithm (1) is used to solve Equation (7).

Algorithm 1. multi graph regularized nuclear norm minimization [5] method combined with proposed WNNM method

Procedure Alg(M,A,S_d^{com}, S_t^{com})
Sparsify: S_d^{com}, S_t^{com}
Initialize: $\lambda, \alpha_1, \alpha_2, v_1, v_2, L_t^{com}, L_d^{com}, Y = M, Z = M^T$

$$AA \leftarrow \begin{pmatrix} A \\ \sqrt{v_1}I \\ \sqrt{v_2}I \end{pmatrix}$$

For loop, iterate (k)

$$YY_k \leftarrow \begin{pmatrix} M \\ \sqrt{v_1}Z^T \\ \sqrt{v_2}Y \end{pmatrix}$$

$X_k \leftarrow \text{WNNM}(YY_k, AA, \lambda)$
$Y_k \leftarrow solve - sylvester(v_1 I, \alpha_1 L_d^{com}, v_1 X_k')$
$Z_k \leftarrow solve - sylvester(v_2 I, \alpha_2 L_t^{com}, v_2 X_k)$
End Loop

In Algorithm 1, $S_d^{com} = S_d + S_d^{cos} + S_d^{cor} + S_d^{ham} + S_d^{jac} = \sum_{i=1}^{nsim} S_d^i$ and $S_t^{com} = S_t + S_t^{cos} + S_t^{cor} + S_t^{ham} + S_t^{jac} = \sum_{i=1}^{nsim} S_t^i$ show the combined similarity for drug and target, $D_d^{com} = diag\left(\sum_j S_d^{com}\right)$ and $D_t^{com} = diag\left(\sum_j S_t^{com}\right)$ show the combined degree matrix for the drug and target also $L_d^{com} = D_d^{com} - D_d^{com}$ and $L_t^{com} = D_t^{com} - D_t^{com}$ show the combined Laplacian matrix for the drug and target, respectively. This equation is solved using the method presented in [5]. Please refer to the mentioned article for more details.

3. Experiments and Analysis of Results

In this section, the experiments and results of the proposed method are analyzed separately.

3. 1. Dataset and Evaluation Criteria The information related to the interactions between drugs and target proteins for public databases of KEGG BRITE, RENDA, SuperTarget and DrugBank have been assessed by Yamanishi et al. [7]. Similar to Yamanishi et al. study, we applied four benchmark datasets from four different classes of target protein. In fact, these criteria are simulated from public databases. The following is a description of these datasets:

• Enzymes (Es): 445 drugs, 664 targets, and 2926 interactions were extracted in this dataset .

•Ion channels (IC): 201 drugs, 204 targets, and 1476 interactions are extracted in this dataset .

• G protein-coupled receptors (GPCRs): 223 drugs, 95 targets, and 635 interactions are extracted in this dataset .

• Nuclear receptors (NRs): 54 drugs, 26 targets, and 90 interactions are extracted in this dataset .

It is notable that the foregoing datasets were simulated from public databases, which are available with the address of http://web.kuicr.kyoto-u.ac.jp/supp/yoshi/drugtarget publicly. In the present research, cross-validation settings of leave-one-out (LOO) were used for data segmentation. Three modes of the dataset were considered in the results section. In addition, CVS for drug prediction, CVS for target prediction, and interaction prediction was introduced with titles of CVS1, CVS2, and CVS3, respectively. This segmentation was based on Mongia et al. study [5], as presented below:

• CVS1/drug prediction: All drug profiles are set aside to be used as the experiment set, which tests the algorithm's ability to predict the interactions of new drugs, that is, drugs for which no cross-information is available.

• CVS2/target prediction: The entire target profiles are set aside to be used as the experiment set to assess the algorithm's ability to predict interactions of new targets .

• CVS3/pair prediction: Random drug-target pairs are set aside as the experiment set for prediction. This a normal adjustment for validation and evaluation.

When at least one DTI is known for d_i and t_j respectively in the training data the CVS1 predicts the unknown pair (d_i,t_j). To prevent using the pairs, CV used the pairs between the drugs having >= 2 targets and the targets interacting with >= 2 drugs, which should be used in three other scenarios. Some of these pairs are selected by random for testing in each round of CV and the union of the rest of them and other entries are used for training.

However, when there are no DTIs for observation of new drugs and new targets in the training data, CVS2 and CVS3 predict new drugs and new targets respectively.

Performance of CV on drugs in CSV2, where the rows corresponding to drugs are randomly blinded for

testing and the resting rows are used for training. Also performance of CV on targets in CSV3 where the columns (accounting for targets) are randomly blinded for testing and the resting columns are used for training as well.

We have made various tasks of CV under 3 scenarios showed in Figure 1 respectively.

In addition, area under the ROC curve (AUC) and area under the Precision-Recall (AUPR) were applied to assess the performance of the proposed method according to Mongia et al. study [5].

3. 2. Analysis of Experiments' Results This section includes a comparison of the proposed method with previous works in recent years. In the current research, we applied the techniques presented in others study for comparison [18, 19]. Notably, all methods are performed from the same data set and the same CVS. In addition, the results of other works were extracted from the articles. The results are shown in the tables below based on the AUC and AUPR criteria .

Table 1 presents a comparison of the methods based on the AUC criterion in four benchmark datasets and various CVSs. According to the results, the proposed method had acceptable performance in the evaluation of DTI in similar datasets, compared to other techniques. A very important point in these tables is related to the prediction of the target-drug pair. In this regard, the rows related to CVS3 are shown in Table 1.

Table 2 compares the methods based on the AUPR criteria and the results obtained from the techniques in four benchmark datasets in three different CVSs. According to the results, the proposed method had acceptable performance in DTI evaluation (CVS3) in all four benchmark datasets, compared to other methods.

This paper presents a new approach based on the known drug-target interactions based on similarity graphs. The weighted nuclear norm minimization method was used to identify the drug-target interactions. Our proposed method encodes the adjacency between the drug and the target by graphs. Also, known drug-

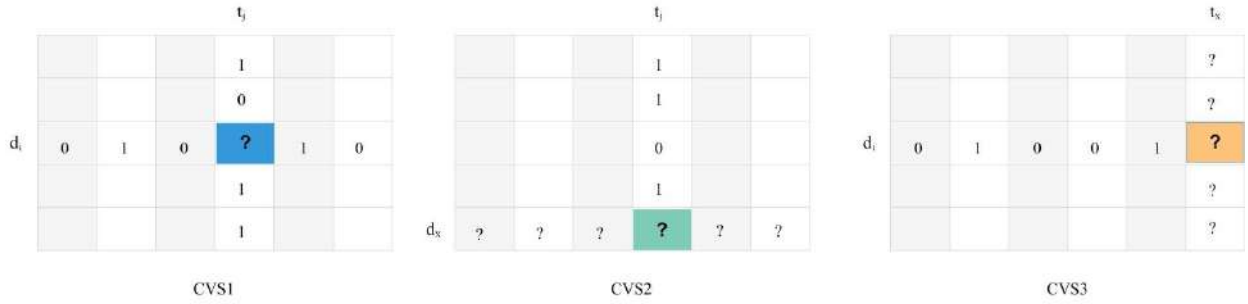


Figure 1. Presentation of cross-validation schemes for three scenarios. Each column represents a scenario. Row includes the DTI matrices, in which the entries marked with "?" are the pairs of interest to be tested

TABLE 1. Comparison of the proposed method with other techniques based on AUC criteria in four datasets in various CVSs

CVS	Dataset	[18]	[19]	[20]	[21]	Proposed method
CSV1	Es	0.9272	0.9067	0.96	0.97	**0.9721**
	IC	0.9368	0.9286	0.97	**0.98**	0.9526
	GPCRs	0.8966	0.8694	0.94	**0.96**	0.9024
	NRs	0.8373	0.8124	0.88	0.92	**0.9421**
CSV2	Es	0.7755	0.7952	0.78	0.84	**0.8512**
	IC	0.7669	0.7576	0.79	**0.94**	0.8013
	GPCRs	0.8800	0.8067	0.88	0.91	**0.9186**
	NRs	0.8615	0.8124	0.86	0.90	**0.9015**
CSV3	Es	0.9705	0.9635	0.93	0.92	**0.9512**
	IC	0.9832	0.9786	0.94	0.97	**0.9969**
	GPCRs	0.9493	0.9458	0.88	0.93	**0.9902**
	NRs	0.8679	0.9329	0.79	0.88	**0.9339**

TABLE 2. Comparison of the proposed method with other techniques based on the AUPR criteria in four databases in different CVSs

CVS	Dataset	[18]	[19]	[20]	[21]	Proposed method
CSV1	Es	0.7808	0.5465	0.87	**0.92**	0.8532
	IC	0.7786	0.7437	**0.92**	0.92	0.8011
	GPCRs	0.5989	0.5397	0.73	0.79	**0.7944**
	NRs	0.4774	0.4907	0.60	0.83	**0.7720**
CSV2	Es	0.3848	0.2409	0.40	0.73	**0.7322**
	IC	0.3538	0.3090	0.36	0.69	**0.6921**
	GPCRs	0.4059	0.3463	0.42	**0.63**	0.5812
	NRs	0.5203	0.5373	0.56	0.71	**0.7366**
CSV3	Es	0.8837	0.8093	0.80	0.82	**0.9055**
	IC	0.9373	0.8459	0.81	0.80	**0.9411**
	GPCRs	0.7543	0.6933	0.60	0.61	**0.7601**
	NRs	0.6383	0.7072	0.46	0.64	**0.7888**

target interaction, drug-drug similarity, target-target and combination of similarities have been used as input. The proposed method was performed on four benchmark based on the AUC and AUPR criteria. Eventually, the results showed an improvement in the performance of the proposed method

4. CONCLUSION

The present research proposed a novel approach to identify drug-target interactions, which applied the drug-drug, target-target, and target-drug interaction similarity graph method. In the current research, the proposed method's performance was improved by using the WNNM in order to eliminate NNM limitations in the DTI use. In addition, the proposed technique was assessed in four benchmark datasets based on the AUC and AUPR criteria. The final results were indicative of the improved performance of the proposed method compared to previous approaches in the field.

5. REFERENCES

1. Sachdev, K., Gupta, M.K., "A comprehensive review of feature based methods for drug target interaction prediction", ***Journal of Biomedical Informatics***, Vol. 93, (2019), 1031-59, doi: 10.1016/j.jbi.2019.103159
2. Sachdev, K., Gupta, M.K., "A hybrid ensemble-based technique for predicting drug–target interactions", ***Chemical Biology & Drug Design***, Vol. 6, (2020), 1447-1455. doi: 10.1111/cbdd.13753
3. Ezzat, A., Wu, M., Li, X.L., Kwoh, C.K., "Computational prediction of drug–target interactions using chemogenomic approaches: an empirical survey", ***Briefings in Bioinformatics***, Vol. 20, (2019), 1337-1357. doi: 10.1093/bib/bby002
4. Keiser, M.J., Roth, B.L., Armbruster, B. N., Ernsberger, P., Irwin, J.J., Shoichet, B.K., "Relating protein pharmacology by ligand chemistry", ***Nature Biotechnology***, Vol. 25, (2007), 197-206, doi: 10.1038/nbt1284
5. Mongia, A., Majumdar, A., "Drug-target interaction prediction using Multi Graph Regularized Nuclear Norm Minimization", ***Plos one***, Vol. 15, (2020), 226-484, doi: 10.1371/journal.pone.0226484
6. Yıldırım, M.A., Goh, K.I., Cusick, M. E., Barabási, A.L., Vidal, M., "Drug—target network," ***Nature Biotechnology***, Vol. 25, (2007), 1119-1126, doi: 10.1038/nbt1338
7. Yamanishi, Y., Araki, M., Gutteridge, A., Honda, W., Kanehisa, M., "Prediction of drug–target interaction networks from the integration of chemical and genomic spaces", ***Bioinformatics***, Vol. 24, (2008), 232-240, doi: 10.1093/bioinformatics/btn162
8. Ezzat, A., Wu, M., Li, X.L., Kwoh, C.K., "Drug-target interaction prediction via class imbalance-aware ensemble learning", ***BMC Bioinformatics***, Vol. 17, (2061), 267-276, doi: 10.1186/s12859-016-1377-y
9. Thafar, M.A., Olayan, R.S., Ashoor, H., Albaradei, S., Bajic, V.B., Gao, X., "DTiGEMS :+drug–target interaction prediction using graph embedding, graph mining, and similarity-based techniques", ***Journal of Cheminformatics***, Vol. 12, (2020), 1-17, doi: 10.1186/s13321-020-00447-2
10. Ding, H ,.et al., Similarity-based machine learning methods for predicting drug–target interactions: a brief review. ***Briefings in Bioinformatics***, Vol. 15, No. 5, (2014), 734-747, doi: 10.1093/bib/bbt056
11. Sun, X., Y. Wang, and X. Zhang. Multi-View Subspace Clustering via Non-Convex Tensor Rank Minimization. in 2020 IEEE International Conference on Multimedia and Expo (ICME). 2020. IEEE, doi: 10.1137/20M1318006
12. Yang, Y., et al., Foreground-Background Separation via Generalized Nuclear Norm and Structured Sparse Norm Based Low-Rank and Sparse Decomposition. ***IEEE Access***, (2020), Vol. 8, 84217-84229, doi: 10.1109/ACCESS.2020.2992132
13. Gu, S., et al. Weighted nuclear norm minimization with application to image denoising. in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2014, doi: 10.1109/CVPR.2014.366
14. D. Rhea, "The case of equality in the von Neumann trace inequality," preprint, 2011, doi: 10.1016/j.laa.2015.05.015
15. Gu, S., Xie, Q., Meng, D., Zuo, W., Feng, X., & Zhang, L. Weighted nuclear norm minimization and its applications to low level vision. ***International Journal of Computer Vision***, Vol. 121, No. 2, (2017), 183-208, doi: 10.1007/s11263-016-0930-5
16. Hattori, M., Tanaka, N., Kanehisa, M., Goto, S., "SIMCOMP/SUBCOMP: chemical structure search servers for network analyses", ***Nucleic Acids Research***, Vol. 38, (2010), 652-656, doi: 10.1093/nar/gkq367
17. AMoZ, J., "Identification of common molecular subsequences", ***Journal of Molecular Biology***, Vol. 147, (1981), 195-197. doi: 10.1016/0022-2836(81)90087-5
18. Zheng, X., Ding, H., Mamitsuka, H., Zhu, S., "Collaborative matrix factorization with multiple similarities for predicting drug-target interactions", in Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining 2013, 1025-1033, (2013). doi : 10.1145/2487575.2487670
19. Van, T., Laarhoven and Marchiori, E., "Predicting drug-target interactions for new drug compounds using a weighted nearest neighbor profile", ***Plos one***, Vol. 8, (2013), 52-566. doi: 10.1371/journal.pone.0066952
20. Ezzat, A., et al., Drug-target interaction prediction with graph regularized matrix factorization. ***IEEE/ACM Transactions on Computational Biology and Bioinformatics***, Vol. 14, No. 3, (2016), 646-656. doi: 10.1109/TCBB.2016.2530062
21. Olayan, R.S., H. Ashoor, and V.B. Bajic, DDR: efficient computational method to predict drug–target interactions using graph mining and machine learning approaches. ***Bioinformatics***, Vol 34, No. 7, (2018), 1164-1173. doi: 10.1093/bioinformatics/btx731

Persian Abstract

چکیده

شناسایی تعامل بین دارو و پروتئین‌های هدف، نقش بسیار مهمی در فرآیند کشف داور دارد. از آنجایی که انجام آزمایشات پیش‌بینی این فرآیند زمانبر، بسیار پرهزینه و خسته کننده می‌باشد. پیش‌بینی محاسباتی می‌تواند یک راهکار مناسبی برای کاهش فضای جستجو برای بررسی تعامل بین دارو و هدف بجای استفاده از آزمایشات پر هزینه باشد. در این مقاله یک راهکار نوین بر اساس، تعامل‌های شناخته شده در بین دارو–هدف برپایه گراف شباهت ارائه شده است. در این مقاله نشان داده شده که این کاربرد جز مسائل با مرتبه پایین می باشند که در این راستا از روش حداقل رساندن نرم هسته‌ای وزنی برای تشخیص تعامل بین دارو و هدف استفاده شده است. همچنین در این مقاله برای نمایش مجاورت بین دارو و هدف از گراف مجاورت استفاده شد که تعامل دارویی–هدف، شباهت دارو–دارو، هدف–هدف و ترکیبی از شباهت‌ها به عنوان ورودی استفاده شد. روش پیشنهادی بر روی چهار مجموعه داده معیار شامل آنزیم‌ها (Es)، کانال‌های یونی(IC)، گیرنده‌های همراه پروتئین G (GPCRs) و گیرنده‌های هسته‌ای (NRs)، بر اساس معیارهای AUC، AUPR مورد ارزیابی قرار گرفته است. نتایج بدست آمده نشان دهنده بهبود عملکرد روش پیشنهادی می‌باشد.

International Journal of Engineering

Journal Homepage: www.ije.ir

Objective Functions for the Optimization of an Ultra Wideband Antenna

A. Contreras*

Department of Circuits and Communications, Faculty of Engineering, Universidad del Zulia, Maracaibo, Venezuela

PAPER INFO

Paper history:
Received 30 March 2020
Received in revised form 16 March 2021
Accepted 06 June 2021

Keywords:
Objective Function
Ultra Wideband Antenna
Particles Swarm Optimization
Bandwidth Adjustment
S_{11} Magnitude

ABSTRACT

This work proposes an objective function to optimize an ultra wideband antenna for adjusting the bandwidth and coupling with other elements, based on the performance comparison of several objective functions from the literature. The optimal dimensions of a printed rectangular monopole antenna were obtained with the Particle Swarm Optimization method to compare such functions. In the results of the comparison, the linear functions had a mean value of S_{11} magnitude near the threshold, but they presented a smaller standard deviation than the rest of the functions. The logarithmic and cubic functions showed a mean value of S_{11} magnitude higher than the double of the threshold, but they had superior standard deviation values, which did not happen with the quadratic function. Hence, the proposed function is the mean of a logarithmic expression with the quadratic argument. With this function, a bandwidth adjustment of 130%, a mean S_{11} magnitude of -22.1 dB and a standard deviation equal to 6.7 dB were obtained on the resonant band for the designed antenna. In this way, the proposed function can be used to avoid interference with other wireless systems and to obtain a uniform coupling of the antenna.

***doi**: 10.5829/ije.2021.34.07a.19*

1. INTRODUCTION

The Ultra Wideband (UWB) antennas have achieved relevance in current wireless communication systems because numerous devices need an antenna to operate at different frequencies for several applications. UWB antennas can substitute multiple narrow band ones, reducing the number of antennas, consequently decreasing costs and power consumption [1-2].

However, the design of UWB antennas presents challenges as spatial limitations, interference, and the gain performance for multiple wireless applications [3]. Moreover, these antennas can turn up physical phenomena as resonance and coupling between components because they have sizes similar to the wavelength [4].

The mentioned challenges can be resolved with the optimization in the performance of UWB antennas to improve the results according to the specific requirement of the systems [5]. Many optimization methods have been implemented in the different shapes of UWB antennas, such as Particle Swarm Optimization (PSO), Genetic Algorithm (GA), Surrogate Based Optimization (SBO) and Fractional Factorial Design (FFD) [6].

These optimization methods have been used because the antennas have a complex model that makes it difficult to apply traditional mathematical methods of optimization. Likewise, objective functions are used as surrogate models for improving the parameters selected by the designers. The most used optimization parameter in antennas is the resonant bandwidth [7], the objective functions mainly depend on S_{11} magnitude and many of them are nonlinear [8-14]. Nevertheless, there is not a comparison between the performance of the objective functions and how these adjust the antenna dimensions for improving its fundamental parameters.

In this work, the performance of several objective functions, which have been employed in the optimization of an UWB antenna, are compared for matching the responses in the adjustment of bandwidth and S_{11} magnitude. In this way, the results support the proposal of an objective function that allows improving such parameters to avoid interference with other wireless systems and to obtain a uniform coupling of the antenna.

*Corresponding Author Institutional Email: acontreras@fing.luz.edu.ve (A. Contreras)

Please cite this article as: A, Contreras, Objective Functions for the Optimization of an Ultra Wideband Antenna, International Journal of Engineering, Transactions A: Basics Vol. 34, No. 7, (2021) 1743-1749

2. METHOD OF OPTIMIZATION

In this work, the PSO method was selected because it has been applied in multiple optimization processes with antennas for reducing the computation time through the prevention of next particle generations equal to previous [15]. The PSO belongs to the family of evolutionary algorithms and it is one of the fastest heuristic approaches for finding the optimal solution of a complex problem [16].

Figure 1 shows the PSO method algorithm, which was programmed in Visual Basic for invoking a Finite Elements Methods (FEM) software, the High-Frequency Structure Simulator (HFSS) from Ansoft. The antenna dimensions are imported to HFSS and the results are exported for each particle. With the results, the objective functions were evaluated and the minimum search mechanism was applied.

Step 1 comprises the initialization of the PSO method parameters, which are: N the number of dimensions, M the number of particles, N_{iter} the number of total iterations, w the inertial weight, c_1 and c_2 cognitive and social parameters, η_1 and η_2 random values between 0 and 1 for each particle and iteration, and V_{max}^n the maximum velocity for each dimension that is calculated with the following equation:

$$V_{\max}^{n} = 0.1(X_{\max}^{n} - X_{\min}^{n}) \quad (1)$$

where X_{max}^n is the maximum position and X_{min}^n is the minimum position of the dimension n.

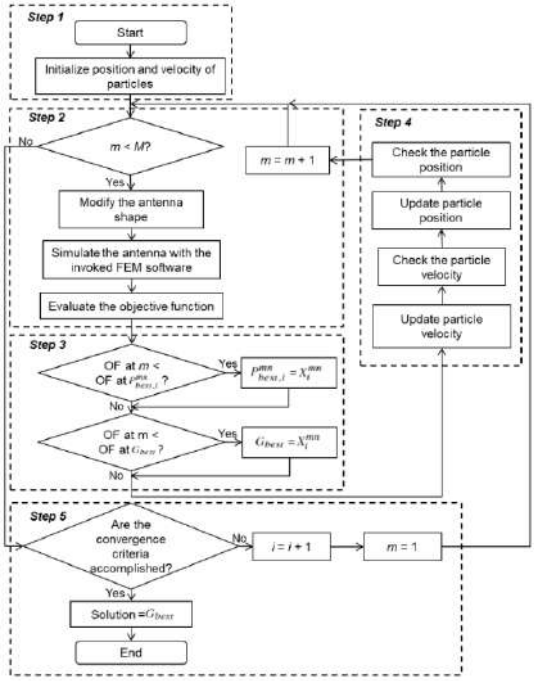


Figure 1. Flow diagram of the PSO method

In step 2, the dimensions of the antenna are modified for the simulation through FEM software, and the objective function is evaluated with the results. Then, the value of the objective function is compared with the best response of the particle ($P_{best,i}^{mn}$) and the swarm (G_{best}) for each iteration i, with the decisions of step 3.

In step 4, the velocity and position are updated. The particle velocity is a real variable that must be between a maximum value (V_{max}^n) and a minimum value ($-V_{max}^n$) [17]. The velocity for the next iteration is determined as follows:

$$V_{i+1}^{mn} = wV_i^{mn} + c_1\eta_1(P_{best,i}^{mn} - X_i^{mn}) + c_2\eta_2(G_{best} - X_i^{mn}) \quad (2)$$

The position must be in a confinement interval, hence it must be verified that its value is in the range $\{Lim_{min}^n, Lim_{max}^n\}$ [18]. The position for the next iteration is calculated as follows:

$$X_{i+1}^{mn} = X_i^{mn} + V_{i+1}^{mn} \quad (3)$$

Finally (in step 5), the convergence criteria are evaluated. These criteria are that the number of iterations is equal to the maximum number of total iterations (N_{iter}), that the difference between the solutions of the best particle is less than a determined value, and that the best value of particle does not change in a consecutive number of iterations (N_{iter}^{cons}). When any of these three conditions is accomplished, the process of optimization ends, and the best position of the swarm is obtained [17].

3. OBJECTIVE FUNCTIONS

There are diverse objective functions to optimize an antenna, most of them search the optimal value of dimensions for improving S_{11} magnitude and bandwidth. In the following, different functions that have been applied in the last five years are described. In this work, linear and nonlinear functions for the optimization of UWB antennas in communication systems with single or multiple bands were selected.

3. 1. Linear Function with the Mean of S_{11} Magnitude This function calculates the average of a set of values for each frequency component, which are the S_{11} magnitude or the value K, and the equation is [8, 9]:

$$OF1 = \min\ \mu\{m(f_k)\}$$
$$m(f_k) = \begin{cases} |S_{11}(f_k)| & if\ |S_{11}(f_k)| \geq |S_{11}|_{th} \\ K & if\ |S_{11}(f_k)| < |S_{11}|_{th} \end{cases} \quad (4)$$

where $\mu\{m(f_k)\}$ is the mean value of $m(f_k)$, $|S_{11}|$ is the S_{11} magnitude for the frequency sample f_k, $|S_{11}|_{th}$ is the threshold of the S_{11} magnitude, and K is equal to -10 dB.

3. 2. Linear Function with Bandwidth and Penalty Factor This function considers the bandwidth plus a discrete variable that depends on the S_{11} magnitude surpassing a threshold value [10]. It is determined as follows:

$$OF2 = \max\ A + 0.5B$$
$$A = \begin{cases} 1 & if\ |S_{11}(f_k)| \le |S_{11}|_{th} \\ 0 & if\ |S_{11}(f_k)| > |S_{11}|_{th} \end{cases}$$
$$B = \begin{cases} \dfrac{f_M - f_m}{1GHz}, & f_M > f_{\max} \quad f_m < f_{\min} \\ 0 & otherwise \end{cases} \tag{5}$$

where B is the bandwidth, A is a discrete value that depends on S_{11} magnitude, f_M is the maximum frequency and f_m is the minimum frequency.

3.3. Nonlinear Function with Sum of Logarithms of the S_{11} A nonlinear function with the sum of logarithms of the S_{11} magnitude with several logarithm bases was evaluated by Chen [11]. While, obtaining that high base values improved the impedance coupling in each frequency sample and the logarithm with base 8 had the best performance. The equation used is stated as follows:

$$OF3 = \max \sum_{k=1}^{N} \log_8 \left(\left\| S_{11}(f_k) \right\| \right) \tag{6}$$

3. 4. Nonlinear Function with Mean of Cubic S_{11} Magnitude This nonlinear function was implemented to improve the impedance coupling of an UWB antenna [12], and it uses the mean of the cubic S_{11} magnitude of the frequency components in the design range as follows:

$$OF4 = \min\ \mu \left\{ \left| S_{11}(f_k) \right|^3 \right\} \tag{7}$$

3. 5. Nonlinear Function with Difference of the Squared Minimum Frequency This function seeks to minimize the maximum reflection level and to control the minimum resonance frequency [13]. It is expressed as follows:

$$OF5 = \max(l_2, ..., l_k) + \beta \left[\frac{\max(f_1 - f_{1\text{target}}, 0)}{f_{1\text{target}}} \right]^2 \tag{8}$$

where $l_k(x) = |S_{11}(f_k)| - |S_{11}|_{th}$, β is the penalty factor equal to 1000, f_1 is the minimum resonant frequency and $f_{1\text{target}}$ is the minimum frequency of the design.

4. COMPARISON OF OBJECTIVE FUNCTIONS

The comparison consisted of applying different objective functions to optimize the design of the same antenna and to analyze the results. In this work, the design optimization problem for UWB antenna aims to obtain an adjustment of bandwidth to avoid interference with other wireless systems, and a smaller value of S_{11} magnitude as uniform as possible for improving the impedance matching in the resonant band of the antenna.

The comparison variables are the mean of S_{11} magnitude, the standard deviation of S_{11} magnitude between samples and the bandwidth adjustment, which was determined with the following equation:

$$B_{adj} = 100\left(\frac{B}{B_d} \right) (\%) \tag{9}$$

where B is the obtained bandwidth and B_d is the design bandwidth. The bandwidth adjustment must be equal to or higher than 100% and it is considered better when its value is near 100%.

A printed rectangular monopole antenna (PRMA) with microstrip feeding method was selected because this has been the most used in applications of UWB wireless communication systems. The microstrip permits to obtain a high level of adaptability with devices that work in the microwaves frequency range. The selected substrate was the FR4-epoxy given its high level of relative permittivity that contributes to reduce the antenna's dimensions [6].

The characteristics of the substrate are: thickness h = 1.6 mm, relative permittivity ϵ_r = 4.4 and tangent of permittivity loss tan(δ) = 0.02. The frequency range of the design was from 1.7 GHz to 3.7 GHz ($|S_{11}| \le |S_{11}|_{th}$), which represents a design bandwidth of 2 GHz considering $|S_{11}|_{th}$ equal to -10 dB.

Before implementing the optimization method with the different objective functions, the PRMA dimensions (Figure 2) were calculated, as they determined the central values of the confinement interval of the optimization. For calculating the resonant patch dimensions, the following equation was used [19]:

$$2\pi W + L = 2\pi \left(\frac{7,2}{1.15 f_L} + p \right) (cm) \tag{10}$$

where L is the length of the patch, W is the width of the patch, p is the separation between the resonant plane and ground plane and f_L is the minimum frequency of the design in GHz.

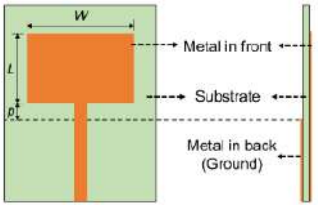


Figure 2. Geometry of the UWB PRMA

In this manner, the dimensions calculated for resonant patch were: L = 22.06 mm, a W = 33 mm and a p = 2 mm. The transmission line of the microstrip had a length of a quarter of the maximum wavelength and a width of 3 mm. With these values, the calculated antenna was simulated, obtaining a mean of S_{11} magnitude in the design range equal to -10.7 dB, a minimum frequency of 1.7 GHz and a bandwidth of 1.1 GHz, which does not comply with the design premise of the bandwidth.

Then, the setting of the PSO algorithm applied for each objective function with the following parameters was: N = 3 dimensions (L, W, and p), M = 10 particles, N_{iter} = 120 iterations, N_{iter}^{cons} = 40 iterations, w varying between 0.9 and 0.4 during the optimization process, c_1 varying from 2.8 to 2, and c_2 from 1.2 to 2. Besides, the boundary conditions were L = 22.06±20 mm, W = 33±20 mm and p = 2±2 mm. For the simulation, the frequency was from 1 GHz to 4.5 GHz, and the frequency step between samples was 100 MHz. In addition, 50 realizations were made for each objective function.

Table 1 shows the obtained values of the mean and standard deviation of each comparison variable for each objective function in the design range. The bandwidth adjustment was over 100% for all objective functions, achieving the design premise. With *OF1*, a low value of bandwidth adjustment was obtained but with a high standard deviation value. This function had a high value of S_{11} magnitude mean near the threshold. The *OF3* had the best performance with 127.5% followed by *OF4* with 134% but both had high deviation. In contrast, *OF5* had an adjustment of 136% with a low deviation.

The highest bandwidth adjustment, equal to 150%, was obtained with *OF2*, and it presented the highest mean of S_{11} magnitude among all functions. Therefore, this function is not recommended in the cases when it is necessary to adjust the bandwidth and to obtain a high level of coupling.

On the other hand, the nonlinear functions had a better performance regarding the mean of S_{11} magnitude, because lower values were obtained near the double of the threshold. The lowest value was achieved by *OF3*, with -22.6 dB.

Concerning the standard deviation of S_{11} magnitude, the linear functions presented low values of this variable, and the *OF2* had the lowest value. The quadratic function was the second with lower deviation, but its mean of S_{11} magnitude was higher than other nonlinear functions, which exhibited better performance.

In general, to accomplish bandwidth adjustment it is necessary to implement linear objective functions that consider the sum or mean of S_{11} magnitude in the design range. Additionally, logarithmic functions can be used to improve this adjustment. The quadratic function had a good performance in the S_{11} magnitude and low standard deviation, which can be employed to improve the coupling performance.

TABLE 1. Results of the UWB antenna optimization with the objective functions

Function	B_{adj}(%)	$\|S_{11}\|_{mean}$(dB)	$\|S_{11}\|_{std}$(dB)
OF1	133.5±21.9	-16.8±2.6	4.9±2.4
OF2	150.0±0.0	-16.6±1.5	4.3±1.8
OF3	127.5±8.6	-22.6±0.6	8.7±1.9
OF4	134.0±10.8	-22.5±1.0	13.1±1.5
OF5	136.0±3.2	-19.8±1.4	4.9±2.6

5. PROPOSAL OF OBJECTIVE FUNCTION

According to the results obtained in section 4, a novel objective function was proposed. This is a nonlinear function that calculates the mean of the base 8 logarithm of the squared difference between S_{11} magnitude and its threshold when the specified criterion is accomplished, as presented in the following equation:

$$OF6 = \max\ \mu\left\{C_k.\log_8\left(\left(|S_{11}(f_k)|-|S_{11}|_{th}\right)^2\right)\right\}$$
$$C_k = \begin{cases} 1 & if\ |S_{11}(f_k)| \le |S_{11}|_{th} \\ 0 & otherwise \end{cases} \quad (11)$$

The purpose of this function is to maximize the performance of the UWB antenna, considering the Equations (4), (6) and (8) for a good bandwidth adjustment, a good coupling and a low standard deviation of S_{11} magnitude, in the design bandwidth.

The UWB antenna was optimized using the PSO method with the proposed function for 50 realizations as it was made with the rest of the functions, and the results are shown in Table 2. Low values of standard deviation were obtained for each one of the comparison variables.

Figure 3 shows the box and whisker plot for the bandwidth adjustment of each objective function with its median and mean, where the proposed function has a central distribution with similar values of mean and median, and with few variations in the results obtained. This performance is also presented by the *OF2* but with a higher mean of bandwidth adjustment.

The proposed function has the second best mean value of adjustment after the *OF3*, which showed the lowest mean value but with high deviation. Regarding the mean of S_{11} magnitude for each objective function, the proposed function presented closer mean and median values in comparison with the rest of the functions, obtaining a central distribution as it is shown in Figure 4.

TABLE 2. Results of the UWB antenna optimization with the proposed objective function

Function	B_{adj}(%)	$\|S_{11}\|_{mean}$(dB)	$\|S_{11}\|_{std}$(dB)
OF6	130.0±3.3	-22.1±0.6	6.7±1.0

Furthermore, this function showed a behavior similar to the other nonlinear functions in this variable, i.e. a mean of S_{11} magnitude lower than the double of the threshold. However, it obtained the smallest standard deviation value of all functions, which implies the optimal responses are very close to each other even though they have different initial points.

Concerning the standard deviation of S_{11} magnitude in the design range, the statistical results obtained are shown in Figure 5. The *OF3* had higher median and mean values than *OF1* and *OF2*, *OF5*, and the proposed function (*OF6*).

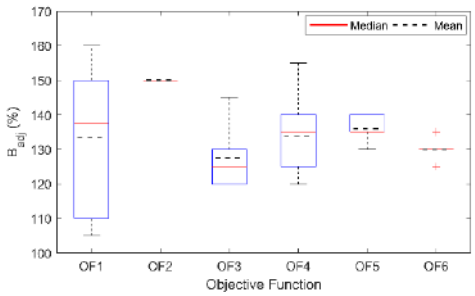


Figure 3. Box and whisker plot for the bandwidth adjustment of each objective function, indicating its median and mean

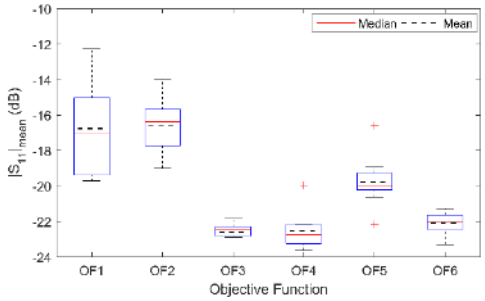


Figure 4. Box and whisker plot for the mean of S_{11} magnitude of each objective function, indicating its median and mean

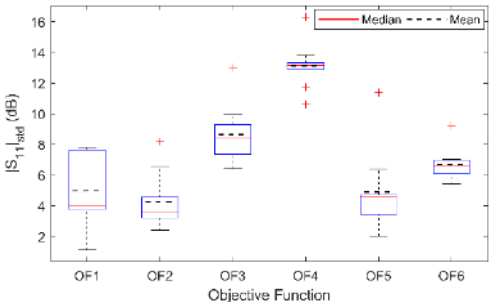


Figure 5. Box and whisker plot for the standard deviation of the S_{11} magnitude of each objective function, indicating its median and mean

This is a disadvantage of the *OF3*, which had the best response in the other comparison variables. Between the nonlinear objective functions, the proposed function had the second best performance, after the *OF5*, but it obtained a lower variation between the results of the realizations. With the function proposed in this research, a more uniform coupling of the antenna in different frequency components in the design range was obtained.

Figure 6 shows the mean of S_{11} magnitude as a function of the standard deviation of S_{11} magnitude resulting from optimization with all the objective functions and for the antenna with the calculated dimensions, which permits to affirm the performance of the proposed objective function is between the best alternatives to minimize both variables.

Moreover, Figure 7 shows the results of optimization with all the objective functions and the initially calculated antenna through the relationship of two comparison variables, the mean of S_{11} magnitude and the bandwidth adjustment. The proposed function had a good adjustment, which permits to avoid interference of other communication systems that are out of the bands.

In summary, with the proposed objective function a good performance of the UWB antenna according to the bandwidth adjustment and mean of S_{11} magnitude can be obtained. With the first one, a better adjustment than most of the objective functions was obtained. While with the second one, better uniformity of the frequency components in the design range was achieved.

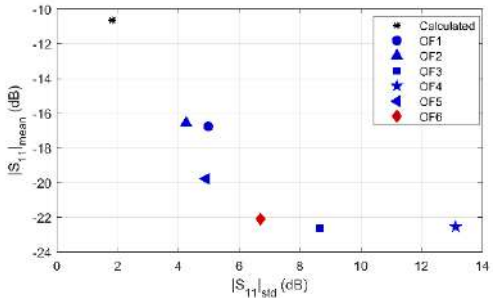


Figure 6. Mean of S_{11} magnitude as a function of the standard deviation of S_{11} magnitude for objective functions and calculated antenna

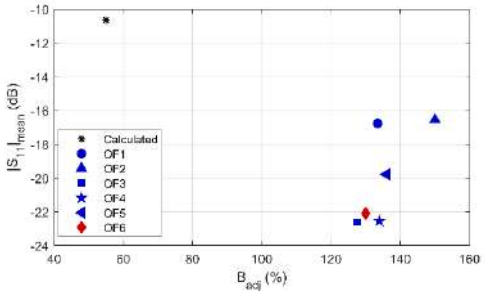


Figure 7. Mean of S_{11} magnitude as a function of bandwidth adjustment for objective functions and calculated antenna

6. CONCLUSIONS

In this work, linear and nonlinear objective functions that have been used in the process of optimization of UWB antennas were compared. The solutions from the linear objective functions had high levels of the mean of S_{11} magnitude that were near the threshold, while the nonlinear functions solutions had levels around the double of the threshold. Therefore, the last ones permit to enhance the coupling of the antenna. Concerning the standard deviation of S_{11} magnitude and the bandwidth adjustment, there was no clear trend of these variables for the different objective functions.

An objective function for the optimization of UWB antennas was proposed with basis on the relevant results of the comparison of diverse parameters. This function improves the performance of the optimized antenna according to the mean and standard deviation of S_{11} magnitude and it has a good bandwidth adjustment. It means that this is a trade-off solution of the three variables. In consequence, the proposed function can be used to avoid interference with other wireless communication systems, whose reception is not necessary, and to obtain a more uniform coupling of the antenna in the frequency components in the whole design range.

It is important to consider that these results have been obtained under a particular set of conditions, which have been described in section 4. Results could be different for PSO with a different number of particles, different dimensions, or another optimization method. Future work can be conducted for the evaluation of different configurations and methods, as well as the multi-objective formulation of the design problem, considering the bandwidth and the S_{11} magnitude simultaneously.

7. REFERENCES

1. Ahmed, F., Feng, Y., and Li, R., "A Multiband Multiple-input Multipleoutput Antenna System for Long Term Evolution and Wireless Local Area Networks Handsets", ***International Journal of Engineering. Transactions B: Applications***, Vol. 29, No. 8, (2016), 1087-1093. DOI: 10.5829/idosi.ije.2016.29.08b.08
2. Ghobadi, Ch., and Majidzadeh, M., "Multi Attribute Analysisofa Novel Compact UWB Antennawith Via-fed Elements for Dual Band Notch Function", ***International Journal of Engineering. Transactions A: Basics***, Vol. 27, No. 10, (2014), 1565-1572. DOI: 10.5829/idosi.ije.2014.27.10a.10
3. Liang, X. L., Ultra Wideband Antennas and Design. Rijeki: InTech, 2012.
4. Sorrentino, R., and Bianchi, G., Microwave and RF engineering. Chichester: John Wiley & Sons Ltd, 2010.
5. Zich, R. E., Mussetta, M., Grimaccia, F., Albi, R., Carbonara, A., Antuono, P. D., Guffanti, T., and Zucchelli, E., "Comparison of different optimization techniques in antenna design – Part I," in IEEE Antennas and Propagation Society International Symposium (APSURSI), Chicago, 7-8, (2012). DOI: 10.1109/APS.2012.6348756
6. Contreras, A., and Urdaneta, M., "Ultra wideband antennas for communications systems on microwave frequency bands: A review", ***Revista Ingeniería UC***, Vol. 25, No. 2, (2018), 134-148.
7. Balanis, C., Antenna theory. Analysis and design. New Jersey: John Wiley & Sons Ltd, 2016.
8. Li, Q., Dong, J., Yang, J., Zhuang, X., Yu, X., Hu, G., and Guo, Y. "Automated Topology Optimization of Internal Antenna Design Using Improved BPSO", in International Applied Computational Electromagnetics Society Symposium (ACES), Suzhou, 4-5, (2017).
9. Jayasinghe, J. M., Anguera, J., Uduwawala, D. N., and Andújar, A., "Nonuniform Overlapping Method in Designing Microstrip Patch Antennas Using Genetic Algorithm Optimization", ***International Journal of Antennas and Propagation***, Vol. 2015, (2015), 1-8. DOI: 10.1155/2015/805820
10. Yu, C., Xu, T., and Liu, C., "Design of a Novel UWB Omnidirectional Antenna Using Particle Swarm Optimization", ***International Journal of Antennas and Propagation***, Vol. 2015, (2015), 1-8. DOI: 10.1155/2015/303195
11. Chen, Y., "A Two-Stage Optimization Technique for Multiband Antenna Design", in 12th European Conference on Antennas and Propagation (EuCAP), London, 2-5, (2018). DOI: 10.1049/cp.2018.1112
12. Chiu, Y., and Chen, Y., "Multi-objective optimization for UWB antennas in impedance matching, gain, and fidelity factor," in IEEE International Symposium on Antennas and Propagation & USNC/URSI National Radio Science Meeting, Vancouver, 1940–1941, (2015). DOI: 10.1109/APS.2015.7305359
13. Koziel, S., and Bekasiewicz, A., "Fast design optimization of UWB antennas using response features," in 21st International Conference on Microwave, Radar and Wireless Communications, Krakow, 4-7, (2016). DOI: 10.1109/MIKON.2016.7492066
14. Radiom, S., Aliakbarian, H., Vandenbosch, G., and Gielen, G., "Optimised small-size tapered monopole antenna for pulsed ultra-wideband applications designed by a genetic algorithm", ***IET Microwaves, Antennas & Propagation***, Vol. 3, No. 4, (2009), 663-672. DOI: 10.1049/iet-map.2008.0195
15. Robinson, J., and Rahmat-Samii, Y., "Particle Swarm Optimization in Electromagnetics", ***IEEE Transactions on Antennas and Propagation***, Vol. 52, No. 2, (2004), 397-407. DOI: 10.1109/TAP.2004.823969
16. Nahvi, H., and Mohagheghian, I., "A Particle Swarm Optimization Algorithm for Mixed Variable Nonlinear Problems", ***International Journal of Engineering. Transactions A: Basics***, Vol. 24, No. 1, (2011), 65-78.
17. Parsopoulos, K. E., and Vrahatis, M. N., Particle Swarm Optimization and Intelligence. Advances and Applications. Hershey: IGI, 2010.
18. Clerc, M., Particle Swarm Optimization. London: ISTE, 2006.
19. Ray, K. P., "Design Aspects of Printed Monopole Antennas for Ultra-Wide Band Applications", ***International Journal of Antennas and Propagation***, Vol. 2008, (2008), 1-8. DOI: 10.1155/2008/713858

Persian Abstract

چکیده

این کار یک عملکرد هدف را برای بهینه سازی آنتن فوق باند پهن برای تنظیم پهنای باند و اتصال با سایر عناصر ، بر اساس مقایسه عملکرد چندین توابع هدف از منابع ، پیشنهاد می کند. ابعاد بهینه یک آنتن تک قطبی مستطیلی چاپ شده با استفاده از روش بهینه سازی ازدحام ذرات برای مقایسه چنین توابع به دست آمد. در نتایج مقایسه ، توابع خطی دارای مقدار میانگین S11 در نزدیکی آستانه بودند ، اما انحراف استاندارد کمتری نسبت به بقیه توابع ارائه دادند. توابع لگاریتمی و مکعبی مقدار میانگین S11 بزرگتر از دو برابر آستانه را نشان دادند ، اما آنها دارای مقادیر انحراف استاندارد برتر بودند ، که با تابع درجه دوم اتفاق نیفتاد. از این رو ، تابع پیشنهادی میانگین بیان لگاریتمی با استدلال درجه دوم است. با استفاده از این عملکرد ، یک پهنای باند ۱۳۰٪ ، یک میانگین اندازه -22.1 -S11 دسی بل و یک انحراف معادل ۶.۷ دسی بل بر روی باند تشدید برای آنتن طراحی شده بدست آمد. به این ترتیب می توان از عملکرد پیشنهادی برای جلوگیری از تداخل در سیستم های بی سیم دیگر و به دست آوردن یک اتصال یکنواخت آنتن استفاده کرد.

IJE TRANSACTIONS A: Basics Vol. 34, No. 7, (July 2021) 1750-1765

International Journal of Engineering

Journal Homepage: www.ije.ir

Stability Analysis of AC/DC Microgrids in Island Mode

M. Khanalizadeh Eini, M. Mirhosseini Moghaddam*, A. Tavakoli, B. Alizadeh

Department of Electrical Engineering, Lahijan Branch, Islamic Azad University, Lahijan, Iran

P A P E R I N F O

Paper history:
Received 14 April 2021
Received in revised form 08 May 2021
Accepted 22 May 2021

Keywords:
AC/DC Microgrid
Islanded Performance
Microgrid Stability
Nonlinear Controller

A B S T R A C T

This study aims to introduce a new structure based on a nonlinear controller for controlling and analyzing the stability of the microgrids. In the proposed model, AC and DC resources and loads are located on two different sides. In addition, an AC/DC bidirectional interface converter is applied to supply loads by AC/DC sources. There are AC/DC products on both sides of the converter and each side can supply the load of the other side via a bidirectional interface converter and its load. Alternatively, an energy storage system is used for the system stability on the DC side. The nonlinear microgrid controller is designed to adjust the AC bus side frequency and the DC bus side voltage properly. In this structure, the coordinated optimal power exchange and precise regulation of control signals lead to constant improvement. Thus, system performance is improved. The results show that the proposed model is efficient for both reduction of the fluctuations and improvement of the system stability

***doi**: 10.5829/ije.2021.34.07a.20*

1. INTRODUCTION

A microgrid (MG) is a combined distributed generation unit (DG), load-bearing unit, and energy storage system operating as connect to the network, island, or transition between two modes. There are some advantages in applying microgrids such as enhancing the level of reliability and customer satisfaction, improvement of the power quality and voltage profile levels, also an increase level of the flexibility while minimizing the energy losses [1, 2]. The structure of existing microgrids is in the form of AC, DC, or combined AC/DC, where the advantages of both AC and DC microgrids in the combined structure are applied. Nevertheless, the growing trend of applying combined AC/DC microgrids has significantly led to increase sustainability problems in these interconnected structures. The problem of sustainability in microgrids is mainly due to the lack of energy resources in the islanded operation mode as well as the inertia required for responding to dynamic frequency variations in manufacturing units. Thus, maintaining stable performance in power changing conditions and occurrence of faults is considered a key issue in controlling microgrids [3, 4]. Accordingly, in order to maintain the system stability, proper control decisions should be made by the control system based on dynamic network changes. Hence, the main goal of the distribution system operator in using microgrids is optimal power-sharing in islanded mode and grid-on, frequency control, and stable operation in sudden power changes conditions as well as during fault occurrence [5]. Different methods have been proposed by Li et al. [6] and Dragičević et al. [7] to analyze and improve the stability of microgrids. These studies have mainly focused on improving control methods. Note that the structure of microgrid and its operational features are very different from those of traditional networks.

The studies related to microgrid stability have mainly focused on the mathematical model of microgrid stability analysis to improve its stability [8]. Microgrid stability is improved through optimizing droop benefits. The microgrid model affects the accuracy and speed of calculation. However, many studies have been conducted on the proposed models to improve the accuracy of microgrid stability analysis. When the load fluctuates, the DGs adjust their output power dynamically and participate in the voltage regulation as well as microgrid frequency. The system stability improvement is achieved

* Corresponding Author Institutional Email: m.mirhosseini@liau.ac.ir
(M. Mirhosseini Moghaddam)

Please cite this article as: M. Khanalizadeh Eini, M. Mirhosseini Moghaddam, A. Tavakoli, B. Alizadeh, Stability Analysis of AC/DC Microgrids in Island Mode, International Journal of Engineering, Transactions A: Basics Vol. 34, No. 7, (2021) 1750-1765

by adding a complementary control loop, creating three-level control, and other optimization methods [9, 10].

Nikos [11] has improved AC/DC hybrid microgrid controller method which was applied to model, control, and simulate the microgrid. Employing a linear droop controller such as voltage feedback led to increased microgrid voltage stability in the islanded mode. Note that numerous converters were employed and their fluctuations were considered. Dheer et al. [12] have introduced a method analogous to the normal droop control in which the feedback was taken from the phase angle instead of sampling the voltage magnitude, and the phasor area was analyzed. The DC side loads and resource had not been considered. Implementing this control method in the grid-on microgrids involving an AC/DC source resulted in more precision in the reactive power distribution to control the voltage. A hierarchical control method with three control levels was suggested by Khorsandi at al. [13], in which droop control was performed at the primary level. Then, the deviations generated in the primary controller were compensated at the secondary level and the load section was managed by the microgrid at the third level. Complex implementation could be considered a disadvantage of this method. A dynamic analysis of the DC link on the energy storage along with the generation resources was performed by Tejwani and Suthar [14]. In that model, the strategy of droop control was used for various operational modes, power distribution between units, and microgrid frequency control. Furthermore, in the storage source, there was a boost converter which would charge and recharge the battery. The droop values of the control loop were set at two levels in order to maintain the small-signal stability and allowable system frequency [15].

Additionally, optimization methods are used to determine droop values. Yu et al. [15] have presented a small signal stability model for voltage control and microgrid current. The equations were based on state-space equations. This model included nonlinear equations converted to a linear equation after simplification. Note that AC/DC coordination was not considered which barely affects the system efficiency. Thale and Agarwal [16] have analyzed the small-signal stability for two parallel inverters connected to a network-independent AC system. Analyzing the small-signal stability helps choose the optimal droop utility and the cutoff frequency of the system. However, it is hard to describe the dynamic behavior of connected parallel inverters which cannot be applied to every system.

In this research, a new structure based on a nonlinear controller is proposed to analyze the stability of microgrids in the presence of renewable energy sources as well as energy storage. Based on this model, several challenges of hybrid microgrids are solved. First, the hybrid microgrid is defined as two independent AC and DC buses. In this way, each bus prevents increase in the number of converters on each side by providing loads on its side. In addition, a bidirectional AC/DC interface converter is employed to connect these two buses, which is the criterion for power exchange between these buses through the interface converter of DC side voltage changes and AC side frequency changes. Meanwhile, independent nonlinear control for each element is defined based on the value of error relative to the reference value. As a result, eliminating the error of all system elements will lead to system stability. Finally, to improve the stability of the system, a suitable algorithm is used to charge and discharge energy storage resources. A battery and a capacitor bank are utilized to model the energy storage resources, which can act as system backups in different charging and discharging modes.

In the following, the dynamic modeling of PV and wind resources as well as energy storage is provided first. Then, the structure of the proposed model for accurate control and coordination of AC/DC sectors is described. Finally, the study system and simulation results are presented.

2. MODELING RESOURCES AND CONVERTERS

The model to be introduced involves production sources and AC/DC loads, and can connect to the network. Similar to solar cells (PVs), batteries, and direct winds, the DC resources connect to a DC link through their interface converter. Using this bus, they feed the loads that require the DC power. As with diesel generators, AC power supplies are connected to the AC bus, through which AC loads are also fed. The DC power generated in the DC bus may be transmitted to the AC bus by a bidirectional converter and vice versa. In the main model, all sources and converters are required to be dynamically modeled.

Figure 1 displays the diagram of the microgrid structure such as two AC and DC sections with interface converter. In such a system, changes in load or generation on one side can affect the other side. The power required to maintain the stability of the system is supplied through the same bus or the other side.

The controllers are designed in a decentralized model to generate their power according to the needs of each part, while stabilizing the DC voltage and AC frequency. The AC/DC hybrid microgrid consists of a PV unit with a DC/DC converter, a wind permanent magnet synchronous generator with an AC/DC converter. The energy storage system is connected to the DC bus via a DC/DC bidirectional converter. On the AC side, a synchronous generator is directly connected to the AC bus, and both buses are connected to each other by a bidirectional VSC converter with an LC filter. The dynamic model of resources and converters are described as follows.

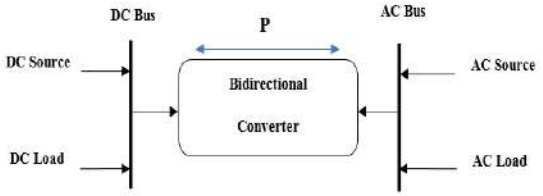


Figure 1. The diagram of AC/DC microgrid structure

2. 1. Modeling PV Power System in DC Microgrids The simulated PV energy system consists of a PV module and a boost converter [17]. Figure 2 reveals the diagram of the converter circuit. This system includes GBT switch, L inductance, Cpv input capacitor, D diode, and Cdc output capacitor. The converter is assumed to operate continuously in the steering mode.

$$\frac{dx_1}{dt}=\frac{1}{L}\left[-R\cdot x_1+x_2-(1-\mu_1)x_3\right] \quad (1)$$

$$\frac{dx_2}{dt}=\frac{I_{pv}-x_1}{C_{pv}} \quad (2)$$

$$\frac{dx_3}{dt}=\frac{1}{C_{dc}}\left[(1-\mu_1)x_1-i_o\right] \quad (3)$$

where, x_1, x_2, x_3, and μ_1 represent the PV current (i_L), mean values of input voltage (V_{pv}), PV output voltage (V_{opv}), and control signal (μ_1), respectively. C_{PV}, C_{dc}, L, I_{PV}, R, and i_o are the PV voltage stabilization capacitor, DC voltage stabilization capacitor, inductance, output current of PV cell, resistance, and output current of PV converter, respectively.

2. 2. Modeling the Wind Power System Figure 3 illustrates the wind power supply mechanism performed by a permanent magnetic synchronous generator. In this figure, an intermediate rectifier is used between the generator and the DC bus plus the boost converter. Since the output power changes according to the wind speed change, the output voltage should be adjusted to the desired level [18]:

$$\frac{dx_4}{dt}=\frac{V_{in}}{L_w}-\frac{R_w}{L_w}x_4-(1-\mu_2)\frac{x_5}{L_w} \quad (4)$$

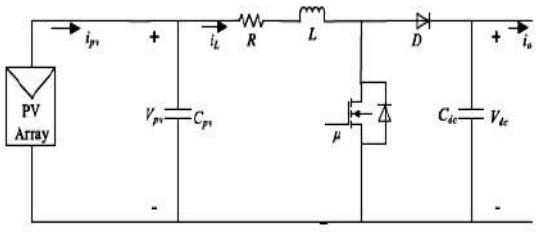


Figure 2. PV source converter circuit diagram

$$\frac{dx_5}{dt}=(1-\mu_2)\frac{x_4}{C_w}-\frac{I_{owind}}{C_w} \quad (5)$$

where, x_4, x_5, and μ_2 are the mean values of wind source current (i_w), wind source output voltage (V_w), and control signal (μ_2), respectively. I_{owind} and V_{in} denote the output current of the converter and voltage generated by the wind source, respectively. In addition, this model includes L_w input inductor with R_w series resistance, D diode (rectifier), C_w output capacitor, and IGBT switch.

2. 3. The Dynamic Models of Diesel Generator and Turbine's Controller In this paper, a two-axis model of the diesel generator is implemented. The dynamic model contains both mechanical and electrical dynamics. The excitation system is the main part of diesel generator where an excitation controller is used to maintain the terminal voltage at the desired level. The turbine control system adjusts the generator power output.

The complete dynamics model of the diesel generator with the presence of mechanical and electrical structures as well as the dynamics of the turbine and excitation system can be represented by the following equations [19]:

$$\frac{dx_6}{dt}=x_7+\omega_0 \quad (6)$$

$$\frac{dx_7}{dt}=-\frac{D}{2H}(x_7-\omega_0)+\frac{\omega_0}{2H}x_{11}-\frac{\omega_0}{2H}\left(x_8I_q+x_9I_d\right) \quad (7)$$

$$\frac{dx_8}{dt}=-\frac{1}{T'_{do}}x_8-\frac{(x_d-x'_d)}{T'_{do}}I_d+\frac{1}{T'_{do}}(x_{10}) \quad (8)$$

$$\frac{dx_9}{dt}=-\frac{1}{T'_{qo}}x_9+\frac{(x_q-x'_q)}{T'_{qo}}I_q \quad (9)$$

$$\frac{dx_{10}}{dt}=-\frac{x_{10}}{T_A}-\frac{K_A}{T_A}(V_{ref}+V_c-V_t) \quad (10)$$

Turbine's controller system:

$$\frac{dx_{11}}{dt}=-\frac{x_{11}}{T_T}+\frac{K_T}{T_T}x_{12} \quad (11)$$

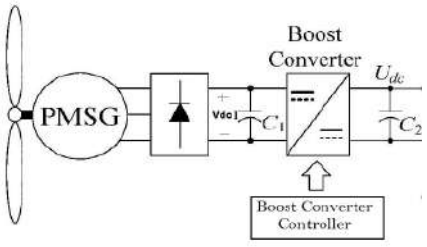


Figure 3. The diagram of wind circuit and its converter

$$\frac{dx_{12}}{dt} = -\frac{x_{12}}{T_G} + \frac{K_G}{T_G}P_C - \frac{K_G}{R_r\omega_0 T_G}x_7 \quad (12)$$

where, x_6, x_7, x_8, x_9, x_{10}, x_{11}, and x_{12} are the values of the angular load (δ), rotor velocity (ω), leakage flux field variable of q axis (E_q), leakage flux variable of d axis (E_d), excitation field voltage (E_{fd}), mechanical power (P_m), and reactance distribution (X_m). In addition, other parameters have their usual definitions as discussed by Abdullah et al. [19].

2. 4. Modeling the Combined Energy Storage System The storage device connected to the DC bus is shown in Figure 4. The combined system includes a battery and a capacitor bank. The battery connects to the DC bus through a boost-buck converter, consisting of an inductor, Rbat series resistor, output filter capacitor, and two IGBT switches [1]. Both switches are controlled through their signal. The converter is used to adjust the DC bus voltage. It can operate in a charge and discharge mode depending on the load demand. Further, the capacitor bank (CB) is connected to the DC bus through the boost converter. These two states are expressed mathematically as follows.

$$Z = \begin{Bmatrix} 1, & if\ (i_{batref} > 0) \\ 0, & if\ (i_{batref} < 0) \end{Bmatrix}, Y = \begin{Bmatrix} 1, & if\ (i_{bcref} > 0) \\ 0, & if\ (i_{bcref} < 0) \end{Bmatrix} \quad (13)$$

where, i_{batref} is the battery's reference current generated through controlling its level based on electricity demand. If $i_{batref} > 0$ and $i_{bc} > 0$, then the battery converter is modeled by the following differential equations:

$$\begin{aligned} \frac{di_{bat}}{dt} &= \frac{V_{bat}}{L_{bat}} - \frac{R_{bat}}{L_{bat}}i_{bat} - (1-\mu_3)\frac{V_{dc}}{L_{bat}} \\ \frac{di_{CB}}{dt} &= \frac{V_{CB}}{L_{CB}} - \frac{R_{CB}}{L_{CB}}i_{CB} - (1-\mu_5)\frac{V_{dc}}{L_{CB}} \end{aligned} \quad (14)$$

where, i_{bat}, i_{CB}, μ, V_{bat}, and V_{dc} represent the battery output current, capacitor bank output current, control signal, battery voltage, and DC bus voltage, respectively. In addition, R_{bat}, L_{bat}, and C_{dc}, are the series resistance with battery source, the inductor series with battery source, the converter output capacitor. If $i_{batref} < 0$ and $i_{CBref} < 0$, then the differential equations are stated as follows:

$$\begin{aligned} \frac{di_{bat}}{dt} &= \frac{V_{bat}}{L_{bat}} - \frac{R_{bat}}{L_{bat}}i_{bat} - (\mu_4)\frac{V_{dc}}{L_{bat}} \\ \frac{di_{CB}}{dt} &= \frac{V_{CB}}{L_{CB}} - \frac{R_{CB}}{L_{CB}}i_{CB} - (\mu_6)\frac{V_{dc}}{L_{CB}} \end{aligned} \quad (15)$$

In order to simplify the system, a virtual control signal is introduced as follows:

$$\begin{aligned} \mu_{Bat} &= [K - (1-\mu_3) + (1-K)\mu_4] \\ \mu_{CB} &= \left[L(1-\mu_5) + (1-L)\mu_6\right] \end{aligned} \quad (16)$$

Finally, Equations (14), (15), and (16) lead to the following equations.

$$\frac{dx_{13}}{dt} = \frac{V_{bat}}{L_{bat}} - \frac{R_{bat}}{L_{bat}}x_{13} - \mu_{Bat}\frac{x_3}{L_{bat}} \quad (17)$$

$$\frac{dx_{14}}{dt} = \frac{V_{CB}}{L_{CB}} - \frac{R_{CB}}{L_{CB}}x_{14} - \mu_{CB}\frac{x_3}{L_{CB}} \quad (18)$$

where, x_{13} and x_{14} denote the mean values of (i_{bat}) battery current and the capacitor bank current (i_{CB}), while μ_{Bat} and μ_{CB} are the mean values of control signals.

2. 5. Bidirectional Converter Model of DC Bus Interface, AC Bus, and Filter The interface converter between AC and DC sections should have both rectifier and inverter capabilities. If the output power of the AC side is less than its power consumption, it would be necessary to receive part of the required power through the DC side. In this case, the converter acts as an inverter. Also, if the output power of the DC side is low, it is necessary to receive part of the required power through the AC side. In this case, the converter functions as a rectifier. The converter receives the changes according to DC voltage and AC voltage as well as frequency and acts accordingly. The following equations indicate the relationships between the various variables of the system [20, 21].

$$\begin{aligned} \frac{dx_{15}}{dt} &= -\frac{R}{L}x_{15} + x_7 x_{16} + u_{d1} \\ \frac{dx_{18}}{dt} &= -\frac{R}{L}x_{18} + x_7 x_{19} + u_{d2} \end{aligned} \quad (19)$$

$$\begin{aligned} \frac{dx_{16}}{dt} &= -x_7 x_{15} - \frac{R}{L}x_{16} - u_{q1} \\ \frac{dx_{19}}{dt} &= -\frac{R}{L}x_{19} - x_7 x_{18} + u_{q2} \end{aligned} \quad (20)$$

$$\begin{aligned} \frac{dx_{17}}{dt} &= \frac{3u_{sq1}x_{16}}{2C\,x_{17}} - \frac{i_L}{C} \\ \frac{dx_{20}}{dt} &= \frac{3u_{sq2}x_{19}}{2C\,x_{20}} + \frac{i_L}{C} \end{aligned} \quad (21)$$

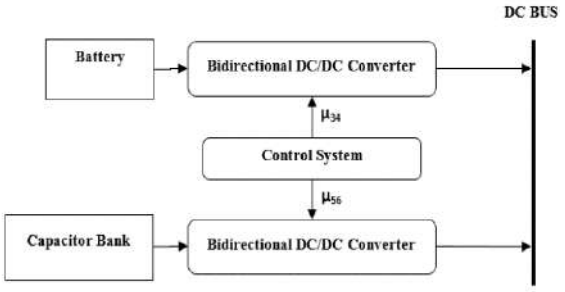


Figure 4. Combined storage system model and control algorithm

where, u_{sd1}, u_{sd2}, u_{sq1}, and u_{sq2} are the components of the d and q axes of the source voltage related to both rectifier and inverter modes, respectively. Further, x_{15}, x_{16}, x_{18}, and x_{19} are the line current components while u_{rd1}, u_{rd2}, u_{rq1}, and u_{rq2} are the input voltage components of the converter in both inverter and rectifier modes.

3. NONLINEAR CONTROL METHOD

The backstepping nonlinear control method is developed based on the dynamic model of each microgrid component. PV, wind, energy storage, and synchronous generator resources along with the corresponding converter are controlled by their respective control system. The most important part of this model is the bidirectional converter. This converter is controlled through the signals of voltage and frequency on both sides. The control method for each section is shown in Figure 5. In this model, the resources are controlled based on backstepping method. Each production source and interface converter will be controlled independently. Specifically, the sum of the error of each unit becomes zero. The error of each unit independently tends to zero. Hence, the derivative of the sum of Lyapunov's functions also tends to be zero. In this way, both the error of each unit and the error of the entire system are zero. This indicates the stability of the system whereby the system will be practically stable. The main role of stability is played by the control variable. Each unit generates a control signal that always sets the system error to zero. To prove the equation of the control variable, the derivative of the Lyapunov energy function is set to zero where the equation of the control variable is obtained. Next, initially, the error of each unit is compared with its reference value after which it is solved using the backstepping method.

In this method, the error of each unit is measured first and then compared with the base value. Thereafter, via the backstepping method, the system output is stabilized by the control and input changes. In the proposed model, the sources are controlled independently and by their own error. Practically, the main criterion is the voltage of DC bus as well as voltage and frequency of the AC bus. The value of error in each unit is measured by its own control system, which tends to zero. It can be stated that the total errors of the entire system will be equal to zero. The design method is summarized as follows:

1) Calculating the errors for each control target or state variable.
2) Calculating the dynamics of the error.
3) Calculating Lyapunov function and its derivatives based on errors.
4) Repeating the first three steps until the control rules appear.
5) Calculating control rules to stabilize the error dynamics in the last step.
6) Stability analysis of the entire system with the derived control rules.

Load supply is the most important issue in hybrid microgrids. Voltage changes are a good operational indicator for controlling load changes in DC bus. To evaluate these changes, the power balance on the DC side must first be checked. Accordingly, the amount of net power (P_{net}) in the microgrid can be defined as follows:

$$P_{net} = P_w + P_{pv} + P_{ESS} - P_{loss} - P_{load,DC} \pm P_{transfer} \quad (22)$$

P_{pv}, P_w, and P_{ESS} are the output power of PV, wind source, and storage system, respectively, $P_{Load,DC}$ is the power required for the DC side load, $P_{transfer}$ denotes the power exchanged between the AC and DC bus, and P_{loss} reflects the system loss. Under these circumstances, P_{ESS} plays an important role in maintaining the microgrid stability. The dynamics of the DC bus voltage in a microgrid can be adjusted based on the following principle of power balance:

$$CV_{dc}\frac{dV_{dc}}{dt} = P_{net} \ , \ P_{DC,TOT} = V_{dc} i_{dc,TOT} \quad (23)$$

where, V_{dc} , $i_{dc,ToT}$, and $P_{dc,ToT}$ represent DC bus voltage, sum of the current injected into the DC bus by the sources, and the sum of DC power, respectively. C capacitance is equivalent to the output for all converters. From the above equation, it can be seen that the DC bus voltage depends on the power balance in this bus and the elevation or reduction of DC bus voltage indicates excess or lack of power. To adjust the DC bus voltage, the power balance in the DC bus must be maintained. PV, wind, and storage sources first receive the error relative to the reference value and then generate a proportional control signal to compensate for this error. If the DC side sources cannot respond to the load increase, the power required from the AC side is provided by a bidirectional interface converter.

3. 1. PV Control The main goal is to control the DC output voltage of the DC-DC converter by changing the duty cycle in relation to the solar radiation variations. First, the error equation of each state variable is calculated with respect to the reference value. The design method of PV control scheme and connected converter goes through the following steps:
Step 1: Depending on the purpose of the design, the first tracking errors are defined as follows:

$$e_1 = x_2 - V_{PV(ref)} \quad (24)$$

The dynamics of e_1 is as follows:

$$\frac{de_1}{dt} = \frac{1}{C_{PV}}(i_{PV} - x_1) - \frac{dV_{PV(ref)}}{dt} \quad (25)$$

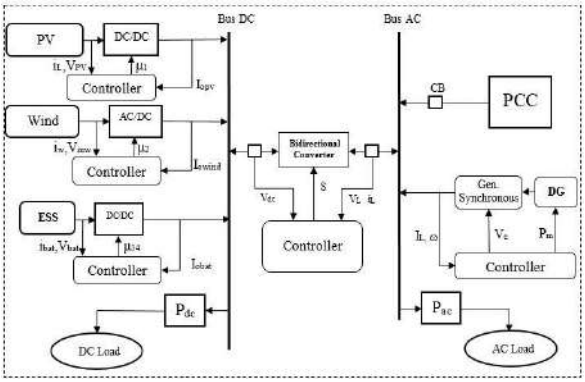


Figure 5. The control system model

where, x_1 is a stabilizing function, for which the Lyapunov function is:

$$V_1 = \frac{1}{2}e_1^2 \tag{26}$$

Incorporating Equation (25) into Equation (26) yields:

$$\frac{dV_1}{dt} = e_1\left(\frac{1}{C_{PV}}\left(i_{PV} - x_1\right) - \frac{dV_{PV\,(ref)}}{dt}\right) \tag{27}$$

To ensure the dynamic stability the derivative of e_1, the derivative of V_1 should be a definite or semi-definite negative. To achieve this, the system parameters are set as follows:

$$\frac{1}{C_{PV}}\left(i_{PV} - x_1\right) - \frac{dV_{PV\,(ref)}}{dt} = -c_1 e_1 \tag{28}$$

Where, c_1 is used to set the output response. Embedding Equation (28) into Equation (27) yields:

$$\frac{dV_1}{dt} = -c_1 e_1^2 \le 0 \tag{29}$$

The combined value α of i_L can be determined from Equation (29) as follows:

$$\alpha = C_{PV}\left(\frac{dV_{PV(ref)}}{dt} - c_1 e_1\right) - i_{PV} \tag{30}$$

Step 2: To achieve zero tracking error, the i_L inductor current should be equal to α. Hence, another error variable is defined as follows:

$$e_2 = x_1 - \alpha \tag{31}$$

Derivation of Equation (31) leads to the dynamics of the error as follows:

$$\frac{de_2}{dt} = \frac{1}{L}\left(-Rx_1 + x_2 - \left(1-\mu_1\right)x_3\right) - \frac{d\alpha}{dt} \tag{32}$$

In this case, Lyapunov second function is equal to:

$$V_2 = V_1 + \frac{1}{2}e_2^2 \tag{33}$$

its derivative is:

$$\frac{dV_2}{dt} = -c_1 e_1^2 + c_2\frac{de_2}{dt} \tag{34}$$

The stability of the dynamic error (derivative of e_2) is obtained only if the derivative of V_2 is less than or equal to zero. Hence,

$$\frac{de_2}{dt} = \frac{1}{L}\left(-Rx_1 + x_2 - \left(1-\mu_1\right)x_3\right) - \frac{d\alpha}{dt} = -c_2 e_2 \tag{35}$$

Using Equations (34) and (35), the main function is determined as follows:

$$\frac{dV_2}{dt} = -c_1 e_1^2 + c_2 e_2^2 \le 0 \tag{36}$$

Which is a definite or semi-definite negative. The proposed control law, along with the sustainability analysis of the entire PV system, is shown in the next step.

Step 3: Since the condition shown by Equation (35) is not correct, it is necessary to analyze the overall stability of the system. For this purpose, the final error based on Equation (36) is defined as follows:

$$e_3 = \frac{1}{L}\left(-Rx_1 + x_2 - \left(1-\mu_1\right)x_3\right) - \frac{d\alpha}{dt} + c_2 e_2 \tag{37}$$

The dynamics of which is written as follows:

$$\frac{de_3}{dt} = \frac{1}{L}\begin{pmatrix} -R\dfrac{dx_1}{dt} + \dfrac{dx_2}{dt} + \dfrac{d\mu_1}{dt}x_3 \\ -\left(1-\mu_1\right)\dfrac{dx_3}{dt} \end{pmatrix} - \frac{d^2\alpha}{dt^2} + c_2\frac{de_2}{dt} \tag{38}$$

At this stage, Lyapunov function is presented as follows:

$$V_3 = V_2 + \frac{1}{2}e_3^2 \tag{39}$$

By substituting the derivative of V_2 of Equation (36) and $\dot{e}_3$ of Equation (38) in Equation (39), the derivative V_3 is stated as follows:

$$\frac{dV_3}{dt} = c_3 e_3\begin{pmatrix} \dfrac{1}{L}\begin{pmatrix} -R\dfrac{dx_1}{dt} + \dfrac{dx_2}{dt} \\ +\dfrac{d\mu_1}{dt}x_3 - \left(1-\mu_1\right)\dfrac{dx_1}{dt} \end{pmatrix} \\ -\dfrac{d^2\alpha}{dt^2} + c_2\dfrac{de_2}{dt} \end{pmatrix} - c_1 e_1^2 - c_2 e_2^2 \tag{40}$$

Now, the final control law is selected as follows:

$$\frac{d\mu_1}{dt}=-\frac{1}{x_3}\left(\begin{array}{l}\frac{1}{L}\left(\begin{array}{l}-R\frac{dx_1}{dt}+\frac{dx_2}{dt}+\frac{d\mu_1}{dt}x_3\\-(1-\mu_1)\frac{dx_3}{dt}\end{array}\right)\\-\frac{d^2\alpha}{dt^2}+c_2\frac{de_2}{dt}+c_2e_2\end{array}\right) \tag{41}$$

where, $0<\mu_1<1$, then the presented control law is used to generate the duty cycle of the PV converter. Similar steps are used to obtain control rules for the bidirectional DC/DC converter of the storage system plus the excitation system of synchronous generators. Accordingly, the derivative of these control laws is discussed in the following sections without repeating these steps.

3. 2. DC/DC Converter Control for Wind Energy Integration The tracking error for the wind converter is expressed as follows:

$$e_4=x_4-I_{ref} \tag{42}$$

$$e_5=x_5-\sigma \tag{43}$$

The first function of the Lyapunov V_4 is selected to analyze the stability of the system as follows:

$$V_4=\frac{1}{2}e_4^2 \tag{44}$$

The derivative of V_4 should be semi-definite for the system to be completely stable ($V_4\leq 0$). To ensure the uninterrupted stability of the wind power system, the second combined set of V_5 sets is written as follows:

$$V_5=V_4+\frac{1}{2}e_5^2 \tag{45}$$

By taking the derivative V_5 and substituting the value V_4 of Equation (41), we have:

$$\frac{dV_5}{dt}=-c_4e_4^2+e_5\left(\frac{de_5}{dt}+\frac{(1-\mu_2)e_4}{L_w}\right) \tag{46}$$

Substituting e_5 of Equation (43) into Equation (46), we have:

$$\begin{aligned}\frac{dV_5}{dt}=&-c_4e_4^2+e_5\left(\frac{(1-\mu_2)x_4}{C_w}\right)\\&+e_5\left(-\frac{1}{(1-\mu_2)}-\frac{I_{owind}}{C_w}\right)\\&+e_5\left(-c_5e_5-\frac{\sigma}{(1-\mu_2)}\cdot\frac{d\mu_2}{dt}+\frac{(1-\mu_2)e_4}{L_w}\right)\end{aligned} \tag{47}$$

To keep the wind system dynamics constant, V_5 should run as follows:

$$\frac{dV_5}{dt}=-c_4e_4^2+c_5e_5^2 \tag{48}$$

$$\sigma=\frac{L_w}{(1-\mu_2)}\left(c_4e_4+\frac{V_{in}}{L_w}-\frac{dI_{ref}}{dt}-\frac{R_w}{L_w}x_4\right) \tag{49}$$

where $c_4>0$ and $c_5>0$ are the constant parameters. The control signal for μ_2 is obtained as follows:

$$\begin{aligned}\frac{d\mu_2}{dt}=&\frac{1}{\sigma}\left(\frac{e_4(1-\mu_2^2)}{L_w}\right)\\&+\frac{1}{\sigma}\left(-e_5(c_4-c_5)(1-\mu_2)+\frac{(1-\mu_2)^2x_4}{C_w}\right)\\&+\frac{1}{\sigma}\left(-\frac{(1-\mu_2)I_{owind}}{C_w}\right)\end{aligned} \tag{50}$$

where, $0<\mu_2<1$. Then the presented control law is employed to generate the time cycle of the wind turbine energy converter. The stability control system ensures asymptotic stability and convergence of error dynamic to zero.

3. 3. Controlling DC/DC Converters for Storage Source The control process of the storage source is presented as follows: 1. Adjusting the direct current bus voltage of the microgrid under the variations of load demand, 2. Tracking battery's fast current and capacitor bank current to their reference values, 3. Analyzing the asymptotic stability of the storage resource.

It is not possible to directly track the V_{dc} voltage from the V_{dcref} value. Thus, the indirect voltage regulation method is used to achieve this goal where I_{bat} is tracked by its reference value, obtained as follows:

$$I_{batref}=\left(\frac{V_{dcref}x_{13}-V_{CB}I_{CBref}}{\Phi V_{bat}}\right) \tag{51}$$

where, Φ represents the switching losses coefficient in the converters whose value should be greater than 1. To achieve control goals, the tracking error is defined as follows:

$$e_6=x_{13}-I_{batref}\ ,\ e_7=\frac{x_{13}}{L_{bat}}-\beta_0 \tag{52}$$

The chosen Lyapunov functions, V_6 and V_7, for analyzing the stability of the system are as follows:

$$V_6=\frac{1}{2}e_6^2\ ,\ V_7=V_6+\frac{1}{2}e_7^2 \tag{53}$$

The virtual control β_0 should be written as follows:

$$\beta_0=\frac{1}{\mu_{bat}}\left(c_5e_6+\frac{V_{bat}}{L_{bat}}-\frac{R_{bat}}{L_{bat}}x_{13}-\frac{dI_{batref}}{dt}\right) \tag{54}$$

The μ_{Bat} control law is defined as follows:

$$\frac{d\mu_{Bat}}{dt}=\frac{1}{\beta_0}\left[e_7\mu_{Bat}^2-\mu_{Bat}^2\frac{R_{bat}}{x_{13}C_{dc}}-\mu_{Bat}\frac{R_{bat}}{C_{dc}}\right]+\frac{1}{\beta_0}\left[\mu_{Bat}\frac{x_{13}}{L_{bat}C_{dc}}-\mu_{Bat}c_6e_6\right] \quad (55)$$

Similarly, the signal control converter is obtained in the capacitive bank system as follows:

$$\frac{d\mu_{CB}}{dt}=\frac{L_{CB}}{x_{14}}\left(\frac{V_{CB}}{L_{CB}}-\frac{R_{CB}}{L_{CB}}x_{14}-c_7e_7-\frac{dI_{CBref}}{dt}\right) \quad (56)$$

Where, $0<\mu_{Bat}<1$ and $0<\mu_{CB}<1$ represent the control law used to create a time cycle for the bidirectional converter connected to the battery. $\beta_0\neq0$ and $c_{6\&7}>0$ are the gain of controller in the controller design, and the μ_{CB} is employed to generate a time cycle of the bidirectional converter connected to the capacitor bank.

3. 4 Excitation Control and Backstepping Valve Steam for Synchronous Generator Applying the proposed backstepping control scheme can contribute to control the excitation system and steam valve. The chosen Lyapunov function, V_8, for analyzing the stability of the system is as follows:

$$V_8=\frac{1}{2}e_8^2 \quad (57)$$

The excitation control input is defined as:

$$V_c=\frac{x_{10}}{K_A}-(V_{ref}-V_t)+\frac{T_A}{K_A}\cdot\frac{d\beta_5}{dt}-\frac{T_A}{K_A}c_8e_8 \quad (58)$$

Also, the steam valve position control input is:

$$P_c=-\frac{T_G}{K_G}\left(-\frac{x_{12}}{T_G}+\frac{K_Gx_7}{R_r\omega_0T_G}-\frac{d\beta_6}{dt}+c_8e_9\right) \quad (59)$$

where:

$$\beta_1=-c_7e_8\ ,\ \beta_2=\frac{T_T}{K_T}\left(\frac{x_{11}}{T_T}-e_8c_8\right),$$

$$\beta_3=\frac{x_{11}}{I_q}-\frac{D}{I_q\omega_0}(x_7-\omega_0)+\frac{2H}{I_q\omega_0}(x_7-\omega_0)+\frac{2H}{I_q\omega_0}\cdot\frac{d\beta_1}{dt},$$

$$\beta_4=\frac{2H}{I_q\omega_0}c_8e_8$$

$$\beta_5=x_8+(x_d-x_d')I_d+T_{do}'\beta_3-T_{do}'c_8e_8\ ,$$

$$\beta_6=\frac{x_{11}}{I_q}-\frac{D}{I_q\omega_0}(x_6-\omega_0)+\frac{2H}{I_q\omega_0}(x_6-\omega_0)+\frac{2H}{I_q\omega_0}\cdot\frac{d\beta_1}{dt}$$

where, c_1-c_8 are positive constant parameters and e_1-e_8 are error variables.

3. 5. Bidirectional Converter Control between Two Sides If the output power of the DC side is less than its power consumption, it is necessary to receive part of the required power through the AC side. In this case, the converter acts as a rectifier. There are two controlled variables: i_{abc} and V_{DC}. Based on Equations (19)-(21), a controlling mathematical model can be expressed. First, the equations for the rectifier mode are described as below:

$$\frac{dx_r}{dt}=f_r(x)+g_ru_r,\quad y=h(x) \quad (60)$$

where, x is the rectifier state vector, u_r denotes the rectifier control input vector, y represents the rectifier output vector, while f and g are smooth vector fields.

$$X_r=\begin{bmatrix}x_{15}\\x_{16}\\x_{17}\end{bmatrix},f_r(x)=\begin{bmatrix}f_{r1}\\f_{r2}\\f_{r3}\end{bmatrix}=\begin{bmatrix}-\frac{R}{L}x_{15}+x_7x_{16}\\-x_7x_{15}-\frac{R}{L}x_{16}\\\frac{u_{sq1}x_{16}}{2Cx_{17}}-\frac{i_1}{C}\end{bmatrix},$$

$$g_r=\begin{bmatrix}\frac{1}{L}&0\\0&\frac{1}{L}\\0&0\end{bmatrix},\ u_r=\begin{bmatrix}u_{r1}\\u_{r2}\end{bmatrix}=\begin{bmatrix}u_{sd1}-u_{rd1}\\u_{sq1}-u_{rq1}\end{bmatrix} \quad (61)$$

The system has two control inputs in this case. Since some variables affect the system performance directly, it is better to use their effects for the feedback control law. The chosen Lyapunov function, V_9, for analyzing the stability of the system is as follows.

$$V_9=\frac{1}{2}e_9^2 \quad (62)$$

Hence, the output is chosen as:

$$y_1=i_{d1}\ ,\ y_2=v_{dc1} \quad (63)$$

The differential equations as well as the inputs and outputs are obtained based on Equations (60) and (62). The direct relationship between the inputs and outputs is easily obtained as below:

$$\begin{bmatrix}\frac{dy_1}{dt}\\\frac{d^2y_2}{dt^2}\end{bmatrix}=B_r+A_r\begin{bmatrix}u_{r1}\\u_{r2}\end{bmatrix} \quad (64)$$

$$B_r=\begin{bmatrix}f_{r1}\\\frac{3u_{sq1}}{2Cx_{17}^2}f_{r2}-\frac{3u_{sq1}x_{16}}{2Cx_{17}^2}f_{r3}-\frac{i_L}{C}\end{bmatrix},$$

$$A_r=\begin{bmatrix}\frac{1}{L}&0\\0&\frac{3u_{sq1}}{2CLx_{17}}\end{bmatrix}. \quad (65)$$

If the AC side output power is less than its power consumption, it is necessary to receive part of the required power through the AC side. In this case, the converter acts as an inverter. Thus, its state-space equations are of the second-order and with two inputs.

$$\frac{dx_i}{dt} = f_i(x) + g_i u_i, \tag{66}$$

Here we have:

$$X_i = \begin{bmatrix} x_{18} \\ x_{19} \end{bmatrix}, \quad f_i(x) = \begin{bmatrix} f_{i1} \\ f_{i2} \end{bmatrix} = \begin{bmatrix} -\frac{R}{L}x_{18} + x_7 x_{19} \\ -x_7 x_{18} - \frac{R}{L}x_{19} \end{bmatrix},$$
$$g_r = \begin{bmatrix} \frac{1}{L} & 0 \\ 0 & \frac{1}{L} \\ 0 & 0 \end{bmatrix}, \quad u_i = \begin{bmatrix} u_{i1} \\ u_{i2} \end{bmatrix} = \begin{bmatrix} u_{sd2} - u_{rd2} \\ u_{sq2} - u_{rq2} \end{bmatrix}, \tag{67}$$

where, u_{sd2}, u_{sq2}, u_{rd2}, u_{rq2}, i_{d2}, and i_{q2} denote the inverter state electrical parameters. Then, the output is chosen as:

$$w_1 = i_{d2}, \quad w_2 = i_{q2}. \tag{68}$$

Control inputs are displayed as follows:

$$\begin{bmatrix} u_{i1} \\ u_{i2} \end{bmatrix} = \begin{bmatrix} w_{1ref} - k_{13}e_{i1} - k_{32}\int e_{i1}dt \\ w_{2ref} - k_{41}e_{i2} - k_{42}\int e_{i2}dt \end{bmatrix} \tag{69}$$

Figure 6 indicates the nonlinear design diagram of the bidirectional converter.

3. 7. Stability Analysis The stability of a system refers to the system's ability to return back to its steady state after disturbance. The stability of isolated hybrid microgrids has been of major concern in electrical distribution systems. The stability of isolated hybrid microgrids determines whether the generation system can settle down to a new or original steady state once transients disappear. The stability of the distributed generation system depends on the loading and sudden changes in load which can cause the instability. Resources are controlled independently by their own error. The value of error in each unit is measured by its own control system, which tends to zero. Thus, practically the sum of the errors of the entire system will be equal to zero. In order to analyze the stability of a hybrid microgrid, the total energy functions of the system must be examined. These functions are as follows:

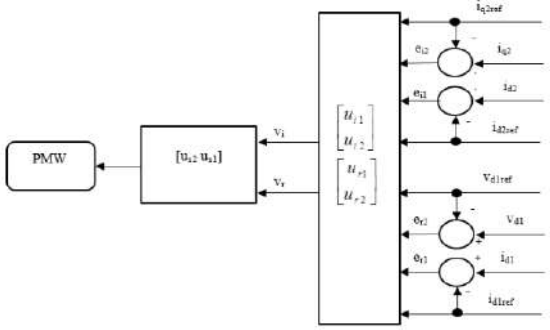


Figure 6. Nonlinear design diagram of the bidirectional converter

$$V_{total} = V_3 + V_5 + V_7 + V_8 + V_9 + e_8c_8 + e_9c_9 + (k_{31} + k_{32})e_{10} + (k_{41} + k_{42})e_{11} \tag{70}$$

$$\dot{V}_{total} = \dot{V}_3 + \dot{V}_5 + \dot{V}_7 + \dot{V}_8 + \dot{V}_9 + \dot{e}_8c_8 + \dot{e}_9c_9 + (k_{31} + k_{32})\dot{e}_{10} + (k_{41} + k_{42})\dot{e}_{11} \le 0 \tag{71}$$

The V_{total} must be negative semi-definite, which is calculated as below:
It can be observed that above equation is negative semi-definite. Thus, the stability of the entire hybrid microgrid system can be guaranteed. In other words, in cause of occurrence of a disturbance, the system will remain stable after the transition.

4. POWER CONTROL ALGORITHM

Power division or control between both sides is considered as the most important issue in AC/DC combined microgrids. Indeed, all resources, load, and bidirectional interface converters between both areas should be involved in this management. It is first assumed that resources, loads, and converters are controlled in a decentralized manner. Thus, a system is practically required to distribute the power. Then, if there is not enough generation on one side, the power should be compensated on the other side, and if the system power is not sufficient, the load must be cut off or reduced. Figure 7 depicts the power control algorithm. This algorithm is compatible with power converters connected to power resources by providing reference signals. The power equation in such a system is defined as below:

$$P_w + P_{pv} - P_{loss} = P_{load} + P_{ESS} \pm P_{transfer} \tag{72}$$

where, P_W, P_{pv}, P_{loss}, P_{load}, P_{ESS}, and $P_{transfer}$ are the wind power, PV power, power loss, load power, energy storage resource power, and diesel generator power, respectively.

Note that load changes on the AC side are practically modelled according to the power transmitted from or to the DC. The resources use a battery and a capacitor bank that can charge and discharge as well as support part of the system in different modes. Thus, both resources are employed to boost stability and provide an algorithm to control power. The predicted modes are as follows:

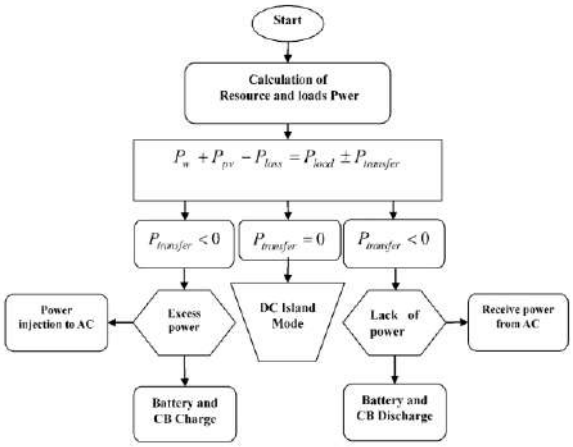


Figure 7. Energy management algorithm

1. The power in DC is surplus.
2. The power in DC is not enough to supply the load.
3. When the DC system gets disconnected due to fault.

In the first mode, the power generated by the resources on DC is greater than that of the load required in this sector. Hence, the surplus power is spent on recharging the energy source. The equations on DC are as follows.

$$P_{transfer} = P_{load} - \left(P_w + P_{pv} - P_{loss}\right) < 0$$
$$P_w + P_{pv} - P_{loss} = P_{load} \pm P_{transfer} \quad (73)$$

$P_{transfer}$ is the power given or taken from the AC. If it is negative, it means the power in DC is surplus. In this mode, both the distributed resources of generation and the energy storage system will share the load demand. The surplus power is utilized to charge the energy storage source or is sent to the other side. The battery does not change until it reaches 80%. In the second mode, the power generated by the resources on DC is less than the power required by the load in this section. Thus, there is a power shortage on this side, which should be supplied through an energy storage source from AC. The condition for power supply by the storage source is that it should be charged. The minimum charging capacity for power injection is 20%. The equations on DC are as follows.

$$P_{transfer\,2} = P_{load} - \left(P_w + P_{pv} - P_{loss}\right) > 0 \quad (74)$$

In the third mode, AC is separated from DC due to a fault or repair. In this case, there is no exchanging power. Thus, the DC system should supply the power required for this side. If the power is not sufficient to supply the load, the storage system will supply at least 20% of the load capacity if it is charged; otherwise, the load should be cut off. If the power is surplus, the storage system can start the charging process for optimal use of the energy, where charging continues up to 80%.

5. NUMERICAL STUDIES

Figure 8 displays the combined microgrid including AC, DC, and DC resources, load, and load bus along with the main interface converter between both sides. The AC voltage is 600 V and the frequency is 60 Hz, while the DC voltage is equal to 400 V. In AC and DC, there are potential resources such as wind and PV. The line resistance and inductance between the buses are considered in AC. In this section, sustainability is analyzed on a combined network. Initially, stability analysis is performed on a hybrid network. AC and DC power output are 3 MW [22].

As mentioned, this study aims to survey if there is a combined network of AC/DC resources and loads that can provide their loads for each separate region. Meanwhile, both regions of this network are connected through a bidirectional interface converter. In addition, the system can operate in the network and islanded mode. These islands occur on any side. Eigenvalue analysis is widely used to evaluate dynamic stability in power systems. To find system eigenvalues, system operating points must be obtained by load distribution analysis or via time domain simulation. The operational points of the stable simulation mode are obtained through the MATLAB Simulink. Figure 9 indicates the diagram of the Eigenvalue of AC/DC microgrid. In addition, the eigenvalue of the AC/DC system is shown in Figures 10 and 11.

Higher droop values in AC microgrids are required to share the appropriate power between resources as well as to improve the system transient response. However, a higher droop in DC microgrid leads to increase power-sharing, voltage drop, and stability reduction. The main goal is to control and maintain the voltage stability and system frequency as fast as possible. The loads on both sides should be supplied with a stable voltage and frequency. In the islanded mode, the system stability should be studied when a fault occurs on the generation side or when the load suddenly changes. In case of reduction of production or increase of load, power exchange occurs in such a way that the network has the minimum voltage and frequency changes. Further, the stability impact on each side is reduced compared to the other side.

In this study, nonlinear control is used for stability analysis in AC/DC hybrid microgrid with bidirectional interface converter. To evaluate the performance of this model, comparisons between the performance of this model and the linear control method proposed by Dheer et al. [12] are presented, in different operating scenarios.

Some scenarios have been explored to evaluate the effectiveness of the proposed control system.

1) Combined microgrid equilibrium mode, where the load power of each side is less than the production power;
2) Changing load and power exchange from AC to DC and vice versa;

Separation of AC side from the DC side.

In the first scenario, the microgrid is initially operated in the islanded mode. In this case, the load power of AC and DC is equal to 2.5 and 1 MW, respectively. In the first second, the load power of each AC and DC side grows by 0.5 MW. The control system of each source raises its output power by changing the voltage on the DC side as well as the voltage and frequency on the AC side to augment the load. Since the generating power capacity of each side is equal to 3 MW and the load power to the AC and DC side has increased by 3 and 1.5 MW, respectively. The load power is supplied by the same side sources and will not pass through the power interface converter. In this state, the hybrid microgrid is normal and the independent control system of each source can boost production and supply load on each side.

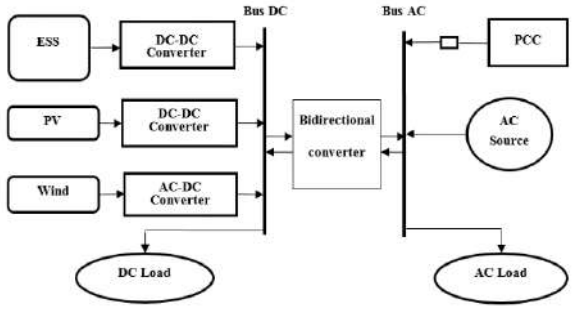


Figure 8. The diagram of AC/DC microgrid model

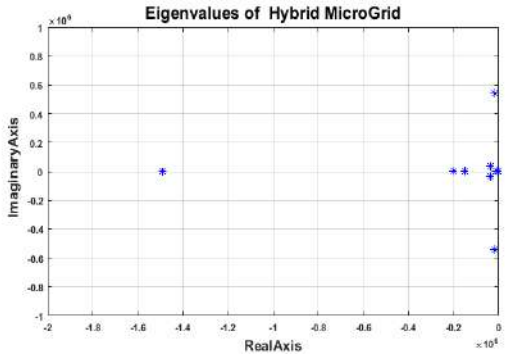


Figure 9. The eigenvalue of hybrid AC/DC microgrid

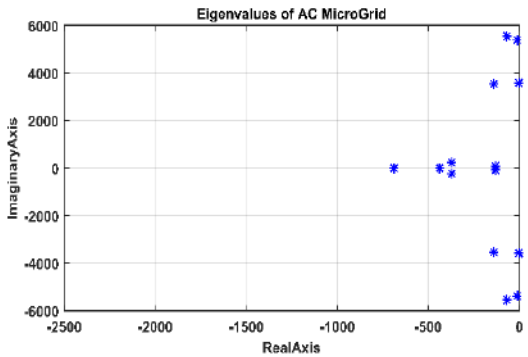


Figure 10. The eigenvalue of AC in the hybrid microgrid

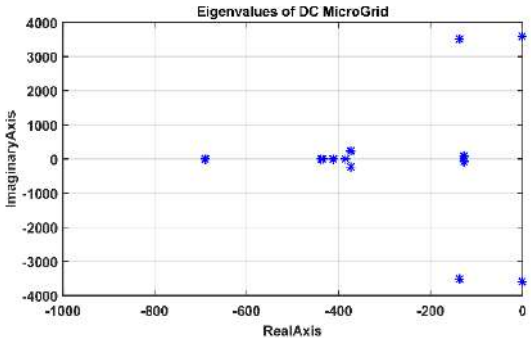


Figure 11. The eigenvalue of DC in the hybrid microgrid

For evaluation and better expression of the performance, the proposed independent nonlinear control method has been compared with the conventional linear control method. Figures 12 and 13 depict the AC and DC power, respectively. As displayed in the figure, the power oscillates in 1 s due to the increase in load and stabilizes in less than 0.2 s with the minimum oscillation. Figures 14 and 15 indicate the AC and DC voltages, respectively. The voltage fluctuations generated in 1 shave been controlled and it can be seen that the independent nonlinear control very well controlled the voltage fluctuations at the same time as the load changes. Figure 16 also reveals the frequency fluctuations per second. The power changes in the bidirectional converter in 1s are also shown in Figure 17. The power passing through the bidirectional converter in this case is zero. The comparison of the proposed method and conventional linear control suggests that the independent nonlinear

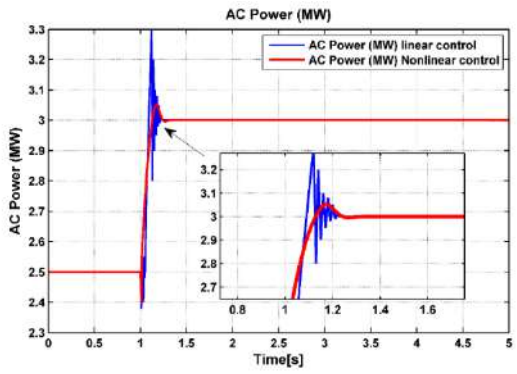


Figure 12. Production power by AC side sources

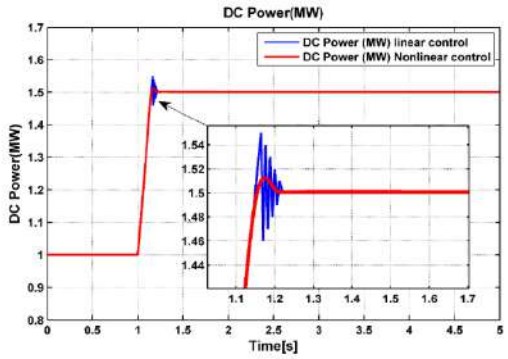


Figure 13. Production power by DC side sources

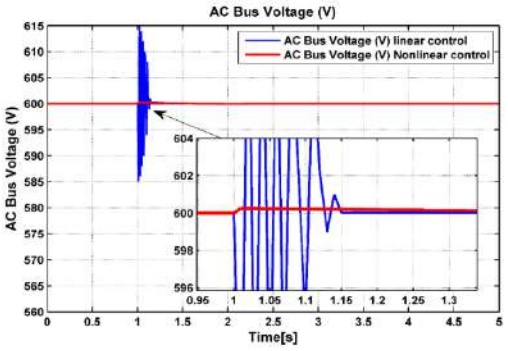


Figure 14. AC bus voltage during load changes

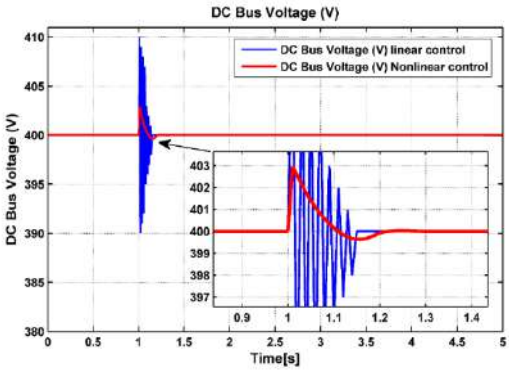


Figure 15. DC bus voltage during load changes

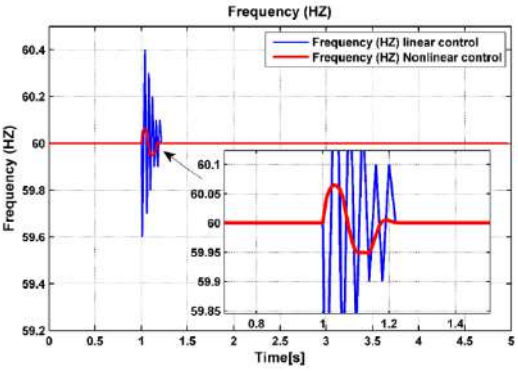


Figure 16. AC bus frequency during load changes

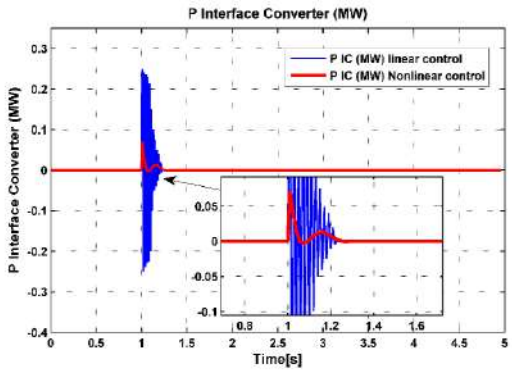


Figure 17. The power flow of the bidirectional converter

control system has followed the load changes within the shortest time and with the minimum fluctuations.

In the second scenario, to express the efficiency of the independent nonlinear control system and to control the bidirectional converter, the power exchange challenge from AC to DC and vice versa is examined. AC and DC side loads are 2.5 and 1 MW, respectively, and the system is in equilibrium. Initially, the DC side load power rises to 2 to 3.2 MW per second. Since the power capacity of each side is equal to 3 MW, the DC side required0.2 MW of power to supply the load. To this aim, the interface converter must transfer a power equivalent to 0.2 MW from AC to DC. The sources on the DC side try to supply the load at their maximum capacity. Figures 18 and 19 depict the AC and DC power output, respectively.

In 2s to supply power to the DC side, the output power of the AC side increases to 2.7 MW. Figure 20 illustrates the transfer power of the converter in rectifier mode. Also, the AC and DC voltage changes are shown in Figures 21 and 22. According to the comparison, the fluctuations due to power changes in the interface converter and DC and AC side are well controlled by the independent nonlinear control system; compared to conventional linear control, the combined microgrid system is stable with minimum time and oscillation.

In the next mode of the second scenario, the AC side power grows from 2.5 to 3.5 MW in 2s. DC microgrid

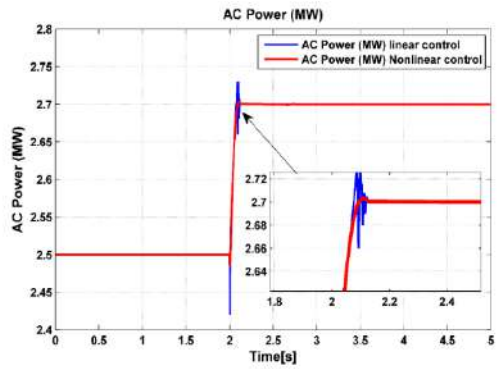


Figure 18. Production power by AC side sources

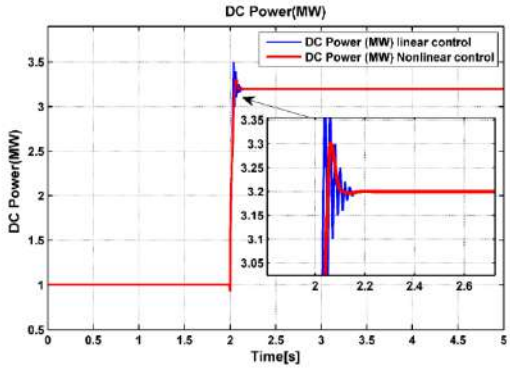


Figure 19. Production power by DC side sources

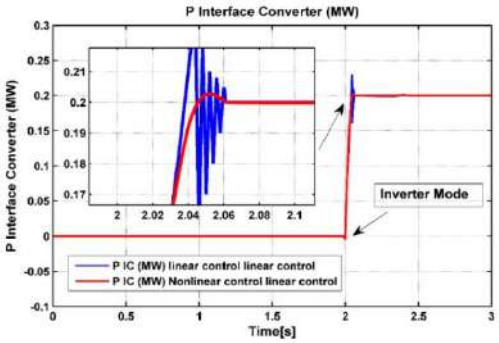


Figure 20. The power flow of the bidirectional converter

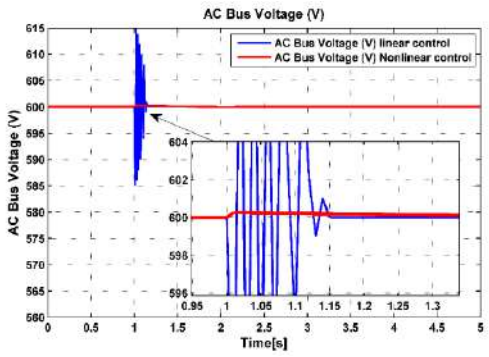


Figure 21. AC bus voltage during load changes

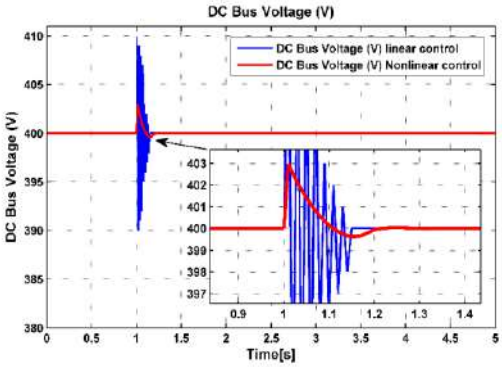


Figure 22. DC bus voltage during load changes

power is 1 MW. Since the load power required by the AC side is less than the output capacity of that side, the interface converter need to be able to transfer 0.5 MW from the DC side to the AC. Figures 23 and 24 show the power changes on the AC and DC sides, respectively. Also, the bidirectional converter operates in the inverter mode and transmits a power equivalent to 0.5 MW to AC (Figure 25). In addition, voltage changes on both sides are shown in Figures 26 and 27. Considering the precise performance of independent nonlinear control and its comparison with the conventional linear control method, it can be stated that changes and power exchange in each direction are handled by the proposed control system in the shortest time, which ensures system stability.

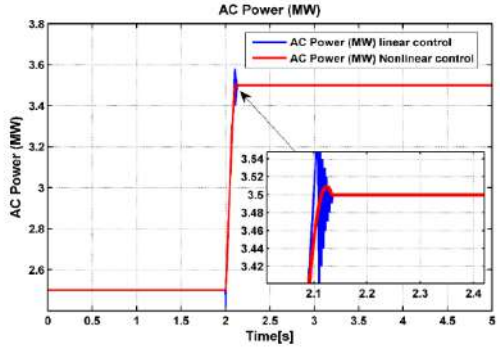


Figure 23. Production power by AC side sources

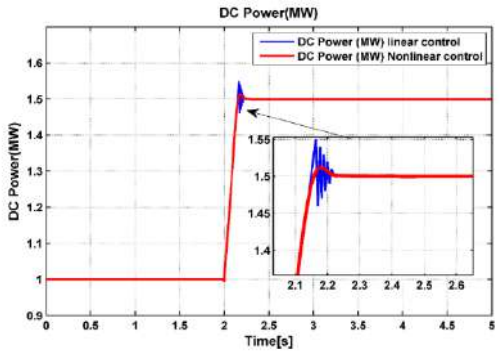


Figure 24. Production power by DC side sources

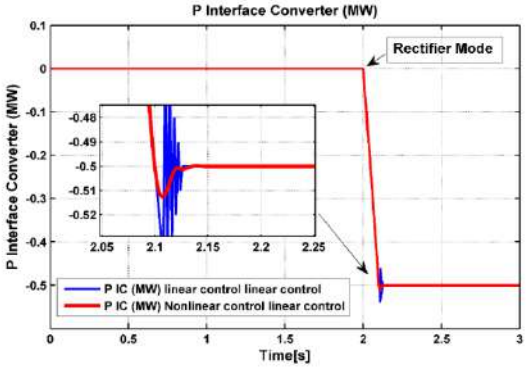


Figure 25. The power flow of the Bidirectional converter

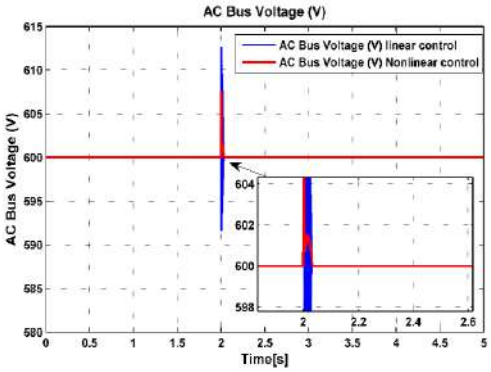


Figure 26. DC bus voltage during load changes

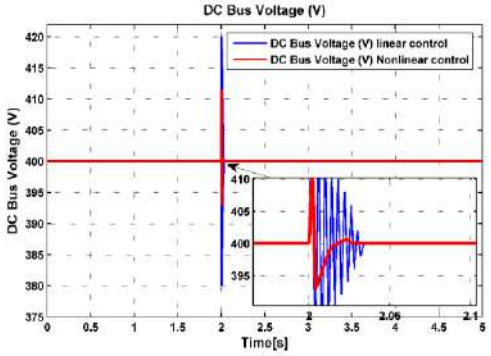


Figure 27. DC bus voltage during load changes

In the third scenario, the bidirectional converter is challenged. In 2s, due to repairs or faults in the system, the bidirectional converter is taken out of the circuit and then returned to the system in 3.5s. Figure 28 reveals the power changes in the bidirectional converter. In this case, it is assumed that the increase in load on each side is controlled and the necessary measures are taken to reduce the load. Figures 29 and 30 indicate the AC and DC voltage changes; while, Figure 31 depicts the frequency changes in 2 and 3.5s.

In this case, the major fluctuations will be on the voltage and frequency of the combined microgrid. This

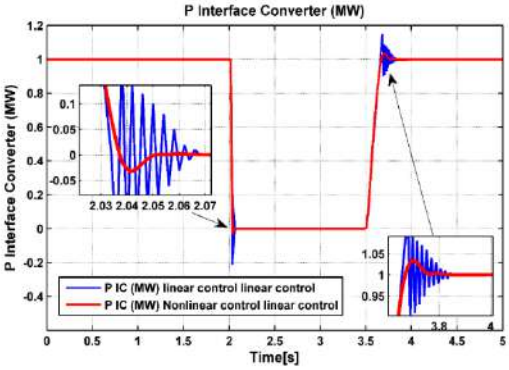


Figure 28. The power flow of the bidirectional converter

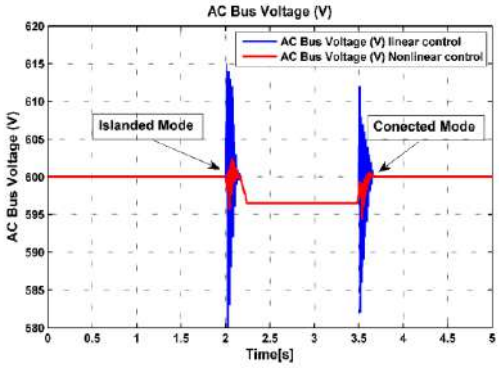


Figure 29. AC bus voltage during the islanded and reconnection mode

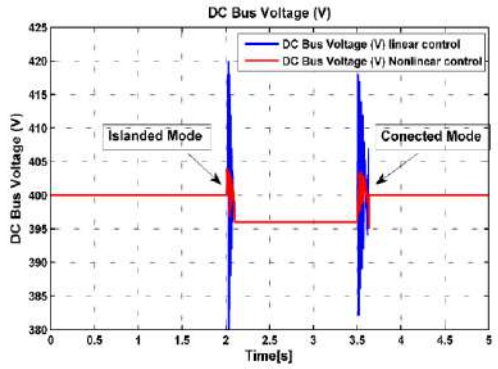


Figure 30. DC bus voltage during the islanded and reconnection mode

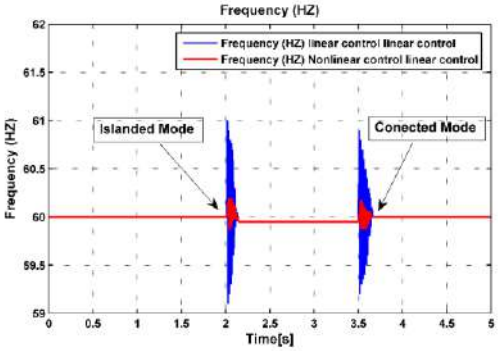


Figure 31. AC bus Frequency during the islanded and reconnection mode

is because the two parts are separated and the independent nonlinear control system must be able to supply the loads on each side as well as keep the system voltage plus frequency constant. by comparison, it can be seen that the independent nonlinear control system follows the changes very accurately. also, by controlling the power changes on both ac and dc sides by independent controllers, it ensures the stability of the voltage and frequency of the combined microgrid.

6. CONCLUSION

In this paper, a nonlinear controller was introduced to improve the stability of microgrids in islanded mode. In the proposed model, AC resources and loads on one side were related to DC resources and loads on the other side through a bidirectional AC/DC interface converter. Also, the hybrid microgrid would maintain its power balance in each part and use the other side for the power supply if there is unbalancing. In this model, each part, in addition to its load, could provide the other side load through a bidirectional interface converter. Accordingly, a nonlinear controller was used to properly stabilize the AC bus frequency and the DC bus voltage proportionate with the load changes on both sides. Also, the coordinated exchange and optimal regulation of control signals in this structure led to improve stability and in turn improved system performance. Here, the control systems existing for each source first increased the load deficiency and changed the system power with voltage and frequency variations while generating the maximum required power. Meanwhile, once changes are received by AC and DC buses on both sides of the converter, the control system would receive load changes, power shortages or faults and act accordingly. In such a system, a central converter was responsible for power exchange instead of using several converters. In addition, an energy storage system was used to improve the stability. With this proposed method, voltage and frequency fluctuations became stable during the first few cycles. What maintains stability in this network has been the correct exchange of power between two parts. This is performed by designing the bidirectional interface converter control system carefully and practically. This nonlinear control system ensures system stability.

7. REFERENCES

1. Armghan, H., Yang, M., Wang, M., Ali, N., Armghan, A.J.I.J.o.E.P. and Systems, E., "Nonlinear integral backstepping based control of a dc microgrid with renewable generation and energy storage systems", ***International Journal of Electrical Power & Energy Systems,*** Vol. 117, No., (2020), 105613, doi: 10.1016/j.ijepes.2019.105613.
2. Gundabathini, R., Pindoriya, N.M.J.E.P.C. and Systems, "Improved control strategy for bidirectional single phase ac-dc converter in hybrid ac/dc microgrid", ***Journal of Electrical Power Components Systystems,*** Vol. 45, No. 20, (2017), 2293-2303, doi: 10.1080/15325008.2017.1402970.
3. Pourbehzadi, M., Niknam, T., Aghaei, J., Mokryani, G., Shafie-khah, M., Catalão, J.P.J.I.J.o.E.P. and Systems, E., "Optimal operation of hybrid ac/dc microgrids under uncertainty of renewable energy resources: A comprehensive review", ***Electronic Power Energy System***, Vol. 109, No., (2019), 139-159, doi: 10.1016/j.ijepes.2019.01.025.
4. Jin, N., Hu, S., Gan, C. and Ling, Z.J.I.T.o.I.E., "Finite states model predictive control for fault-tolerant operation of a three-phase bidirectional ac/dc converter under unbalanced grid voltages", ***IEEE Transactions Industrial Electronics***, Vol. 65, No. 1, (2017), 819-829, doi: 10.1109/TIE.2017.2686342.
5. Issa, W., Al-Naemi, F., Konstantopoulos, G., Sharkh, S. and Abusara, M.J.E.P., "Stability analysis and control of a microgrid against circulating power between parallel inverters", ***Energy Procedia***, Vol. 157, No., (2019), 1061-1070, doi: 10.1016/j.egypro.2018.11.273.
6. Li, P., Yan, S., Yu, X. and Zhang, J., "The h∞ control method of bidirectional converter in hybrid ac/dc microgrid", in 2016 IEEE Power and Energy Society General Meeting (PESGM), IEEE. Vol., No., (2016), 1-5.
7. Dragičević, T., Lu, X., Vasquez, J.C. and Guerrero, J.M.J.I.t.o.p.e., "Dc microgrids—part ii: A review of power architectures, applications, and standardization issues", ***IEEE Transactions on Power Electronics,*** Vol. 31, No. 5, (2015), 3528-3549, doi: 10.1109/TPEL.2015.2464277.
8. Malik, S.M., Ai, X., Sun, Y., Zhengqi, C., Shupeng, Z.J.I.G., Transmission and Distribution, "Voltage and frequency control strategies of hybrid ac/dc microgrid: A review", ***IET Generation, Transmission & Distribution,*** Vol. 11, No. 2, (2017), 303-313, doi: 10.1049/iet-gtd.2016.0791.
9. Zubieta, L.E.J.I.E.M., "Are microgrids the future of energy?: Dc microgrids from concept to demonstration to deployment", ***IEEE Electrification Magazine,*** Vol. 4, No. 2, (2016), 37-44, doi: 10.1109/MELE.2016.2544238.
10. Han, H., Hou, X., Yang, J., Wu, J., Su, M. and Guerrero, J.M.J.I.T.o.S.G., "Review of power sharing control strategies for islanding operation of ac microgrids", ***IEEE Transactions on Smart Grid***, Vol. 7, No. 1, (2015), 200-215, doi: 10.1109/TSG.2015.2434849.
11. Hatziargyriou, N., "Microgrids: Architectures and control, John Wiley & Sons, (2014).
12. Dheer, D.K., Soni, N., Doolla, S.J.S.E., Grids and Networks, "Improvement of small signal stability margin and transient response in inverter-dominated microgrids", ***Sustainable Energy Grids and Networks***, Vol. 5, (2016), 135-147, doi: 10.1016/j.segan.2015.12.005.
13. Khorsandi, A., Ashourloo, M., Mokhtari, H., Iravani, R.J.I.G., Transmission and Distribution, "Automatic droop control for a low voltage dc microgrid", ***IET Generation, Transmission and Distribution,*** Vol. 10, No. 1, (2016), 41-47, doi: 10.1049/iet-gtd.2014.1228.
14. Tejwani, V.S. and Suthar, B.N., "Control strategy for utility interactive hybrid pv hydrogen system", in 2016 IEEE Power and Energy Society General Meeting (PESGM), IEEE. (2016), 1-5.
15. Yu, K., Ai, Q., Wang, S., Ni, J. and Lv, T.J.I.T.o.S.G., "Analysis and optimization of droop controller for microgrid system based on small-signal dynamic model", ***IEEE Transactions on Smart Grid***, Vol. 7, No. 2, (2015), 695-705, doi: 10.1109/TSG.2015.2501316.
16. Thale, S.S. and Agarwal, V.J.I.T.o.S.G., "Controller area network assisted grid synchronization of a microgrid with renewable energy sources and storage", ***IEEE Transactions on Smart Grid,*** Vol. 7, No. 3, (2015), 1442-1452, doi: 10.1109/TSG.2015.2453157.
17. Alnejaili, T., Drid, S., Mehdi, D., Chrifi-Alaoui, L., Belarbi, R., Hamdouni, A.J.E.C. and Management, "Dynamic control and advanced load management of a stand-alone hybrid renewable power system for remote housing", ***Energy Conversion and Management,*** Vol. 105, (2015), 377-392, doi: 10.1016/j.enconman.2015.07.080.
18. Dragičević, T., Guerrero, J.M., Vasquez, J.C. and Škrlec, D.J.I.T.o.p.E., "Supervisory control of an adaptive-droop regulated dc microgrid with battery management capability", ***IEEE Transactions on Power Electronics,*** Vol. 29, No. 2, (2013), 695-706, doi: 10.1109/TPEL.2013.2257857.

19. Abdullah, M.A., Yatim, A., Tan, C.W., Saidur, R.J.R. and reviews, s.e., "A review of maximum power point tracking algorithms for wind energy systems", ***Renewable and Sustainable Energy Reviews,*** Vol. 16, No. 5, (2012), 3220-3227, doi: 10.1016/j.rser.2012.02.016.

20. Du, C., Agneholm, E. and Olsson, G.J.I.T.o.P.D., "Comparison of different frequency controllers for a vsc-hvdc supplied system", ***Transactions on Power Transmission & Distribution***, Vol. 23, No. 4, (2008), 2224-2232, doi: 10.1109/tpwrd.2008.921130.

21. Ruan, S.-Y., Li, G.-J., Peng, L., Sun, Y.-Z., Lie, T.J.I.J.o.E.P. and Systems, E., "A nonlinear control for enhancing hvdc light transmission system stability", ***International Journal of Electrical Power & Energy Systems,*** Vol. 29, No. 7, (2007), 565-570, doi: doi: 10.1016/j.ijepes.2007.01.008.

22. Yang, P., Xia, Y., Yu, M., Wei, W. and Peng, Y.J.I.T.o.I.E., "A decentralized coordination control method for parallel bidirectional power converters in a hybrid ac–dc microgrid", ***IEEE Transactions Industrial Electronics***, Vol. 65, No. 8, (2017), 6217-6228, doi: 10.1109/TIE.2017.2786200.

Persian Abstract

چکیده

این مطالعه، ساختار جدیدی مبتنی بر کنترل‌کننده غیرخطی برای کنترل و تجزیه و تحلیل ثبات ریزشبکه‌های ترکیبی ارائه می‌دهد. در مدل پیشنهادی، منابع و بارهای AC و DC در دو طرف مختلف قرار دارند. علاوه بر این ، یک مبدل رابط دوطرفه AC / DC برای تأمین بار توسط منابع AC / DC استفاده شده است. تولیدات AC / DC در هر دو طرف مبدل وجود دارد و هر طرف می‌تواند بار طرف دیگر را از طریق مبدل رابط دو طرفه تأمین کند. همچنین، یک سیستم ذخیره انرژی برای پایداری سیستم در سمت DC استفاده شده است. کنترل‌کننده غیرخطی ریزشبکه برای تنظیم صحیح فرکانس سمت باس AC و ولتاژ سمت باس DC طراحی شده است. در این ساختار، تبادل هماهنگ و بهینه توان همراه با تنظیم دقیق سیگنال‌های کنترل منجر به بهبود پایداری می‌شود. بنابراین، عملکرد سیستم بهبود می‌یابد. نتایج نشان می‌دهد که مدل پیشنهادی برای کاهش نوسانات و بهبود پایداری سیستم کارآمد است.

IJE TRANSACTIONS A: Basics Vol. 34, No. 7, (July 2021) 1766-1777

International Journal of Engineering

Journal Homepage: www.ije.ir

A Bi-objective Robust Optimization Model for an Integrated Production-distribution Problem of Perishable Goods with Demand Improvement Strategies: A Case Study

A. Aazami, M. Saidi-Mehrabad*, S. M. Seyedhosseini

School of Industrial Engineering, Iran University of Science and Technology, Tehran, Iran

PAPER INFO

Paper history:
Received 21 January 2021
Received in revised form 02 April 2021
Accepted 07 April 2021

Keywords:
Integrated Production-Distribution Planning
Perishable Goods
Bi-objective Optimization
Robust Optimization
Environmental Considerations

ABSTRACT

This paper develops a bi-objective optimization model for the integrated production-distribution planning of perishable goods under uncertainty. The first objective seeks to maximize the profit in a specific supply chain with three levels: plants, distribution centers, and in the last level, customers. Since transportation is one of the major pollution sources in a distribution problem, the second objective is to minimize their emission. In the considered problem, the decisions of production, location, inventory, and transportation are made in an integrated structure. In developing the demand function, the effect of the product freshness and the price is formulated. Besides, to encourage customers, three strategies, including perished product return, discount, and credit policies, are proposed. Also, robust optimization is utilized to cope with the operational uncertainty of some cost parameters. To prove the applicability of this research and the feasibility of the environmental aspect, a case study is conducted. Finally, the numerical computations on the case study provide a trade-off between the environmental and economic goals and indicate a 37.5 percent increase in the profit using the developed model.

doi: *10.5829/ije.2021.34.07a.21*

1. INTRODUCTION

Production and distribution problems play a vital role in supply chain management [1]. Many companies are seeking to optimize their production and distribution planning (PDP) simultaneously and in an integrated framework to achieve the highest profit as well as the highest satisfaction of customer demand [2]. In this article, the first objective aims to maximize the production and distribution (P-D) profit considering a green supply chain (SC). The considered SC has three levels, including plants, distribution centers (DCs), and customers, where the manufacturer and the distributor make their decisions in an integrated structure. On the other hand, the second objective seeks to reduce the emission of environmental pollutions. Therefore, we develop a multi-period model for the Green PDP (GPDP) problem regarding perishable goods.

Perishable goods are damaged and decayed during their shelf life [3]. The freshness of a group of perishable goods, such as dairy and packaged vegetables, decreases during their shelf life, and they will be useless after a fixed lifetime [4, 5]. Customers decide to buy these goods regarding their most common visual cue, i.e., the best-before-date (BBD) [4]. The product we study in this paper, denoted by $\mathcal{P}$, is perishable with a specific BBD and a fixed lifetime. The more time left until the BBD, the fresher are the $\mathcal{P}$ goods, and consequently, the more willing the customers are to buy them. Therefore, for $\mathcal{P}$ goods, apart from the price and level of advertisement [6, 7], the freshness factor plays a vital role in the demand function [4, 8]. In addition to this, in developing the demand function, we apply three customer encouragement strategies (the return, discount, and crediting policies), which distinguishes it from the literature.

Researchers concluded that the effect of integrated P-D decisions on the profit would be much more

* Corresponding Author's Email: mehrabad@iust.ac.ir (M. Saidi-Mehrabad)

Please cite this article as: A. Aazami, M. Saidi-Mehrabad, S. M. Seyedhosseini, A Bi-objective Robust Optimization Model for an Integrated Production-distribution Problem of Perishable Goods with Demand Improvement Strategies: A Case Study, International Journal of Engineering, Transactions A: Basics Vol. 34, No. 7, (2021) 1766-1777

remarkable for perishable goods because they are not delivered on time in the non-integrated approach [4]. Therefore, the freshness of this product reduces also during transportation [9]. Besides, some cost parameters in the PDP are uncertain in reality. To deal with the uncertainty, robust optimization is applied regarding the availability of their interval values. In summary, we develop a two-objective robust model for $\mathcal{P}$ goods' GPDP problem ($\mathcal{P}-G$PDP). Also, the augmented Epsilon constraint (AEC) method is used to make a trade-off between the environmental and economic goals.

The major contributions of our article are summarized as follows. First, although many researchers studied the PDP problems [2, 10], as far as we deeply investigated, no research is found that develops a bi-objective multi-period optimization model with integrating P-D decisions for a specific $\mathcal{P}$ product. In fact, in addition to maximizing the profit, the proposed model also addresses the vital goal of reducing environmental impacts. The key motivation for developing this model is to get closer to real-world conditions. Second, we develop a new demand function considering three key factors, including price, advertisement, and product freshness, as well as three encouraging strategies, including credit period, discount, and perished goods return policies. Third, we employ robust optimization to cope with the operational uncertainty of the cost parameters. After our extensive studies, robust optimization has not been used to control the uncertainty of $\mathcal{P}-G$PDP specific problem parameters. Also, a particular case study is conducted to present the industrial application of this study.

The rest of this research is arranged as follows. Section 2 represents a summary of the PDP problem background for perishable goods. In section 3, we develop the bi-objective optimization model. Section 4 presents a robust optimization approach. In section 5, the method for creating a trade-off between the environmental and economic aims is discussed. In section 6, the performance of the expanded model is examined using a specific case study. Finally, the conclusions and future suggestions are presented.

2. LITERATURE REVIEW

We reviewed articles related to the $\mathcal{P}-G$PDP. As the literature on the PDP problem is very extensive [11], we focused more on providing PDP studies for perishable goods. Ahumada and Villalobos [12] presented a model for the operational decision-making for producing, harvesting, and distributing tomato goods. Their operational decisions included packaging, warehousing, transportation, as well as multiple harvests. To consider the effect of harvest decisions on product quality and freshness, they estimated the color change distribution. Amorim et al. [4] presented an integrated PDP model considering a loose and fixed lifetime for perishable goods. Their goals were the minimization of the P-D costs and maximization of the delivered product's shelf life. Fahimnia et al. [13] added a PDP model to the literature considering a two-level SC. Our model is closely similar to that presented by Fahimnia et al. [13] in terms of problem dimensions. However, they did not consider the effect of perishability and many realities. Amorim et al. [14] published a survey on the PDP models that had considered perishability. They provided a new framework categorizing the models of perishability in terms of three criteria: a) Physical deterioration, b) authority boundaries, and c) customer value. Bilgen and Celebi [15] considered the PDP in a multi-site mode in a yogurt production line. Their objective function was to maximize profit where pricing was dependent on the shelf life. By presenting a review study, Diaz et al. [16] categorized PDP published papers according to criteria including production, inventory, routing, modeling method, type of objective function, and solution method. Makui et al. [17] addressed the PDP for goods with a very limited expiration date, such as calendars, using postponement policy. They employed a robust optimization to cope with the uncertainty.

Fattahi et al. [18] considered a two-level SC in which a manufacturer of perishable goods produces and distributes its products among several customers. The first aim was to reduce production, inventory, and distribution costs, while the second goal was to minimize lateral transshipment costs such that no shortage occurs for the SC. Devapriya et al. [19] developed a PDP model concerning perishability in which they paid more attention, how to distribute the product. Ensafian and Yaghoubi [20] developed an integrated model for platelet SC with two types of production methods. They took into account the PDP and presented a bi-objective robust model, maximizing the delivered platelet units' freshness while minimizing the cost. Guarnaschelli et al. [21] considered the PDP for a two-level dairy SC. They presented an integrated two-stage stochastic model. Biuki et al. [22] modeled a specific type of integrated PDP for perishable goods with inventory, location, and routing decisions by real-world data. Liu et al. [23] recently considered the integrated PDP in a blood SC containing one supplier and some blood centers while the transshipment among the blood centers was allowed.

In the perishability context, Bakker et al. [24] reviewed the developments in inventory control and distribution of perishable goods. Coelho and Laporte [25] reviewed different policies for inventory management and reprocessing of perishable goods. They analyzed three policies, including fresh-first, old-

first, and optimized priority. A few years later, Janssen et al. [26] completed Bakker et al.'s survey of perishable inventory models using key topics. Gharehyakheh et al. [27] presented an integrated model to minimize transportation costs and CO_2 emissions as well as maximize product freshness. They addressed a routing problem considering shelf life, temperature, and energy consumption prediction. Recently, Navazi et al. [28] developed a three-objective model to distribute perishable goods and gathered the remained perished goods for recycling while distributing fresh goods.

According to our in-depth studies, no article considered two-objective optimization to model GPDP for perishable goods considering the freshness factor and uncertainty. Each new contribution described in Section 1 differentiates our research from the literature.

3. MATHEMATICAL MODEL

3. 1. Problem Description This section describes our $\mathcal{P}-G$PDP problem for an SC with three levels: plants, DCs, and customers. Consider a P-D company that has many customers in different locations. The company's first goal is to maximize its profit by integrating production, inventory, and distribution decisions for product $\mathcal{P}$. The product flow in the SC network is displayed in Figure 1. After identifying several potential locations for the DCs, the P-D company establishes optimal DCs with different capacities. It is not possible to relocate the DCs during the planning periods. Also, the plants and customers have pre-determined locations. Since transportation is one of the major environmental pollution sources, the P-D company's second goal is to minimize the environmental impact of transportation, regarding the vital importance of paying attention to green issues.

To reduce the risk of perishing, customers are looking to buy a product with a longer lifetime. However, the P-D company must first increase each period's production by taking into account the relevant capacity to minimize setup costs. Second, considering the transportation capacity, it must deliver more frequently and larger amounts of the product to decrease transportation costs. In developing demand function, in addition to the effect of price, product freshness, and advertisement, we apply three other strategies to attract customers and improve market share, including 1) Perished product return policy: customers can return their purchase at a specific rate, where for each unit of the returned product, a salvage value is considered, 2) Discount policy: different discounts are considered for the product with different shelf lives, 3) Credit policy:

the P-D company provides a specific credit period for the payment deadline. Therefore, under each of these policies, a special rate is considered, which is the same for different customers.

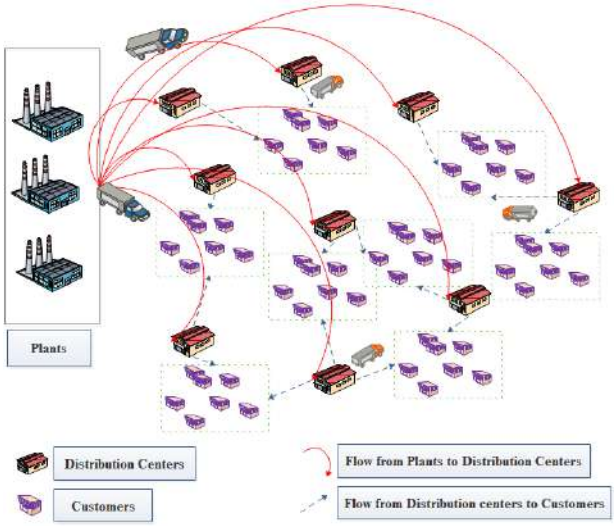


Figure 1. The three-level SC network for $\mathcal{P}-G$PDP

Moreover, the parameters of P-D costs (production and transportation) are uncertain, and in the real world, the data are available only in interval form. To deal with this operational uncertainty, we use Bertsimas and Sim's robust optimization [29]. In general, we look to reply to the following inquiries in the considered $\mathcal{P}-G$PDP problem: 1) How to control the inventory level and plan the optimum production, such as the production periods and lot size? 2) How to determine the optimal DCs location and the optimum product flow in the network? 3) How much are the advertisement costs and customer encouragement strategies? To find an answer to these inquiries, we extend a robust optimization model to maximize the P-D profit and minimize transportation's environmental impact.

3. 2. Notations In this sub-section, the notations are defined.

Sets

$\mathcal{F}$	Set of plants, indexing with *f*
$\mathcal{D}$	Set of potential DCs, indexing with *d*
$\mathcal{K}$	Set of the DCs capacities, indexing with *k*
$\mathcal{C}$	Set of customers, indexing with *c*
$\mathcal{T}$	Set of periods, indexing with *t*
$\mathcal{H}$	Set of consumable periods for the product, indexing with *h*

Parameters

fc_f	The fixed cost of setting up plant *f* in every period
qc_f	Unit production cost in plant *f* (**uncertain: interval**)
fdc_{dk}	Fixed establishing cost of the DC with capacity *k* at location *d*
$tc^{f\to d}$	Transportation cost between plant *f* and DC *d* (**uncertain: interval**)
$tc^{d\to c}$	Transportation cost between DC *d* and customer *c* (**uncertain: interval**)

hc_d^h	Holding cost of each product unit with age *h* during a period in DC *d*
cap^{f-}	Capacity of each vehicle to travel from plant *f* to DC *d*
cap^{d-}	Capacity of each vehicle to travel from DC *d* to customer *c*
$e^{f\to d}$	Environmental effects (CO_2 production) caused by each vehicle traveling from plant *f* to DC *d*
$e^{d\to c}$	Environmental effects (CO_2 production) caused by each vehicle traveling from DC *d* to customer *c*
cap_f	The capacity of production in plant *f* in every period
cap_k	The capacity of storing the product in a DC with capacity *k*
dem_c	Nominal demand of customer *c* for the fresh product
pr	Unit selling price for the fresh product
Ω	Cost of considering an entire credit for customers
Ψ	Salvage value for each unit perished product

Variables

Q_{ft}	Production amount for plant *f* in the period *t*
sz_{ft}	1, if plant *f* is set up in the period *t; 0*, otherwise
dz_{dk}	1, if a specific DC with capacity *k* is established at location *d*; 0, otherwise
A	Marketing cost
$fw_t^{f\to d}$	The flow of the goods from plant *f* to DC *d* for period *t*
$fw_{ht}^{d\to c}$	The flow of the goods with shelf life *h* from DC *d* to customer *c* for period *t*
I_{dt}^h	Product inventory with shelf life *h* in DC *d* for period *t*
s_1^h	The return rate of the sold product with age *h*
s_2^h	The discount rate of the sold product with age *h*
s_3^h	Credit rate of the sold product with age *h*
r_{ct}	The amount of product returned from customer *c* for period *t*
π^{PD}	The total profit
E^{PD}	The total amount of environmental effects (CO_2 production resulted from transportation)

3. 3. Problem Modeling In this sub-section, we first develop the new demand function affected by price, product freshness, advertisement, and the encouragement strategies. Then, based on this demand function, we develop a bi-objective optimization model to maximize the P-D Company's profitability and minimize the environmental impacts.

3. 3. 1. Development of Demand Function According to previous research, the potential demand depends on price as stated in Equation (1) [7, 30, 31]:

$$\mathfrak{D}(pr) = \mathscr{b}(1-\alpha.pr)^{\frac{1}{\upsilon}} \tag{1}$$

where $\mathscr{b}$ is the fresh goods' demand at the lowest price, and α and υ are non-negative parameters that indicate the price elasticity of demand ($pr \le \frac{1}{\alpha}$). Now suppose that the amount of demand $\mathfrak{D}(pr)$ in Equation (1) needs the advertisement cost $\mathcal{A}$. To include the impact of advertisement into the demand function, Equation (2) is obtained:

$$\mathfrak{D}(pr,A) = \mathfrak{D}(pr).\left(1-\frac{\mathcal{A}-A}{\mathcal{A}}\right) = \mathscr{b}(1-\alpha.pr)^{\frac{1}{\upsilon}}.\left(\frac{A}{\mathcal{A}}\right) \tag{2}$$

where $1-\frac{\mathcal{A}-A}{\mathcal{A}}$ shows the demand reduction factor due to inadequate advertisement.

$\mathfrak{D}(pr,A)$ is the demand function for fresh goods ($h=1\in\mathscr{H}$). Besides, if demand decreases linearly due to the lack of product freshness, the maximum market share is as Equation (3):

$$\mathcal{S}^h(pr,A) = \begin{cases} \eta.\mathfrak{D}(pr,A) & \forall h\in\mathscr{H} \\ 0 & \forall\, h>|\mathscr{H}| \end{cases} \tag{3}$$

where $\eta = 1-\frac{h-1}{|\mathscr{H}|}$ is the freshness factor, i.e., the ratio of the remaining product shelf life.

It is evident that $\mathcal{S}^h(pr,A) < \mathfrak{D}(pr,A)\ \forall h\in \mathscr{H}\backslash\{1\}$. In fact, due to the lack of product freshness, some potential demand ($LS = \mathfrak{D}(pr,A) - \mathcal{S}^h(pr,A)$) is lost, which is compensated as much as possible by using customer encouragement strategies.

The customer encouragement strategies (the credit, return, discount policies) are applied to demand function and, accordingly, to the amount of supply. Equation (4) presents the suggested demand function as a result of these policies:

$$\mathfrak{D}^h(pr,A,s_1,s_2,s_3) = \frac{\eta}{1-(s_1+s_2+s_3)}.\mathfrak{D}(pr,A) = \frac{|\mathscr{H}|+1-h}{|\mathscr{H}|\left(1-(s_1+s_2+s_3)\right)}.\mathscr{b}(1-\alpha.pr)^{\frac{1}{\upsilon}}.\left(\frac{A}{\mathcal{A}}\right) \quad \forall\, h\in\mathscr{H} \tag{4}$$

where $\frac{\eta}{1-(s_1+s_2+s_3)} = \frac{|\mathscr{H}|+1-h}{|\mathscr{H}|\left(1-(s_1+s_2+s_3)\right)}$ is the adjustment coefficient of potential demand ($\mathfrak{D}(pr,A)$) regarding the four key elements of the freshness (η), return (s_1), discount (s_2) and crediting (s_3) rates ($0\le s_1+s_2+s_3<1$). After applying the policies, it is concluded that $\mathfrak{D}^h(pr,A,s_1,s_2,s_3)\ge\eta.\mathfrak{D}(pr,A)$. Now if $\mathcal{S}^h(pr,A,s_1,s_2,s_3)$ is the maximum market share with applying the encouragement strategies, it will be clear that $\mathcal{S}^h(pr,A,s_1,s_2,s_3)\le\mathfrak{D}^h(pr,A,s_1,s_2,s_3)$. Thus, if $LS<\mathfrak{D}^h(pr,A,s_1,s_2,s_3)-\eta.\mathfrak{D}(pr,A)$, not only lost sales in the current period are compensated, but it is also possible to supply some demand of the subsequent periods. In this case, regarding the encouragements, the customer will be more inclined to purchase and stock for the future.

Consequently, in every period of production and supply, the maximum demand/sale for the product with shelf life h is $\frac{|\mathscr{H}|+1-h}{|\mathscr{H}|\left(1-(s_1+s_2+s_3)\right)}.\mathfrak{D}(pr,\mathrm{A}) = \frac{|\mathscr{H}|+1-h}{|\mathscr{H}|\left(1-(s_1+s_2+s_3)\right)}.\mathscr{b}(1-\alpha.pr)^{\frac{1}{\upsilon}}.\left(\frac{A}{\mathcal{A}}\right)$.

3. 3. 2. Proposed Formulation After defining the demand function, the objective functions are presented as follows:

$$\begin{aligned}&\boldsymbol{max}\,\pi^{PD}=\\&pr.\sum_{d\in\mathcal{D}}\sum_{c\in\mathcal{C}}\sum_{h\in\mathcal{H}}\sum_{t\in\mathcal{T}}fw_{ht}^{d\to c}-\\&\Big(\Big[\sum_{f\in\mathcal{F}}\sum_{t\in\mathcal{T}}fc_f.sz_{ft}\Big]+\\&\Big[\sum_{f\in\mathcal{F}}\sum_{t\in\mathcal{T}}qc_f.Q_{ft}\Big]+\\&\Big[\sum_{d\in\mathcal{D}}\sum_{k\in\mathcal{K}}fdc_{dk}.dz_{dk}\Big]+\\&\Big[\sum_{d\in\mathcal{D}}\sum_{h\in\mathcal{H}}\sum_{t\in\mathcal{T}}hc_d^h.I_{dt}^h\Big]+\\&\Big[\sum_{f\in\mathcal{F}}\sum_{d\in\mathcal{D}}\sum_{t\in\mathcal{T}}tc^{f\to d}\left\lceil\frac{fw_t^{f\to d}}{cap^{f\to d}}\right\rceil+\\&\sum_{d\in\mathcal{D}}\sum_{c\in\mathcal{C}}\sum_{h\in\mathcal{H}}\sum_{t\in\mathcal{T}}tc^{d\to c}\left\lceil\frac{fw_{ht}^{d\to c}}{cap^{d\to c}}\right\rceil\Big]+\\&\Big[\sum_{d\in\mathcal{D}}\sum_{c\in\mathcal{C}}\sum_{h\in\mathcal{H}}\sum_{t\in\mathcal{T}}\big(s_2^h.pr+\\&\Omega.s_3^h\big)fw_{ht}^{d\to c}\Big]+\Big[\sum_{c\in\mathcal{C}}\sum_{t\in\mathcal{T}}(pr-\Psi)r_{ct}\Big]+\\&[A]\Big)\end{aligned}\tag{5}$$

$$\begin{aligned}&\boldsymbol{min}\;E^{PD}=\\&\Big[\sum_{f\in\mathcal{F}}\sum_{d\in\mathcal{D}}\sum_{t\in\mathcal{T}}e^{f\to d}\left\lceil\frac{fw_t^{f\to d}}{cap^{f\to d}}\right\rceil+\\&\sum_{d\in\mathcal{D}}\sum_{c\in\mathcal{C}}\sum_{h\in\mathcal{H}}\sum_{t\in\mathcal{T}}e^{d\to c}\left\lceil\frac{fw_{ht}^{d\to c}}{cap^{d\to c}}\right\rceil\Big]\end{aligned}\tag{6}$$

Equation (5) shows the objective function of profit maximization for the P-D company. Its first section is the income from the sale of the product. The second section shows the total costs, including nine subsections. The first one is the setup cost for the plants. The second one addresses the variable cost of production. The third one shows the fixed cost for establishing the DCs. In the fourth subsection, the inventory cost is calculated. Subsections 5 and 6 are the costs of transporting the product. $\left\lceil\frac{fw_t^{f\to d}}{cap^{f\to d}}\right\rceil$ and $\left\lceil\frac{fw_{ht}^{d\to c}}{cap^{d\to c}}\right\rceil$ are the number of vehicles for transporting from a specific plant to the DCs and then to the customers. The expressions $\left\lceil\frac{fw_t^{f\to d}}{cap^{f\to d}}\right\rceil$ and $\left\lceil\frac{fw_{ht}^{d\to c}}{cap^{d\to c}}\right\rceil$ are linearized according to Equations (7) and (8).

$$\begin{cases}n_t^{f\to d}\geq\dfrac{fw_t^{f\to d}}{cap^{f\to d}}\\ n_t^{f\to d}<\dfrac{fw_t^{f\to d}}{cap^{f\to d}}+1\\ n^{f\to d}\in\mathbb{Z}^+\end{cases}\tag{7}$$

$$\begin{cases}n_t^{d\to c}\geq\displaystyle\sum_{h\in\mathcal{H}}\frac{fw_{ht}^{d\to c}}{cap^{d\to c}}\\ n_t^{d\to c}<\displaystyle\sum_{h\in\mathcal{H}}\frac{fw_{ht}^{d\to c}}{cap^{d\to c}}+1\\ n_t^{d\to c}\in\mathbb{Z}^+\end{cases}\tag{8}$$

Subsection 7 is the costs of applying discount and credit policies to customers. In the eighth subsection, the cost resulted from the product return policy is subtracted, and the salvage value of the perished goods is added to the profit. Finally, the ninth subsection shows the cost of advertisement. Equation (6) shows the second objective function, minimizing the environmental effects of transportation. In the following, the constraints of the developed model are given:

$$\begin{aligned}&\sum_{d\in\mathcal{D}}fw_{ht}^{d\to c}=\frac{|\mathcal{H}|+1-h}{|\mathcal{H}|\left(1-\left(s_1^h+s_2^h+s_3^h\right)\right)}.dem_c(1-\\&\alpha.pr)^{\frac{1}{v}}.\left(\frac{A}{\mathcal{A}}\right)\quad\forall\,c\in\mathcal{C},h\in\mathcal{H},t\in\mathcal{T}\end{aligned}\tag{9}$$

$$Q_{ft}\leq cap_f.sz_{ft}\;\;\forall\,f\in\mathcal{F},t\in\mathcal{T}\tag{10}$$

$$Q_{ft}=\sum_{d\in\mathcal{D}}fw_t^{f\to d}\quad\forall\,f\in\mathcal{F},t\in\mathcal{T}\tag{11}$$

$$I_{dt}^h=\sum_{f\in\mathcal{F}}fw_t^{f\to d}-\sum_{c\in\mathcal{C}}fw_{ht}^{d\to c}\quad\forall\,d\in\mathcal{D},t\in\mathcal{T},h=1\tag{12}$$

$$I_{dt}^h=I_{dt-1}^{h-1}-\sum_{c\in\mathcal{C}}fw_{ht}^{d\to c}\;\;\forall\,d\in\mathcal{D},t\in\mathcal{T},h\in\mathcal{H}\backslash\{1\}\tag{13}$$

$$\sum_{f\in\mathcal{F}}fw_t^{f\to d}\leq\sum_{k\in\mathcal{K}}cap_k.dz_{dk}\;\;\forall\,d\in\mathcal{D},t\in\mathcal{T}\tag{14}$$

$$\sum_{h\in\mathcal{H}}I_{dt}^h\leq\sum_{k\in\mathcal{K}}cap_k.dz_{dk}\;\;\forall\,d\in\mathcal{D},t\in\mathcal{T}\tag{15}$$

$$\begin{cases}A,Q_{ft},fw_t^{f\to d},fw_{ht}^{d\to c},I_{dt}^h,s_1^h,s_2^h,s_3^h,r_{ct}\geq 0\\ sz_{ft},dz_{dk}\in\{0,1\}\end{cases}\tag{16}$$

Equation (9), is based on the developed demand function, shows the maximum demand/sale of goods with shelf life h in each period. In Equation (10), subject to the plant setup, the maximum production in each plant is restricted to its capacity. Equation (11) is an equilibrium constraint for each plant's production amount and the supply amount to the DCs. According to Equation (12), the inventory for a quite fresh product ($h=1$) in every DC is equal to the amount taken from the plants minus the supply to the customers. Equation (13) relates to the inventory of non-fresh goods. In Equation (14), subject to a DC's establishment, the transportation amount from the plants to the DC is limited to its capacity. Equation (15) refers to the bounded capacity of the Dcs to store goods with different shelf life. Finally, Equation (16) describes the range of the decision variables.

4. UNCERTAINTY CONTROL

In our problem, the parameters of production and transportation costs have constraint-wise uncertainty and are considered as an interval. Some other parameters may also be uncertain. However, we assume that some of them, such as the demand, are in the worst-case situation. Besides, the uncertainty of the other parameters, such as capacity, is not to the extent to be considered; thus, we assume their uncertainty is negligible. In this paper, Bertsimas and Sim (B&S) method [29] is used to achieve solution robustness. The

B&S method is employed because it works right for the constraint-wise uncertainty, and it can control the conservatism level. Some researchers used this method to deal with this kind of operational uncertainty, as well [32]. To explain the B&S method, first consider the following optimization model in general:

$$\min_{x} \sum_{j\in J} a_{0j}x_j \;\; s.t. \;\; Ax \le b \tag{17}$$

where a_{0j} are uncertain parameters. Based on robust programming (RP), the above optimization problem is solved under uncertainty in such a way that the solution will always be feasible, and the amount of the objective function is optimal in the strict case [33]–[35]. Therefore, if the uncertain parameters are assumed to be as $a_{0j} \in \left[a_{0j}^{L}, a_{0j}^{U}\right]$, then the robust counterpart (RC) of the above uncertain model is as follows:

$$\min_{x} \max_{a_{0j}\in\left[a_{0j}^{L},a_{0j}^{U}\right]} \left\{\sum_{j\in J} a_{0j}x_j\right\} \;\; s.t. \;\; Ax \le b \tag{18}$$

If we consider the nominal value of the parameter as $\overline{a_{0j}} = \frac{a_{0j}^{L}+a_{0j}^{U}}{2}$ and the deviation of each parameter from the nominal value as $\widehat{a_{0j}} = a_{0j}^{U} - \overline{a_{0j}}$, in an RP approach proposed by B&S [29], the conservatism level is controlled by defining a parameter $0 \le \Gamma \le |J|$ and presenting a new RC. Then, the RC of the above objective function is as follows:

$$\begin{gathered}\min_{x:\, Ax\le b} t \\ s.t. \\ \sum_{j\in J} \overline{a_{0j}}x_j + \\ \max_{\{S\cup\{t\}|S\subseteq J,|S|=\lfloor\Gamma\rfloor,t\in S\setminus J\}} \left\{\begin{matrix} \sum_{j\in S} \widehat{a_{0j}}|x_j| + \\ (\Gamma - \lfloor\Gamma\rfloor)\widehat{a_{0t}}|x_t| \end{matrix}\right\} \le t\end{gathered} \tag{19}$$

In the B&S article, based on the strong duality theorem, it is proved that the above model is equivalent to the following linear model. It should be explained that the considered problem meets the condition of having non-negative decision variables.

$$\begin{gathered}\min_{x:\, Ax\le b} t \\ s.t. \\ \sum_{j\in J} \overline{a_{0j}}x_j + \Gamma.q_0 + \sum_{j\in J} p_{0j} \le t \\ q_0 + p_{0j} \ge \widehat{a_{0j}}x_j \;\; \forall j \in J \\ q_0, p_{i0} \ge 0 \; \forall j \in J\end{gathered} \tag{20}$$

In the above robust model, known as the B&S method, if $\Gamma = 0$, only the nominal values of the parameters are considered. In this case, the conservatism is very insignificant due to ignoring the parameters perturbation. The more $\Gamma \to |J|$, the more conservative the model.

Regarding the uncertainty of the unit production cost in plant f (qc_f), transportation cost between plant f and DC d ($tc^{f\to d}$), and transportation cost between DC d and customer c ($tc^{d\to c}$), uncertain quantities are replaced with the production and transportation costs in the objective function. Also, Equation (21) is added to the optimization problem.

$$\begin{gathered}\theta \ge \sum_{f\in\mathcal{F}}\sum_{t\in\mathcal{T}} qc_f.Q_{ft} + \\ \sum_{f\in\mathcal{F}}\sum_{d\in\mathcal{D}}\sum_{t\in\mathcal{T}} tc^{f\to d}n_t^{f\to d} + \\ \sum_{d\in\mathcal{D}}\sum_{c\in\mathcal{C}}\sum_{t\in\mathcal{T}} tc^{d\to c}n_t^{d\to c}\end{gathered} \tag{21}$$

Finally, regarding the general form (20), the robust counterpart of the above equations replaces.

5. TRADE-OFF OBJECTIVE FUNCTIONS

To solve two/multi-objective optimization problems (MODM), various approaches were proposed, including the weighted sum method (WSM), Epsilon constraint (EC), augmented Epsilon constraint (AEC), goal programming (GP), and lexicographic (Lex). The general MODM problem is formulated as model (22).

$$\begin{cases} Min\ (f_1(x), f_2(x), \dots, f_n(x)) \\ x \in X \end{cases} \tag{22}$$

Suppose the first goal is the main goal, and the other objectives are restricted to a higher bound and are applied as the constraints of the problem. If the EC method is used, the following single-objective model is obtained:

$$\begin{cases} Min\ f_1(x) \\ f_i(x) \le \varepsilon_i \; i = 2,3,..,n \\ x \in X \end{cases} \tag{23}$$

where the second to nth targets are limited to the maximum value of ε_i. In model (23), different solutions are achieved by changing the values of ε_i, which may fail to be efficient. By partially amending model (23), known as the AEC method, the mentioned demerit can be solved. To better implement the AEC method, the appropriate initial range of ε_i can be obtained from Lex [36]. In the AEC method, first, a suitable range of ε_i changes must be determined. Then, for different values of ε_i, the Pareto front must be obtained. The AEC model is as follows:

$$\begin{cases} Min\ f_1(x) - \sum_{i=2}^{n} \phi_i s_i \\ f_i(x) + s_i = \varepsilon_i \; i = 2,3,..,n \\ x \in X \\ s_i \ge 0 \end{cases} \tag{24}$$

where s_i is a non-negative variable for slack, and ϕ_i is a parameter for normalizing the first objective function value relative to objective i ($\phi_i = \frac{R(f_1)}{R(f_i)}$). In the proposed AEC method, we first set the range $\varepsilon_i \in [\mathrm{Min}(f_i), \mathrm{Max}(f_i)]$ based on the Lex method for the objectives, and after setting ε_i, we solve the proposed robust model. Therefore, the final robust and single-objective model is as follows:

$$\begin{gathered}\max \pi^{PD}=pr.\sum_{d\in\mathcal{D}}\sum_{c\in\mathcal{C}}\sum_{h\in\mathcal{H}}\sum_{t\in\mathcal{T}}fw_{ht}^{d\to c}-\\ \left(\begin{array}{c}\left[\sum_{f\in\mathcal{F}}\sum_{t\in\mathcal{T}}fc_f.sz_{ft}\right]+\\ \left[\sum_{d\in\mathcal{D}}\sum_{k\in\mathcal{K}}fdc_{dk}.dz_{dk}\right]+\\ \left[\sum_{d\in\mathcal{D}}\sum_{h\in\mathcal{H}}\sum_{t\in\mathcal{T}}hc_d^h.I_{dt}^h\right]+\\ \left[\sum_{d\in\mathcal{D}}\sum_{c\in\mathcal{C}}\sum_{h\in\mathcal{H}}\sum_{t\in\mathcal{T}}\left(s_2^h.pr+\Omega.s_3^h\right)fw_{ht}^{d\to c}\right]+\\ \left[\sum_{c\in\mathcal{C}}\sum_{t\in\mathcal{T}}(pr-\Psi)r_{ct}\right]+[A]+\theta+\phi.\mathrm{w}\end{array}\right)\\ s.t.\end{gathered} \tag{25}$$

$$\begin{gathered}\left[\begin{array}{c}\sum_{f\in\mathcal{F}}\sum_{d\in\mathcal{D}}\sum_{t\in\mathcal{T}}e^{f\to d}\left[\frac{fw_t^{f\to d}}{cap^{f\to d}}\right]+\\ \sum_{d\in\mathcal{D}}\sum_{c\in\mathcal{C}}\sum_{h\in\mathcal{H}}\sum_{t\in\mathcal{T}}e^{d\to c}\left[\frac{fw_{ht}^{d\to c}}{cap^{d\to c}}\right]\end{array}\right]+\\ w=\varepsilon\ (Corresponding\ to\ E^{PD})\\ RC\left(\begin{array}{c}\theta\geq\sum_{f\in\mathcal{F}}\sum_{t\in\mathcal{T}}qc_f.Q_{ft}+\\ \sum_{f\in\mathcal{F}}\sum_{d\in\mathcal{D}}\sum_{t\in\mathcal{T}}tc^{f\to d}n_t^{f\to d}+\\ \sum_{d\in\mathcal{D}}\sum_{c\in\mathcal{C}}\sum_{t\in\mathcal{T}}tc^{d\to c}n_t^{d\to c}\end{array}\right)\\ \mathrm{x}\in\mathrm{X}\\ w\geq 0\end{gathered} \tag{26}$$

where $\mathrm{x}\in\mathrm{X}$ represents all the constraints of the developed model in the previous sections. In the next section, model (25) is run for a case study to evaluate and analyze the developed model's output.

6. NUMERICAL RESULTS

In this section, a case study is carried out to prove the extended model's applicability. It should be noted that we used GAMS software to implement the model.

6. 1. Case Study Saida Company in Isfahan is a manufacturer and distributor of ready-to-eat foodstuffs as perishable goods. The company's plants are located on Isfahan's outskirts, and the company does not have DCs inside the city. Therefore, to distribute its perishable goods inside the city, it requires establishing DCs in appropriate locations and pursuing its PDP in an integrated manner. By investigating the sales amount and longevity of the goods, we concluded that optimizing the P-D decisions of one certain product, lettuce, is more important at present. In the following, we name Saida Company as S company. The price and freshness of lettuce have a notable effect on sales. S company applies the three proposed strategies to increase its market share.

S company has established two plants ($\mathcal{F}=\{1,2\}$) located outside the city of Isfahan. After meetings with the managers, it was concluded that the number of candidate locations for the DCs is 12 locations ($\mathcal{D}=\{1,2,\dots,12\}$) with three types of capacity ($\mathcal{K}=\{1,2,3\}$). The capacity of these DCs is 1500, 750, and 500, respectively. According to the clustering, customers are centralized in 20 points of the city ($\mathcal{C}=\{1,2,\dots,20\}$). The product has a shelf life of 4 days ($\mathcal{H}=\{1,2,\dots,4\}$) and the P-D decisions are made in a 30-day cycle ($\mathcal{T}=\{1,2,\dots,30\}$). S company gives each unit of the product at a price of 120,000 Rials to the market. To facilitate calculations, we consider each unit's price equal to 12,000 Tomans and express every thousand Tomans as a unit in all the income and costs.

6. 2. Result Analysis

6. 2. 1. Determining the Pareto Front and the Best Solution After running the developed robust model, the Pareto optimal solutions are obtained regarding the model's objectives and the used trade-off. Given the uncertainty of some cost parameters and the robust approach, we assume that $\Gamma=6800$. We should note that the sensitivity analysis on the amount and how to adjust this parameter is presented in the following subsections.

Table 1 shows the corresponding payoff matrix. The Pareto front is also obtained, as shown in Figure 2. According to Figure 2, as the second objective, i.e., the minimization of the environmental impacts, gets worse, the first objective, i.e., the profit, improves. In other words, to the extent that an increase in greenhouse gas emissions is allowed (the second goal gets worse), there is much more profit. But from a specific value onwards, although the second objective worsens, not much profit is made for the first objective. Therefore, that point would be a practical solution for reporting to the management. Consequently, we select the Pareto solution of 110,000 and 9,050 units, respectively, for the first and second objectives and present the subsequent analysis based on this solution.

TABLE 1. The payoff matrix

	π^{PD}	E^{PD}
π^{PD}	112,300	67,000
E^{PD}	10,000	8,150

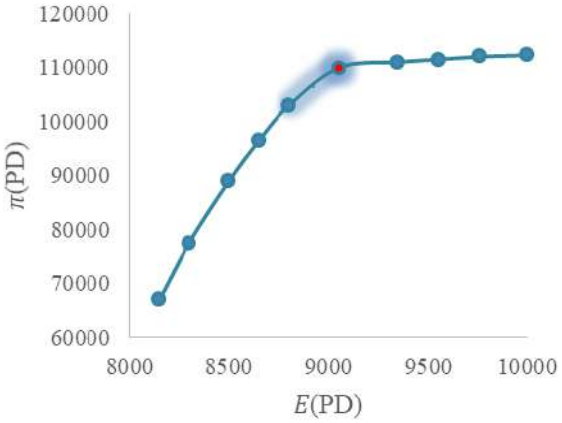


Figure 2. The Pareto front

In the following, some of the most important outputs of the robust model are reported based on Pareto's best-selected solution. The optimal amount of advertisement costs is 43,940 units. The optimal amount of production in the first seven periods is given in Table 2. Table 3 reports the optimal rates. Because only the perished goods are accepted for return, the return rate on the fourth day of a product age is obtained ten percent ($s_1^4 = 0.1$).

After receiving the location result, it is determined that four DCs out of 12 potential locations will be established in the optimal solution. The location of the activated DCs is shown in Figure 3. One large-capacity DC, two medium-capacity DCs, and one small-capacity DC should be established. Additionally, the S company's profit in the best-selected solution is 110,000 units, 30,000 units more than when it does not use our integrated GPDP model.

6. 2. 2. Robustness and Conservatism Analysis

We analyzed model robustness and how to adjust the conservatism coefficient (Γ). For this purpose, the conservatism coefficient value has changed from the maximum to minimum, and it is assumed that the value of the first objective, i.e., the profit, is considered. According to the uncertain parameters and their dimensions, the maximum value for Γ is 7,980, indicating that the uncertainty of the parameters is completely controlled. When one hundred percent conservatism is applied, the profit will be at its lowest value. As the conservatism decreases, the profit definitely increases. Figure 4 shows the changes in the profit over the percentage changes in conservatism. The percentage of conservatism is actually the percentage of uncertainty that is controlled. This figure illustrates well the effect of conservatism on the profit. In fact, by reducing the conservatism from the specified amount onwards, there is no significant profit increase. Therefore, that point is a proper solution to the conservatism coefficient we considered in the previous subsections.

TABLE 2. The optimal amount of production in the first week

f	t	1	2	3	4	5	6	7
1		2000	1500	0	1000	2000	1500	0
2		1500	1000	0	0	1500	0	0

TABLE 3. The optimal discount and credit rates

h	s_2^h	s_3^h
1	0	0
2	0.1	0.05
3	0.2	0.1
4	0.45	0.25

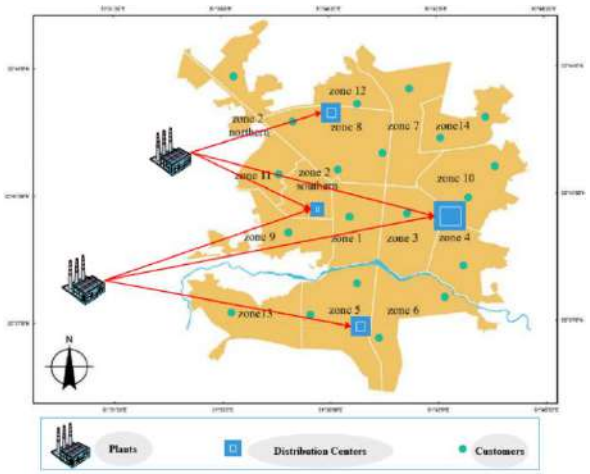


Figure 3. The established DCs

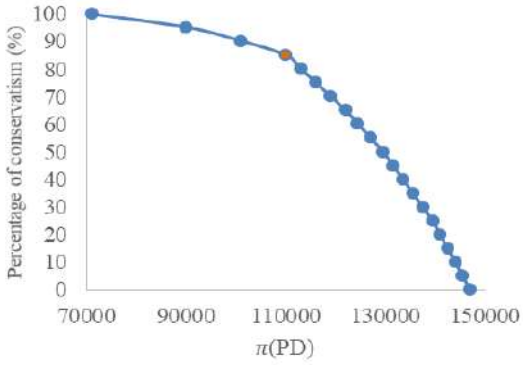


Figure 4. The profit changes affected by the conservatism

6. 3. Sensitivity Analysis

6. 3. 1. The Encouragement Strategies and P-D Parameters We analyzed the effect of unit production, transportation, setup costs and the impact of applying the encouragement strategies on the profit. In Figure 5, the profit decreases due to the rise in the unit production cost. Figure 6 shows an almost linear decline in the profit as the transportation costs increased. Finally, Figure 7 shows the trend of declining profit in exchange for an increase in fixed setup costs, which has less effect on the profit than does the unit production cost. Besides, in Figure 8, it is seen that the profit increases by considering each of the proposed strategies. The maximum profit is obtained when all three encouragement strategies were applied.

6. 3. 2. The Demand Function In addition to previous sensitivity analysis, we investigate here the impact of some parameters from the developed demand function, i.e., α, υ, and $\mathscr{b}$, on the potential demand and supply, income, and the encouragement strategies.

Figures 9, 10, and 11 analyze the change in the coefficients α and υ and parameter $\mathscr{b}$.

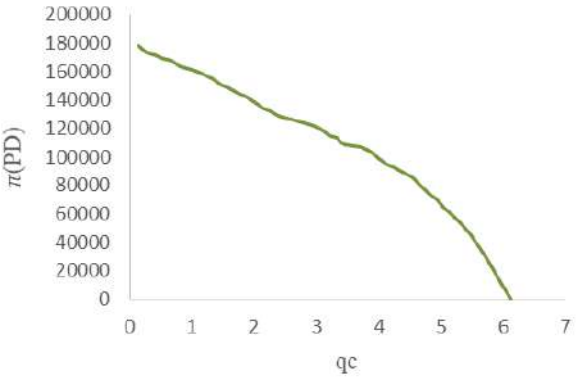


Figure 5. The impact of the unit production cost changes

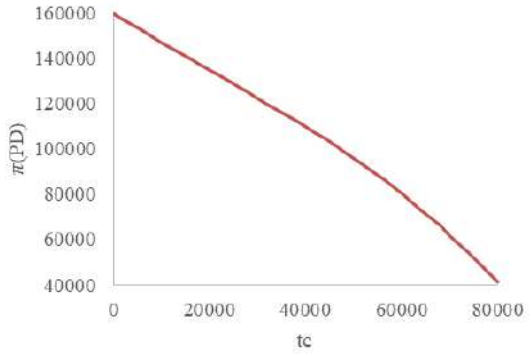


Figure 6. The impact of the transportation cost changes

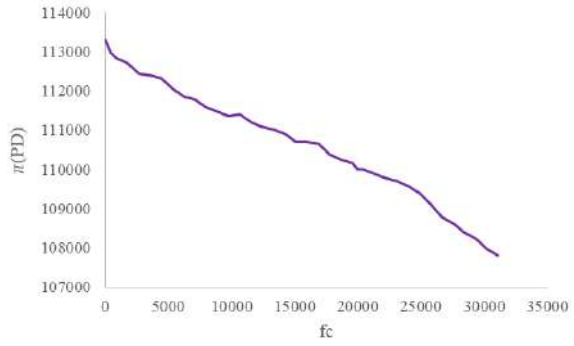


Figure 7. The impact of the setup cost changes

Considering Figure 9a, the potential demand and the supply decrease by an increase in α. From $\alpha = 0.14$ onwards, step by step, S company meets the potential demand. Figure 9b depicts an increase in the potential demand and the supply as υ increases. The growing trend for the supply continues to the point that S company fails to supply more due to its limited capacity. Indeed, from roughly $\upsilon = 1$, S company cannot follow the potential demand. From Figure 9c, it is seen that the potential demand has an increasing linear trend by an increase in $\mathscr{b}$ and the supply goes up, as well. It can be understood that S company fails to meet the potential demand from $\mathscr{b} = 168{,}000$ onwards.

Figure 8. The impact of the proposed strategies on the profit

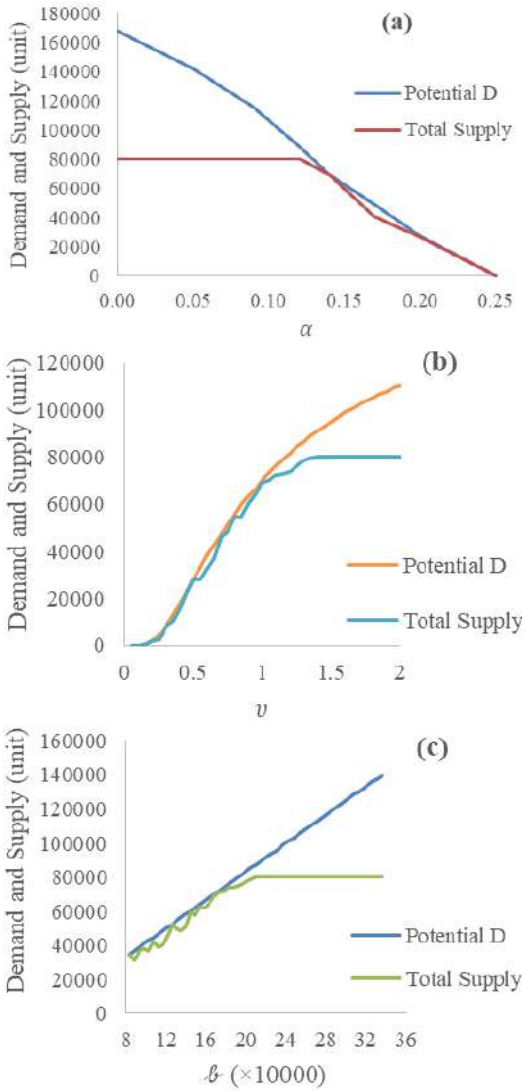


Figure 9. The analysis of α, υ, and $\mathscr{b}$ on the demand/supply

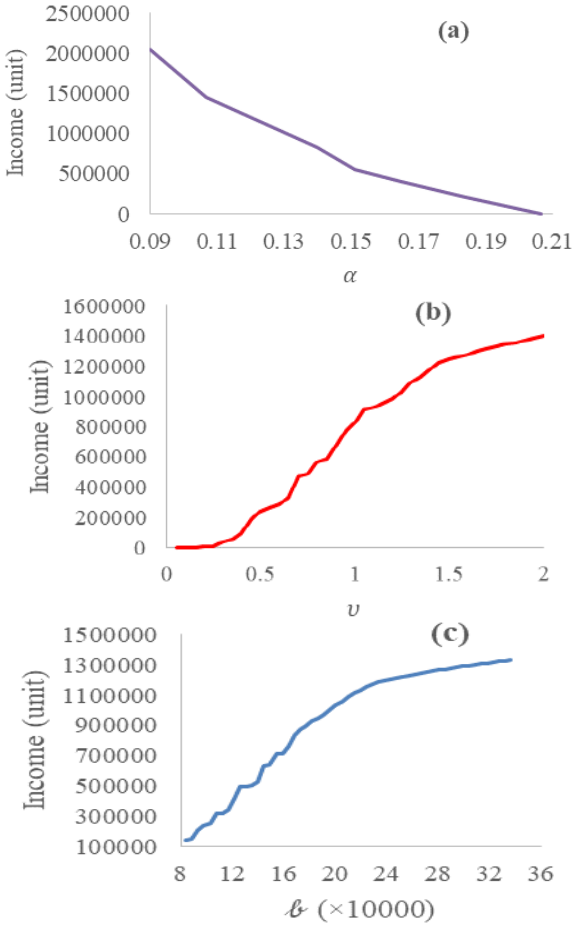
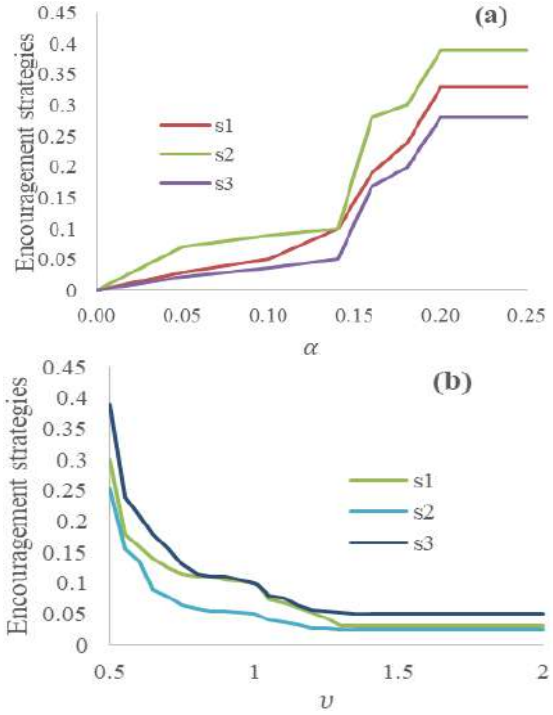


Figure 10. The analysis of α, υ, and $\mathcal{B}$ on the income



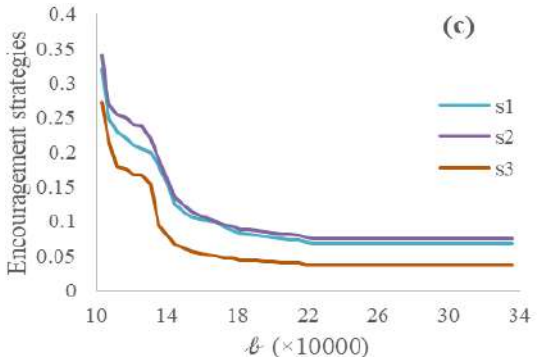


Figure 11. The analysis of α, υ, and $\mathcal{B}$ on the strategies

Figure 10a shows the downtrend of the income for an increase in α. Figures 10b and 10c show the increasing trend of the income by an increase in υ and the parameter $\mathcal{B}$, respectively. Figure 11a indicates that the encouragement strategies have a rising trend by an increase in α. Indeed, the encouragement strategies must compensate for the decreased potential demand observed in Figure 9a. As demonstrated in Figure 11b, the amount of encouragement strategies goes down when coefficient υ is increased. As the potential demand has an increasing manner with the rise in υ, the encouragement strategies are not required to be increased. For the same reason, in Figure 11c, the encouragement strategies show a reducing trend in exchange for the rise in parameter $\mathcal{B}$.

7. CONCLUSIONS

In this work, we focused on a bi-objective P-D optimization problem for perishable goods under uncertainty. The first goal was to maximize the P-D company's profit in an SC with three levels: plants, DCs, and customers. After considering the economic dimension, regarding the importance of green issues in distribution problems, the second goal was to minimize the environmental impact. In fact, for the P-D company, which owns its plants and DCs, we pursued the optimum solution for its location, inventory, production, and distribution problems. To make a trade-off between the environmental and economic goals and obtain the Pareto front, we used the augmented Epsilon constraint method. Then, we conducted a real case study to verify the application of our developed model. Concerning the control of uncertainty through Bertsimas and Sim's approach, we analyzed model robustness and how properly the conservatism coefficient was adjusted. We also performed other sensitivity analyses on the problem's main parameters to achieve some significant management results. Finally, the numerical results using the case study data showed the developed model's efficiency and performance. With the aim of industrial

application, this research brings optimization models close to real-world conditions, especially by considering the operational uncertainty. Some managerial insights can be obtained from this research as follows:

- Considering the product freshness in developing the demand function and modeling the PDP in an integrated structure can prevent the goods' perishability to an acceptable extent.
- Employing the perished product return, discount, and crediting strategies can significantly compensate for the lost market share.
- Considering the P-D decisions in an integrated framework can increase the profit, in addition to a decrease in the environmental effects.
- Coping with the operational uncertainty using the B&S robust optimization creates rational and realistic conservatism, model robustness and prevents risk in the decision-making.
- Using the developed two-objective model, we obtained solutions with a significant reduction in environmental effects at the cost of a slight reduction in the final profit.

In summary, the key contributions are as follows. First, we developed a two-objective multi-period optimization model that for the first time addresses the integrated P-D decisions for specific $\mathcal{P}$ goods. The objectives were to maximize the profit and minimize the emission of environmental pollution. Second, a new demand function was introduced considering three key factors, including price, advertisement, and product freshness, as well as three customer encouraging strategies, i.e., the discount, return, and credit policies. Third, we utilized robust optimization to cope with the operational uncertainty of the production and transportation costs. There are some suggestions to develop this paper. For instance, one can model the freshness factor impact on demand function in a nonlinear approach. Other methods for controlling the uncertainty, such as stochastic programming, can be used. It will also be valuable to provide meta-heuristic approaches to solve large-scale problems. Finally, more SC levels can be considered, and the model can be extended for multi-product situations.

8. REFERENCES

1. Kazemi, A., Fazel Zarandi, M. H. and Moattar Husseini, S. M., "A Multi-Agent System to Solve the Production-Distribution Planning Problem for a Supply Chain: A Genetic Algorithm Approach." ***International Journal of Advanced Manufacturing Technology***, Vol. 44, No. 1-2, (2009), 180-193, doi: 10.1007/s00170-008-1826-5.
2. Wei, W., Guimarães, L., Amorim, P. and Almada-Lobo, B., "Tactical Production and Distribution Planning with Dependency Issues on the Production Process." ***Omega***, Vol. 67, (2017), 99-114, doi: 10.1016/j.omega.2016.04.004.
3. Chen, Z., "Optimization of Production Inventory with Pricing and Promotion Effort for a Single-Vendor Multi-Buyer System of Perishable Products." ***International Journal of Production Economics***, Vol. 203, (2018), 333–49, doi: 10.1016/j.ijpe.2018.06.002.
4. Amorim, P., Günther, H.O. and Almada-Lobo, B., "Multi-Objective Integrated Production and Distribution Planning of Perishable Products." ***International Journal of Production Economics***, Vol. 138, No. 1, (2012), 89-101, doi: 10.1016/j.ijpe.2012.03.005.
5. Nahmias, S., "Perishable Inventory Theory: A Review." ***Operations Research***, Vol. 30, No. 4, (1982), 680-708, doi: 10.1287/opre.30.4.680.
6. Xie, J. and Wei, J.C., "Coordinating Advertising and Pricing in a Manufacturer–Retailer Channel." ***European Journal of Operational Research***, Vol. 197, No. 2, (2009), 785-791, doi: 10.1016/j.ejor.2008.07.014.
7. Amirtaheri, O., Zandieh, M., Dorri, B. and Motameni, A.R., "A Bi-Level Programming Approach for Production-Distribution Supply Chain Problem." ***Computers and Industrial Engineering***, Vol. 110, (2017), 527-537, doi: 10.1016/j.cie.2017.06.030.
8. Avinadav, T., Herbon, A. and Spiegel, U., "Optimal Inventory Policy for a Perishable Item with Demand Function Sensitive to Price and Time." ***International Journal of Production Economics***, Vol. 144, No. 2, (2013), 497-506, doi: 10.1016/j.ijpe.2013.03.022.
9. Esmaili, M. and Sahraeian, R., "A New Bi-Objective Model for a Two-Echelon Capacitated Vehicle Routing Problem for Perishable Products with the Environmental Factor." ***International Journal of Engineering, Transactions A: Basics***, Vol. 30, No. 4, (2017), 523-531, doi: 10.5829/idosi.ije.2017.30.04a.10.
10. Raa, B., Dullaert, W. and Aghezzaf, E.H., "A Matheuristic for Aggregate Production–Distribution Planning with Mould Sharing." ***International Journal of Production Economics***, Vol. 145, No. 1, (2013), 29-37, doi: 10.1016/j.ijpe.2013.01.006.
11. Fahimnia, B., Farahani, R.Z., Marian, R. and Luong, L., "A Review and Critique on Integrated Production–Distribution Planning Models and Techniques." ***Journal of Manufacturing Systems***, Vol. 32, No. 1, (2013), 1-19, doi: 10.1016/j.jmsy.2012.07.005.
12. Ahumada, O. and Villalobos, J.R., "Operational Model for Planning the Harvest and Distribution of Perishable Agricultural Products." ***International Journal of Production Economics***, Vol. 133, No. 2, (2011), 677-687, doi: 10.1016/j.ijpe.2011.05.015.
13. Fahimnia, B., Luong, L. and Marian, R., "Genetic Algorithm Optimisation of an Integrated Aggregate Production–Distribution Plan in Supply Chains." ***International Journal of Production Research***, Vol. 50, No. 1, (2012), 81-96, doi: 10.1080/00207543.2011.571447.
14. Amorim, P., Meyr, H., Almeder, C. and Almada-Lobo, B., "Managing Perishability in Production-Distribution Planning: A Discussion and Review." ***Flexible Services and Manufacturing Journal***, Vol. 25, No. 3, (2013), 389-413, doi: 10.1007/s10696-011-9122-3.
15. Bilgen, B. and Çelebi, Y., "Integrated Production Scheduling and Distribution Planning in Dairy Supply Chain by Hybrid Modelling." ***Annals of Operations Research***, Vol. 211, No. 1, (2013), 55-82, doi: 10.1007/s10479-013-1415-3.
16. Díaz-Madroñero, M., Peidro, D. and Mula, J., "A Review of Tactical Optimization Models for Integrated Production and Transport Routing Planning Decisions." ***Computers & Industrial Engineering***, Vol. 88, (2015), 518-535, doi: 10.1016/j.cie.2015.06.010.

17. Makui, A., Heydari, M., Aazami, A., & Dehghani, E., "Accelerating Benders decomposition approach for robust aggregate production planning of products with a very limited expiration date." ***Computers & Industrial Engineering***, Vol. 100, (2016), 34-51, doi: 10.1016/j.cie.2016.08.005.
18. Fattahi, P., Bashiri, M. and Tanhatalab, M., "Bi-Objectives Approach for a Multi-Period Two Echelons Perishable Product Inventory-Routing Problem with Production and Lateral Transshipment." ***International Journal of Engineering, Transactions C: Aspects***, Vol. 30, No. 6, (2017), 876-886, doi: 10.5829/ije.2017.30.06c.08.
19. Devapriya, P., Ferrell, W. and Geismar, N., "Integrated Production and Distribution Scheduling with a Perishable Product." ***European Journal of Operational Research***, Vol. 259, No. 3, (2017), 906-916, doi: 10.1016/j.ejor.2016.09.019.
20. Ensafian, H., and Yaghoubi, S., "Robust optimization model for integrated procurement, production and distribution in platelet supply chain." ***Transportation Research Part E: Logistics and Transportation Review***, Vol. 103, (2017), 32-55, doi: 10.1016/j.tre.2017.04.005.
21. Guarnaschelli, A., Salomone, H.E. and Méndez, C.A., "A Stochastic Approach for Integrated Production and Distribution Planning in Dairy Supply Chains." ***Computers and Chemical Engineering***, Vol. 140, (2020), doi: 10.1016/j.compchemeng.2020.106966.
22. Biuki, M., Kazemi, A. and Alinezhad, A., "An Integrated Location-Routing-Inventory Model for Sustainable Design of a Perishable Products Supply Chain Network." ***Journal of Cleaner Production***, Vol. 260, (2020), doi: 10.1016/j.jclepro.2020.120842.
23. Liu, P., Hendalianpour, A., Razmi, J., and Sangari, M. S., "A solution algorithm for integrated production-inventory-routing of perishable goods with transshipment and uncertain demand." ***Complex & Intelligent Systems***, (2021), 1-17, doi: 10.1007/s40747-020-00264-y.
24. Bakker, M., Riezebos, J. and Teunter, R.H., "Review of Inventory Systems with Deterioration since 2001." ***European Journal of Operational Research***, Vol. 221, No. 2, (2012), 275-284, doi: 10.1016/j.ejor.2012.03.004.
25. Coelho, L.C. and Laporte, G., "Optimal Joint Replenishment, Delivery and Inventory Management Policies for Perishable Products." ***Computers and Operations Research***, Vol. 47, (2014), 42-52, doi: 10.1016/j.cor.2014.01.013.
26. Janssen, L., Claus, T. and Sauer, J., "Literature Review of Deteriorating Inventory Models by Key Topics from 2012 to 2015." ***International Journal of Production Economics***, Vol. 182, (2016), 86-112, doi: 10.1016/j.ijpe.2016.08.019.
27. Gharehyakheh, A., Krejci, C. C., Cantu, J., and Rogers, K. J., "A Multi-Objective Model for Sustainable Perishable Food Distribution Considering the Impact of Temperature on Vehicle Emissions and Product Shelf Life." ***Sustainability***, Vol. 12, No. 16, (2020), doi: 10.3390/su12166668.
28. Navazi, F., Sazvar, Z., & Tavakkoli-Moghaddam, R., "A sustainable closed-loop location-routing-inventory problem for perishable products." ***Scientia Iranica***, (2021), doi: 10.24200/SCI.2021.55642.4353.
29. Bertsimas, D. and Sim, M., "The Price of Robustness." ***Operations Research***, Vol. 52, No. 1, (2004), 35-53, doi: 10.1287/opre.1030.0065.
30. Aazami, A., and Saidi-Mehrabad, M., "Benders decomposition algorithm for robust aggregate production planning considering pricing decisions in competitive environment: A case study." ***Scientia Iranica***, Vol. 26, No. 5, (2019), 3007-3031, doi: 10.24200/SCI.2018.5563.1346.
31. Aazami, A., & Saidi-Mehrabad, M., "A production and distribution planning of perishable products with a fixed lifetime under vertical competition in the seller-buyer systems: A real-world application." ***Journal of Manufacturing Systems***, Vol. *58*, (2021), 223-247, doi: 10.1016/j.jmsy.2020.12.001.
32. Kaviyani-Charati, M., Heidarzadeh Souraki, F. and Hajiaghaei-Keshteli, M., "A Robust Optimization Methodology for Multi-Objective Location-Transportation Problem in Disaster Response Phase under Uncertainty." ***International Journal of Engineering, Transactions B: Applications***, Vol. 31, No. 11, (2018), 1953-1961, doi: 10.5829/ije.2018.31.11b.20.
33. Yanıkoğlu, İ., Gorissen, B.L. and den Hertog, D., "A Survey of Adjustable Robust Optimization." ***European Journal of Operational Research***, Vol. 277, No. 3, (2019), 799-813, doi: 10.1016/j.ejor.2018.08.031.
34. Beyer, H.G. and Sendhoff, B., "Robust Optimization-a Comprehensive Survey." ***Computer Methods in Applied Mechanics and Engineering***, Vol. 196, No. 33-34, (2007), 3190-3218, doi: 10.1016/j.cma.2007.03.003.
35. Ben-Tal, A., El Ghaoui, L. and Nemirovski, A., Robust Optimization., Vol. 28, **Princeton University Press**, (2009), ISBN-13: 978-0691143682.
36. Nikas, A., Fountoulakis, A., Forouli, A. and Doukas, H., "A Robust Augmented -Constraint Method (AUGMECON-R) for Finding Exact Solutions of Multi-Objective Linear Programming Problems." ***Operational Research***, (2020), 1-42, doi: 10.1007/s12351-020-00574-6.

Persian Abstract

چکیده

این مقاله، یک مدل بهینه‌سازی دوهدفه را برای برنامه‌ریزی یکپارچه تولید–توزیع چنددوره‌ای محصولات فسادپذیر تحت شرایط عدم قطعیت، توسعه می‌دهد. هدف اول، سود را در یک زنجیره تأمین سه سطحی شامل کارخانه‌ها، مراکز توزیع و مشتریان، حداکثر می‌نماید. ازآنجایی‌که موضوع حمل‌ونقل در مسئله توزیع، یکی از مهم‌ترین منابع آلودگی‌های زیست‌محیطی است؛ هدف دوم در پی کمینه‌سازی انتشار آن‌هاست. در مسئله موردنظر، تصمیمات تولید، مکان‌یابی، موجودی و توزیع به‌صورت یکپارچه، اتخاذ می‌شوند. علاوه بر قیمت، با توجه به فسادپذیر بودن محصول، اثر تازگی محصول نیز در توسعه تابع تقاضا، اعمال می‌گردد. همچنین، جهت تشویق مشتریان، سه استراتژی شامل سیاست‌های بازگشت محصول فاسد، تخفیف، و دوره اعتباری، پیشنهاد می‌شوند. به‌منظور مقابله با عدم قطعیت درونی برخی پارامترهای هزینه، بهینه‌سازی استوار اعمال می‌شود. به‌منظور نشان دادن کاربردپذیری مدل توسعه‌یافته و نیز عملی بودن لحاظ جنبه زیست‌محیطی، یک مطالعه موردی، انجام می‌شود. محاسبات عددی روی مطالعه موردی، موازنه‌ای را بین اهداف زیست‌محیطی و اقتصادی، ارائه می‌کند و ۳۷.۵ درصد افزایش سود را با استفاده از مدل توسعه داده‌شده، نشان می‌دهد.

IJE TRANSACTIONS A: Basics Vol. 34, No. 7, (July 2021) 1778-1784

International Journal of Engineering

Journal Homepage: www.ije.ir

System of Comprehensive Assessment of Project Risks in Energy Industry

G. A. Stroykov[a], N. V. Babyr*[b], I. V. Ilin[c], R. S. Marchenko[c]

[a] *Organization and Management Department, Saint-Petersburg Mining University, Russian Federation*
[b] *Saint-Petersburg Mining University, St. Petersburg, Russian Federation*
[c] *Higher school of Business and Management, Peter the Great St. Petersburg Polytechnic, University Federal State Autonomous Educational, St. Petersburg, Russian Federation*

PAPER INFO

Paper history:
Received 19 April 2021
Received in revised form 08 May 2021
Accepted 06 June 2021

Keywords:
Project Risks
Energy Industry
Risk Assessment
Simulation Modeling
Monte Carlo Method
Investment Projects

ABSTRACT

The article proposes to consider the problem of comprehensive assessment of project risks as applied to the energy industry. The authors of the research focused on the description of the applied solution. A real investment project on replacement of a bark boiler at Mondi Syktyvkar enterprise was chosen as an object for testing the results. We proposed to divide the risks accompanying the project into 2 categories: risks for which there is necessary and statistical information for their quantitative assessment and risks for which this information is absent. As a technique of a quantitative assessment of risks from the first category it is expedient to apply a method of Simulation modeling of Monte Carlo. In this case, the authors of the article conducted a significant analysis of existing methods for assessing project risks and the choice of the Monte Carlo methodology is due to the practical orientation of the study. In practice, the real enterprise is quite problematic to use more complex methods of assessment, such as methods of Real Options or methods of fuzzy logic, neural networks, etc. As a method of qualitative risk assessment (from the second category) the method of expert evaluation with subsequent calculation of risk premium in the discount rate was chosen. This method is common in practice and easy enough to implement. According to the results of the analysis (statistical and expert) the most dangerous risks of energy projects were identified: Production and technological risks (the risk of choosing the wrong technological scheme, the risk of reducing the quality of internal controls, the risk of incorrect calculation of the design capacity of energy production, the risk of industrial safety), security risks (the risk of hacking attacks on information systems of energy enterprises), as well as country risks. Among the most influential risks (based on the analysis of their impact on the main technical and economic indicators of the project) are: the risk of rising prices for purchased gas (fuel), the risk of high volatility of the dollar exchange rate. The results of the study were used in a real project and the risk assessment methodology was implemented in the project activities of Mondi Syktyvkar enterprise.

***doi**: 10.5829/ije.2021.34.07a.22*

1. INTRODUCTION

Major investments in energy projects are fraught with a large number of risks that are both common to all investment projects and specific to energy industry. Lack of a comprehensive risk assessment methodology can result in negative consequences for the company.

As energy generating equipment that is operated by electricity providers (power grids, substations, and process control systems) and most manufacturing companies is always subject to wear, it becomes necessary to invest a large amount of money in equipment upgrades and projects connected with equipment upgrades in the energy industry. There are a lot of risks associated with such investments, including those specific to industry, that need to be correctly assessed at the stage of conducting a feasibility study for each investment project in energy sector. Companies usually do not use complex risk assessment methods; as a result, the quality of risk management deteriorates and they cannot reach the same quality level as some competitors boast. In fact, manufacturing companies

*Corresponding Author Institutional Email: babyr_nv@pers.spmi.ru
(N. V. Babyr)

Please cite this article as: G. A. Stroykov, N. V. Babyr, I. V. Ilin, R. S. Marchenko, System of Comprehensive Assessment of Project Risks in Energy Industry, International Journal of Engineering, Transactions A: Basics Vol. 34, No. 7, (2021) 1778-1784

either do not assess risks at all or do it in a very formal way that does not take into account the specific features of the energy industry.

Researchers study such aspects as risk-return analysis in application to investments in renewable energy sources [1], the importance of project finance in low-risk projects [2], the advantages of the Monte Carlo modeling method in evaluating public-private partnership projects [3], and conducting feasibility studies based on the Monte Carlo method using multi-energy balance financial equations that take into account the uncertainties and risks associated with different variables in the design and construction of solar thermal power stations [4].

Recently, the issue of risk assessment has been raised by many researchers around the world, in particular such issue of project risk assessment applied to public-private partnership projects using the example of waste incineration in energy industry [5]. Assessement of the effect of external risks on the success of oil and gas construction projects [6], study and classify structures, methods, and models of in-project quantitative risk analysis [7], discuss the issue of risk perception in the integrated design and construction project delivery [8], and analyze performance risks [9].

For many researchers, the most pressing issue remains the problem the issues of integrating risk management systems in project decision making and those of improving project effectiveness by assessing project risks [10-12]. Some researchers discussed the problem of involving experts in project risk analysis [13]. A number of researchers analyze project risks based on the characteristics of a project [14], and some researchers have set themselves the task of developing models and methods for managing supply chain risks and delays in construction projects [15].

Russian researchers discussed investment risk management in the mining industry, the use of a risk-based approach to safety issues at coal deposits [16-18]. The strategic risk analysis have implemented investment projects [19-21], economic assessment of heat and electricity generation [22-24], anthropogenic hazard assessment [25], organizational and economic mechanisms for implementing strategic innovation projects [26-29], and the development of a stationary intelligent system for assessing and monitoring power quality indicators [30].

The main purpose of this study is to develop a system for a comprehensive assessment of risks associated with energy investment projects based on a combination of simulation modeling methods and discount rate calculation. The results of the study were tested on the example of the company Mondi Syktyvkar. One of the company's activities is the generation of electricity as a result of burning wood waste from the main production and natural gas in measles boilers. The company uses the received electricity for its production needs, and also sells it to third-party consumers.

2. THEORY AND EXPERIMENTAL

The reliability and validity of the statements, conclusions, and recommendations presented in the article stem from the fact that a significant body of theory on the topic was studied and numerous documents and statistical data on various manufacturing and energy companies were analyzed, including methodological recommendations for assessing project cost effectiveness and reports produced by rating and consulting agencies. To achieve the aim of the study, probability and statistical analysis, expert evaluation, and simulation methods were used. The methodological foundation of the study consists of works by leading Russian and foreign researchers in such fields as project risk assessment, project risk management, and simulation modeling, data provided by consulting and analytical agencies, and publicly available business reports [31].

The algorithm for conducting a qualitative risk analysis using the questionnaire method is shown in Figure 1:

Based on the results of the expert opinion survey and the statistical analysis, we identified critical risks (risk level> 31), dangerous risks (risk level from 21 to 30), moderate risks (risk level from 11 to 20), and also low risks (from 0 to 10).

Based on the results of the qualitative analysis (the choice between quantitative and qualitative analysis was made based on the availability of sufficient statistical data for each of the parameters), it can be concluded that the most critical risks in developing a standard energy project are the risk of underestimating capital costs for equipment, the risk of cyber attacks, and the risk of mistakes in electricity price planning. The greatest emphasis in project development should therefore be placed on organizational, marketing, and security risks. Application of the simulation modeling method (Monte

Figure 1. Algorithm of questionnaire method application for expert assessment of project risks

Carlo method) in combination with other statistical methods of qualitative and quantitative risk assessment provides the most realistic picture of the probability distribution of various risks in the aggregate affecting the investment project (provided that a qualitative assessment has been made, the risks have been identified, the probabilities and the degree of impact of risks have been determined) [32].

The Monte Carlo method used in mathematical modeling is applicable to managing uncertainty in some technical and economic parameters of the projecta. The resulting variables of the equation describing the project model and including these parameters are NPV, ID, and IRR, i.e. the main indicators of project performance, based on which strategic decisions concerning the project will be made. The Monte Carlo method makes it possible to take into account the uncertainty of the variables in the equation which is connected with a probability distribution.

In addition, the method does not take into account the presence of correlations or other relationships between the parameters of the model; as a result, a large number of invalid scenarios are simulated. It follows from the above that the approach under consideration needs to be refined in order to improve the reliability of simulation results.

Taking into account correlations between input parameters is a very important task in making the model being used as objective as possible but it cannot be solved using the existing tools. It is advisable to use the following algorithm for taking into account correlations between parameters (Figure 2).

3. RESULTS AND DISCUSSION

The study is based on a model that was developed by the authors using the feasibility study for the project aimed at installing a new energy generating device (a bark boiler) at Mondi Syktyvkar. The result of developing an economic and mathematical model is project performance indicators that are used by investors and company directors to decide whether the project is feasible or compare it with other projects.

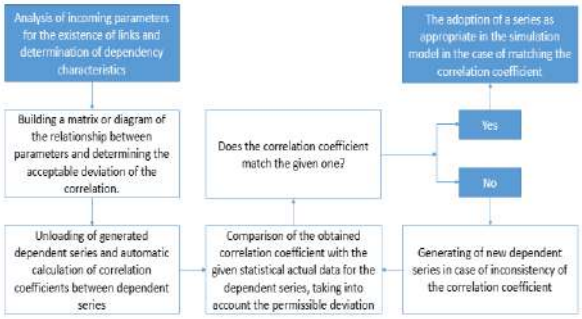


Figure 2. Flowchart of accounting for correlation between technical and economic parameters of the project model in Monte Carlo simulation. Source: developed by the authors

3. 1. Country Risk Premium Calculation The country risk premium was calculated based on data provided by Expert RA, a rating agency. It deals with analytics (mainly regarding the credit and insurance industry); since 1996, it has been compiling annual rankings of Russian regions by their attractiveness for investors. The methodology of country (and regional) risk rating is presented in researches.

Based on the scale presented, the country risk premium for the project aimed at installing a new bark boiler at a facility located in Syktyvkar (Komi Republic) is 3%.

3. 2. Risk Premiums for Other Types of Risks The discount rate, as mentioned earlier, takes into account all other important risks except for those that are covered in simulation modeling. According to the methodology, the least dangerous risks are not taken into account (natural hazards, environmental risks, and risks connected with infrastructure and logistics). The country risk premium in the project under consideration is 3%.

To find final risk values, the sum of the products of risk occurrence probabilities and numerical values of damage is calculated (in shares; the scale of damage corresponds to the scale in the questionnaire). The possible damage caused by the production risks is 26.8%, and that caused by the safety risks is 18.9%. To find risk premiums for these risks, it is necessary to add them up and analyze the sensitivity of the project's net present value (NPV) to changes in the discount rate (relative to net cash flows). The sum of the possible damage caused by the two types of risks being considered is 45%, so it is necessary to find the value of the discount rate at which the NPV deviates from the sum of cash flows by 45%. The results of the sensitivity analysis are shown in Table 1.

Based on the sensitivity analysis, the risk premium will be 4.37%. The value of the discount rate for the project aimed at installing a new bark boiler at Mondi Syktyvkar is taken to be 4.37% + 1.88% + 3% = 9.25%.

3. 3. Installing a New Bark Boiler at Mondi Syktyvkar: Simulation Modeling Simulation modeling was carried out using @RISK software and MS Excel[2]. The number of iterations was 5,000, and Latin hypercube sampling was the statistical method used. The main simulation modeling parameters are shown in Table 2.

[2] http://www.palisade.com/risk/ru/

TABLE 1. Analysis of NPV sensitivity to changes in the discount rate

Increase in the discount rate	NPV, mln RUB	Change in NPV, mln RUB
+4.37%	3931	-45%
+3.5%	4613	-20%

3. 4. Model Structure For the project under consideration, an economic model was developed using data provided by the company. According to this data, the estimated NPV is 692 million rubles, PI is 2.5, IRR is 28%, the payback period is 8.02 years, and the discounted payback period is 13.59 years.

3. 5. Choosing Distributions for the Input Parameters of the Model The type of probability distribution of a random value (as a characteristic of the input parameter of a simulation model) is determined by analyzing the distribution of a random value using special software (statistical analysis function @RISK). For this purpose, it is necessary to have a sufficient statistical base on the input parameter series. If such a database was available (e.g., by the values of actual and planned Capital costs of energy projects), the probability distribution of this value was analyzed. For Capital costs, the analysis showed the Exponential Distribution. In case there is no sufficient basis for the analysis, the normal distribution is accepted.

3.6 Input And Output Parameters Of The Model
Input model parameters selection is determined in each new project separately. Several technical and economic parameters used as input parameters have been selected in this study. The main principle of input parameters selection is influence of the revealed risks on the corresponding project model parameters. The following is a description of the input parameters of the energy project model, for which there was a sufficient basis for determining their probabilistic and statistical characteristics. For the other input parameters, the law of normal distribution of a random variable and a deviation from the mathematical expectation of about 10% of the studied parameter of the deterministic model were applied. The values of NPV, PI, and IRR are the output parameters of the simulation model.

TABLE 2. Simulation modeling parameters for installing a new bark boiler at Monti Syktyvkar

Number of simulations	1
Number of iterations	5,000
Number of inputs	56
Number of outputs	3
Random number generator	Mersenne Twister
RNG seed	1776983321

3. 6. 1. Price for Gas In the project under study, the main cost advantage of installing new energy-producing equipment was gained due to savings on fuel (gas). An analysis of the daily prices for gas over a 10-year period showed a triangular distribution of gas prices. The average value for this parameter was 3.9 rubles/m^3.

3. 6. 2. USD Exchange Rate It is proved that the hypothesis that the USD exchange rate is normally distributed. Based on this, normal distribution was adopted for the random variable of the USD/RUB exchange, and the main distribution characteristics (the mean value and the standard deviation) were taken based on an analysis of a sample of daily exchange rates for the last four years.

3. 7. Simulation Results
3. 7. 1. Net Present Value The main input parameter in the financial model of the investment project under consideration is the NPV. The simulation results are shown in Figure 4.

Simulation modeling was carried out using @RISK and MS Excel. As the simulation results showed, the most expected NPV value adjusted for risks is 587.65 million rubles, which is 15% less than that value that was calculated using the deterministic model (692 million rubles). The probability of having a non-negative NPV is 80%, which is a quite good result. Also, the probability of having a minimum NPV of -2,124.01 million rubles does not exceed 5%, with the same probability for having a maximum NPV of 6,945.67 million rubles.

At a confidence interval of 95% under the optimistic scenario, the NPV is 1,469 million rubles. The pessimistic scenario at a confidence interval of 5% reflects a negative NPV with a loss of 645 million rubles. The expected risk-adjusted PI is 1.9 (Figure 5).
The expected risk-adjusted IRR is 23%. In addition, a statistical analysis of the distribution of the risk-adjusted IRR was carried out.

Table 3 compares the two options for calculating the performance indicators of the project for installing new energy generating equipment (a bark boiler) at Mondi Syktyvkar, i.e. using a deterministic model and using Monte Carlo simulation modeling.

As a result of the study, the main project performance indicators that were adjusted for general risks, industry-specific risks, and correlations between input parameters turned out to be approximately 15% lower than the project values that were not adjusted for risks by simulation modeling.

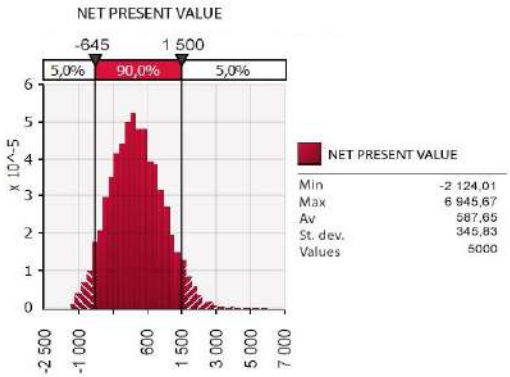


Figure 4. NPV distribution obtained by simulation modeling. Source: developed by the authors

Figure 5. PI distribution. Source: developed by the authors

TABLE 3. Model output parameters of the project for installing a new bark boiler at Mondi Syktyvkar

	Deterministic model	Monte Carlo simulation modeling
NPV (expected), mln RUB	692	587
PI (expected)	2.5	1.9
IRR (expected)	28%	23%

4. CONCLUSIONS

The findings of the study resulted in the following conclusions:

1. In order to apply a complex and comprehensive approach to investment project evaluation, a methodology was developed for assessing risks and risk types that companies face when implementing investment projects connected with energy producing equipment.When the methodology was tested on the project for installing a new bark boiler at Mondi Syktyvkar, the project performance indicators fell by almost 15%. This suggests that if risks are not properly assessed, project performance indicators may be exaggerated, which can lead to negative economic consequences for the company if a decision is made to invest in the project.

2. The authors proposed a methodology for calculating risk premiums which should be taken into account when calculating the discount rate for the project. As it is necessary to factor in important risks that cannot be covered by simulation modeling due to the fact that there are no input parameters in the model structure that can be directly affected by these risks, the risk premium should be calculated based on an analysis of the project's NPV sensitivity to changes in the discount rate, which is found by analyzing possible damage from the occurrence of these risks. Based on the results of the qualitative analysis, it was concluded that the most critical risks in developing a standard energy project are the risk of underestimating capital costs for equipment, the risk of cyber-attacks, and the risk of mistakes in electricity price planning.

3. To improve simulation modeling results, it is advisable to use a modification of the Monte Carlo simulation method that takes into account correlations between risks. If possible, correlations are not factored in the simulation results can become distorted. The modification of the simulation method has the form of an addition to the simulation algorithm (step 4) that includes six steps:

1. Analyzing the NPV calculation model and finding input and output parameters
2. Identifying and analyzing possible risks
3. Finding the type of probability distribution and the statistical characteristics
4. Finding correlations between model parameters
5. Carrying out NPV simulation modeling, generating scenarios that take into account the correlations, and checking scenarios for compliance with a given correlation coefficient
6. A statistical analysis of the resulting values

The methodology was tested using as a case study the project for installing a new bark boiler at Mondi Syktyvkar.

Besides, the expert methods applied for qualitative estimation of design risks can be not absolutely reliable. To level out this possibility, it is necessary to choose the right experts in accordance with their competence. Or a possible solution may be to apply certain weighting coefficients for evaluation of each expert. However, this methodology requires additional description and justification, which should be developed as a continuation in future studies of authors.

By combining qualitative methods of risk assessment and modern quantitative ones, it is possible to achieve really excellent results. In addition, it is necessary to consider the implementation of the developed risk assessment system in the business processes of the company, as well as to offer an application toolkit in the form of an information system.

These are the questions that lie in the plane of our research, which we are going to continue within the framework of this topic.

5. REFERENCES

1. Krömer, S., Gatzert, N. "Renewable energy investments with storage: a risk-return analysis", ***International Journal of Energy Sector Management***, Vol. 12, No. 4, (2018), 714-736. DOI: 10.1108/IJESM-02-2018-0009
2. Steffen, B. "The importance of project finance for renewable energy projects", ***Energy Economics***, Vol. 69, (2018), 280-294. DOI: 10.1016/j.eneco.2017.11.006
3. Berk, A.S., Podhraski, D. "Superiority of Monte Carlo simulation in valuing real options within public-private partnerships" ***Risk Management***, Vol. 20, No. 1, (2018), 1-28. DOI: 10.1057/s41283-017-0025-9
4. Gu, Y., Zhang, X., Are Myhren, J., Han, M., Chen, X., Yuan, Y. "Techno-economic analysis of a solar photovoltaic/thermal (PV/T) concentrator for building application in Sweden using Monte Carlo method", ***Energy Conversion and Management***, Vol. 165, (2018), 8-24. DOI: 10.1016/j.enconman.2018.03.043
5. Wu, Y., Xu, C., Li, L., Wang, Y., Chen, K., Xu, R. "A risk assessment framework of PPP waste-to-energy incineration projects in China under 2-dimension linguistic environment", ***Journal of Cleaner Production***, Vol. 183, (2018), 602-617. DOI: 10.1016/j.jclepro.2018.02.077
6. A Kassem, M., Khoiry, M.A., Hamzah, N. "Assessment of the effect of external risk factors on the success of an oil and gas construction project", ***Engineering, Construction and Architectural Management***, (2020). DOI: 10.1108/ECAM-10-2019-0573
7. Hartono, B. "From project risk to complexity analysis: a systematic classification", ***International Journal of Managing Projects in Business***, Vol. 11, No. 3, (2018), 734-760. DOI: 10.1108/IJMPB-09-2017-0108
8. Wu, P., Xu, Y., Jin, R., Lu, Q., Madgwick, D., Hancock, C.M. "Perceptions towards risks involved in off-site construction in the integrated design & construction project delivery", ***Journal of Cleaner Production***, Vol. 213, (2019), 899-914. DOI: 10.1016/j.jclepro.2018.12.226
9. Lee, P., Lam, P.T.I., Lee, W.L. Performance risks of lighting retrofit in Energy Performance Contracting projects. ***Energy for Sustainable Development***, Vol. 45, (2018), 219-229. DOI: 10.1016/j.esd.2018.07.004
10. Enshassi, M.S.A., Walbridge, S., West, J.S., Haas, C.T. Integrated Risk Management Framework for Tolerance-Based Mitigation Strategy Decision Support in Modular Construction Projects. ***Journal of Management in Engineering***, Vol. 35, No. 4. (2019), DOI: 10.1061/(ASCE)ME.1943-5479.0000698
11. Chen, C.C., Nakayama, M., Shou, Y., Charoen, D. (2018), Increasing project success in China from the perspectives of project risk, methodology, tool use, and organizational support. ***International Journal of Information Technology Project Management***, Vol. 9, No. 1, 40-58. DOI: 10.4018/IJITPM.2018010103
12. da Rocha, J.E.N., Pacheco, M.A.C. "Energy efficiency risk analysis and policy in Brazil", ***Energy Efficiency***, Vol. 12, No. 8, (2019), 2227-2239. DOI: 10.1007/s12053-019-09815-w
13. Baptistucci, C.B., Pech, G., Carvalho, M.M. "Experts' engagement in risk analysis: A model merging analytic hierarchy", ***Journal of Modern Project Management***, Vol. 6, No. 1, (2018), 6-17. DOI: 10.19255/JMPM01601
14. Shishodia, A., Dixit, V., Verma, P. "Project risk analysis based on project characteristics", ***Benchmarking***, Vol. 25, No. 3, (2018), 893-918. DOI: 10.1108/BIJ-06-2017-0151
15. Panova, Y., Hilletofth, P. "Managing supply chain risks and delays in construction project", ***Industrial Management and Data Systems***, Vol. 118, No. 7, (2018), 1413-1431. DOI: 10.1108/IMDS-09-2017-0422
16. Frolova, V., Dolina, O., Shpilkina, T. "Investment risk management at mining enterprises", ***E3S Web of Conferences***, Vol. 105, (2019), DOI: 10.1051/e3sconf/201910501054
17. Korshunov, G.I., Rudakov, M.L., Kabanov, E.I. "The use of a risk-based approach in safety issues of coal mines", ***Journal of Environmental Management and Tourism***, Vol. 9, No. 1, (2018), 181-186. DOI: 10.14505/jemt.v9.1(25).23
18. Semykina, I.Y., Skrebneva, E.V. "Problems and solutions of reliability issues for external power supply in the coal mines", ***Journal of Mining Institute***, Vol. 226, (2017), 452-455. DOI: 10.25515/pmi.2017.4.452
19. Gasparian, M.S., Kiseleva, I.A., Korneev, D.G., Lebedev, S.A., Lebedev, V.A. "Strategic analysis of risks when implementing investment projects", ***Espacios***, Vol. 39, No. 27, (2018).
20. Dmitrieva, D., Ilinova, A. "Application of strategic analysis methods and tools in Russian mining and chemical complex", ***International Journal of Applied Engineering Research***, Vol. 11, No. 8, (2016), 5567-5572.
21. Nedosekin, A.O., Rejshahrit, E.I., Kozlovskiy, A.N. "Strategic approach to assessing economic sustainability objects of mineral resources sector of Russia", ***Journal of Mining Institute***, Vol. 237, (2019), 354-360. DOI: 10.31897/PMI.2019.3.354
22. Tcvetkov Pavel, Cherepovitsyn, A.E., Makhovikov Aleksei. "Economic assessment of heat and power generation from small-scale liquefied natural gas in Russia", ***Energy Reports***, Vol. 6, (2020), 391-402. DOI: 10.1016/j.egyr.2019.11.093.
23. Ptashkina-Girina, O.S., Nizamytdinova, N.S., Guseva, O.A. "Technical-Economic Assessment of Small Hydro-Power Units" Proceedings - 2018 International Ural Conference on Green Energy, UralCon, article number 8544277, (2018), 101-106. DOI: 10.1109/URALCON.2018.8544277
24. Bozhkov, M.I., Kostin, V.N. "Technocenological approach to managing power supplying substation loads", ***International Journal of Applied Engineering Research***, Vol. 11, No. 9, (2016), 6736-6739
25. Carranza, J.R.Z., Kovshov, S., Lyubin, E. "Assessment of anthropogenic factor of accident risk on the main oil pipeline pascuales-cuenca in Ecuador", ***Journal of Applied Engineering Science***, Vol. 16, No. 3, (2018), 307-312. DOI: 10.5937/jaes16-17019
26. Nikulina, A.Y., Kruk, M.N. "Organizational and economic mechanism of oil and gas projects in the Russian arctic shelf", ***Journal of Internet Banking and Commerce***, Vol. 21 (Special Issue 6), (2016).
27. Ponomarenko, T.V., Cherepovitsyn, A.E., Fedoseev, S.V., Sidorov, D.E. "Organizational-economic mechanism of financing strategic investment projects at the regional level in regions with poor infrastructure", ***International Journal of Applied Engineering Research***, Vol. 11, No. 16, (2016), 9007-9013.
28. Kremcheeva, D., Kremcheev, E. "Use of a quality management system at the iron and steel enterprise", ***Journal of Mechanical Engineering Research and Developments***, Vol. 41, No. 1, (2018), 151-155. DOI: 10.7508/jmerd.2018.01.018
29. Cherepovitsyn, A.E., Ilinova, A.A. "Methods and tools of scenario planning in areas of natural resources management", ***European Research Studies Journal***, Vol. 21, No. 1, (2018), 434-446.

30. Zamyatin, E.O., Shklyarskiy, Y.E., Yakovleva, E.V. "Concept for electric power quality indicators evaluation and monitoring stationary intellectual system development", ***International Journal of Applied Engineering Research***, Vol. 11, No. 6, (2016), 4270-4274.
31. Kryzia, D., Kopacz, M., & Kryzia, K. "The Valuation of the Operational Flexibility of the Energy Investment Project Based on a Gas-Fired Power Plant" ***Energies***, Vol. 13, No. 7, (2020), 1567.
32. Platon, Victor & Constantinescu, Andreea. "Monte Carlo Method in Risk Analysis for Investment Projects", ***Procedia Economics and Finance***, Vol. 15, (2014), 393-400. 10.1016/S2212-5671(14)00463-8.

Persian Abstract

چکیده

در این مقاله پیشنهاد شده است که مسئله ارزیابی جامع ریسک های پروژه در صنعت انرژی مورد استفاده قرار گیرد. نویسندگان تحقیق بر توصیف راه حل کاربردی متمرکز شده اند. یک پروژه سرمایه گذاری واقعی برای جایگزینی دیگ بخار پوست در شرکت Mondi Syktyvkar به عنوان یک هدف برای آزمایش نتایج انتخاب شد. ما پیشنهاد کردیم ریسک های همراه پروژه را به ۲ دسته تقسیم کنیم: ریسک هایی که اطلاعات لازم و آماری برای ارزیابی کمی آنها وجود دارد و خطراتی که این اطلاعات برای آنها وجود ندارد. به عنوان یک روش ارزیابی کمی خطرات از گروه اول ، استفاده از روشی برای مدل سازی شبیه سازی مونت کارلو به مصلحت است. در این مورد ، نویسندگان مقاله تجزیه و تحلیل قابل توجهی از روش های موجود برای ارزیابی خطرات پروژه انجام داده اند و انتخاب روش مونت کارلو به دلیل جهت گیری عملی مطالعه است. در عمل ، شرکت واقعی برای استفاده از روش های پیچیده تر ارزیابی ، مانند روش های گزینه های واقعی یا روش های منطق فازی ، شبکه های عصبی و غیره ، کاملاً مشکل ساز است. به عنوان یک روش ارزیابی ریسک کیفی (از دسته دوم) ، روش ارزیابی کارشناس با محاسبه بعدی حق بیمه در نرخ تخفیف انتخاب شد. این روش در عمل معمول است و اجرای آن به اندازه کافی آسان است. با توجه به نتایج تجزیه و تحلیل (آماری) خطرناکترین خطرهای پروژه های انرژی مشخص شد: خطرات تولید و فن آوری (خطر انتخاب طرح اشتباه فن آوری ، خطر کاهش کیفیت کنترل های داخلی ، خطر نادرست محاسبه ظرفیت طراحی تولید انرژی ، خطر ایمنی صنعتی) ، خطرات امنیتی (خطر حملات هک به سیستم های اطلاعاتی شرکت های انرژی) و همچنین خطرات کشور. از جمله مهمترین خطرات (بر اساس تجزیه و تحلیل تأثیر آنها بر شاخص های اصلی فنی و اقتصادی پروژه) عبارتند از: خطر افزایش قیمت برای گاز خریداری شده (سوخت) ، خطر نوسان زیاد نرخ ارز دلار. نتایج این مطالعه در یک پروژه واقعی مورد استفاده قرار گرفت و روش ارزیابی ریسک در فعالیت های پروژه شرکت Mondi Syktyvkar اجرا شد.

IJE TRANSACTIONS A: Basics Vol. 34, No. 7, (July 2021) 1785-1798

International Journal of Engineering

Journal Homepage: www.ije.ir

Multi-objective Optimal Design of Gas-fired Heater Based on Modified Design Model of Fired Heater Taking into Account Exergy, Economic and Environmental Factors

S. M. Ebrahimi Saryazdi*[a,b], F. Rezaei[a], Y. Saboohi[a], F. Sassani[b]

[a] *Energy Engineering Department, Sharif University of Technology, Azadi Ave., Tehran, Iran*
[b] *Mechanical Engineering Department, University of British Columbia, Vancouver, BC, Canada*

PAPER INFO

Paper history:
Received 15 February 2021
Received in revised form 31March 2021
Accepted 01 April 2021

Keywords:
Mathematical Model
Gas Fired Heater
Multi-objective Optimization
Exergy Analysis
Environment
Gas Reduction Station

ABSTRACT

Heaters are one of the central parts of natural gas reduction stations using turboexpanders to prevent the formation of hydrate and corrosion failure. This study intends to design a fired heater by applying a combustion sub-model to derive an optimal model for this kind of application. This model is developed to accurately consider all subsections of the fired heater namely radiation, convection, and shield sections, as well as flue gas composition, and its volume. Within this context, a multi-objective optimization is employed to identify the optimal design of the gas-fired heater in the natural gas reduction station for the Ramin power plant case study. The total economic and environmental costs, together with modified exergy efficiency, are selected as objective functions. Multi-criteria-decision-making-method is employed on Pareto frontiers optimal curve to suggest the optimal solution. Results show that the developed model can outperform previous models in thermal efficiency with relatively similar costs. Besides, the optimal point in Pareto suggested by the decision-making-method accounts for a higher modified exergy efficiency (1.3%) than the counterpart, which thermal efficiency is regarded as an objective function. At the same time, its total cost remained almost constant. The effects of changes in each of the design parameters on the objective functions are also evaluated.

***doi**: 10.5829/ije.2021.34.07a.23*

NOMENCLATURE

a_1, a_2	Volume of air components	*TAD*	Adiabatic temperature of combustion [K]
A	Heat transfer area [m^2]	*TRE*	Real temperature of combustion [K]
AAS	Air to fuel ratio [–]	U_c	Overall heat transfer coefficient of convection section [w/m^2k]
$Air_{\min}$	Minimum air volume for fuel combustion	v_w	Wind speed [m/s]
B	Constant parameter	*V*	Volume of flue gases [m^3]
c_{cost}	Unitary cost [$/$m^2$]	x_k	Molar fraction of component k in the flue gas [%]
c_{ele}	Electricity price [$/kWh]	*W*	Width of fired heater [m]
c_{CO_2}	Penalty cost of CO_2 emission [$/kg]	**Abbreviations**	
c_{SO_2}	Penalty cost of SO_2 emission [$/kg]	CRF	Capital Recovery Factor
C	Carbon mass percentage in fuel [ppm], Cost [$], Constant parameter	GA	Genetic Algorithm
CO	CO mass percentage in flue gas [ppm]	HTRI	Heat Transfer Research Inc.

* Corresponding Author Institutional Email: ebrahimi@energy.sharif.edu (S. M. Ebrahimi Saryazdi)

Please cite this article as: S. M. Ebrahimi Saryazdi, F. Rezaei, Y. Saboohi, F. Sassani, Multi-objective Optimal Design of Gas-fired Heater Based on Modified Design Model of Fired Heater Taking into Account Exergy, Economic and Environmental Factors, International Journal of Engineering, Transactions A: Basics Vol. 34, No. 7, (2021) 1785-1798

CO_2	CO_2 mass percentage in flue gas [ppm]
$C_{p,dg}$	Heat capacity of fuel [J/kg]
d	Distance $[m]$, Diameter $[m]$
d_{i+}	Distance of point ith from the ideal point
D	Constant parameter
e	Standard molar exergy [J/mole]
ex	Specific exergy [J/kg]
Ex	Exergy [W]
EC	Equilibrium constant of reaction
f	Fanning friction factor [-]
$f_1, f_2, f_3, f_4, f_5, f_6, f_7, f_8, f_9, f_{10}, f_{11}, f_{12}, f_{13}$	Volume of fuel components
F_{ij}	Points of the Pareto frontier
F_{ij}^{n}	Non-dimensionalized objective
F_j^{ideal}	Ideal value for jth objective
G	Mass velocity of hot water
GCV	Gross Calorific Value of fuel [J/kg]
h	Annual operating time [h/y], Specific enthalpy [kJ/kg]
HF	Humidity factor
H_2	Hydrogen mass percentage in fuel [%]
H	Height of fired heater [m]
i	Annual discount rate [%]
k_1	Numerical constant
k_2	Numerical constant
kk_1	Distance between the centers of the end tube and the wall [–]
$L_1, L_2, L_3, L_4, L_5, L_6$	Losses parameters [W]
L	Length [m]
LHV	Lower heating value of fuel [J/kg]
$LMTD$	Logarithmic mean temperature difference [K]
m	Mass flow rate [kg/s]
M_{dg}	Mass of flue gases in kg/kg fuel
M_{moist}	Moisture mass percentage [%]
n	Number of years, Number of point on Pareto frontier, Mole of flue gas [mole]
N_t	Number of tubes
o_{cost}	Fuel unit cost
Q_A	Heat absorbed by the process fluid [W]
Q_{conv}	Heat absorbed in convection section [W]
Q_f	Heat liberated by combustion [W]
Q_{rad}	Heat absorbed in radiation section [W]
Q_{shld}	Heat absorbed in shield section [W]
Q_L	Heat duty [W]

LINMAP	Linear Programming Technique for Multidimensional Analysis of Preference
MINLP	Mixed-Integer Nonlinear Programming
MCDM	Multi-Criteria Decision-Making
REFPROP	Reference Fluid Thermodynamic and Transport Properties Database

Greek Letters

α	Relative effectiveness factor of the tubes bank [-]
η_{exe}	Modified exergy efficiency [-]
η_{comb}	Combustion efficiency [-]
η_{FH}	Fired heater thermal efficiency [-]
η_p	Pump efficiency [-]
v	Mean specific volume of hot water [m^3/kg]
ξ	Constant value
ρ	Density [kg/m^3]
σ	Stefan-Boltzman constant $= 2.041\times10^{-7}\left[kJ/m^2\cdot K^4\cdot h\right]$

Subscripts

air	Air
cc	Distance between the centers of the two tubes side by side
$conv$	Convection section
cp	Plane area of tubes bank
e	Equivalent length
env	Environmental emission
$firebox$	Firebox section
$flue$	Flue gases
$fuel$	Fuel
F	Feed
FH	Fired heater
g	Effective gas temperature in firebox
i	Inlet, Inside, ith point
k	Component k of flue gases
ng	Natural gas
o	Outlet, Outside
opr	Operation
P	Product
rad	Radiation section
$radc$	Ceiling tubes in radiation section
s	Stack
$shld$	Shield section
$total$	Total
w	Water
W	Wall

r_{cost}	Unitary cost [\$/m^2]	0	Ambient state
$\bar{R}$	Universal gas constant, 8.314 [J/kg·K]	**Superscripts**	
s	Specific entropy [J/kg·K]	ch	Chemical
S_i	Heat surface area of each tube in the convection section[m^2]	ph	Physical
T	Temperature [K]	UN	Unavoidable

1. INTRODUCTION

Over 69% of natural gas is transported from wells to consumers through transmission pipelines [1], known as an economical way to transport natural gas across large distances [2]. Besides that, Iran has 11th place globally in the total length of natural gas pipelines [1]. Natural gas pressure should be increased in gas compression stations to overcome friction and heat losses through the natural gas pipeline, and then decreased in gas reduction stations to the desired values [3]. Iran has 2500 gas reduction stations [3], which traditionally consist of a gas-fired heater, throttling valve, heat exchanger, and control system. The pressure reduction in the gas reduction station results in a drop in temperature because of the positive Joule-Thompson coefficient [4] and consequent formation of hydrates. Thus, the preheating system regulates the outlet temperature of natural gas from the turboexpander or throttling valve to provide appropriate means of preventing the formation of hydrate and corrosion failure [5].

As mentioned earlier, gas-fired heaters are one of the central parts of natural gas reduction stations to prevent the formation of hydrate and corrosion failure, especially when a turbo-expander is implemented instead of a pressure reduction valve. Gas-fired heaters [6] or bath-type heaters [7] are commonly employed in natural gas reduction stations. These two types of preheaters burn a remarkable portion of the passing natural gas and have low energy efficiency [7, 8].

With regard to gas reduction station preheating system, the focus of recent research has been on replacing the conventional preheating system with novel preheating systems such as solar systems [3, 9], geothermal systems [10], combined heat and power systems (with gas turbine [11], internal combustion engines [12, 13], and Molten-carbonate fuel cells [14] as a prime mover), and thermoelectric generator [15].

Darabi et al. [14] indicated that employing a hybrid turboexpander-fuel cell in gas pressure reduction stations is presently uneconomical. Farzaneh-Gord et al. [16] analyzed the impact of a number of solar collectors with a storage tank on gas reduction stations. According to their economic assessment, the payback ratio was 11 years. Ghezelbash et al. [10] studied the use of vertical ground-coupled heat pumps in gas reduction stations equipped with turboexpander. In this configuration, the ground heat pump preheats the natural gas to reduce fuel consumption of bath-type heaters up to 45.8% annually with a payback period of 6 years.

To the authors' best of knowledge, in recent years, there has been growing interest in the economic and exergy assessment of novel preheating configurations replacing fired heaters with renewable heating systems and novel-heating technologies. Results of previous studies indicated that renewable heating systems and novel-heating technologies are not cost-effective solutions. However, previous studies failed to address the optimal detailed design of the fired heater. Therefore, this paper will focus on the optimal design of fired heaters in gas reduction stations.

A fired heater or tubular heater is a combustion furnace where combustion gases heat the fluids inside the tube. The main advantages of this equipment are having continuous performance and a reduction in foam formation. The oil, gas, and chemical process industries have widely applied this equipment for the heating purposes [17]. Fired heaters consist of three sections, namely radiation (which is also called combustion chamber), convection, and shield section, see Figure 1. The hot combustion gases produced in the radiation section pass through the shield and then convection sections. The shield section, which consists of two/three rows of bare tubes, is located between the other two parts and prevents direct radiation to the convection tubes. The flue gases circulate through a staggered convection tube bundle to increase the heat transfer rate before discharging to the atmosphere. Corbels fill the space between the convection tube and sidewall to prevent flue gas bypass flow [18]. Regarding the geometrical configuration of the radiation section, this equipment is classified as vertical cylindrical or box-type heaters. The last one, which is considered in this study, is typically applied in enormous heat demands. It should be mentioned that tube size and the number of passes of radiation, shield, and convection sections are determined based on the application and fluid flow rate [19].

Various mathematical models have been proposed to simulate the performance of fired heaters. In this regard, Ebrahimi et al. [20] developed the multi-zone mathematical model to simulate the performance of the furnace. Previous studies indicated that solving the matrices in calculating the total heat exchange area limited the application of this model [21]. Ibrahim and Al-Qassimi [22] developed a model to simulate a box-

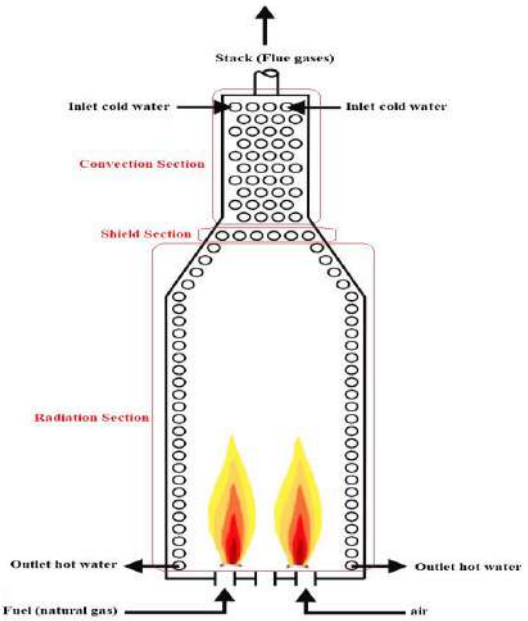


Figure 1. Schematic of fired heater sections

type fired heater in a crude oil atmospheric topping unit. In this study, temperature profile and heat abortion per layer in the convection section were calculated and used to assess this section. Ibrahim and Al-Qassimi [17] studied direct and indirect methods to calculate heater efficiency and indicated that the direct approach has less computational time. Recently, findings regarding fired heaters have led to improve the process operations; however, little attention has been paid to optimize their design. Limitations such as the maximum allowable absorption rate in the radiation section, the maximum permissible pressure drop of heater fluid flow, and so on considerably affect optimal geometry and combustion conditions of the fired heater. Concerning the optimal design of fired heaters, Heat Transfer Research Inc. (HTRI) and Aspen Technology Inc. proposed comprehensive optimal design producers [23]. Moreover, there are century-old standards, such as the standards of the American Petroleum Institute (API) standard 530 [24] and 560 [25], which describe design calculation procedures of fired heaters. Mussati et al. [26] proposed a mathematical model to obtain the optimal design of the fired heater using Mixed-Integer Nonlinear Programming (MINLP). In this study, sensitivity analysis of fuel cost and capital investment cost is carried out. Several case studies were considered, and their optimal design results indicate improvements in economic and operative performances. Haratian et al. [27] optimized box-type fired heater by developing a mathematical model associated with the genetic algorithm (GA). They applied the modified total cost as an objective function by including the pumping cost. Their results have an acceptable accuracy with literature and showed that total cost could decrease up to nearly 2.5% compared to the original design.

Previous studies have not considered the combustion process in their mathematical model of the fired heater, and have assumed the combustion efficiency to be constant [26, 27]. Moreover, these studies have employed a direct method in their design model to calculate thermal efficiency, which depends on stack temperature and excess air fraction. The accuracy of mentioned methods for predicting thermal efficiency subsides with increased stack temperature and excess air fraction [28]. Therefore, the novelty of this work is to propose a new design model of fired heaters associated with the combustion sub-model. An indirect method (heat loss method) integrated with the combustion model is also employed to predict the thermal efficiency of the fired heater through an iterative design process. The optimal design of fired heaters, especially for oil industry applications using total economic cost as an objective function, has attracted much attention in recent years. However, the optimal design of the fired heater of the gas reduction station has not been investigated. With this in mind, the gas reduction station of the Ramin power plant in Iran was selected as a case study. Previous studies also indicated that exergy analysis could be a powerful tool for designing and optimizing processes [29-32]. Therefore, multi-objective optimization of fired heaters applying total economic and environmental cost along with modified exergy efficiency as objective functions are proposed in this study. The environmental cost is integrated with economic costs of fired heater to include emissions of CO_2 and SO_2 pollutants in the fired heater design process. For decision variables, the outlet hot water temperature, the external diameter of tube in radiation as well as convection sections, number of wall-side tubes as well as the ceiling in radiation section, and excess air ratio are considered for the first time in optimal design progress of fired heater due to natural-gas reduction station conditions.

As the initial step, a mathematical model of the box-type fired heater and the combustion process sub-model are developed (section 2.1.1). In this study, the combustion sub-model is applied to obtain the composition and volume of flue gases (section 2.1.2). Then, the fired heater's optimal design is derived using a multi-objective GA (section 2.2). The multi-criteria-decision-making method (MCDM) is employed to obtain an optimal point of the Pareto-frontier and facilitate decision-making, which is generally dependent upon engineering experiences and objective function. Eventually, the variations of objective functions with decision variables are presented in section 3. Given the above, the main contributions of this study are:

- Employing a multi-objective optimization for the optimal design of fired heater in pre-heating natural gas entering the gas-pressure reduction station. Total cost and modified exergy efficiency are both considered objective functions. The environmental cost is also added to the total cost of the fired heater.
- Proposing a modified design model of the fired heater and its diverse sections, including radiation, convection, and shield, etc. Moreover, the geometric structure, dimensions, the number of required tubes, thermal efficiency, and losses are estimated by applying the indirect method and using the combustion sub-model.

2. MODELLING

The flow chart for the optimal design of the fired heater is shown in Figure 2. In the present project, the genetic algorithm (GA) is used to optimize fired heater design. In the first step, GA parameters (Table 1), decision variables (Table 2), and objective functions are determined (which are presented in section 2.2). Then, a mathematical model of the fired heater and sub-model of combustion, which are developed in MATLAB environment, are employed to design all subsections of the fired heater (radiant, shield, convection, and chamber) and to achieve performance parameters of fired heater. The combustion sub-model is provided in section 2.1.2. Inputs of combustion model include fuel and air composition, excess air, fuel and air temperature, air humidity, and air pressure. Combustion sub-model is employed to calculate thermal efficiency (section 2.1.1) and environmental cost (section 2.2.1). Reference Fluid Thermodynamic and Transport Properties Database (REFPROP) database is applied to achieve thermodynamic properties of fluids in these sub-models [33]. A new generation is produced by employing selection, mutation, and crossover operators to evaluate each generation's objective functions. Eventually, the MCDM method (Linear Programming Technique for Multidimensional Analysis of Preference (LINMAP)) is applied to suggest an optimal Pareto Front solution [34]. The formulation of the LINMAP method is presented in Appendix A.

2. 1. Mathematical Model

2. 1. 1. Fired Heater Sub-model The total heat transferred from the radiant and convection sections is calculated from Equation (1).

$$Q_A = m_w \cdot \left(h_{w,o} - h_{w,i}\right) = Q_{rad} + Q_{conv} + Q_{shld} \quad (1)$$

where, Q_{rad}, Q_{conv}, and Q_{shld} refer to the heat transfer rate absorbed in the radiation, convection, and shield sections, respectively.

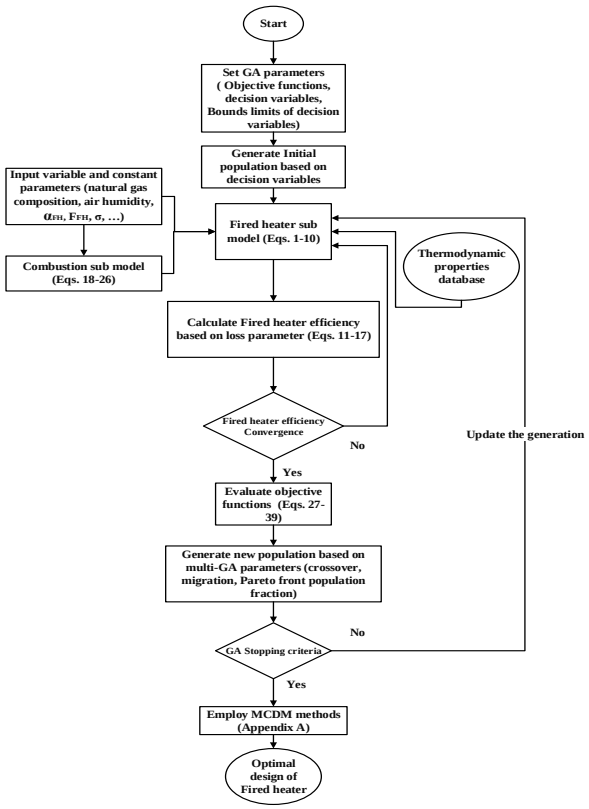


Figure 2. Flow chart of fired heater design and performance model

The amount of heat transferred in the radiation section is the function of the plane area of tubes bank (A_{cp}), effective gas temperature in firebox (T_g), and wall temperature (T_W) [17]:

$$Q_{rad} = \sigma \cdot \left(\alpha \cdot A_{cp}\right) \cdot F \cdot \left(T_g^4 - T_W^4\right) \quad (2)$$

$$A_{cp} = N_{t,rad,total} \cdot d_{cc} \cdot L_{rad} \quad (3)$$

The amount of heat transferred in the shield section is presented in Equation (4) [17]:

$$Q_{shld} = \sigma \cdot \left(\alpha \cdot A_{cp,shld}\right) \cdot F \cdot \left(T_g^4 - T_W^4\right) \quad (4)$$

$$A_{cp,shld} = N_{t,shld} \cdot d_{cc} \cdot L_{rad} \quad (5)$$

Then, the heat balance equation is derives based on Equations (1), (2), and (4). The Newton-Raphson method was used to solve this equation and determine the effective gas temperature (T_g).

$$F\left(T_g\right) = C \cdot T_g^4 + D \cdot T_g - B \quad (6)$$

The constants *C*, *D*, and *B* are determined by fuel type, percentage of excess air, operating conditions, and the fired heater's geometric characteristics. The number of tubes required in the convection section is obtained by Equation (7) [17]:

$$N_{t,conv} = \left(\frac{Q_{conv}}{U_c \cdot LMTD \cdot S_i} \right) \tag{7}$$

The geometric dimension of the heater is defined as Equations (8) and (9):

$$H = \left(\frac{N_{t,rad}}{2} - 1 \right) \cdot d_{cc} + kk_1 \tag{8}$$

$$W \geq \left(\frac{N_{t,radc}}{2} - 1 \right) \cdot d_{cc} + kk_1 \tag{9}$$

The net released heat required in the combustion section is obtained by the heater efficiency (Equation (10)):

$$\eta_{FH} = Q_A / Q_L \tag{10}$$

The difference between the amount of heat absorbed and released by combustion is equal to the sum of the heat losses (Equation (11)):

$$Q_L = Q_A + \sum_{i=1}^{6} L_i \tag{11}$$

where *L* refers to different types of losses that are represented from Equations (12)-(17) [22]:

$$L_1 = 100 \cdot M_{dg} \cdot C_{p,dg} \cdot (T_s - T_{air}) \tag{12}$$

$$L_2 = 900 \cdot H_2 \cdot \left(2445.21 + \left(1.88 \cdot (T_s - T_{air}) \right) \right) \tag{13}$$

$$L_3 = 188 \cdot HF \cdot AAS \cdot (T_s - T_{air}) \tag{14}$$

$$L_4 = 100 \cdot M_{moist,fuel} \cdot \left(2445.21 + \left(1.88 \cdot (T_s - T_{air}) \right) \right) \tag{15}$$

$$L_5 = 100 \cdot \left(\frac{CO}{CO + CO_2} \cdot 24656000 \cdot C \right) \tag{16}$$

$$L_6 = 54.8 \cdot \left(\left(T_W / 55.55 \right)^4 - \left(T_{air} / 55.55 \right)^4 \right) + \left(1.957 \cdot \left(\frac{T_W - T_{air}}{55.55} \right)^{1.25} \cdot \sqrt{\frac{(196.8 \cdot \upsilon_w) + 68.9}{68.9}} \right) \tag{17}$$

2. 1. 2. Combustion Sub-model According to the formulation of the indirect method (Equation (10)-(17)), it is indicated that composition and volume of flue gases were used to calculate the thermal efficiency of the fired heater. Therefore, a combustion sub-model is applied to obtain these parameters. In the combustion model, the model of complete and incomplete combustion of natural gas is proposed. Complete and incomplete combustion of natural gas can be written as Equations (18) and (19), respectively [35, 36].

$$\begin{aligned} & f_1 \cdot N_{2,fuel} + f_2 \cdot O_{2,fuel} + f_3 \cdot CO_{2,fuel} + \\ & f_4 \cdot H_2O_{fuel} + f_5 \cdot SO_{2,fuel} + f_6 \cdot CO_{fuel} + \\ & f_7 \cdot H_{2,fuel} + f_8 \cdot CH_{4,fuel} + f_9 \cdot C_2H_{4,fuel} + \\ & f_{10} \cdot C_2H_{6,fuel} + f_{11} \cdot C_3H_{8,fuel} + f_{12} \cdot C_4H_{10,fuel} + \\ & f_{13} \cdot H_2S_{fuel} + a_1 \cdot N_{2,air} + a_2 \cdot O_{2,air} \rightarrow \\ & 0.01 \cdot \left(f_1 + \left(a_1 \cdot AAS \right) \right) N_2 + 0.01 \cdot \left(a_2 \cdot \left(AAS - Air_{\min} \right) \right) O_2 \\ & + 0.01 \cdot \left(f_3 + f_6 + f_8 + 2 \cdot f_9 + 2 \cdot f_{10} + 3 \cdot f_{11} + 4 \cdot f_{12} \right) CO_2 \\ & + 0.01 \cdot \left(f_4 + f_7 + 2 \cdot f_8 + 2 \cdot f_9 + 3 \cdot f_{10} + 4 \cdot f_{11} + 5 \cdot f_{12} + f_{13} \right) H_2O \\ & + 0.01 \cdot \left(f_5 + f_{13} \right) H_2S \end{aligned} \tag{18}$$

$$\begin{aligned} & f_1 \cdot N_{2,fuel} + f_2 \cdot O_{2,fuel} + f_3 \cdot CO_{2,fuel} + f_4 \cdot H_2O_{fuel} + \\ & f_5 \cdot SO_{2,fuel} + f_6 \cdot CO_{fuel} + f_7 \cdot H_{2,fuel} + f_8 \cdot CH_{4,fuel} + \\ & f_9 \cdot C_2H_{4,fuel} + f_{10} \cdot C_2H_{6,fuel} + f_{11} \cdot C_3H_{8,fuel} + \\ & f_{12} \cdot C_4H_{10,fuel} + f_{13} \cdot H_2S_{fuel} + a_1 \cdot N_{2,air} + a_2 \cdot O_{2,air} \rightarrow \\ & 0.01 \cdot \left(f_1 + \left(a_1 \cdot AAS \right) \right) N_2 + 0.01 \cdot \left(f_5 + f_{13} \right) SO_2 + \\ & \left(\left(-AB \pm \sqrt{AB^2 - 4 \cdot AA \cdot AC} \right) / 2 \cdot AA \right) CO + \\ & \left(2 \cdot \left(a_2 \cdot \left(Air_{\min} - AAS \right) / 100 \right) \right) H_2 - \\ & \left(\left(\left(-AB \pm \sqrt{AB^2 - 4 \cdot AA \cdot AC} \right) / 2 \cdot AA \right) \right) H_2 + \\ & \left(\left(0.01 \cdot \left(f_3 + f_6 + f_8 + 2 \cdot f_9 + 2 \cdot f_{10} + 3 \cdot f_{11} + 4 \cdot f_{12} \right) \right) \right) CO_2 - \\ & \left(\left(\left(-AB \pm \sqrt{AB^2 - 4 \cdot AA \cdot AC} \right) / 2 \cdot AA \right) \right) CO_2 + \\ & \left(\left(0.01 \cdot \left(f_4 + f_7 + 2 \cdot f_8 + 2 \cdot f_9 + 3 \cdot f_{10} + 4 \cdot f_{11} + 5 \cdot f_{12} + f_{13} \right) \right) \right) H_2O - \\ & \left(\left(2 \cdot \left(a_2 \cdot \left(Air_{\min} - AAS \right) / 100 \right) \right) \right) H_2O + \\ & \left(\left(\left(-AB \pm \sqrt{AB^2 - 4 \cdot AA \cdot AC} \right) / 2 \cdot AA \right) \right) H_2O \end{aligned} \tag{19}$$

$$\begin{aligned} AB = & \left(0.01 \cdot \left(f_3 + f_6 + f_8 + 2 \cdot f_9 \right) \right) + \\ & \left(0.01 \cdot \left(2 \cdot f_{10} + 3 \cdot f_{11} + 4 \cdot f_{12} \right) \right) - \\ & \left(2 \cdot \left(a_2 \cdot \left(Air_{\min} - AAS \right) / 100 \right) \right) - \\ & \left(EC_{@TRE} \cdot \left(0.01 \cdot \left(f_4 + f_7 + 2 \cdot f_8 + 2 \cdot f_9 \right) \right) \right) - \\ & \left(EC_{@TRE} \cdot \left(0.01 \cdot \left(3 \cdot f_{10} + 4 \cdot f_{11} + 5 \cdot f_{12} + f_{13} \right) \right) \right) \\ & - \left(EC_{@TRE} \cdot \left(2 \cdot \left(a_2 \cdot \left(Air_{\min} - AAS \right) / 100 \right) \right) \right) \end{aligned} \tag{20}$$

$$AA = (1 - EC_{@TRE}) \tag{21}$$

$$AC = \left(0.01\cdot\left(f_3 + f_6 + f_8 + 2\cdot f_9\right)\right) \cdot\left(0.01\cdot\left(2\cdot f_{10} + 3\cdot f_{11} + 4\cdot f_{12}\right)\right) \cdot\left(2\cdot\left(a_2\cdot\left(Air_{\min} - AAS\right)/100\right)\right) \quad (22)$$

$$Air_{\min} = 0.01\cdot\left(0.5\cdot f_6 + 0.5\cdot f_7 + 2\cdot f_8\right) \cdot\left(3\cdot f_9 + 3.5\cdot f_{10} + 5\cdot f_{11} + 6.5\cdot f_{12} + 1.5\cdot f_{13} - f_2\right) \quad (23)$$

Moreover, EC is an equilibrium constant of following reaction and its value [35, 36].

$CO + H_2O \rightarrow CO_2 + H_2$

If surplus air is greater than or equals unity, the combustion will be complete; otherwise, combustion will be incomplete (Equation (18)).

$$Surplus = AAS\ /\ Air_{\min} \quad (24)$$

Adiabatic temperature of combustion and the actual temperature of combustion are computed as follows:

$$TAD = \frac{GCV + \left(C_{P,fuel}\cdot T_{fuel}\right) + \left(AAS\cdot C_{P,air}\cdot T_{air}\right)}{V_{fluegas}\cdot C_{P,fluegas}} \quad (25)$$

$$TRE = TAD\cdot\eta_{comb}\ /100 \quad (26)$$

2. 2. Optimization Regarding Figure 2, the GA is used to optimize the fired heater design in the present project. In this study, selected GA optimization parameters are shown in Table 1. Optimization is done with eight decision variables for the selected case study presented in Table 2. These decision variables and their range are determined based on previous research and sensitivity analysis results (section 3).

TABLE 1. Optimization conditions

Parameter	Value
crossover fraction	0.8
migration fraction	0.2
Pareto front population fraction	0.35
population size	200

TABLE 2. Decision variables for optimum design

Variables	From	To
$T_{w,o}$ [℃]	180	210
$d_{o,rad}$ [m]	0.04	0.2
$d_{o,conv}$ [m]	0.04	0.2
Nt_{rad} [-]	40	80
Nt_{radc} [-]	10	30
Excess air [-]	0.18	0.27

2. 2. 1. Objective Function As noted at the beginning of this section, the total economic and environmental cost is regarded as one of the objective functions and can be calculated as Equation (27):

$$C_{FH} = C_{rad} + C_{conv} + C_{firebox} + C_{opr} + C_{env} \quad (27)$$

where C_{rad}, C_{conv}, $C_{firebox}$, C_{opr}, and C_{env} correspond to the cost of radiation section, convection section, firebox, operation, and environmental emission, respectively.

$$C_{rad} = r_{cost}\cdot A_{cp}\cdot CRF \quad (28)$$

$$C_{conv} = c_{cost}\cdot A_{conv}\cdot CRF \quad (29)$$

where c_{cost} refers to unitary cost.

$$C_{firebox} = \left[k_1 + k_2\left(A_{conv} + A_{cp}\right)\right]\cdot CRF \quad (30)$$

where k_1 and k_2 correspond to constants required for computing costs.

$$C_{opr} = \left(o_{\text{cost}}\cdot m_{fuel}\right)\cdot h + \left(\frac{c_{ele}}{\eta_p}\left(\frac{m_w}{\rho_w}\cdot\frac{0.00517\cdot f\cdot v\cdot G^2\cdot L_e}{d_i}\right)\right)\cdot h \quad (31)$$

where c_{ele}, o_{cost}, and h are considered to be €0.01/kWh [37], €0.01/m^3 [38], and 7000 h, respectively. Finally, the cost of emission is calculated as follows:

$$C_{env} = \left[\left(m_{CO_2}\cdot c_{CO_2}\right) + \left(m_{SO_2}\cdot c_{SO_2}\right)\right]\cdot h \quad (32)$$

The values of m_{CO_2} and m_{SO_2} are obtained using the combustion sub-model (Section 2.1.2). The values of c_{CO_2} and c_{SO_2} are \$0.032/kg, and \$2.227/kg, respectively [39].

Modified exergy efficiency is considered to be another objective function and is obtained as Equation (33):

$$\eta_{exe} = \frac{Ex_P}{Ex_F - Ex_D^{UN}} \quad (33)$$

where Ex_P and Ex_F refer to product exergy (absorbed exergy from water stream) and fuel exergy, respectively and are computed as Equations (34) and (35) [40]:

$$Ex_P = m_w\left[\left(h_{w,o} - h_{w,i}\right) - T_0\left(S_{w,o} - S_{w,i}\right)\right] \quad (34)$$

$$Ex_F = \xi\cdot m_{fuel}\cdot LHV - Ex_{flue} \quad (35)$$

ξ is a constant value to calculate the exergy efficiency of fuel. The value of this parameter for gas fuel is 0.98

[40]. Equation (36) is applied to achieve exergy of flue gas. The combustion sub-model was implemented to obtain the composition and molar fraction of flue gases.

$$Ex_{flue} = m_{flue} \cdot (ex^{ph}_{flue} + ex^{ch}_{flue}) \quad (36)$$

$$ex^{ph}_{flue} = \sum_k (h_k - h_{0,k}) - T_0 (s_k - s_{0k}) \quad k = N_2, O_2, CO_2, CO, SO_2, H_2O \quad (37)$$

$$ex^{ch}_{flue} = n.\left(\sum_k x_k \cdot e^{ch}_k + \bar{R} T_0 \sum_k x_k \ln(x_k) \right) \quad k = N_2, O_2, CO_2, CO, SO_2, H_2O \quad (38)$$

The standard mole chemical exergy of components is presented in Table 3. Unavoidable exergy destruction of the fired heater is calculated as follows:

$$Ex^{UN}_{D,FH} = Ex_{P,FH} \left(\frac{Ex_{D,FH}}{Ex_{P,FH}} \right)^{UN} \quad (39)$$

To determine the value of $\left(Ex_{D,FH} / Ex_{P,FH}\right)^{UN}$, it is assumed that the minimum temperature difference of HE is 5 K, isentropic efficiency of the pump is 0.95, air and fuel temperature are 673 K, and excess air is 0.12.

2. 3. Case study The heater of the Ramin Power Plant gas reduction station in Iran was selected as a case study in the present research work [41]. As already stated earlier, this heater is used to preheat the natural gas before entering the expansion turbine, which is parallel to the pressure regulator whose purpose is power generation. Heat transfer value, required on the coldest day of the year to preheat the natural gas, is considered the heat transfer rate. The hot fluid is water. Figure 3 depicts a configuration of natural gas pressure reduction station applying turboexpander as well as a fired heater. The properties of the natural gas in natural gas reduction station of the Ramin power plant, regarded as input variables in the mathematical model, are presented in Table 4.

TABLE 3. Standard chemical molar exergy of flue gas components [40]

Substance	e^{ch} [J/mol]
N_2	584
O_2	3869
CO_2	19870
H_2O	9490
CO	275100
SO_2	313400

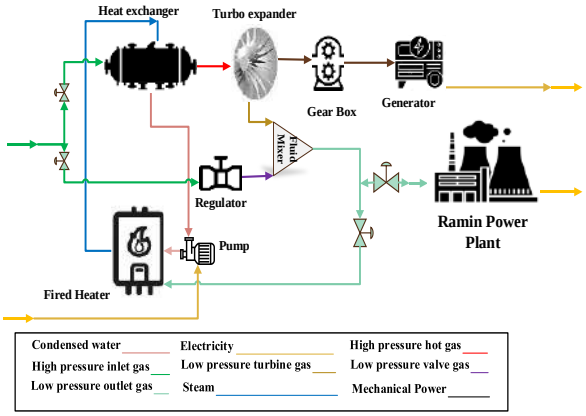


Figure 3. Configuration of power plant natural gas pressure reduction station using turboexpander and hot water GFH

TABLE 4. Dataset for the design of heater [41]

Parameter		Value
m_{ng} [kg/s]		32.83
$T_{ng,i}$ [K]		288.8
$T_{ng,o}$ [K]		388.2
Compositions (volume fraction) [%]	N_2	1.735
	CO_2	3.52
	CH_4	80.43
	C_2H_4	1.5
	C_2H_6	9.02
	C_3H_8	4.54
	C_4H_{10}	0.58
	H_2S	0.09

3. RESULTS AND DISCUSSION

3. 1. Model Validation In this section, the validity and accuracy of the developed model are considered to be examined. The answer reached for the model is compared with the Ibrahim and Al-Qassimi Design [17]. To do so, input variables of design model such as outlet hot water temperature, the external diameter of tube in radiation as well as convection sections, number of the wall-side tube as well as the ceiling in radiation section, excess air ratio, natural gas composition, air humidity ratio, relative effectiveness factor of the tubes bank, and air as well as fuel temperature are collected form this study [17]. As shown in Table 5, there is a plausible difference between the two results, so the proposed model can be successfully used with satisfactory

accuracy. The heat liberated by combustion has the maximum deviation. It results from considering the indirect method of estimating efficiency due to the capability of the combustion model to obtain the composition of flue gases. Design models of fired heaters include thermodynamic and geometrical equations and performance prediction equations (section 2.1.1). These models yield design, thermodynamic and performance parameters of the fired heater through an iterative process (see Figure 2). To evaluate applying indirect method for estimating heater efficiency, the thermal efficiency of mentioned gas-fired heater [17] is calculated and compared with that of experimental methods proposed by previous design models [26, 28], which is presented in Table 6. In these studies [26, 28], the thermal efficiency of the fired heater is calculated using the direct method, which depends on stack temperature and excess air fraction. To do so, the mathematical model of fired heater (section 2.1.1) and direct method equations for calculating thermal efficiency are used to obtain the design and performance parameters of the fired heater. Compared to previous estimation methods [26, 28], the indirect method gives the best prediction with a deviation of 2.22% from experimental values.

TABLE 5. Model validation

Variables	Ref. [17]	This study	Deviation [%]
T_g [K]	1073	1070	**0.28**
T_s [K]	673	675	**0.3**
$T_{w,i,rad}$ [K]	516	523	**1.4**
Q_f [kW]	31091.6	34816	**11.9**
Q_{rad} [kW]	17172.2	16813	**2**
Q_{conv} [kW]	6388.8	6693.2	**4.76**
$N_{t,total}$ [-]	100	110	**10**
$N_{t,shld}$ [-]	8	8	**0**
W [m]	4.8	5.1	**6**
L [m]	20	19.3	**3.5**

TABLE 6. Comparison of various methods to estimate thermal efficiency of fired heater

Variables	Ibrahim Design [17]	Mussati Method [26] (Dev. From [17])	Bahadori Method [28] (Dev. From [17])	This study (Dev. From [17])
η_{FH} [%]	0.7578	0.815 (**5.72 %**)	0.722 (**3.58 %**)	**0.78 (2.22 %)**

3. 2. Optimization Results and Parametric Analysis The Pareto front curve obtained from the optimization using total economic and environmental cost, and modified exergy efficiency (optimal design I) is depicted in Figure 4. According to this figure, an increase in the modified exergy efficiency can lead to a rise in the total cost. The highest modified exergy efficiency is 0.5213, resulting in a total economic and environmental cost of \$587053.56, which has the highest value. The lowest modified exergy efficiency belongs to 0.487, with the minimum cost of \$ 529919.9. To provide valuable insight into the fired heater's multi-optimal design, Pareto-front fitted curve is shown in Figure 4. The LINMAP optimal point marks in Figure 4.

To evaluate proposed objective functions (optimal design I), optimization using thermal efficiency with total economic and environmental cost (optimal design II) as objective functions was implemented, see Figure 5. Characteristics values of LINMAP suggested point of optimal design I is represented and compared with those of optimal design II in Table 7.

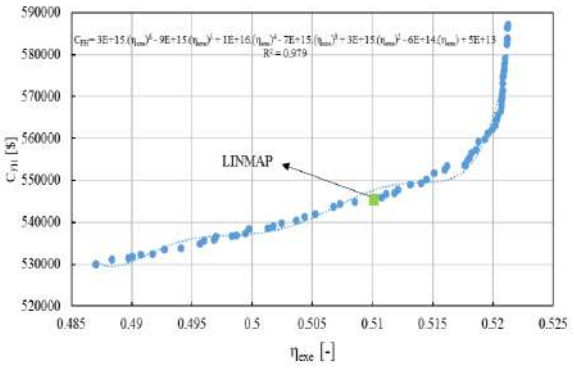


Figure 4. Pareto-front curve for optimization of modified exergy efficiency versus total cost (optimal design I)

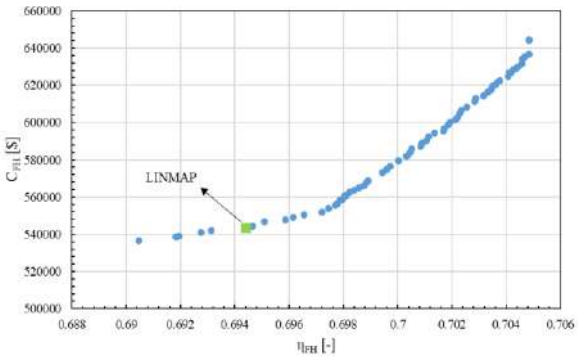


Figure 5. Pareto-front curve for optimization of energy efficiency versus total cost (optimal design II)

TABLE 7. Specifications of the LINMAP suggested points indicated in the Pareto front curves

	Parameters	LINMAP	
		Optimal design I	Optimal design II
Decision variables	$T_{w,o}$ [°C]	195.4	180.1
	$d_{o,rad}$ [m]	0.04	0.0401
	$d_{o,conv}$ [m]	0.1128	0.1317
	$N_{t,rad}$ [-]	40	63
	$N_{t,radc}$ [m]	10	13
	Excess air [-]	0.2227	0.2442
Objective functions	C_{FH} [$]	545370.7	543245.9
	η_{exe} [-]	0.5101	0.4971
	η_{FH} [-]	0.731	0.6944
Design parameters	Q_{rad} [MW]	3.788	4.223
	Q_{conv} [MW]	1.957	1.739
	Q_{shld} [MW]	2.109	1.8929
	A_{cp} [m^2]	45.3612	64.9357
	A_{conv} [m^2]	69.4182	77.5283
	$N_{t,conv}$ [-]	24	23
	ΔP_t [Pa]	31374	17864.5
	T_g [K]	1172.5	1103.4
	$V_{[m^3]}$	76.46	152.56
	m_{fuel} [kg/s]	0.2448	0.2574

According to Table 7, LINMAP suggested point in optimal design I has a higher total cost (0.4%) and modified exergy efficiency (1.3%) than LINMAP suggested point in optimal design II. It results from a considerable reduction in total heat exchanger area (because of reduction in tube length in radiation section, external tube diameter in convection section, number of tubes in radiation section, and number of tubes in convention section), and a rise in effective gas temperature in firebox (T_g).

The heat exchanger area in LINMAP suggested optimal design I decreases in contrast to optimal design II due to reduction in the number of tubes. Furthermore, employing modified exergy efficiency, and total economic and environmental cost as objective functions (optimal design I) results in smaller heat transfer area of radiation zone with a lower number of tubes compared to results of optimal design II, which causes a decrease in investment cost [26]. As mentioned in the methodology section, heat demand is assumed to be constant; therefore, excess air decreases and effective gas temperature in the firebox increases to compensate for the reduction in total heat exchanger area compared with optimal design II. Moreover, a decrease in excess air of LINMAP suggested points of optimal design I leads to improvement in energy efficiency in contrast to that of optimal design II [42]. The total pressure drop of the tube-side of the fired heater increases significantly due to a reduction in external tube diameter in the convection section, which is in line with previous studies [27]. It should be mentioned that allowable pressure drop of tube-side is considered1000 kPa [43].

For sensitivity analysis, the effects of key parameters, namely hot water outlet temperature, the external diameter of the tube in radiation as well as convection sections, number of side-wall as well as ceiling tube in radiation section, and excess air ratio are investigated on objective functions and presented in Figures 6 to 10, respectively. In this study, the LINMAP optimal point considered as a reference point. According to these figures, hot water outlet temperature is employed along with other decision variables due to the remarkable effects of this parameter on objective functions. According to Figure 6, an increase in the tube's external diameter in the radiation section raises total cost due to an increase in the total heat exchanging area. However, with increasing this parameter, modified exergy efficiency remains unchanged at low values of hot water outlet temperature. At high values of hot water outlet temperature, modified exergy efficiency first increases then decreases. An increase in hot water outlet temperature leads to rise in modified exergy efficiency at low values of external diameters of the tube. However, it fluctuates slightly at high values of external diameter, which results from large changes in the mean heat flux of large diameter tubes because of a lower degree of shadowing in these tubes [44].

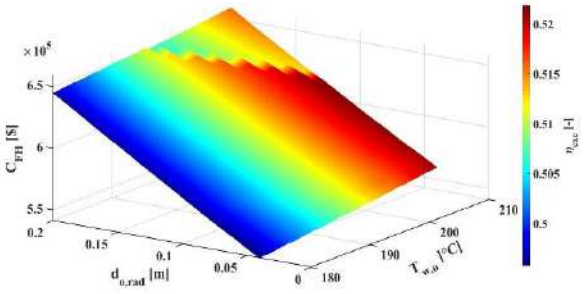


Figure 6. The effect of water outlet temperature and external diameter of tube in radiation section on modified exergy efficiency and total cost

Figure 7 reveals a surface plot of modified exergy, and total economic and environmental cost of fired heater versus external diameter of tube in convention section and hot water outlet temperature. An increase in external diameter of tube in convention section increases total cost, but it causes an initial climb in modified exergy efficiency, decreasing afterwards. It can be seen that a rise in hot water outlet temperature increases total economic and environmental cost at low values of the external diameter of the convention tube. However, it leads to a reduction in total cost at high values of the external diameter of the convention tube. Moreover, it is indicated that the diameter of tube in the radiation section (Figure 6) has a greater impact on total cost than the diameter of tube in the convection section, which concurs with previous research [27].

Figure 8 indicates that the total cost of the fired heater increases gradually by increasing either the number of the side-wall tubes in the radiation section or hot water outlet temperature. The number of the side-wall tubes has no considerable effect on modified exergy efficiency.

According to Figure 9, the effects of the number of the ceiling tubes in the radiation section on total cost and modified exergy efficiency bear a close resemblance to those of the number of the side-wall tubes (Figure 8).

Figure 10 provides an assessment of the total economic and environmental cost, and modified exergy efficiency of the fired heater for excess air and hot water outlet temperature. An increase in excess air leads to a slight rise in the total cost and a small drop in modified exergy efficiency. These results match those observed in previous studies [28, 45]. The latter result is a narrower recirculation zone and a larger vortex shedding upward in the combustion zone with an increase in excess air [45].

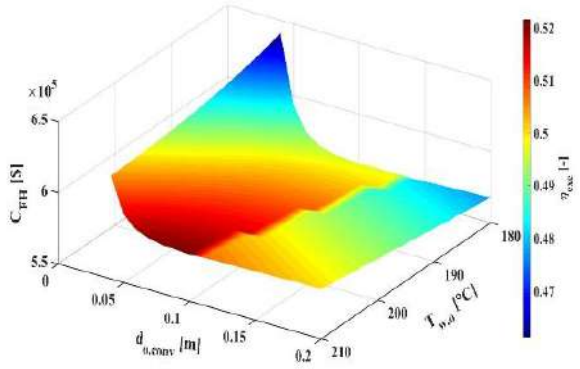


Figure 7. The effect of water outlet temperature and external diameter of tube in convection section on modified exergy efficiency and total cost

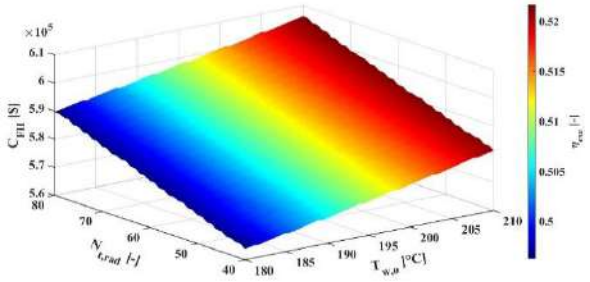


Figure 8. The effect of water outlet temperature and number of side-wall tube in radiation section on modified exergy efficiency and total cost

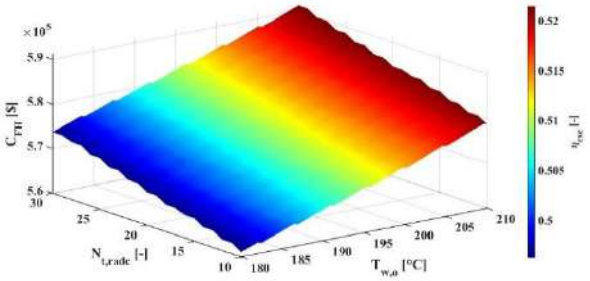


Figure 9. The effect of water outlet temperature and number of ceiling tube in radiation section on modified exergy efficiency and total cost

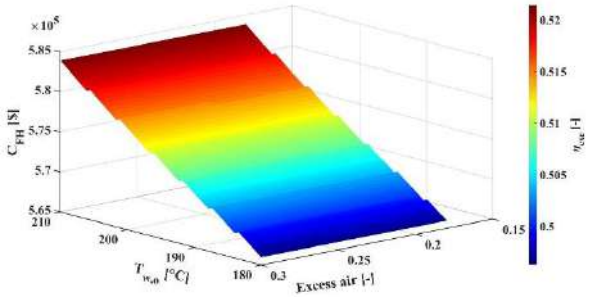


Figure 10. The effect of water outlet temperature and excess air on modified exergy efficiency and total cost

4. CONCLUSIONS

The present study puts forward a fired heater design model integrated with a combustion sub-model to optimize the fired heater. A multi-objective optimization of the fired heater in a natural gas pressure reducing station is carried out using a genetic algorithm. In this regard, the mathematical model of fired heater and sub-model of combustion are developed to achieve design and performance parameters of the fired heater. LINMAP method is applied to suggest an optimal solution between Pareto-set points. The hot water outlet temperature, the external diameter of tubes in radiation

and convection sections, the number of side-wall and ceiling tubes in the radiation section, and excess air ratio are selected as design parameters. Finally, sensitivity analysis of design parameters and their effects on objective functions are performed individually. According to the results, the developed model enables a more accurate prediction of heaters' thermal efficiency compared with previous models. LINMAP yielded a solution for optimization using modified exergy efficiency and total cost as objective functions leading to an increase of 0.4% in total cost and 1.3% in modified exergy efficiency compared with that of optimization applying thermal efficiency and total cost as objective functions, respectively.

As noted in the introduction section, Iran has 2500 gas reduction stations, which traditionally consist of the gas-fired heater, throttling valve, heat exchanger, and control system. Moreover, fired heaters burn a significant portion of the passing natural gas, and have low energy efficiency. Therefore, the results of this study can be applied to assess the potential of energy and natural gas consumption savings in gas reduction stations in Iran.

It is also indicated in the introduction section that renewable heating systems and novel-heating technologies are not cost-effective solutions. Considering pollution cost is known as one of the main solutions to face mentioned challenges, and it should be considered in the optimal design of gas-fired heaters to move toward and develop efficient and low emission heating systems. Therefore, this study can be compared with future studies on applying renewable heating systems in natural gas stations in Iran. In our future research, we intend to focus on the optimal design of the gas reduction station system considering all components of the system, including heat exchanger, heater, turbo-expander, and throttle valve. System restrictions will also be taken into account to determine and propose a set of practical recommendations.

5. APPENDIX A

At first, objective functions should be non-dimensionalized, applying Euclidian non-dimensionalization, due to different dimensions of objective functions on the Pareto frontier [34].

$$F_{ij}^{n} = \frac{F_{ij}}{\sqrt[2]{\sum_{i=1}^{m}\left(F_{ij}\right)^{2}}} \quad \text{(A.1)}$$

Then the distance of each point on the Pareto frontier from ideal point (F_j^{ideal}) in which each objective has its best value, is calculated as follows [34]:

$$d_{i+} = \sqrt[2]{\sum_{j=1}^{n}\left(F_{ij} - F_j^{ideal}\right)^{2}} \quad \text{(A.2)}$$

LINMAP suggested solution is a point with a minimum distance from the ideal point.

6. REFERENCES

1. Dudley, B., *Bp statistical review of world energy*, in *BP Statistical Review, London, UK, accessed Aug*. 2018.116.
2. Daneshi, H., Zadeh, H.K. and Choobari, A.L., "Turboexpander as a distributed generator", in 2008 IEEE Power and Energy Society General Meeting-Conversion and Delivery of Electrical Energy in the 21st Century, Pittsburgh, PA, USA, IEEE., (2008), 1-7.
3. Farzaneh-Gord, M., Arabkoohsar, A., Dasht-bayaz, M.D., Machado, L. and Koury, R., "Energy and exergy analysis of natural gas pressure reduction points equipped with solar heat and controllable heaters", ***Renewable Energy***, Vol. 72, (2014), 258-270.
4. Olfati, M., Bahiraei, M., Heidari, S. and Veysi, F., "A comprehensive analysis of energy and exergy characteristics for a natural gas city gate station considering seasonal variations", ***Energy***, Vol. 155, (2018), 721-733.
5. Bloch, H. and Soares, C., "Turboexpanders and process applications, Gulf Professional Publishing, (2001).
6. Požívil, J., "Use of expansion turbines in natural gas pressure reduction stations", ***Acta Montanistica Slovaca***, Vol. 3, No. 9, (2004), 258-260.
7. Ashouri, E., Veysi, F., Shojaeizadeh, E. and Asadi, M., "The minimum gas temperature at the inlet of regulators in natural gas pressure reduction stations (CGS) for energy saving in water bath heaters", ***Journal of Natural Gas Science and Engineering***, Vol. 21, (2014), 230-240.
8. Soleimani, P., Khoshvaght-Aliabadi, M., Rashidi, H. and Bahmanpour, H., "Performance enhancement of water bath heater at natural gas city gate station using twisted tubes", ***Chinese Journal of Chemical Engineering***, Vol. 28, No. 1, (2020), 165-179.
9. Arabkoohsar, A., Farzaneh-Gord, M., Deymi-Dashtebayaz, M., Machado, L. and Koury, R., "A new design for natural gas pressure reduction points by employing a turbo expander and a solar heating set", ***Renewable Energy***, Vol. 81, (2015), 239-250.
10. Ghezelbash, R., Farzaneh-Gord, M., Behi, H., Sadi, M. and Khorramabady, H.S., "Performance assessment of a natural gas expansion plant integrated with a vertical ground-coupled heat pump", ***Energy***, Vol. 93, No., (2015), 2503-2517.
11. Bargiel, P., Kostowski, W. and Usón, S., "An approach to enhance combined cycle performance by integration with a gas pressure reduction station", ***Journal of Power Technologies***, Vol. 95, No. 1, (2015).
12. Farzaneh-Kord, V., Khoshnevis, A., Arabkoohsar, A., Deymi-Dashtebayaz, M., Aghili, M., Khatib, M., Kargaran, M. and Farzaneh-Gord, M., "Defining a technical criterion for economic justification of employing chp technology in city gate stations", ***Energy***, Vol. 111, (2016), 389-401.
13. Arabkoohsar, A., Gharahchomaghloo, Z., Farzaneh-Gord, M., Koury, R. and Deymi-Dashtebayaz, M., "An energetic and economic analysis of power productive gas expansion stations for employing combined heat and power", ***Energy***, Vol. 133, (2017), 737-748.

14. Darabi, A., Shariati, A., Ghanaee, R. and Soleimani, A., "Economic assessment of a hybrid turboexpander-fuel cell gas energy extraction plant", ***Turkish Journal of Electrical Engineering & Computer Sciences***, Vol. 24, No. 3, (2016), 733-745.
15. Saadat-Targhi, M. and Khanmohammadi, S., "Energy and exergy analysis and multi-criteria optimization of an integrated city gate station with organic rankine flash cycle and thermoelectric generator", ***Applied Thermal Engineering***, Vol. 149, (2019), 312-324.
16. Farzaneh-Gord, M., Arabkoohsar, A., Dasht-bayaz, M.D. and Farzaneh-Kord, V., "Feasibility of accompanying uncontrolled linear heater with solar system in natural gas pressure drop stations", ***Energy***, Vol. 41, No. 1, (2012), 420-428.
17. Ibrahim, H.A.-H. and Al-Qassimi, M., "Simulation of heat transfer in the convection section of fired process heaters", ***Periodica Polytechnica Chemical Engineering***, Vol. 54, No. 1, (2010), 33-40.
18. Association, G.P.S., "Engineering data book, Gas Processors Suppliers Association, (2004).
19. Sinnott, R., "Chemical engineering design, Elsevier, (2014).
20. Ebrahimi, H., Mohammadzadeh, J.S.S., Zamaniyan, A. and Shayegh, F., "Effect of design parameters on performance of a top fired natural gas reformer", ***Applied Thermal Engineering***, Vol. 28, No. 17-18, (2008), 2203-2211.
21. Hottel, H. and Cohen, E., "Radiant heat exchange in a gas-filled enclosure: Allowance for nonuniformity of gas temperature", ***AIChE Journal***, Vol. 4, No. 1, (1958), 3-14.
22. Ibrahim, H.A.-H. and Al-Qassimi, M., "Matlab program computes thermal efficiency of fired heater", ***Periodica Polytechnica Chemical Engineering***, Vol. 52, No. 2, (2008), 61-69.
23. Towler, G. and Sinnott, R., "Chemical engineering design: Principles, practice and economics of plant and process design, Elsevier, (2012).
24. Institute, A.P., *Standard 530: Calculation of heater tube thickness in petroleum refineries*. 2008, Washington, DC: American Petroleum Institute.58.
25. Institute, A.P., *Standard 560: Fired heaters for general refinery services*. 2001, Washington, DC: American Petroleum Institute.
26. Mussati, S., Manassaldi, J.I., Benz, S.J. and Scenna, N.J., "Mixed integer nonlinear programming model for the optimal design of fired heaters", ***Applied Thermal Engineering***, Vol. 29, No. 11-12, (2009), 2194-2204.
27. Haratian, M., Amidpour, M. and Hamidi, A., "Modeling and optimization of process fired heaters", ***Applied Thermal Engineering***, Vol. 157, No., (2019), 113722.
28. Bahadori, A. and Vuthaluru, H.B., "Novel predictive tools for design of radiant and convective sections of direct fired heaters", ***Applied Energy***, Vol. 87, No. 7, (2010), 2194-2202.
29. Abdolalipouradl, M., Khalilarya, S. and Jafarmadar, S., "Energy and exergy analysis of a new power, heating, oxygen and hydrogen cogeneration cycle based on the sabalan geothermal wells", ***International Journal of Engineering*** ***Transactions C: Aspects,*** Vol. 32, No. 3, (2019), 445-450.
30. Dincer, I. and Rosen, M.A., "Exergy: Energy, environment and sustainable development, Newnes, (2012).
31. Kazemzadeh Hannani, S., Saidi, M. and Jafarian, A., "Second law based analysis of fluid flow in the regenerator of pulse tube refrigerators", ***International Journal of Engineering*** ***Transactions A: Basics,*** Vol. 21, No. 2, (2008), 181-194.
32. Rath, M. and Acharya, S., "Exergy and energy analysis of diesel engine using karanja methyl ester under varying compression ratio", ***International Journal of Engineering*** ***Transactions B: Applications,*** Vol. 27, No. 8, (2014), 1259-1268.
33. Lemmon, E.W., Huber, M.L. and McLinden, M.O., *Nist standard reference database 23: Reference fluid thermodynamic and transport properties-refprop*, in *NIST standard reference database*. 2013.
34. Sayyaadi, H. and Mehrabipour, R., "Efficiency enhancement of a gas turbine cycle using an optimized tubular recuperative heat exchanger", ***Energy***, Vol. 38, No. 1, (2012), 362-375.
35. Ismail, O.S. and Umukoro, G.E., "Modelling combustion reactions for gas flaring and its resulting emissions", ***Journal of King Saud University-Engineering Sciences***, Vol. 28, No. 2, (2016), 130-140.
36. Turns, S.R., "Introduction to combustion, McGraw-Hill Companies, (1996).
37. Ministry, P., *Electricity tariffs and their general conditions from the beginning of may, 2019*. 2019.11.
38. Ministry, P., *Natural gas tariffs*. 2020.2.
39. He, L., Lu, Z., Pan, L., Zhao, H., Li, X. and Zhang, J., "Optimal economic and emission dispatch of a microgrid with a combined heat and power system", ***Energies***, Vol. 12, No. 4, (2019), 604.
40. Frangopoulos, C.A., "Exergy, energy system analysis and optimization-volume iii: Artificial intelligence and expert systems in energy systems analysis sustainability considerations in the modeling of energy systems, EOLSS Publications, Vol. 3, (2009).
41. Pakzad, A., "Turboexpanders and energy recovery", 1 st ed, Tehran, Iran, Afrooz Publishing, (2010), 350.
42. Stehlik, P., Kohoutek, J. and Jebáček, V., "Simple mathematical model of furnaces and its possible applications", ***Computers & Chemical Engineering***, Vol. 20, No. 11, (1996), 1369-1372.
43. Shekarchian, M., Zarifi, F., Moghavvemi, M., Motasemi, F. and Mahlia, T., "Energy, exergy, environmental and economic analysis of industrial fired heaters based on heat recovery and preheating techniques", ***Energy Conversion and Management***, Vol. 71, (2013), 51-61.
44. Hájek, J. and Jegla, Z., "Standards for fired heater design: Analysis of two dominant heat flux variation factors", ***Applied Thermal Engineering***, Vol. 125, (2017), 702-713.
45. Khodabandeh, E., Pourramezan, M. and Pakravan, M.H., "Effects of excess air and preheating on the flow pattern and efficiency of the radiative section of a fired heater", ***Applied Thermal Engineering***, Vol. 105, (2016), 537-548.

Persian Abstract

چکیده

در این پژوهش، مدلی جدید به منظور طراحی بهینه گرمکن آب داغ با لحاظ کردن مدل احتراق ارائه شد. مدل توسعه یافته توانایی طراحی تمامی بخش های گرمکن آب داغ از قبیل محفظه تشعشعی، بخش همرفت و بخش محافظ و همچنین محاسبه ترکیبات و حجم گازهای حاصل از احتراق را دارد. طراحی بهینه چندهدفه گرم کن آب داغ ایستگاه تقلیل فشار گاز نیروگاه رامین بر مبنای مدل توسعه یافته انجام شد. هزینه کل اقتصادی و زیست محیطی همراه با بازدهی اگزرژی اصلاحی به عنوان توابع هدف در نظر گرفته شدند. علاوه براین، روش تصمیم گیری چند معیاره به منظور پیشنهاد نقطه بهینه از بین نقاط بهینه نمودار پرتو بهینه سازی چند هدفه استفاده شد. مدل توسعه یافته جدید عملکرد بهتری در پیش بینی بازدهی گرم کن آب داغ نسبت به مدل های پیشین دارد. نتایج بهینه سازی نشان داد که نقطه بهینه پیشنهادی روش تصمیم گیری چندمعیاره لینمپ بر مبنای توابع هدف پیشنهادی نسبت به نقطه مشابه بهینه سازی بر مبنای توابع هدف اقتصادی و بازدهی انرژی، بازدهی اگزرژی بیشتری دارد (۱/۳ ٪) در حالی که هزینه کل ثابت باقی می ماند. در نهایت، تأثیر هر یک از متغیرهای تصمیم گیری بهینه سازی بر توابع هدف مورد ارزیابی قرار گرفت.

IJE TRANSACTIONS A: Basics Vol. 34, No. 7, (July 2021) 1799-1803

International Journal of Engineering

Journal Homepage: www.ije.ir

Experimental Analysis for Determination of Longitudinal Friction Coefficient Function in Braking Tractor Semi-trailer

N. Thanh Tung*

Vinh Long University of Technology Education, Nguyen Hue street, Vinh Long city, Vietnam

P A P E R I N F O

Paper history:
Received 26 April 2021
Received in revised form 26 May 2021
Accepted 28 May 2021

Keywords:
Longitudinal Friction Coefficient
Longitudinal Slip Coefficient
Vehicle Dynamics
Tractor Semi-trailer

A B S T R A C T

The longitudinal friction coefficient is a very important parameter to calculate vehicle dynamics. The theoretically longitudinal friction coefficient is often used to investigate the vehicle dynamics. However, the longitudinal friction coefficient depends on many factors and changes when the vehicle moves on the actual road. This paper presents the experimental method to determine the longitudinal friction coefficient function when braking the tractor semi-trailer on the road. The results of this study can be used as an input to dynamic survey model for the tractor semi-trailer and to verify the theoretical model. The experimental results showed that when braking the tractor semi-trailer on the dry asphalt, the maximum longitudinal friction coefficient is φ_{xmax}=0.89 and the minimum longitudinal friction coefficient is φ_{xmin}=0.72. Matlab-Simulink software was used to investigate and compare the longitudinal friction coefficient determined by experiment and the theoretically longitudinal friction coefficient according to Ammon tire model. The survey results showed that the average error between experiment and theory was about 17%.

***doi**: 10.5829/ije.2021.34.07a.24*

1. INTRODUCTION

The movement of the vehicle depends on the tire-road forces F_x, F_y, F_z. The tire-road forces F_x, F_y, F_z are usually determined by the tire model and they depend on the tire-road friction coefficient [1-3].

$$F_x(t) = F_z(t)\varphi_x(t) \quad (1)$$

$$F_y(t) = F_z(t)\varphi_y(t) \quad (2)$$

Therefore, the longitudinal friction coefficient function by time can be determined:

$$\varphi_x(t) = \frac{F_x(t)}{F_z(t)} \quad (3)$$

Thus, in order to determine the longitudinal friction coefficient function $\varphi_x(t)$, $F_x(t)$ and $F_z(t)$ should be known [4].

*Corresponding Author Institutional Email: tungnt@vlute.edu.vn (N. Thanh Tung)

2. THEORETICAL BASES

The method of separating structure of the multi-object system for the vehicle dynamic model ¼ is used as shown in Figure 1, the equation of vehicle movement is as follows [5]:

$$\begin{cases} m\ddot{z} = F_C + F_K \\ m_A\ddot{\xi} = F_{CL} - (F_C + F_K) \\ (m + m_A)\ddot{x} = F_X + F_W \\ F_Z = F_G + F_{CL} \end{cases} \quad (4)$$

$$\text{Or} \quad \begin{cases} F_X = (m + m_A)\ddot{x} - F_W \\ F_Z = F_G + m\ddot{z} + m_A\ddot{\xi} \end{cases} \quad (5)$$

The longitudinal slip coefficient of the wheel by time is determined by the following Equations (6) and (7).

The longitudinal slip coefficient of the wheel when braking [6, 7]:

$$s_x(t) = \frac{r_d\dot{\varphi}(t) - \dot{x}(t)}{\dot{x}(t)}; r_d\dot{\varphi}(t) < \dot{x}(t); -1 < s_x < 0 \quad (6)$$

Please cite this article as: N. Thanh Tung, Experimental Analysis for Determination of Longitudinal Friction Coefficient Function in Braking Tractor Semi-trailer, International Journal of Engineering, Transactions A: Applications, Vol. 34, No. 7, (2021) 1799-1803

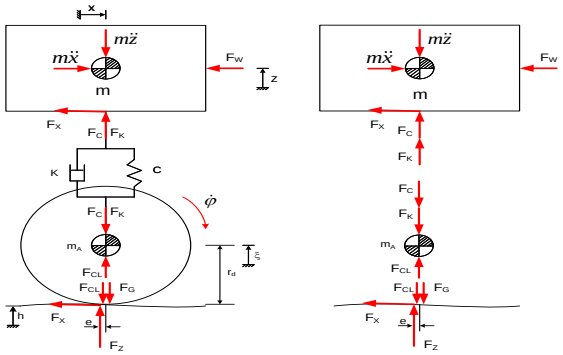


Figure 1. Model of vehicle dynamics ¼

The longitudinal slip coefficient of the wheel when accelerating [5, 6, 8]:

$$s_x(t) = \frac{r_d\dot{\varphi}(t) - \dot{x}(t)}{r_d\dot{\varphi}(t)}; r_d\dot{\varphi}(t) > \dot{x}(t); 0 < s_x < 1 \qquad (7)$$

So in order to determine the longitudinal friction coefficient function of the wheel according to the longitudinal slip coefficient $\varphi_x(s_x)$, the longitudinal friction coefficient function by time $\varphi_x(t)$ should be known as in formula (3) and the longitudinal slip coefficient function by time $s_x(t)$ as in formula (6,7). The longitudinal friction coefficient function according to the longitudinal slip coefficient is shown in Figure 2 [9, 10].

3. EXPERIMENT AND SIMULATION

3. 1. Experiment In this paper, a 6-axle tractor semi-trailer was chosen for experiment. During the experiment, the braking systems of the axles I, II, III, IV, V were adjusted to stop working. The diagram for installation of the experimental sensor is shown in Figure 3.

To determine the longitudinal friction coefficient function of the wheel according to the longitudinal slip coefficient $\varphi_x(s_x)$, the following 6 parameters were experimentally determined: Measurement of the longitudinal velocity of the tractor semi-trailer ($\dot{x}$) by

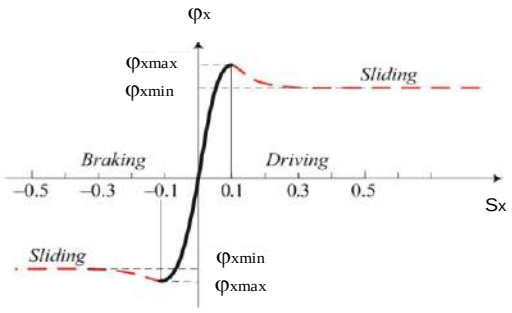
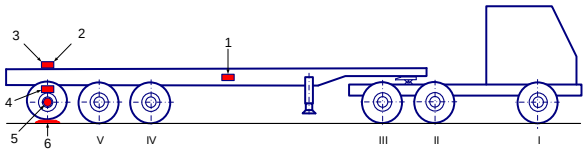


Figure 2. Longitudinal friction coefficient as a function of slip ratios



1. Sensor for measuring the body longitudinal velocity;
2. Sensor for measuring the body longitudinal acceleration;
3. Sensor for measuring the body vertical acceleration;
4. Sensor for measuring the axle vertical acceleration;
5. Sensor for measuring the wheel angular velocity;
6. Weigh the axle weight

Figure 3. Sensor installation diagram on the experimental tractor semi-trailer

Kistler GPS sensor (1); measurement of the longitudinal acceleration ($\ddot{x}$) and vertical acceleration ($\ddot{z}$) of the tractor semi-trailer by MMA7361LC-XYZ sensor (2,3); measurement of the vertical acceleration ($\ddot{\xi}$) of the axle by MMA7361LC-XYZ sensor (4); measurement of the wheel angular velocity ($\dot{\varphi}$) by Sharp Rotary Encoder sensor (5); weighing the un-sprung mass (m_A) and the sprung mass (m) with ULSTRALIM electronic balance (6). Diagram of the measurement system, signal reception and braking experimental result processing of the tractor semi-trailer are shown in Figure 4 [11, 12].

Let the tractor semi-trailer moved steadily on the dry asphalt at a speed of 50 km/h and then braked to let the tractor came to a complete stop. Based on 6 parameters determined by the experiment, the computer processed and output graphs $F_z(t)$, $F_x(t)$, $\varphi_x(t)$, $s_x(t)$ as shown in Figures 5, 6, 7 and 8. The longitudinal friction coefficient function of the wheel according to the longitudinal slip coefficient $\varphi_x(s_x)$ is shown in Figure 9.

The experimental results show that the value and shape of the graphs $F_z(t)$, $F_x(t)$, $\varphi_x(t)$, $s_x(t)$, $\varphi_x(s_x)$ are consistent with the theoretical rules. When braking the tractor semi-trailer on the dry asphalt, the maximum longitudinal friction coefficient is φ_{xmax}=0.89 and the minimum longitudinal friction coefficient is φ_{xmin}=0.72,

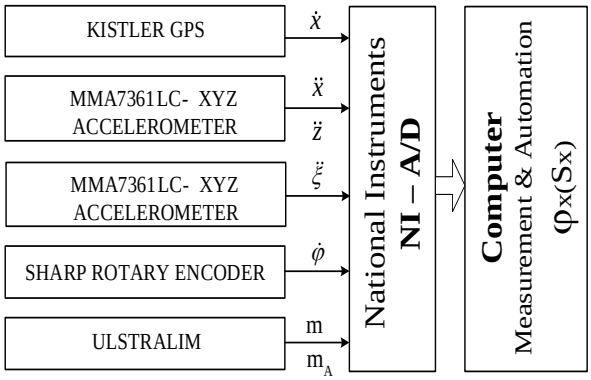


Figure 4. The longitudinal friction coefficient measurement system

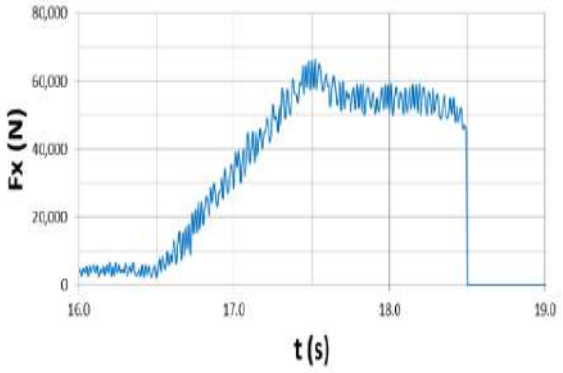


Figure 5. Tire longitudinal force

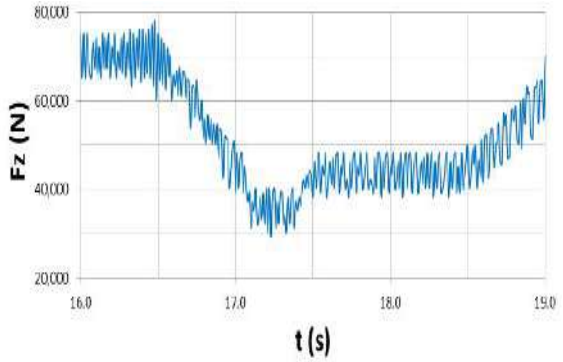


Figure 6. Tire vertical force

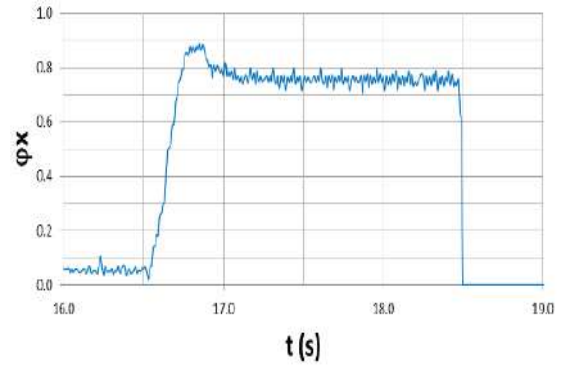


Figure 7. Longitudinal friction coefficient

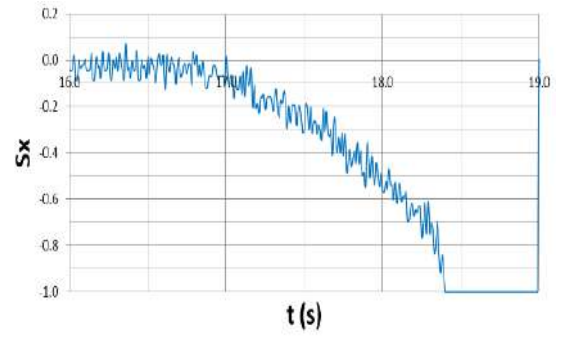


Figure 8. Longitudinal slip ratio

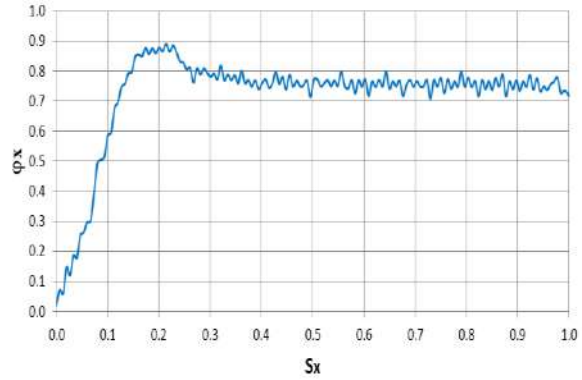


Figure 9. Function of longitudinal friction coefficient

the value of the longitudinal friction coefficient was determined by experiment equivalent to theoretical value.

3. 2. Simulation Matlab-Simulink software was used to investigate and compare the longitudinal friction coefficient function of the wheel according to the longitudinal slide coefficient $\boldsymbol{\varphi_x(s_x)}$ determined by experiment and the function $\boldsymbol{\varphi_x(s_x)}$ simulated according to Ammon tire model on the same type of road with the maximum longitudinal friction coefficient φ_{xmax}=0.89 and the minimum longitudinal friction coefficient φ_{xmin}=0.72 [4, 5]. The results of function simulation $\boldsymbol{\varphi_x(s_x)}$ by experiment and according to Ammon tire model are shown in Figure 10. The survey results showed that the average error between experiment (the longitudinal friction coefficient function $\boldsymbol{\varphi_x(s_x)}$ determined by experiment) and theory (the longitudinal friction coefficient function $\boldsymbol{\varphi_x(s_x)}$ determined by Ammon tire model) was about 17%.

The function $\boldsymbol{\varphi_x(s_x)}$ determined by experiment and function $\boldsymbol{\varphi_x(s_x)}$ determined by Ammon tire model were used as an input to the dynamic survey model of the tractor semi-trailer as in Figures 11 and the system of dynamic equations of the tractor semi-trailer as fomulas (8-49).

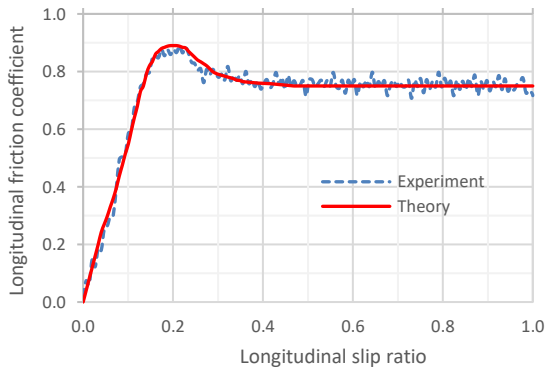


Figure 10. Function $\boldsymbol{\varphi_x(s_x)}$ determined by experiment and Ammon tire model

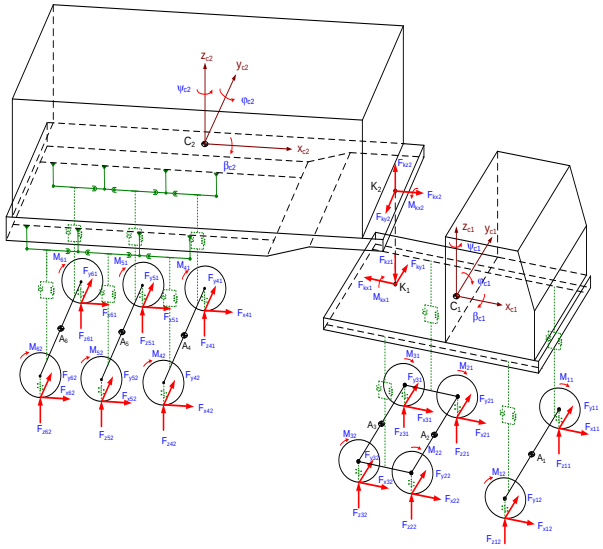


Figure 11. Dynamic model of the tractor semi-trailer

The system of dynamic equations of the tractor semi-trailer are as follows [5]:

$$m_{c1}\ddot{x}_{c1} = F_{xij} - F_{wx1} - F_{kx1}(i = 1 - 3) \quad (8)$$

$$m_{c1}\ddot{y}_{c1} = F_{yij} - F_{wy1} - F_{ky1}(i = 1 - 3) \quad (9)$$

$$J_{zc1}\ddot{\psi}_{c1} = F_{x1j}l_1 + F_{x1j}b_1 + F_{ky1}l_{k1} - F_{yij}l_i + (F_{xi2} - F_{xi1})b_i \quad (10)$$

$$m_{c2}\ddot{x}_{c2} = F_{xij} + F_{kx2}(i = 4 - 6) \quad (11)$$

$$m_{c2}\ddot{y}_{c2} = F_{ky2} + F_{yij}(i = 4 - 6) \quad (12)$$

$$J_{zc2}\ddot{\psi}_{c2} = (F_{xi2} - F_{xi1})b_i + F_{ky2}l_{k2} - F_{yij}l_i(i = 4 - 6) \quad (13)$$

$$m_{c1}\ddot{z}_{c1} = F_{Cij} + F_{Kij} - F_{kz1}(i = 1 \div 3) \quad (14)$$

$$J_{yc1}\ddot{\varphi}_{c1} = (F_{C1j} + F_{K1j})l_i + F_{kz1}l_{k1} - F_{kx1}(h_{c1} - h_{k1}) + M_{ij}(i = 1 \div 3) \quad (15)$$

$$m_{c2}\ddot{z}_{c2} = F_{Cij} + F_{Kij} + F_{kz2}(i = 4 \div 6) \quad (16)$$

$$J_{yc2}\ddot{\varphi}_{c2} = -(F_{Cij} + F_{Kij})l_i + F_{kx2}(h_{c2} - h_{k2}) + F_{kz2}l_{k2} + M_{ij}(i = 4 \div 6) \quad (17)$$

$$J_{xc1}\ddot{\beta}_{c1} = (F_{Ci2} + F_{Ki2} - F_{Ci1} - F_{Ki1})w_i + M_{kx1} \quad (i = 1 \div 3) \quad (18)$$

$$J_{xc2}\ddot{\beta}_{c2} = (F_{Ci2} + F_{Ki2} - F_{Ci1} - F_{Ki1})w_i - M_{kx2}(i = 4 \div 6) \quad (19)$$

$$m_{Ai}\ddot{z}_{Ai} = F_{CLij} + F_{KLij} - F_{Cij} - F_{Kij}(i = 1 \div 6) \quad (20\text{-}25)$$

$$m_{Ai}\ddot{y}_{Ai} = F_i + F_{yij}(i = 1 \div 6) \quad (26\text{-}31)$$

$$J_{Axi}\ddot{\beta}_{Ai} = (F_{Ci1} + F_{Ki1} - F_{Ci2} - F_{Ki2})w_i + (F_{CLi2} - F_{CLi1})b_i - F_{yij}(r_{ij} + \xi_{Aij}) \quad (32\text{-}37)$$

$$J_{Ayij}\ddot{\varphi}_{ij} = M_{Aij} - M_{Bij} - F_{xij}r_{dij} \quad (38\text{-}49)$$

Graphs of force F_x, F_z surveyed by Matlab-Simulink software with the input function $\varphi_x(s_x)$ determined by experiment and Ammon function are presented in Figures 12 and 13. The survey results showed that when the input function $\varphi_x(s_x)$ determined by experiment and Ammon function were used, the average error of force F_x was 11,81% and the average error of force F_z is 19 and 25%.

The obtained results determined by the experiment had shapes and values consistent with theoretical rules as well as research results of many other authors [8, 11, 12].

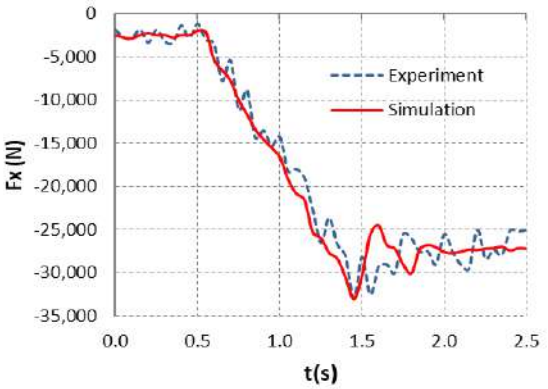


Figure 12. Tire longitudinal force F_x

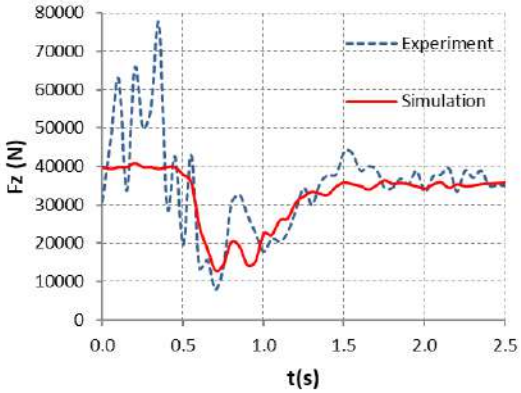


Figure 13. Tire vertical force F_z

4. CONCLUSION

The longitudinal friction coefficient function $\varphi_x(s_x)$ can be determined by measuring the longitudinal velocity of the vehicle ($\dot{x}$), the longitudinal acceleration of the vehicle ($\ddot{x}$), the vertical acceleration of the body ($\ddot{z}$), the vertical acceleration of the axle ($\ddot{\xi}$), the wheel angular velocity and weighing the vehicle weight. The experimental results can be used to study the dynamics of the tractor semi-trailer and other types of vehicles on actual roads.

5. REFERENCES

1. Rajamani, R., "Vehicle dynamics and control, Springer Science & Business Media, (2011).
2. Jazar, R.N., "Vehicle dynamics: Theory and application, Springer, (2017).
3. Van Huong, V., Vehicle dynamics. 2014, Vietnam Education Publishing House, Vietnam.
4. Tung, N.T., Huong, V.V. and Kiet, P.T.J.I.J.o.M.P.B., "Experimental research on determining the vertical tyre force of a tractor semi-trailer", Vol. 34, No. 22n24, (2020), 2040163, doi: 10.1142/S0217979220401633.
5. Tung, N.T. and Van Huong, V., "The effect of the wheel rotation angle on the braking efficiency of the tractor semi-trailer on the wet roundabout route", in International Conference on Engineering Research and Applications, Springer. Vol., No., (2020), 798-804. doi.org/10.1007/978-3-030-64719-3_87
6. Jazar, R.N., "Advanced vehicle dynamics, Springer, Springer Nature Switzerland AG, (2019).
7. Chen, W., Xiao, H., Wang, Q., Zhao, L. and Zhu, M., "Integrated vehicle dynamics and control, John Wiley & Sons, (2016).
8. Bouteldja, M., Jacob, B. and Dolcemascolo, V., "Optimized design of weigh-in-motion multiple-sensors array by an energetic approach", in International Conference on Heavy Vehicles HVParis 2008: Weigh-In-Motion (ICWIM 5), Wiley Online Library. (2009), 187-198.
9. Schiehlen, W., *Dynamical analysis of vehicle systems. Cism courses and lectures*. 2007, Springer, Wien, NewYork.
10. Mitschke, M. and Wallentowitz, H., "Dynamik der kraftfahrzeuge, Springer, Vol. 4, (1972).
11. Emery, L., "Design and construction of a variable dynamic vehicle", in 16th International Technical Conference on the Enhanced Safety of VehiclesNational Highway Traffic Safety AdministrationTransport CanadaTransport Canada. No. DOTHS808759, (1998).
12. Sampson, D.J.M., "Active roll control of articulated heavy vehicles", Citeseer, (2000),

Persian Abstract

چکیده

ضریب اصطکاک طولی پارامتر بسیار مهمی برای محاسبه دینامیک خودرو است. از ضریب اصطکاک طولی نظری اغلب برای بررسی پویایی خودرو استفاده می شود. با این حال ، ضریب اصطکاک طولی به فاکتورهای زیادی بستگی دارد و هنگام حرکت وسیله نقلیه در جاده واقعی تغییر می کند. در این مقاله روش آزمایشی برای تعیین عملکرد ضریب اصطکاک طولی هنگام ترمزگرفتن نیمه تریلر تراکتور در جاده ارائه شده است. نتایج این مطالعه می تواند به عنوان ورودی به مدل بررسی پویا برای نیمه تریلر تراکتور و برای تأیید مدل نظری مورد استفاده قرار گرفته است. نتایج تجربی نشان داد که هنگام ترمز نیمه تریلر تراکتور روی آسفالت خشک ، حداکثر ضریب اصطکاک طولی φxmax = 0.89 و حداقل ضریب اصطکاک طولی minxmin = 0.72 می باشد. از نرم افزار Matlab-Simulink برای بررسی و مقایسه ضریب اصطکاک طولی تعیین شده با آزمایش و ضریب اصطکاک طولی نظری با توجه به مدل لاستیک آمون استفاده شد. نتایج بررسی نشان داد که میانگین خطای بین آزمایش و نظریه حدود ۱۷٪ بوده است.

International Journal of Engineering

Journal Homepage: www.ije.ir

Evaluating the Efficiency of Coal Loading Process by Simulating the Process of Loading onto the Face Conveyor with a Shearer with an Additional Share

N. K. Linh*[a], V. V. Gabov[b], Y. V. Lykov[b], R. Y. Urazbakhtin[b]

[a] *Ha Noi University of Mining and Geology, Vietnam*
[b] *Saint-Petersburg Mining University, Saint-Petersburg, Russia*

PAPER INFO

Paper history:
Received 07 July 2020
Received in revised form 23 April 2021
Accepted 13 May 2021

Keywords:
Coal
Mining
Shearer
Actuator
Loading
Modeling

ABSTRACT

This paper analyses the possibility of increasing the efficiency of loading the destroyed rock mass into the face scraper conveyor by the lagging screw actuator of the shearer in the process of coal mining in the complex mechanized treatment faces of coal mines. It was taken into account that the most significant influence on the loading efficiency of coal is exerted by the dimensions of the cross-sectional area of the loading window, the distance between the screw and conveyor and the height of the bottom of the conveyor. Non-traditional technical solutions are proposed that reduce the negative impact of the gap between the screw and the conveyor on the efficiency of coal loading by the lagging screw actuator of shearers and increase the degree of filling of the conveyor groove. Technical solutions contribute to the formation of a rational section of the cargo flow in the trough of the downhole conveyor and, therefore, increase its productivity. The results of modeling the process of loading coal onto the face conveyor by an auger actuator with an additional loading device are presented. Evaluation of the effectiveness of the proposed constructive technical solutions for the interface unit in the loading area confirmed an increase of 2.94 times the maximum capacity of the screw executive body for loading coal onto the face conveyor while 2.7 times less specific energy consumption during loading.

***doi**: 10.5829/ije.2021.34.07a.25*

1. INTRODUCTION

The mining of high-tech coal seams in countries with a developed mining industry is carried out mainly by long complex mechanized treatment faces equipped with dual auger shearers [1-5] with symmetrical layout. Such shearers can work on one-sided with stripping and shuttle schemes of coal extraction, with self-cutting into a seam on a new strip of excavation by oblique drives. Such combines are technological, highly productive, and their screw augers are simple in design, reliable in operation, combined with the operations of separating coal from the massif, unloading it from the destruction zone and loading it onto the face conveyor [6-8]. According to the efficiency of the process of coal mining on medium and powerful formations, there is currently no alternative [9-11].

However, the auger actuators are also characterized by significant disadvantages associated, in particular, with the operating conditions of the lagging auger actuator. These conditions are characterized by the following features:

- destroying a pack of coal along the soil, the lagging auger actuator loads it and part of the coal mass destroyed by the leading auger actutator onto the armoured conveyer;
- the mass of coal during loading moves "passively" along the channel formed by the interfaced equipment (becoming a conveyor, the gearbox actuator, auger actuator, retaining plate or blade), only under pressure created by the auger;
- the loading channel is structurally saturated with local resistances that cause circulation of coal in the flow, enhance its crushing, dust formation and increase energy consumption;
- the channel is open to move part of the destroyed mass to the bottomhole soil.

*Corresponding Author Email: khaclinhhumg@gmail.com (Nguyen Khac Linh)

Please cite this article as: N. K. Linh, V. V. Gabov, Y. V. Lykov, R. Y. Urazbakhtin, Evaluating the Efficiency of Coal Loading Process by Simulating the Process of Loading onto the Face Conveyor with a Shearer with an Additional Share, International Journal of Engineering, Transactions A: Basics Vol. 34, No. 7, (2021) 1804-1809

The problem also lies in the fact that the gap between the auger actuator and the conveyor and the cross-sectional area of the loading window are formed in the process of assembling equipment units without taking into account the special requirements for the loading channel.

To date, there are no technical solutions and proven methods for choosing the structure and parameters of loading devices for shearers that provide a given level of efficiency of the coal loading process according to the criteria of the maximum possible productivity, minimizing the yield of small classes of coal and specific energy consumption energy taking into account additional resistance to the movement of coal flow in the loading channel [1, 12-14].

Over the last few decades, the shearer drum's coal loading performance has not been adequately studied. In fact, many conventional studies have either overlooked it or obtained it empirically. The main factors affecting it were studied by the end of the 1980s and earlier of the 1990s [6, 15, 16]. In these studies, the authors showed the impact of some factors, such as mining conditions, the shearer's operating parameters, and mining equipment on the shearer drum's loading rate using empirical, theoretical analysis and computer technology. As a result, some key factors affecting the shearer drum's coal loading performance have been noted and recommended to consider for future studies. Having considered those key factors, a Turkish company also stated that the distance of the drum from the scraper conveyor might influence the shearer drum's coal loading performance [6].

Moreover, it has been shown that there is a difference between drums with conical hub and cylindrical hub regarding the coal loading performance [16]. Furthermore, by utilizing a helical vane in the form of a varying pitch and a curved generatrix, it is possible to improve the shearer drum's coal loading performance. The influences of the pulling speed, rotation speed on the coal loading rate of the shearer drum were also presented by Liu and Gao [12]. Nevertheless, the abovementioned models have not shown a solution to completely reduce the amount of coal left on the cleaning conveyor and shearer drum [14].

Hence, in this study, we want to develop a technical solution to improve the efficiency of the shearer's loading process and then compare its performance with the one being used by DEM software (Discrete Element Method) to validate the accuracy and reasonableness of the proposed technology. The DEM software is a computer simulation that has been widely used in several fields, such as geotechnical engineering, geology, and machinery fields [12–17], and especially the screw conveying field [12–14].

Therefore, this article proposes the installation of an additional share in the transition zone between the auger and the conveyor and evaluates the efficiency of the coal loading process by a shearer for two options: with an installed share and without a share.

Taking into account the insufficient study of the influence of the distance L from the auger actuator to the conveyor (Figure 1) and the configuration of the loading channel, the efficiency assessment was carried out using the modeling method. This made it possible to take into consideration the influence of the peculiarities of the nodes arrangement of the equipment in the considered zone on the process of passive movement of the destroyed mass to the conveyor.

The efficiency of the loading process also depends on the distance L_m, by which the coal flow during loading fills the trough of the downhole conveyor. This distance can be determined by the formula:

$$L_m = \frac{h_z + h_k}{\mathrm{tg}(\rho + \beta)} - L, \text{ mm} \quad (1)$$

where β is the angle of inclination of the conveyor to the horizontal, degrees; ρ - depositional gradient, degrees; h_z - excess of the upper boundary of the surface of the cargo flow of the bottom face of the conveyor, mm; h_k - conveyor bead height, mm; L is the distance along the soil between the auger and the face conveyor, mm.

The shorter the displacement mass displacement path (L) to the downhole conveyor, the lesser the resistance to the movement of the coal flow during loading onto the conveyor and the lesser the volume of unloaded coal and the higher the efficiency of the loading process. However, in modern designs of shearers and conveyors, this distance remains substantially significant - 300 mm or more.

2. TECHNICAL SOLUTION

As a result of the analysis of known designs of shearer units in the loading zone, the installation of an additional share in the transition zone between the auger and the

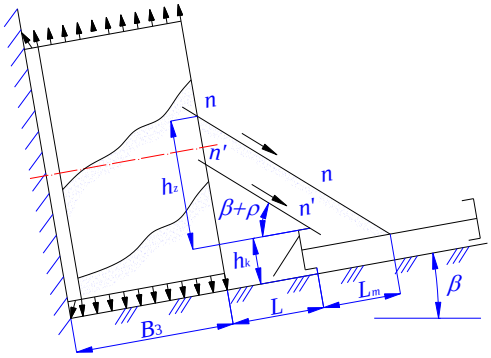


Figure 1. Scheme of the formation of cargo flow on the face conveyor

face conveyor, which may deviate from the initial position of its installation under the influence of oversize during its transportation, is proposed. The loading plate with additional share 1 is shown in Figure 2 in relation to a shearer with two screw actuators symmetrically located along its length and regulated by the thickness of the formation 7. The retaining flap 4 with an additional share 1 provide an increase in the efficiency of coal unloading from the transition zone, loading and formation of coal flow 10 on the bottomhole conveyor.

The loading device consists of a share 1, made in the form of a part of a truncated cone, a support bracket 6, a retaining flap 4, while the axis of symmetry of the working surface of the loading flap is parallel to the axis of rotation of the screw actuator 7. The share 1 is installed between the shield 4 and the face conveyor 9 at an angle less than 90 ° to its board and connected by a hinge 2 and rod 3 with a spring 11 with a bracket 5, which is attached

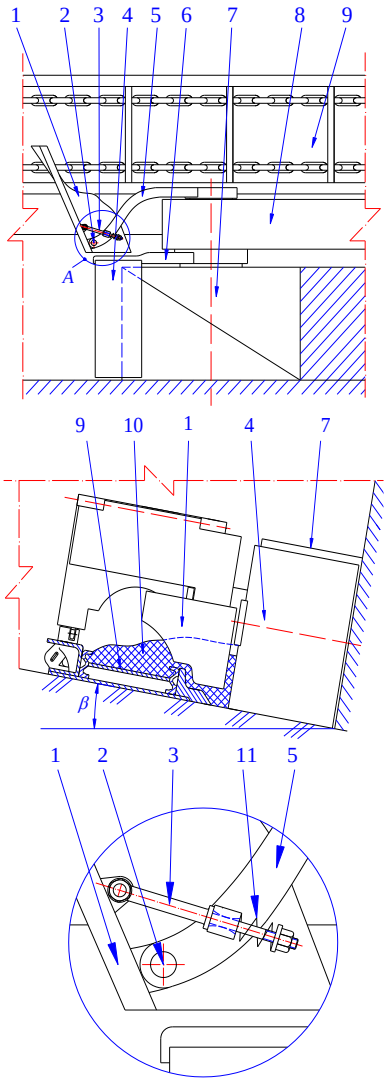


Figure 2. Loading device of a mining combine

to the housing 8 of the drive of the auger actuator. The lower edge of the share 1 is equidistant to the mating surface of the face of the conveyor 9 and is installed with a gap.

The auger actuator 7 (Figure 2), lagging along the direction of the combine harvester, destroys the lower layer of the coal seam, moves the broken mass by the auger blades with support on the loading plate 4 and ploughshare 1 to the downhole conveyor. In this case, the ploughshare reduces the resistance to movement of the flow of the destroyed mass to the downhole conveyor, reduces the circulation of coal in the flow and contributes to the formation of a rational cross-section of the flow.

The gap between the screw actuator and the bottomhole conveyor and the cross-sectional area of the loading window are limited in size due to the peculiarities of the layout of the equipment nodes, which creates increased resistance to the movement of the coal flow onto the conveyor. A loading plate with a ploughshare increase the efficiency of loading coal onto the face conveyor, as they provide:

- a more complete cleaning of the soil and loading of coal on the downhole conveyor;
- reducing the resistance to movement of the coal flow and, consequently, increasing the completeness of unloading of coal located in the zone between the conveyor side and the screw executive body, to the downhole conveyor;
- an increase in the fill factor of coal in the trough of the downhole conveyor;

2. 1. Simulation of the Process of Loading Coal with an Assessment of the Effectiveness of the Proposed Technical Solution Modeling is carried out to evaluate the efficiency of loading coal with a shearer with an additional share in the transition zone between the screw and the face conveyor (Figure 3). When modeling, the EDEM 3D program was used. There are two forms, unit-wall, and particles, in EDEM. The wall is only represented by the surface. Particles are generally only represented by a sphere. In order to simulate the coal loading process of the drum, the drum had to be simplified (Figure 3). In this model, the tube hub consisted of the cylindrical surface, the blade comprised the helical surface, and the rear shield.

Figure 3 was a simulation model of the combined drum and coal wall. The coal wall was filled with identical diameter particles and bonded together using a common bond between the particles in the model. Since the simulation's focus was the coal loading of the drum, the cutting force of picks was not in the scope of the research. Meanwhile, to reduce the simulation time and improve simulation efficiency, the small values of adhesive bond strength of the normal direction and tangential were taken in the model. As long as the particles did not collapse before cut by the picks, the

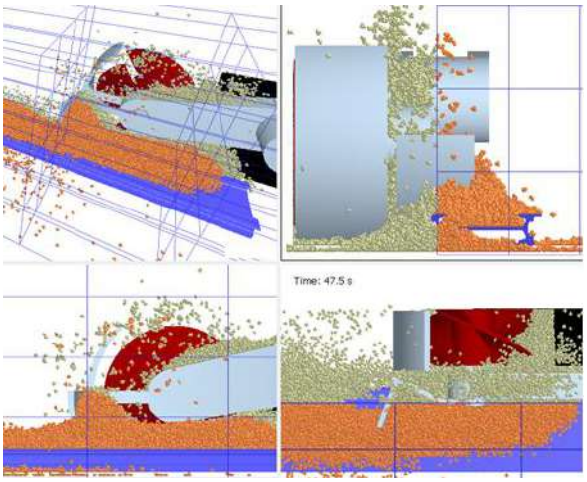


Figure 3. Feature of the process of loading coal with a lagging screw with an additional share

bonding strength of normal direction was 8 Pa and the adhesive shear strength was 8 Pa in this paper. Statistical area of effective output particles can be set according to the simulation requirement, and statistics area was the outer side of coal wall in particle output direction in this study. Particles on the cleaning conveyor are colored orange for easy observation.

The process of loading coal onto the face conveyor is considered for two options: without a share and with an additional share. The following indicators were adopted as criteria for the process efficiency: productivity ($Q_к$), specific energy consumption (H_w) and secondary grinding (particle size distribution, W_{-d}).

When simulating the process of loading coal onto the downhole conveyor, the following parameter values were adopted: screw diameter - 1800 mm, hub diameter - 600 mm, screw angle of the screw - $\alpha = 19$ °, number of screw blades - 3, blade thickness - 50 mm, working width - 800 mm, auger rotation frequency - 60 rpm, conveyor bottom face height - h_k = 350 mm, conveyor pit width - 800 mm, combine harvester gearbox wall height - 350 mm, combine feed speed - 4 m / min; simulation time - t = 18 s.

Based on the simulation results of the process of loading coal onto the downhole conveyor, the dependences of the change in loading mass and the moment of resistance on the screw (Figures 4 and 5) over time are constructed.

When modeling the energy intensity of coal transportation by a screw was determined by the formula:

$$H_W = \frac{P}{Q_K} = \frac{M \cdot n}{9550 \cdot Q_K}, \text{kW·h/t} \quad (2)$$

where: M_{cp} is the average value of the torque on the screw, Nm; Q_K - coal loading capacity on the conveyor by screw, t/h; n_{ob} - rotational speed of the auger actuator, rpm.

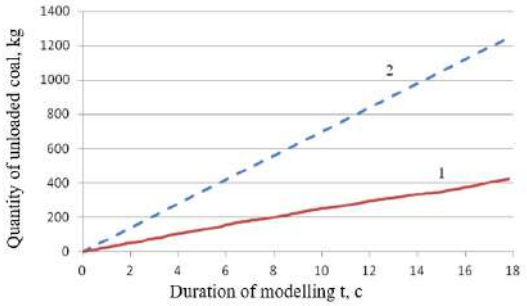


Figure 4. The dependence of the amount of loaded coal on the conveyor on time: 1. The amount of coal loaded onto the conveyor by a basic shearer;
2. The amount of coal loaded onto the conveyor by a combine with an additional share.

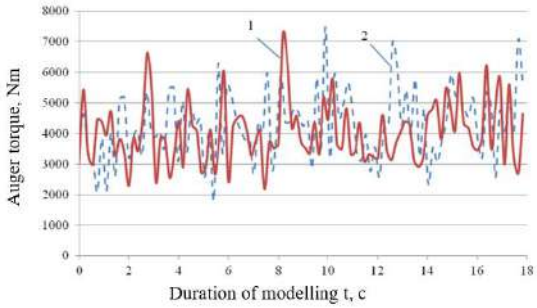


Figure 5. Change in torque on the screw: 1. Torque on the screw of the base shearer; 2. Torque on the auger of a shearer with an additional share.

The average torque on the screw is determined by the formula:

$$M_{cp} = \frac{1}{n} \cdot \sum_{i=1}^{n} M_i, \text{ Nm} \quad (3)$$

where: M_i is the current value of the torque on the screw, Nm.

The productivity of the lagging screw executive body for loading coal onto the conveyor was determined by the following expression:

$$Q_K = 3,6 \cdot \frac{m}{t}, \text{ t/h} \quad (4)$$

where m is the mass of coal loaded onto the conveyor, kg; t - loading time, s.

As a result of modeling the process, taking into account the proposed design changes of the interface unit, it was found that the productivity of loading coal by the screw executive body on the downhole conveyor increases by 2.94 times, while the specific energy consumption during loading of coal by the combine decreases by 2.7 times.

TABLE 1. The results of modeling the process of loading coal onto the face conveyor using the proposed technical solution

Index	Basic shearer	Shearer with additional share
The average torque on the screw, N·m	4075.0	4333.7
Capacity for loading coal nto a conveyor, t/h	85.8	252.0
Specific energy consumption, kW·h/t	0.25	0.09

2. 2. The Discussion of the Results As a result of the analysis of known designs of shearer units, a comprehensive solution is proposed aimed at increasing the efficiency of loading coal onto the face conveyor. This goal is achieved by the fact that in the auger actuator of the shearer containing the auger and loading guard, a share is installed between the guard and the face conveyor at an angle less than 90 degrees to its board. The share is installed above the conveyor with a gap. The lower edge of the share repeats the profile of the bead conveyor. The ploughshare is connected via an arm to the drive housing of the screw actuator. A loading plate with a ploughshare forms a flow without circulation of coal from the destruction zone to the downhole conveyor. Improving the efficiency of coal loading is achieved by increasing the completeness of soil cleaning, reducing grinding, dust formation and energy consumption.

3. CONCLUSIONS

The proposed technical solution is a lagging screw actuator with a loading shield and a share, provide:

- directivity of the flow with a decrease in resistance to the movement of coal from the destruction zone to the downhole conveyor;
- decrease in circulation and additional grinding of coal;
- increase the completeness of soil cleaning;
- reduction of dust formation and specific energy consumption.

4. REFERENCES

1. Nguyen, K. L., Gabov, V. V. and Zadkov, D. A., "Improvement of drum shearer coal loading performance", ***Eurasian mining***, No. 2, (2018), 22-25, DOI: 10.17580 / em.2018.02.06.
2. Zvonarev, I. E. and Shishlyannikov, D. I., "Efficiency improvement of loading of potassium ore by means of "Ural-20R" heading-and-winning machine", IOP Conference Series: Earth and Environmental Science, Vol. 87, (2017), 022025, DOI:10.1088/1755-1315/87/2/022025.
3. Voronova, E. Y. and Khazanovich, G. Sh., "General methodology for the optimization of the parameters of aggregated tunneling systems", IOP Conference Series: Materials Science and Engineering, Vol. 560, (2019), 012048, DOI: 10.1088 / 1757-899X / 560/1/012048.
4. Ayhan, M. and Eyyuboglu, E. M., "Comparison of globoid and cylindrical shearer drums' loading performance", ***The Journal of The South African Institute of Mining and Metallurgy***, Vol. 106, (2006), 51-56.
5. Zagrivnyi, E. A. and Basin, G. G. "External dynamics formation in mining machines", ***Journal of Mining Institute*** Vol. 217, (2016), 140-149.
6. Gabov, V. V., Zadkov, D. A. and Nguyen, K. L., "Features of Elementary Burst Formation During Cutting Coals and Isotropic Materials with Reference Cutting Tool of Mining Machines", ***Journal of Mining Institute***, Vol. 236, (2019), 153-161, DOI: 10.31897 / PMI.2019.2.153.
7. Nguyen, K. L., Gabov, V. V. and Lykov, Y.V., "Substantiation of the parameters of coal unloading process onto the conveyor using shearer drums", IOP Conference Series: Earth and Environmental Science, Vol. 194, 042019, 2018, 1-5, DOI:10.1088/1755-1315/194/4/042019.
8. Gao, K., Changlong, D., Jianghui, D. and Qingliang, Z., "Influence of the drum position parameters and the ranging arm thickness on the coal loading performance", ***Minerals***, Vol 5, 2015, 723-36, DOI:10.3390/min5040520.
9. Piotr, G., "Modeling and simulation of coal loading by cutting drum in flat seams", ***Archives of Mining Sciences***, Vol. 61, No. 2, (2016), 365–379, DOI: 10.1515/amsc-2016-0027.
10. Sriradha S., Vladislav K. and Tadeusz K., "Advance design of lacing and breakout patterns for shearer drums", ***Transactions of the Institution of Mining and Metallurgy***, Vol 114, (2005), 118-124, DOI: 10.1179/037178405X44548.
11. Zakharov, V. N., Linnik, V. Yu., Linnik, Yu. N. and Zhabin, A. B., "Mathematical apparatus to calculate shear drum parameters for specific mining conditions", ***Gornyi Zhurnal***, No. 11, 2019, 73-76, DOI: 10.17580/gzh.2019.11.13.
12. Liu, S., Du, C., Zhang, J. and Jiang, H., "Parameters analysis of shearer drum loading performance", ***Mining Science and Technology (China)***, Vol. 21, (2011), 621-24. DOI: 10.1016/j.mstc.2011.10.005
13. Hoseinie, S. H., Ghodrati, B. and Kumar, U., "Assessment of Reliability-Related Measures for Drum Shearer Machine, a Case Study", ***International Symposium High Performance Mining***, (2014), 55-62.
14. Gao, K., Liping, W., Changlong, D., Jiannan, L. and Jianghui, D., "Research on the effect of dip angle in mining direction on drum loading performance: a discrete element method", ***The International Journal of Advanced Manufacturing Technology***, Vol. 89, (2016), 23-34, DOI: 10.1007/s00170-016-9251-7.
15. Jonathon, C. R., David, C. R., Mark, T. D. and David, W. H., "Longwall automation: Delivering enabling technology to achieve safer and more productive underground mining", ***International Journal of Mining Science and Technology***, Vol. 25, (2015), 865-876, https://doi.org/10.1016/ j.ijmst.2015.09.001.
16. Seyed H. H., Mohammad A., Mohammad A. K., Behzad G. and Uday K., "Reliability analysis of drum shearer machine at mechanized longwall mines", ***Journal of Quality in Maintenance Engineering***, Vol. 18, (2012), 98-119, DOI: 10.1108/13552511211226210
17. Nguyen, V. T., Rogachev, M. K. and Aleksandrov, A. N., "A new approach to improving efficiency of gas-lift wells in the conditions of the formation of organic wax deposits in the Dragon field", ***Journal of Petroleum Exploration and Production Technology,*** Vol. 10, (2020), 3663–3672. DOI: 10.1007/s13202-020-00976-4.

Persian Abstract

چکیده

در این مقاله امکان افزایش کارایی بارگیری توده سنگ تخریب شده در نوار نقاله خراشنده توسط محرک پیچ عقب برش در فرآیند استخراج زغال سنگ در تصفیه مکانیزه پیچیده معادن ذغال مورد بررسی قرار گرفته است. در نظر گرفته شد که مهمترین تأثیر در بازده بارگیری زغال سنگ توسط ابعاد سطح مقطع پنجره بارگیری ، فاصله بین پیچ و نوار نقاله و ارتفاع پایین نوار نقاله اعمال می شود. در این مقاله راه حل های فنی غیر سنتی ارائه شده است که تأثیر منفی شکاف بین پیچ و نوار نقاله را بر روی کارایی بارگیری زغال سنگ توسط محرک پیچ عقب برش کاهش می دهد و درجه پر شدن شیار نوار نقاله را افزایش می دهد. راه حل های فنی به تشکیل یک بخش منطقی از جریان بار در دهانه نوار نقاله زیر چاهی کمک می کنند و بنابراین ، بهره وری آن را افزایش می دهند. نتایج مدلسازی فرآیند بارگیری زغال سنگ بر روی نوار نقاله توسط یک محرک اوگر با یک دستگاه بارگیری اضافی ارائه شده است. ارزیابی اثربخشی راه حل های فنی سازنده پیشنهادی برای واحد رابط در منطقه بارگیری ، افزایش ۲.۹۴ برابر حداکثر ظرفیت دستگاه اجرایی پیچ برای بارگذاری زغال سنگ بر روی نوار نقاله را تأیید میکند. در حالی که ۲.۷ برابر مصرف انرژی در هنگام بارگیری کمتر است.

AIMS AND SCOPE

The objective of the International Journal of Engineering is to provide a forum for communication of information among the world's scientific and technological community and Iranian scientists and engineers. This journal intends to be of interest and utility to researchers and practitioners in the academic, industrial and governmental sectors. All original research contributions of significant value focused on basics, applications and aspects areas of engineering discipline are welcome.

This journal is published in three quarterly transactions: Transactions A (Basics) deal with the engineering fundamentals, Transactions B (Applications) are concerned with the application of the engineering knowledge in the daily life of the human being and Transactions C (Aspects) - starting from January 2012 - emphasize on the main engineering aspects whose elaboration can yield knowledge and expertise that can equally serve all branches of engineering discipline.

This journal will publish authoritative papers on theoretical and experimental researches and advanced applications embodying the results of extensive field, plant, laboratory or theoretical investigation or new interpretations of existing problems. It may also feature - when appropriate - research notes, technical notes, state-of-the-art survey type papers, short communications, letters to the editor, meeting schedules and conference announcements. The language of publication is English. Each paper should contain an abstract both in English and in Persian. However, for the authors who are not familiar with Persian, the publisher will prepare the latter. The abstracts should not exceed 250 words.

All manuscripts will be peer-reviewed by qualified reviewers. The material should be presented clearly and concisely:

- *Full papers* must be based on completed original works of significant novelty. The papers are not strictly limited in length. However, lengthy contributions may be delayed due to limited space. It is advised to keep papers limited to 7500 words.
- *Research* notes are considered as short items that include theoretical or experimental results of immediate current interest.
- *Technical notes* are also considered as short items of enough technical acceptability with more rapid publication appeal. The length of a research or technical note is recommended not to exceed 2500 words or 4 journal pages (including figures and tables).

Review papers are only considered from highly qualified well-known authors generally assigned by the editorial board or editor in chief. Short communications and letters to the editor should contain a text of about 1000 words and whatever figures and tables that may be required to support the text. They include discussion of full papers and short items and should contribute to the original article by providing confirmation or additional interpretation. Discussion of papers will be referred to author(s) for reply and will concurrently be published with reply of author(s).

INSTRUCTIONS FOR AUTHORS

Submission of a manuscript represents that it has neither been published nor submitted for publication elsewhere and is result of research carried out by author(s). Presentation in a conference and appearance in a symposium proceeding is not considered prior publication.

Authors are required to include a list describing all the symbols and abbreviations in the paper. Use of the international system of measurement units is mandatory.

- On-line submission of manuscripts results in faster publication process and is recommended. Instructions are given in the IJE web sites: www.ije.ir-www.ijeir.info
- Hardcopy submissions must include MS Word and jpg files.
- Manuscripts should be typewritten on one side of A4 paper, double-spaced, with adequate margins.
- References should be numbered in brackets and appear in sequence through the text. List of references should be given at the end of the paper.
- Figure captions are to be indicated under the illustrations. They should sufficiently explain the figures.
- Illustrations should appear in their appropriate places in the text.
- Tables and diagrams should be submitted in a form suitable for reproduction.
- Photographs should be of high quality saved as jpg files.
- Tables, Illustrations, Figures and Diagrams will be normally printed in single column width (8cm). Exceptionally large ones may be printed across two columns (17cm).

PAGE CHARGES AND REPRINTS

The papers are strictly limited in length, maximum 6 journal pages (including figures and tables). For the additional to 6 journal pages, there will be page charges. It is advised to keep papers limited to 3500 words.

Page Charges for Papers More Than 6 Pages (Including Abstract)	
For International Author ***	**$55 / per page**
For Local Author	**100,000 Toman / per page**

AUTHOR CHECKLIST

- Author(s), bio-data including affiliation(s) and mail and e-mail addresses).
- Manuscript including abstracts, key words, illustrations, tables, figures with figure captions and list of references.
- MS Word files of the paper.

International **J**ournal of **E**ngineering, P.O. Box 13145-1659, Tehran, I.R. Iran
Materials and Energy Research Center, 5 Ahuramazda St., Alvand Ave., Argentina Sq., Tehran, Islamic Republic of Iran
www.merc.ac.ir - www.emerc.ac.ir

THOMSON REUTERS

WEB OF SCIENCE

Scopus®

CAS®
A DIVISION OF THE
AMERICAN CHEMICAL SOCIETY

Google
Scholar