Journal of Engineering Science and Technology
Vol. 15, No. 3 (2020) 1652 - 1669

GLOBAL SOLAR RADIATION PREDICTION USING A COMBINATION OF SUBTRACTIVE CLUSTERING ALGORITHM AND ADAPTIVE NEURO-FUZZY INFERENCE SYSTEM: A CASE STUDY

HISHAM S. HUSSEIN

Department of Electrical Engineering, University of Hail, Hail, Saudi Arabia
E-mail: h.hussein@uoh.edu.sa

Abstract

It is well known that the kingdom of Saudi Arabia is a vast natural potential for developing solar energy, there so solar power generation is growing rapidly. Solar energy depends on different weather and meteorological factors. Moreover, solar radiation variations throughout the year are considered an obstacle for predicting the solar energy. There so it is needed to develop predictive model that gives a precise estimation of the solar energy. The paper proposes effective model to give a reliable prediction of the global horizontal irradiance for the next day in Hail city. Estimating global horizontal irradiance has its influence on different outcomes for solar energy generation companies. It helps in reducing electricity production costs. The developed model applies subtractive clustering algorithm for generating initial model. Afterwards, it applies adaptive neuro-fuzzy inference system for tuning the parameters of the membership functions of the generated model. The simulation results reveal that subtractive clustering algorithm made initially an acceptable solution for adaptive neuro-fuzzy inference system to initiate the search with, instead of starting the search from the beginning. The performance of the final model developed by subtractive clustering algorithm and tuned by adaptive neuro-fuzzy inference system is much better than the initial model.

Keywords: Adaptive neuro-fuzzy inference system, Fuzzy inference system, Global horizontal irradiance, Solar radiation prediction, Subtractive clustering algorithm.

1. Introduction

It is well known that solar energy is considered a promising energy, as it has the capability to produce reliable, accessible and clean energy. The main disadvantages with solar energy are, that it greatly depends on different weather and geographical factors. In addition, solar radiation variations throughout the year are considered a real obstacle for predicting the solar energy in the future [1]. There so it is essential to develop a proper forecasting model that would ultimately predict global horizontal irradiance (GHI) for the next day in Hail city. Estimating global horizontal irradiance has its influence on different outcomes for solar energy generation companies. It helps in reducing electricity production costs and saves other conventional energy generation resources such as fossil fuel energy.

It also helps utility technicians and engineers working in different solar energy generation and monitoring stations to make proper decisions on solar energy generation management system, and it gives them a preliminary insight on the appropriate electricity demand capacity required from solar energy. They will have also the ability to examine the present and the future predicted solar GHI values by comparing it with the past trends of the same variable, and taking into consideration also other factors that could affect it such as weather conditions [2, 3].

Unfortunately, the prediction of solar GHI greatly depends on weather conditions, furthermore, short-term estimation of different solar radiation related variables depends greatly on daily weather conditions, such as air temperature, wind speed and direction, relative humidity, and other weather indicators [4] which therefore makes the prediction process more complicated. Different solar radiation predictive models were suggested in literature, such as, numerical weather prediction [5, 6], and artificial intelligence (AI) models [7, 8].

Fortunately, fuzzy logic has been applied for different physical systems and in many different applications [9, 10]. This is related to its simple structure and its capability of approximating and expressing human knowledge of different applications using very simple rules as would be further explained. In addition, fuzzy logic is capable to handle different kinds of nonlinearity and uncertainty of different dynamical control systems.

In this work, it is suggested to implement computational intelligence techniques. To accomplish this task, we will apply subtractive clustering algorithm for generating initial fuzzy logic systems [11, 12]. Usually, the main purpose behind fuzzy clustering algorithms is to classify separate groupings of data from a given considerably big data set. There so we can get an accurate insight on the given data. Fuzzy clustering enables us to have a minimum number of rules for a Sugeno-type or Mamdani-type fuzzy inference system (FIS) [13, 14]. Furthermore, the rules are portioned based on the fuzzy qualities related with each of the data groupings.

We will apply subtractive clustering algorithm for identifying clusters based on the input/output data given from the training data set, as will be later explained. We will then generate different FISs based on the resulting clusters' information. Finally, we will apply adaptive neuro-fuzzy inference system (ANFIS) structure [15, 16] to tune the parameters related to the membership functions (MFs) in such a way to adapt the MFs to the given input/output data for the purpose of considering any changes in the given data set.

The objective of the proposed research work is to give an acceptable prediction of the next-day GHI for Hail city, using different solar measurements and meteorological observations. Instead of going through manual design of fuzzy logic system, we propose an automatic design and generation of different FISs using a special clustering algorithm and then evaluating the generated FISs to come out with the best FIS. After choosing the best FIS, we run an optimization routine for the purpose of tuning the MFs shapes and its related parameters.

The main case study would be Hail city case study (Hail college of technology site station). We will also take a second case study, King Abdullah city for atomic and renewable energy (KACARE) City Site station to prove the reliability of the developed model. The main issue in this research is how to work with the available solar and meteorological data obtained from different resources.

Therefore, there is a deep necessity to develop a proper workflow technique that would ultimately give us a suitable solar radiation model to predict future GHI parameters. A typical workflow technique has the following steps:

i. obtain the available data from different offline and/or online resources such as databases, and spreadsheets of different solar energy generation and monitoring stations,
ii. preprocessing and data analysis techniques,
iii. and developing predictive models using different nonlinear methods based on the manipulated data from step 2.

In this paper, it is suggested to implement computational intelligence methods; more specifically, the ANFIS structure shall be implemented for the purpose of developing an appropriate predictive model of the next-day GHI. We used five data sets as input predictors to build the model [17]: air temperature (*T*) (^{o}C), Diffuse Horizontal Irradiance (DHI) (Wh/m^2), Direct Normal Irradiance (DNI) (Wh/m^2), relative humidity (RH) (%), and GHI of the current day (Wh/m^2).

The data sets for Hail city case study were collected from the solar monitoring station located in Hail college of technology and prepared by KACARE [18]. 695 days were recorded from the 6th of January- 2015 to the end of November- 2016. The target of the predicted model (model output) would be the next day of GHI (ND-GHI) (Wh/m^2). A sample of the used data is shown in Table 1, that corresponds to Hail monitoring station.

Section 2 gives an overview of subtractive clustering algorithm and ANFIS structure which were proposed for generating an initial FIS model for solar radiation GHI prediction and then further tuning its related parameters. The following sections give a detailed explanation of the related work and the adopted workflow approach for the present work.

2. The proposed Computing Algorithms

The following subsections describe in detail, the proposed artificial intelligence algorithms which are adopted for the current study.

Table 1. A sample of the used data for Hail monitoring station.

Month	*T*	DHI	DNI	GHI	RH	ND-GHI
1	14	930	7305	4477	35.7	4733
2	16.3	1144	7251	5071	33.8	4710
3	17.7	1724	6980	6110	27.9	6025
4	18.6	4125	3289	5991	25.9	7461
5	30.4	3028	5782	7553	8.3	4436
6	33.6	3672	4405	7414	13.5	4172
7	32.2	2587	7469	8342	11.1	8353
8	36.1	3600	4462	7254	8.3	7762
9	30.7	1862	7812	7390	15.7	7561
10	31.9	2624	4821	5949	12.2	6384
11	19.1	2355	2584	4025	50.5	4697
12	20.4	1128	5947	4151	34.4	2906
1	12.8	978	7231	4427	48.9	4352
2	11.3	1284	7041	5096	28.8	4202
3	15.5	955	9391	6584	37.4	6596
4	16.4	848	10852	7888	22.7	7814
5	27.6	2299	7313	7289	27.3	8141
5	27.8	3284	876	3915	24.6	6416
6	28.7	1089	11133	8945	11.7	8771
7	33.6	1420	10275	8802	10.9	8792
8	37.1	2156	7251	7630	6.9	8092
9	35.2	3099	4769	6651	6.3	6009
10	24.4	1811	7132	6415	15.3	6336
11	25.3	2730	2365	3763	14.5	5134

2.1. Subtractive clustering algorithm

The subtractive clustering algorithm is an extension of the mountain clustering algorithm presented in [11]. Mountain algorithm, gives an automatic generation of the clusters and their centre positions, respectively. It finds a potential value for each grid point in the given data space. The potential value is computed based on a function of the distance measure. Consequently, a grid point with the highest potential value would be considered as a cluster center and the remaining grid points located in the area of the first cluster center are controlled based on the emergence of new cluster centers. Unfortunately, its computation grows with the increase of the data dimensions.

Subtractive clustering [12] eliminates the restriction imposed by mountain clustering. It searches for a data point which is capable of being a cluster centre. Consequently this data point would be considered as a first cluster centre, and to determine the next cluster centre, the data points located in the area of the first cluster centre, determined by cluster radii (cluster influence range), would be subtracted. This process continues until all the data points are within the cluster radii. On the contrast to mountain clustering, Subtractive clustering is independent of the data dimensions because each data point is capable of being a potential cluster centre. The cluster radii determination is considered a very crucial issue in subtractive clustering, as it affects the generalization capabilities of the generated FIS structure during the iteration process. In general, a large radii value would give less number of clusters and hence less number of FIS rules. Vice versa, small radii value would give more clusters and consequently more FIS rules.

2.2. Adaptive neuro-fuzzy inference system structure

Because solar radiation has a lot of different dynamical variables, it is essential to implement a modeling algorithm that would considerably account for nonlinearity, seasonality, and complexity of the physical system. ANFIS structure is one of the most popular and yet powerful artificial intelligence structures that would resolve all the aforementioned technical problems of a physical system [19-23]. Usually, FIS structure is applied to systems where the rules are determined by the operator's experience of the model and its related variables. However, in ANFIS, FI is applied to separate data sets.

If there is an input-output data set pattern for a system that a model based on the features of the variables in the system is not available. Also it is not known what is the suitable shape of the MF should look like, then ANFIS structure can be implemented for this task. ANFIS would choose the parameters of the MFs in such a way that would account for any changes and variations in the current data set of the modelled system. Fortunately, ANFIS has the same working principle of neural networks (NN), it maps the given input data set through input MFs, and then through output MFs to the given output data set. It provides a certain method of fuzzy modeling procedure to learn from the given input-output data sets. So ANFIS would tune the MFs parameters that would best allow the generated FIS to track the given input-output data set. The parameters of the MFs changes automatically through the learning process.

ANFIS applies different optimization algorithms to fine-tune the MFs parameters and to improve the performance function indicator. The performance function is the root mean squared error between the real and estimated outputs. Accordingly, ANFIS implements backpropagation (gradient descent method) and least-squares estimation method for MF parameters tuning. The least-squares method identifies the resulting linear parameters using the given training data, whereas gradient descent method is used to update the premise parameters through minimizing the overall quadratic cost function. Figure 1 shows the structure of ANFIS design [24] with two inputs, two rules and one output for the first order Takagi-Sugeno model.

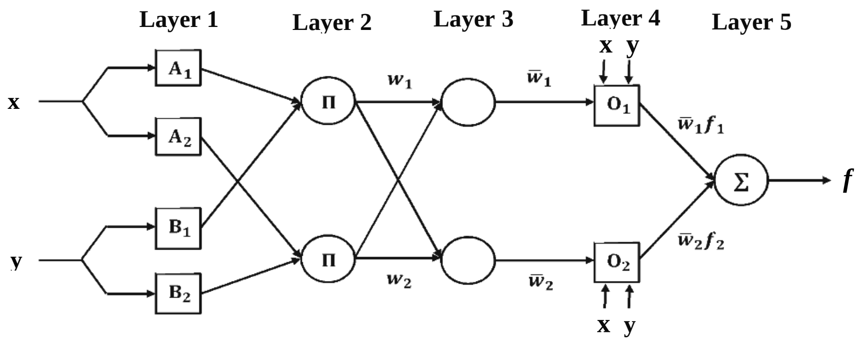

Fig. 1. Two input T-S ANFIS model structure.

Each input will have two MFs as shown in the following fuzzy rules:

Rule No. 1: if x is A_1 and y is B_1 then $z = p_1 x + q_1 y + r_1$

Rule No.2: if x is A_2 and y is B_2 then $z = p_2 x + q_2 y + r_2$

where A_1, A_2, B_1, and B_2 are linguistic values defined as fuzzy sets on the ranges x and y, respectively. The if-part of the rule "x is A" is called the antecedent or premise, while the then-part of the rule "y is B" is called the consequent or conclusion. The Sugeno output membership functions are linear. An example of these rules could be a part of the proposed case problem. We can define the following:

Air temperature and DHI could be x and y, respectively. A_1, A_2, B_1, and B_2 are the membership functions for Inputs x and y. The next-day GHI (Wh/m^2) would be the output of the predicted model, z. When z is a first-order polynomial, the resulting fuzzy inference system is called a first-order Sugeno fuzzy model, as would be given in the following sections.

The structure of ANFIS consists of different layers; the first layer is a fuzzy layer, the second is a product layer, the third is a normalized layer, the fourth is a de-fuzzified layer, and the fifth is the summation layer. We shall describe the differences between these layers and their associated nodes: The first layer which consists of square nodes formulates the MF grades for the input variables. Consequently, the parameters of the input variables are called premise parameters. The second layer multiplies the input signals received from the previous layer and calculates the firing strength for each rule. The third layer generates the normalized firing strengths by calculating the ith rule strength and dividing by the sum of the all the rules' firing strengths.

The nodes of the fourth layer multiply the normalized firing strengths with the matching rule. The parameters associated with this layer are called consequent parameters. The fifth layer which consists of one node computes the overall output by summing up all the input signals. The following section gives a detailed explanation of the adopted workflow approach for the present work.

3. The Proposed Workflow Approach

It is worth mentioning that an intensive work on solar radiation prediction is available [25-28]. However, to the best of the author knowledge, no papers have addressed the same issue for Hail city in Saudi Arabia using artificial intelligence algorithms and more specifically, subtractive clustering algorithm and ANFIS structure.

The adopted workflow approach for generating the best FIS model with optimal parameters and shapes of the generated MFs is shown below:

i. load the following input/output data into Matlab workspace: T (^{0}C), DHI (Wh/m^2), DNI (Wh/m^2), RH (%), and GHI of the current day (Wh/m^2), respectively. The ND-GHI (Wh/m^2) would be the output of the predicted model,

ii. divide the loaded data into three different kinds of data sets, training, validation and testing data sets using specified indices,

iii. run subtractive clustering algorithm to know the sensitivity of the data to subtractive clustering,

iv. iterate on running subtractive clustering algorithm until we get a suitable number of clusters,

v. generate a Sugeno type FIS model using subtractive clustering algorithm with the optimal number of clusters given from point 4.
vi. train the generated FIS with ANFIS structure using the training data set,
vii. evaluate the output of the generated FIS model from the previous step with the training data sets, testing data sets, and compare with the target variable.

Matlab & Simulink programming software (2014a) was used for implementing the suggested steps shown above, and the workflow of the data sets pre-processing technique is shown in Fig. 2. 695 data points were divided into training, validation and testing data sets using specified indices, 70% of the whole data dimensions are used for training the model, generating the initial FIS model and tuning the parameters associated with the chosen MFs. The remaining 30% of the whole data set are divided equally into validation and testing data sets [29, 30].

Consequently, 487 training data points, and 104 validation data points have been used to train and monitor the learning process of the developed ANFIS model. 104 data points were used as testing data set. The experimental results are given in the next section. Finally, section 5 gives the conclusion of the present work.

4. Results and Discussion

In the next subsections, we will demonstrate the effectiveness of the developed FIS model by Subtractive clustering algorithm and further tuned by ANFIS structure.

4.1. ANFIS training structure

To train the generated FIS model, the input/output data set for the ANFIS structure were extracted. The lowest and the largest values of the input/output data set will be applied for the divided data sets. As a result, the samples of the input/output data set obtained from the workflow shown in Fig. 2, were divided into training data set, validation data set, and testing data set for Hail city case study.

The training data set will be used by ANFIS structure for optimizing the MFs parameters and shapes associated with the assumed inputs and output of the generated FIS model using subtractive clustering algorithm. The validation data set are used to make sure that the developed ANFIS structure model does not overfit the data, i.e. the validation data set error is usually monitored during the training process of the ANFIS structure, and ultimately the ANFIS learning routine returns the tuned FIS model for which the validation data set error is at its minimum value.

The testing data set are not used during the training process, instead, they will be used to test the developed FIS models, compare between them, and choose the best FIS model with the least testing error. Eventually, testing data set are badly needed to demonstate the generalization capability and performance of the developed FIS model. The author used fuzzy logic toolbox of MATLAB software (2014a). During the modelling scenario, we used subtractive clustering algorithm with the best number of clusters determined by radii, to generate an initial Sugeno-type FIS structure with initial MFs parameters. The generalized MF type is Gaussian MF (gaussmf).

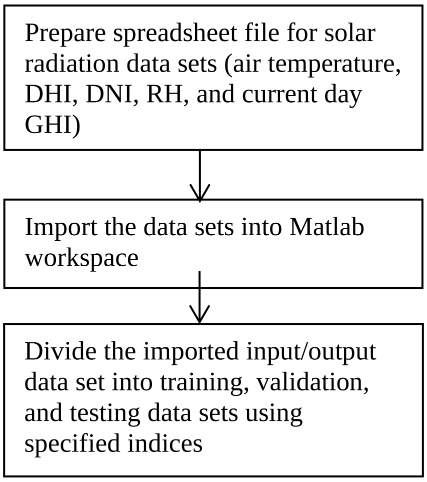

Fig. 2. The workflow of the solar radiation data pre-processing technique.

The hybrid optimization algorithm has been subsequently applied to adjust the MF parameters of the initial FIS generated by subtractive clustering algorithm. This algorithm integrates least-squares and backpropagation techniques. The training error is eventually reduced to approximately 150.024, and the training error plot is shown in Fig. 3.

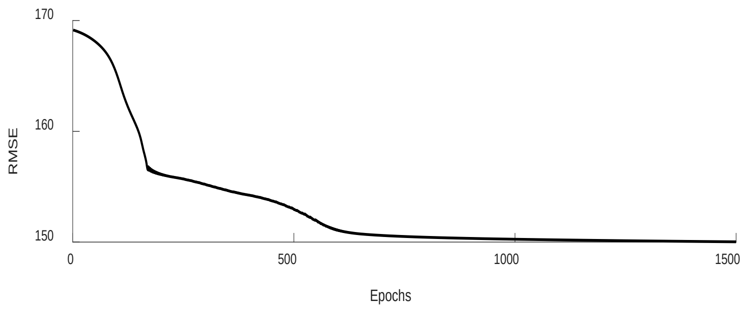

Fig. 3. Training error versus epochs for ANFIS structure (Hail case study).

The training data and FIS output are illustrated in Fig. 4. The training error shows that the FIS output goes well with the training data, even at 150.024. It is worth trying to have a separate data set (testing data set) against the FIS trained system. The testing data set would significantly improve FIS model generalization performance. The testing data and FIS output are illustrated in Fig. 5.

Here, it is illustrated that the testing data set presented to ANFIS was satisfactory to some extent, like the training data set. Figure 5 indicates that there is a very small difference between the testing data set and the FIS output. The structure of ANFIS is a five layer network. It has five inputs, one output, and seven membership functions for each input. In view of that, seven fuzzy rules were developed. Also, these rules are tuned according to the input and output mapping, in such away GHI prediction for each value of the five inputs would be obtained. The generated FIS model properties are shown in Table 2.

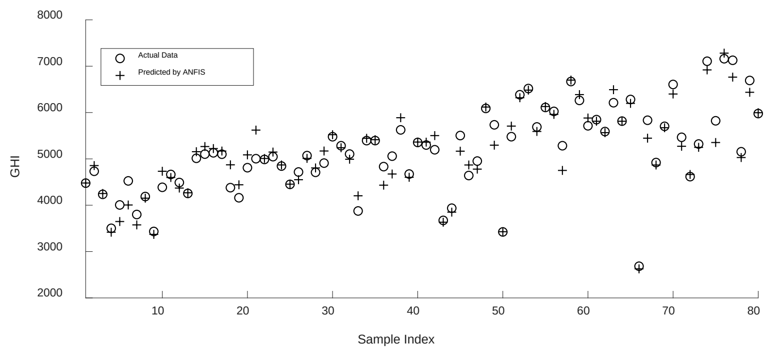

Fig. 4. Training data (o) and FIS output (+) (Hail case study).

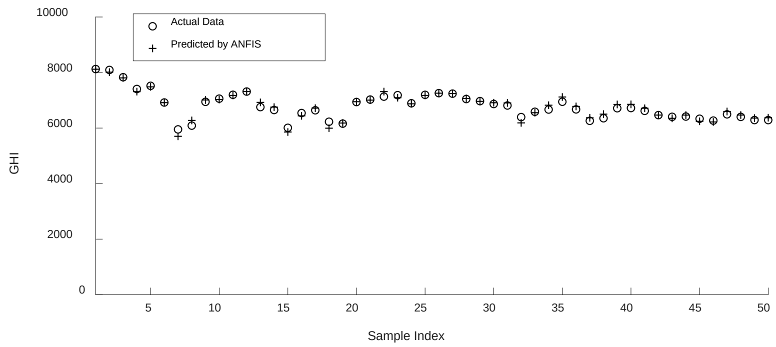

Fig. 5. Testing data (o) and FIS output (+) (Hail case study).

4.2. Generated FIS model

Table 2 shows the parameters of the generated FIS model by implementing initially subtractive clustering algorithm and further tuned by ANFIS structure. It should be noted that the generated FIS model with gaussmf for all the fuzzy sets produces much less root mean squared error (RMSE) when compared with other MF types. This is related to the properties of the gaussmf. The following equation shows the Gaussian function.

$$f(x,\sigma,c) = e^{-\frac{(x-c)^2}{2\sigma^2}} \tag{1}$$

Fortunately, there are only two parameters to be controlled, the mean, c (centre) and the standard deviation, σ: Therefore, the developed FIS model with gaussmf provides the best accuracy for GHI prediction.

As mentioned above, we used subtractive clustering algorithm to initially set the values of the centres and standard deviations for all the fuzzy sets associated with all the inputs. Eventually, it resulted in forming 7 clusters which were converted to 7 rules, and ultimately 7*6=42 fuzzy sets. A sample of the developed FIS model rules which were constructed initially by subtractive clustering and further tuned by ANFIS structure is shown in Table 3.

Table 2. Generated FIS model properties.

FIS type	Sugeno
Number of input variables	5
Number of output variables	1
Number of training data samples	487
Number of testing data samples	104
Number of validation data samples	104
Number of generated fuzzy rules	7
Membership function type	(gaussmf)

Table 3. The fuzzy inference system rules constructed initially by subtractive clustering algorithm and tuned by ANFIS structure.

1	If (in1 is in1mf1)	(in2 is in2mf1)	(in3 is in3mf1)	(in4 is in4mf1)	(in5 is in5mf1)	Then (GHI is out1cluster1)
2	If (in1 is in1mf2)	(in2 is in2mf2)	(in2 is in3mf2)	(in4 is in4mf2)	(in5 is in5mf2)	Then (GHI is out1cluster2)
3	If (in1 is in1mf3)	(in2 is in2mf3)	(in2 is in3mf3)	(in4 is in4mf3)	(in5 is in5mf3)	Then (GHI is out1cluster3)
4	If (in1 is in1mf4)	(in2 is in2mf4)	(in2 is in3mf4)	(in10 is in4mf4)	(in5 is in5mf4)	Then (GHI is out1cluster4)
5	If (in1 is in1mf5)	(in2 is in2mf5)	(in2 is in3mf5)	(in10 is in4mf5)	(in5 is in5mf5)	Then (GHI is out1cluster5)
6	If (in1 is in1mf6)	(in2 is in2mf6)	(in2 is in3mf6)	(in10 is in4mf6)	(in5 is in5mf6)	Then (GHI is out1cluster6)
7	If (in1 is in1mf7)	(in2 is in2mf7)	(in2 is in3mf7)	(in10 is in4mf7)	(in5 is in5mf7)	Then (GHI is out1cluster7)

The FIS model was designed by applying the fuzzification, implication and de-fuzzification processes of a typical fuzzy logic Sugeno model. The fuzzification and implication functions are both specified as the product of fuzzified input values. The aggregation process is specified as the sum of premise fuzzy sets operation while the weighted average of all rule outputs process is chosen for the de-fuzzification process. The developed FIS model rules are illustrated in Table 3 shown, and the predicted errors for the training and testing samples using the developed FIS are respectively evaluated.

As mentioned above, we have used training data for the sake of optimizing the generated FIS model using ANFIS structure. In order to demonstrate the generalization capability and performance of the ANFIS structure, we investigated the predicted GHI using new data, i.e., testing data sets. We have used the same inputs as training data set. Table 4 shows a summary of the actual data which were used as training and testing data sets.

Table 4. Actual GHI data summary (Hail city case study).

Data	$GHI_{average}$	GHI_{min}	GHI_{max}	GHI_{std}
Training Data	5963.8	339	8804	1676.7
Testing Data	5762.6	549	8126	1295.7

The estimation performance of the adopted model was evaluated using RMSE performance function. The formula of the RMSE performance function is shown below,

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i^a - y_i^p)^2} \quad (2)$$

where N is the number of the training, validation, or testing values. y_i^a, and y_i^p are the actual, and predicted values, respectively. The RMSE results are shown in Table 5, and the performance for the training and testing parts are shown in Figs 6 and 7.

Table 5. Experimental Results using subtractive clustering algorithm and ANFIS structure in training and testing phases (Hail city case study).

Runs	Data set	$RMSE_{Average}$
FLS by SUB CLUST	Training	179.15
FLS by ANFIS (SUBCLUST)	Training	150.024
FLS by SUBCLUST	Testing	167.37
FLS by ANFIS (SUBCLUST)	Testing	155.6

Figure 6 shows a sample plot of the actual GHI of the training data set and the predicted GHI by the proposed model. The training results proved that ANFIS has performed much better than subtractive clustering since it reduces the training RMSE from 179.15 to 150.024.

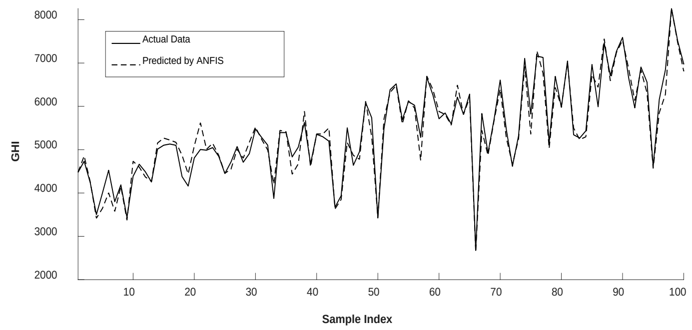

Fig. 6. Actual training data set and predicted GHI (Hail city).

The FIS generated by subtractive clustering reached an average RMSE of 179.15 for the training set samples which represent 3% from the average of the real values. While, The FIS generated by subtractive clustering algorithm and fine-tuned by ANFIS achieved an average RMSE of 150.024 for the same training set which represent 2.52% from the average of the real values Therefore, subtractive clustering managed to reach an average prediction accuracy of 97%, while the FIS generated by subtractive clustering algorithm and fine-tuned by ANFIS has managed to get 97.48% prediction accuracy. It is worth noting that subtractive clustering made a preliminary acceptable result for ANFIS structure to begin the search with instead of starting the search from nothing.

Figure 7 shows a sample plot of the predicted GHI and the actual GHI during the testing phase. The same results were observed as with the training results. The testing results once again proved that ANFIS has performed significantly better since it reduces the testing RMSE from 167.37 to 155.6.

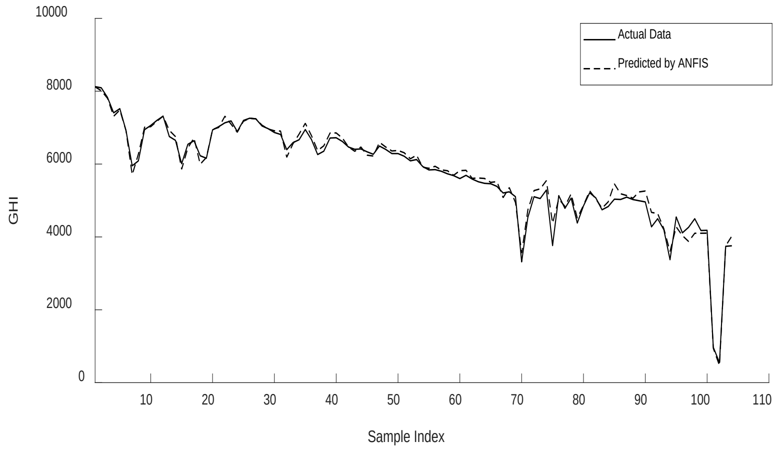

Fig. 7. Actual and predicted GHI for the testing data set (Hail city)

The FIS generated by subtractive clustering achieved an average RMSE of 167.37 for the testing set which represent 2.9% from the average of the real values. While, The FIS generated by subtractive clustering algorithm and fine-tuned by ANFIS, attained an average RMSE of 155.6 for the same testing set which represent 2.7% from the average of the real values. Therefore, subtractive clustering has managed to obtain an average estimation accuracy of 97.1%, while the FIS generated by subtractive clustering algorithm and fine-tuned by ANFIS has achieved an average estimation accuracy of 97.3%. It is worth noting that, we can reduce the error by further ANFIS training as long as the error decreases. We can also do the following steps, if we desire to reduce the error further:

i. adjust the design of ANFIS structure as more training may not help, by trying different types of MF,

ii. add or reduce the number of MFs per input,

iii. and design different linear Sugeno models, considering that linear models tend to give less error and much better performance measure.

Figure 7 shown above, proves that the difference between the actual data and the one predicted by ANFIS is very small, which again demonstrates the effectiveness of the developed ANFIS model. As mentioned before, the validation set error was monitored during the training process of the generated FIS model. Fortunately, it was decreasing along with the training set error which proves that the FIS model has not over-fitted the data.

For further investigation of the developed model, we provide another case study to prove the reliability of the developed model. The data sets for KACARE city site were collected and prepared by KACARE city [18]. 56 months were recorded from February- 2013 to December- 2017. The target of the predicted model (model output) would be the next month of GHI (NM-GHI) (Wh/m^2). A sample of the used data is shown in Table 6. It corresponds to KACARE city site. We have used the same FIS model generated before for Hail city case study, and the FIS model properties are all the same.

Table 6. A sample of the used data for KACARE city site station.

DHI	DNI	GHI	RH	NM-GHI
1760.3	6730	5783.7	26.3	6258.3
3014.2	6355.7	7995.3	9.4	7701.6
1705.1	6753.8	5975.7	16.8	4506.5
1440.6	5319.8	4176.7	39.7	4226.3
3225.5	4966.2	7029.4	19.5	7554.9
2906.6	6854.5	8202.5	9.3	8100.2
2206.3	7298.5	7640.6	12.2	6714.5
1182.3	6614.9	4512.5	43.4	4918.4
2095	5766	5605.2	27.3	5900
3990.5	3705.9	6902.8	16.3	7273.9
2764.3	6218.2	7503.8	11.2	6940.3
2412.2	5045.7	5751	22.5	4579.2
1649.5	7044	5761.5	33	5999.7
3412.4	5626.8	7803.9	17.5	8179.8
2546.6	7320.3	8131	12.2	7465.4

Table 7 shows a summary of the actual data which were used as training and testing data sets.

Table 7. Actual GHI data summary (KACARE city site).

Data	$GHI_{average}$	GHI_{min}	GHI_{max}	GHI_{std}
Training Data	6477.1	4176.7	8202.5	1276.9
Testing Data	6493.8	4150.4	8179.8	1344.5

The RMSE results are shown in Table 8, and the performance for the training and testing parts are shown in Figs 8 and 9.

Table 8. Experimental Results using subtractive clustering and ANFIS structure in training and testing phases (KACARE city site).

Runs	Data set	$RMSE_{Average}$
FLS by SUB CLUST	Training	79
FLS by ANFIS (SUBCLUST)	Training	36.55
FLS by SUBCLUST	Testing	187.34
FLS by ANFIS (SUBCLUST)	Testing	79.21

Figure 8 shows a sample plot of the actual GHI of the training data set and the predicted GHI by the proposed model for KACARE City Site monthly data. The training results proved that ANFIS has performed once again much better than subtractive clustering, since it reduces the training RMSE from 79 to 36.55. The FIS generated by subtractive clustering reached an average RMSE of 79 for the training set samples which represent 1.22% from the average of the real values.

While, The FIS generated by subtractive clustering algorithm and fine-tuned by ANFIS achieved an average RMSE of 36.55 for the same training set which represent 0.56% from the average of the real values Therefore, subtractive clustering managed to reach an average prediction accuracy of 98.8%, while the

FIS generated by subtractive clustering algorithm and fine-tuned by ANFIS has managed to get 99.44% prediction accuracy.

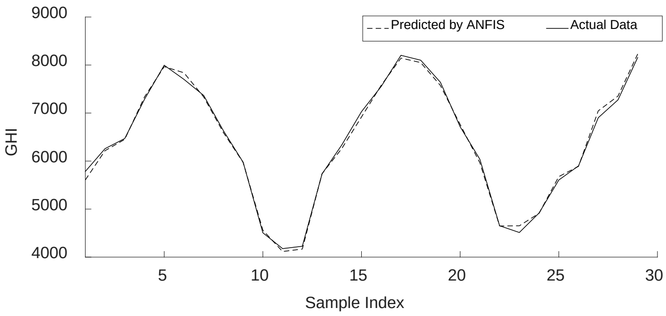

Fig. 8. Actual training data set and predicted GHI (KACARE site).

Figure 9 shows a sample plot of the predicted GHI and the actual GHI during the testing phase. The same results were observed as with the training results. The testing results once again proved that ANFIS has performed significantly better since it reduces the testing RMSE from 187.34 to 79.21. The FIS generated by subtractive clustering achieved an average RMSE of 187.34 for the testing set which represent 2.88% from the average of the real values.

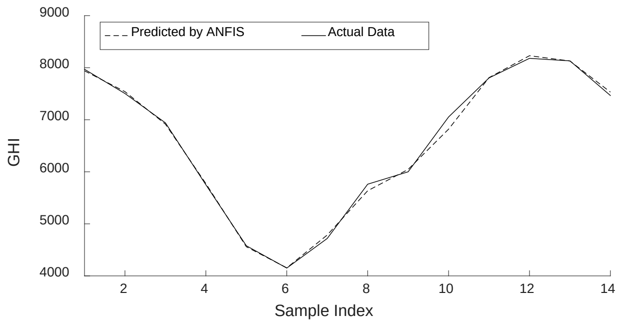

Fig. 9. Actual testing data set and predicted GHI (KACARE site).

While, The FIS generated by subtractive clustering algorithm and fine-tuned by ANFIS, attained an average RMSE of 79.21 for the same testing set which represent 1.22% from the average of the real values. Therefore, subtractive clustering has managed to obtain an average estimation accuracy of 97.12%, while the FIS generated by subtractive clustering algorithm and fine-tuned by ANFIS has achieved an average estimation accuracy of 98.78%.

Table 9 compares between the developed method with other adopted methods from literature. The results reveal that the proposed model of the generated FIS by subtractive clustering algorithm and optimized by ANFIS structure reached a better performance when compared with other models from literature, such as conventional neural network models, Bayesian Neural Networks models, Different empirical models and other proposed methods from literature.

Table 9. Comparison of the proposed method (Hail city case study) with other methods from the literature.

METHODS	RMSE	NRMSE
Neural Networks [31]	6.002%	-
Bayesian Neural Networks [31]	7.5% −15%	-
Empirical models [32]	7.49−9.23%	-
Nonlinear model [32]	8.73%	-
Fuzzy logic model [32]	8.80%	-
Fuzzy C-means and simulated annealing [33]	11.78%	-
Exponential smoothing [34]	357.55	-
ARIMA-GARCH [34]	351.14	-
Two stage model [35]	-	0.048
Subtractive clustering and ANFIS (proposed method)	155.6 (2.7%)	0.0205

The obtained RMSE is quite low relative to other works [31-35]. This is related to the improved generalization capability, performance and faster learning speed of the generated ANFIS model.

5. Conclusions

In this paper, a new solar radiation model has been developed using artificial intelligence methods, more specifically, subtractive clustering algorithm and ANFIS structure. To validate the newly developed FIS model, a separate data set were tested for two site stations in the kingdom of Saudi Arabia, namely, Hail college of technology Station and KACARE city site station, and fortunately, the generated FIS model by subtractive clustering algorithm and fine-tuned by ANFIS structure has proved its effectiveness as there was a very small difference between the testing data output and FIS output.

It is well known that KACARE and other local solar energy generation and monitoring stations rely on accurate solar radiation prediction to optimize operational efficiency and reliability. There so, the proposed model which has been developed using fuzzy logic system, fuzzy clustering, and neuro-fuzzy inference system structure would be of great benefit to these companies. In addition, this work will serve as a benchmark for scientists, graduates and undergraduate students who are involved in applying fuzzy logic, neural networks, fuzzy clustering algorithms, and adaptive neuro-fuzzy inference system structure in their research projects.

Nomenclatures

c	Mean value for the Gaussian function
N	The number of training, validation, or testing values
x	Gaussian membership function Input values
y_i^a	Actual training, validation, or testing values
y_i^p	Predicted training, validation, or testing values

Greek Symbols

σ	Standard deviation for the Gaussian function

Abbreviations

AI	Artificial Intelligence
ANFIS	Adaptive Neuro-Fuzzy Inference System
DHI	Diffuse Horizontal Irradiance
DNI	Direct Normal Irradiance
FIS	Fuzzy Inference System
GARCH	Generalized conditional heteroskedasticity autoregressive models
gaussmf	Gaussian Membership Function
GHI	Global Horizontal Irradiance
KACARE	King Abdullah City for Atomic And Renewable Energy
MFs	Membership Functions
NN	Neural Networks
NRMSE	Normalized Root Mean Square Error
RH	Relative Humidity
RMSE	Root Mean Squared Error
TWO STAGE MODEL	Two cascaded stages based on nonlinear autoregressive neural Network and the autoregressive moving average with exogenous input

References

1. Ehsanul, K.; Pawan, K.; Sandeep, K.; Adedeji, A.A.; and Ki-Hyun, K. (2017). Solar energy: Potential and future prospects. *Renewable and Sustainable Energy Reviews*, 82(1), 894-900.
2. Sen, Z. (2004). Solar energy in progress and future research trends. *Progress in Energy and Combustion Science*, 30(4), 367-416.
3. Muneer, T.; Younes, S.; and Munawwar, S. (2007). Discourses on solar radiation modeling. *Renewable and Sustainable Energy Reviews*, 11(4), 551-602.
4. Sen, Z. (2008). *Solar energy fundamentals and modeling techniques: atmosphere, environment, climate change and renewable energy* (1st ed.). Springer-Verlag London: Springer Science & Business Media.
5. Muneer, T.; Younes, S; and Munawwar, S. (2007). Discourses on solar radiation modeling. *Renewable and Sustainable Energy Reviews*, 11(4), 551-602.
6. Voyant, C.; Muselli, M.; Paoli, C.; and Nivet, M.L. (2012). Numerical weather prediction (NWP) and hybrid ARMA/ANN model to predict global radiation. *Energy*, 39(1), 341-355.
7. Lam, J.C.; Wan, K.K.W.; and Yang, L. (2008). Solar radiation modelling using ANNs for different climates in China. *Energy Conversion and Management*, 49(5), 1080-1090.
8. Zeng, J; Qiao, W. (2011). Short-term solar power prediction using an RBF neural network. *Power and Energy Society General Meeting, IEEE*. Detroit, MI, USA, 1-8.
9. Jafarzadeh, S.; Etezadi-Amoli, M.; and Fadali, M.S.(2012). Fuzzy type-1 and type-2 Takagi-Sugeno-Kang modeling with application to solar power

prediction. IEEE *Power and Energy Society General Meeting*. San Diego, CA, USA, 1-6.

10. Mendel, J.M. (2001). *Uncertain rule-based fuzzy logic systems: introduction and new directions* (1st ed.). Los Angeles: Prentice Hall.
11. Yager, R.; and D. Filev. (1994). Generation of Fuzzy Rules by Mountain Clustering. *Intelligent and Fuzzy Systems Journal*, 2(3), 209-219.
12. Chiu, S. (1994). Fuzzy Model Identification Based on Cluster Estimation. *Journal of Intelligent & Fuzzy Systems*, 2(3), 267-278.
13. Mamdani, E.H.; and Assilian, S. (1975). An experiment in linguistic synthesis with a fuzzy logic controller. *Man-Machine Studies, International Journal*, 7(1), 1-13.
14. Sugeno, M. (1985). *Industrial applications of fuzzy control* (1st ed.). Amsterdam: Elsevier Science publication Inc.
15. Jang, J.-S.R. (1993). ANFIS: Adaptive-network-based Fuzzy Inference Systems. *IEEE Transactions on Systems, Man, and Cybernetics*, 23(3), 665-685.
16. Jang, J.-S.R. (1991). Fuzzy modeling using generalized neural networks and Kalman filter algorithm. *Proceedings of the Ninth National Conference on Artificial Intelligence, AAAI*. Anaheim, CA, USA, 762-767.
17. Khatib, T.; Mohamed A.; and Sopian, K.A. (2012). Review of solar energy modeling techniques. *Renewable and Sustainable Energy Reviews*, 16(5), 2864-2869.
18. King Abdullah City for Atomic and Renewable Energy (KACARE) (2018). Hail College of Technology Station. Solar Resource reports for King Abdullah City for Atomic and Renewable Energy: as part of the Renewable Resource Monitoring and Mapping (RRMM) Program. Retrieved August 10, 2018, from http://www.kacare.gov.sa.
19. Ali, H.; and Seyede, S.H. (2016). Ranking bank customers using Neuro-Fuzzy network and optimization algorithms. *International Journal of Advanced and Applied Sciences*, 3(2), 40-44.
20. Hussein, H.; Aloui, A.; and AlShammari, B. (2018). ANFIS-based PI controller for maximum power point tracking in PV systems. *International Journal of Advanced and Applied Sciences*, 5(2), 90-96.
21. Hussein, H. (2018). An Optimal Design Methodology of ANFIS System for energy load forecasting - Hail city case study (Saudi Arabia). *Fourth International Conference on Engineering & MIS. ACM*. Istanbul, Turkey. 1-7.
22. Guesmi, T. (2017). Neuro-fuzzy and nondominated sorting genetic algorithm based power system stabilizer design. *Int. Journal of advanced and applied sciences*, 4(6), 19-27.
23. Haris, N.A.; and Rahman, F.A. (2017). A study on application of fuzzy methods in entrepreneurship domain. *Int. Journal of advanced and applied sciences*, 4(12), 206-211.
24. Mollaiy-Berneti, S. (2016). Optimal design of adaptive neuro-fuzzy inference system using genetic algorithm for electricity demand forecasting in Iranian industry. *Soft Comput.* 20(12), 4897-4906.
25. Ahmad, M.J.; and Tiwari, G. (2011). Solar radiation models - A review. *International Journal of Energy Research*, 5(4), 271-290.

26. Zell, E.; Gasim, S.; Wilcox, S.; Katamoura, S.; Stoffel T.; Shibli, H.; Engel-Cox, J.; and Al Subie, M. (2015). Assessment of solar radiation resources in Saudi Arabia. *Solar Energy*, 119, 422-438.
27. Hedar, A.R.; Abdel-Hakim, A.E.; and Almaraashi, M. (2016). Granular-Based Dimension Reduction for Solar Radiation Prediction Using Adaptive Memory Programming. *Genetic and Evolutionary Computation Conference Companion*. *ACM*. Denver, CO, USA, 929-936.
28. Hepbasli, A.; and Alsuhaibani, Z.A. (2011). Key review on present status and future directions of solar energy studies and applications in Saudi Arabia. *Renewable and Sustainable Energy Reviews*, 15(9), 5021-5050.
29. Jang, J.-S.R.; Sun, C.-T.; and Mizutani, E. (1997). *Neuro-Fuzzy and Soft Computing: A Computational Approach to Learning and Machine Intelligence* (1st ed.). USA: Prentice Hall.
30. Wang, L.X. (1994). *Adaptive fuzzy systems and control: design and stability analysis* (1st ed.). New Jersey: Prentice Hall.
31. Yacef, R.; Benghanem, M.; and Mellit, A. (2012). Prediction of daily global solar irradiation data using Bayesian neural network: A comparative study. *Renewable Energy*, 48,146-154.
32. Khatib, T.; Mohamed, A; Mahmoud, M.; and Sopian, K. (2011). Modeling of Daily Solar Energy on a Horizontal Surface for Five Main Sites in Malaysia. *International Journal of Green Energy*, 8(8), 795-819.
33. Almaraashi, M. (2017). Short-term prediction of solar energy in Saudi Arabia using automated-design fuzzy logic systems. *PLOS one*, 12(8), 1-16.
34. Ettayyebi, H.; El Himdi, K. (2018). Artificial Neural Network for Forecasting One Day Ahead of Global Solar Irradiance. The second International Conference on smart applications and data analysis for smart cities (SADASC'18). Casablanca, Morocco, 1-10.
35. Alanazi, M.; Mahoor, M., Khodaei, A. (2017). Two-stage hybrid day-ahead solar forecasting. *North American power symposium (NAPS)*. Morgantown, USA, 1-6.