AIChE

Membership Function Comparative Investigation on Productivity Forecasting of Solar Still Using Adaptive Neuro-Fuzzy Inference System Approach

Ahmed F. Mashaly[a] and **A. A. Alazba**[a,b]

[a]Alamoudi Chair for Water Researches, King Saud University, Riyadh, Saudi Arabia; mashaly.ahmed@gmail.com (for correspondence)

[b]Department of Agricultural Engineering, King Saud University, Riyadh, Saudi Arabia

Published online 00 Month 2017 in Wiley Online Library (wileyonlinelibrary.com). DOI 10.1002/ep.12664

Modeling solar still productivity (SSP) is one of the most studied topics in solar desalination due to it having essential applications in the design of solar still systems. This study applied an adaptive neuro-fuzzy inference system (ANFIS) and different membership functions (MFs) to predict the SSP required by designers, operators, and beneficiaries of solar stills. The output of this research can be used as a reference for designing and managing solar stills that could lead to optimizing the performance. The modeling process was based on real-field experimental data. The model considers the solar radiation, relative humidity, total dissolved solids of the feed, total dissolved solids of the brine, and feed flow rate as the input variables. The results show that ANFIS forecasting with generalized bell MF (GBELLMF) produced the highest correlation coefficient (CC) and the smallest root mean square error (RMSE) when compared with other MF types. Thus, the ANFIS model with GBELLMF (CC = 0.99; RMSE = 0.03 L/m^2/h) provides the best SSP prediction accuracy, which is better than other models with MFs. In addition, the statistical indicators demonstrate that the ANFIS model is better for predicting the SSP than multiple linear regressions. These findings demonstrate that ANFIS can be applied to forecast the SSP using weather and operational data as inputs with the best membership function (which is GBELLMF). © 2017 American Institute of Chemical Engineers Environ Prog, 00: 000–000, 2017

Keywords: adaptive neuro-fuzzy inference system, membership function, solar desalination, solar still, modeling

INTRODUCTION

Solar still is a very simple solar device used for converting seawater, brackish water, or wastewater into fresh water and can be fabricated easily with locally available materials. The maintenance is also inexpensive, and no skilled labor is required. Besides, it can be a suitable solution to solve drinking and irrigation water problem [1]. However, it is not popularly used because of its low productivity [2]. In other words, the main problem of the solar desalination using solar stills is their low productivity. Usually, a solar still can produce 2.5–5 L/m^2/day of distillate water [3,4].

However, even though solar stills have low productivities, they are being a sustainable water production technique. Solar stills continue to attract wide research attention that is targeted to increase their productivity. Many experimental and theoretical investigations have been done by many researchers to increase the distillate water by various design parameters like different shapes [5], different materials [6], different depths of water [7], and heat absorbers [8]. Other investigators worked on solar stills integrated with solar heaters [9], solar collectors [10], and solar concentrators [11]. Others researchers used multiple-effect solar stills [12]. While previous design, testing, and experimental work have emphasized methods to increase the productivity of the still, there is still a need to develop a new predictive model that would be able to accurately predict the productivity [13].

Soft computing techniques involving artificial neural network (ANN), fuzzy logic (FL), and adaptive neuro-fuzzy inference system (ANFIS) are widely used for predicting purposes. Comparative studies among ANN, FL, and ANFIS techniques show that the use of ANFIS produces a lower error rate than ANN and FL [14–17]. ANFIS has been used for a wide variety of applications. In the solar field alone, they have been used to model the performance of solar thermal energy system [14], model the solar PV module and maximum power point tracking [18], predict the performance of solar chimney power plants [19], determine the most important variables for diffuse solar radiation [20], and estimate photovoltaic power supply [21].

Generally, predicting is a process that is done by knowing the pattern data based on the existing data. One technique that can be used is ANFIS (adaptive neuro-fuzzy inference system). ANFIS has the ability of ANN and FL. However, there are studies documented in the literature about the application of ANN and FL techniques for solar still performance modeling [22–30]. On the contrary, to the best knowledge of the authors, there is no study dedicated to the investigation of solar still productivity (SSP) modeling and also applying the ANFIS modeling tool. However, the problem of ANFIS that must be considered is the training, testing, and validation data for updating the membership function parameter.

In this research, we tried to find the best type of membership function that has the smallest error between the

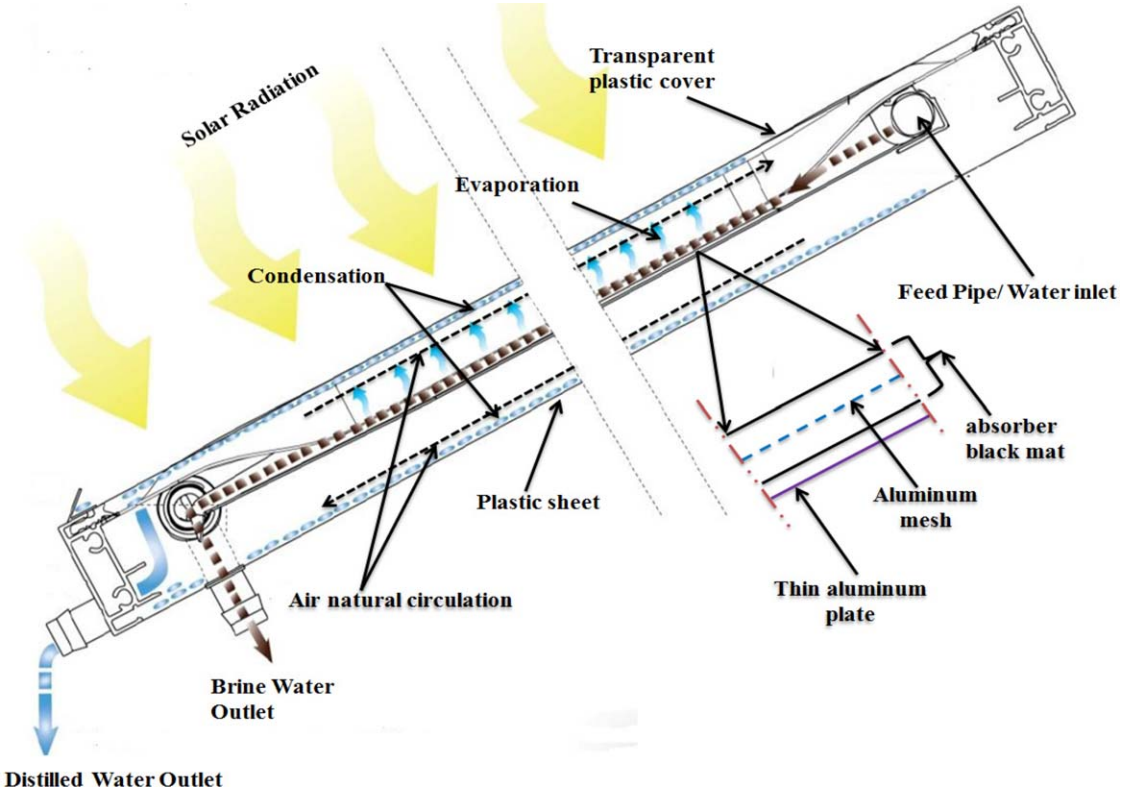

Figure 1. Cross-sectional view of the solar still. [Color figure can be viewed at **wileyonlinelibrary.com**]

observed data and predicted data. Therefore, the objectives of this study were to (1) develop ANFIS models to estimate SSP (solar still productivity) using the ANFIS method and different membership functions; (2) evaluate the performance of the developed ANFIS models using a statistical comparison between the SSP results obtained from the models and observed experimental results; (3) compare the accuracy of the ANFIS models with different membership functions; and (4) compare the selected/best ANFIS model with a multiple linear regression (MLR) model in terms of their suitability for predicting the SSP.

MATERIALS AND METHODS

Experimental Study

The experiments were conducted at the Agricultural Research and Experiment Station—at the Department of Agricultural Engineering, King Saud University, Riyadh, Saudi Arabia (24°44′10.90″N, 46°37′13.77″E) between February and April 2013. The weather data were obtained from a weather station (model: Vantage Pro2; manufacturer: Davis) close by the experimental site (24°44′12.15″N, 46°37′14.97″E). The solar-still system used in the experiments was constructed from a 6 m^2 single stage of C6000 panel (F Cubed, Carocell Solar Panel, Australia). The solar-still panel manufactured using modern, cost-effective materials such as coated polycarbonate plastic. When heated, the panel distilled a film of water that flowed over the absorber mat of the panel. The panel was fixed at angle of 29° to horizontal. The basic construction materials were galvanized steel legs, an aluminum frame, and polycarbonate covers. The transparent polycarbonate was coated on the inside with a special material to prevent fogging (patented by F Cubed, Australia). Cross-sectional view of the solar still is illustrated in Figure 1.

The water was fed to the panel using centrifugal pump (model: PKm 60, 0.5 HP, Pedrollo, Italy) with a constant flow rate was 10.74 L/h. The feed was supplied by eight drippers/nozzles, creating a film of water that flowed over the absorbent mat. Underneath the absorbent mat was an aluminum screen that helped to distribute the water across the mat. Beneath the aluminum screen was an aluminum plate. Aluminum was chosen for its hydrophilic properties, to assist in the even distribution of the sprayed water. Water flowed through and over the absorbent mat, and solar energy was absorbed and partially collected inside the panel; as a result, the water was heated and hot air circulated naturally within the panel. First, the hot air flowed toward the top of the panel, then reversed its direction to approach the bottom of the panel. During this process of circulation, the humid air touches the cooled surfaces of the transparent polycarbonate cover and the bottom polycarbonate layer, causing condensation. The condensed water flowed down the panel and was collected in the form of a distilled stream. Seawater was used as a feed-water input to the system. The system was run from 02/23/2013 to 04/23/2013. Raw seawater was obtained from the Gulf, Dammam, in eastern Saudi Arabia (26°26′24.19″N, 50°10′20.38″E). The initial concentrations of the total dissolved solids (TDS), pH, density (ρ), and electrical conductivity (EC) are 41.4 ppt, 8.02, 1.04 g cm^{-3}, and 66.34 mS cm^{-1}, respectively. The solar still productivity or the amount of distilled water produced (SSP) by the system in a given time, was obtained by collecting and measuring the amount of water cumulatively produced over time. The temperature of the feed water (T_F) was measured by using thermocouples (T-type, UK). Temperature's data for feed brine water was recorded on a data logger (model: 177-T4, Testo, UK) at 1 min intervals. The amount of feed water (M_F) was measured by calibrated digital flow meter was mounted on the feed water line (micro-flo, Blue-White). The amount of brine water and distilled water were measured by a graduated cylinder. TDS concentration and EC were measured using a TDS-calibrated meter (Cole-Parmer Instrument,

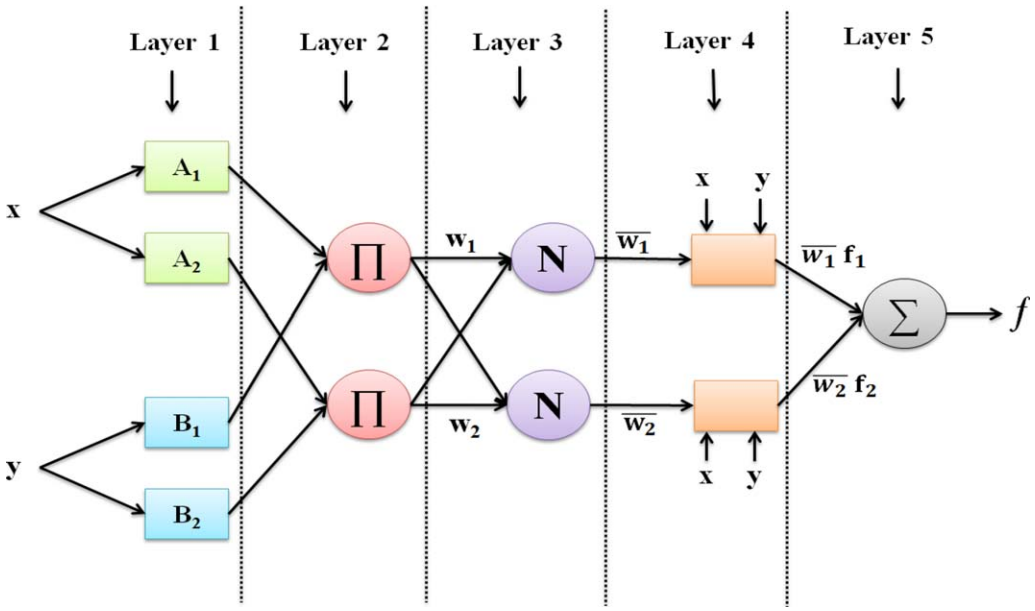

Figure 2. General ANFIS architecture of two-input first-order Sugeno-type model with two rules. [Color figure can be viewed at **wileyonlinelibrary.com**]

Vernon Hills). A pH meter (model: 3510 pH meter; Jenway, UK) was used to measure pH. A digital-density meter (model: DMA 35_N, Anton Paar) was used to measure ρ. The seawater was fed separately to the panel using the pump described above. The residence time—the time taken for the water to pass through takes the still/panel—was approximately 20 min. Therefore, the flow rate for the feed water, the distilled water and the brine water was measured each 20 min. Also, the total dissolved solids of feed water (TDS_F) were measured every 20 min. The weather data, such as ambient temperature (T_o), relative humidity (RH), wind speed (WS), and solar radiation (SR), were obtained from a weather station near the experimental site. In the current study, there is one dependent variable (SSP) and seven independent variables (T_o, RH, WS, SR, TDS_F, M_F, and T_F).

Adaptive Neuro-Fuzzy Inference System (ANFIS)

The adaptive neuro-fuzzy inference system (ANFIS), which incorporates the best features of fuzzy logic (FL) and of the artificial neural network (ANN) systems, is defined by Jang [31]. In terms of architecture, the ANFIS is composed of "if-else" rules and "input–output" data couples of fuzzy numbers and it uses the neural network's learning algorithms for training. Additionally, the ANFIS is an approach to simulating complex nonlinear mappings using the neural network learning and fuzzy inference methodologies and has the capability of working with uncertain noisy and inaccurate environments. The ANFIS utilizes the ANN training process in order to adjust the membership function and the associated parameter that approaches the desired data sets. The learning algorithm of the ANFIS is a hybrid learning algorithm comprising the use of the "back-propagation learning algorithm" and "least squares method" together [32–35]. In order to understand and simplify the process, a sample including two inputs and an output is considered. Five layers are used in order to build the ANFIS architecture of the first-order Sugeno-type inference system, as presented in Figure 2. Two inputs, x and y, and one output, f, along with two fuzzy "if-then" rules are taken into account as an example. In Figure 2 the circle and square show a fixed node and an adaptive node, respectively. The function of each layer of the five layers was explained in the following sections. For a first-order Sugeno fuzzy model, two fuzzy "if-then" rules are formulated as follows:

$$\text{Rule } 1: \text{ If } x \text{ is } A1 \text{ and } y \text{ is } B1, \text{ then } f1 = p1x + q1y + r1, \tag{1}$$

$$\text{Rule } 2: \text{ If } x \text{ is } A2 \text{ and } y \text{ is } B2, \text{ then } f2 = p2x + q2y + r2, \tag{2}$$

where x and y are the inputs and A_1, B_1, A_2, and B_2 are the fuzzy sets p_1, p_2, q_1, q_2, r_1, and r_2 are the coefficients of the output function that are determined during the training.

Layer 1 is the fuzzification layer (layer of input nodes). Every node i is an adaptive node with a node output expressed by:

$$O_i^1 = \mu_{A_i}(x), \text{ for } i = 1, 2, \tag{3}$$

$$O_i^1 = \mu_{B_{i-2}}(y), \text{ for } i = 3, 4, \tag{4}$$

where μ_{A_i} and $\mu_{B_{i-2}}$ are the fuzzy membership functions.

Layer 2 is the rule layer (layer of rule nodes). Every node i in this layer is a fixed node, marked by a circle and labeled Π, representing the simple multiplication. The output of this layer is the product of all the incoming signals and can be formulated as:

$$O_i^2 = w_i = \mu_{A_i}(x) \quad \mu_{B_i}(y) \text{ for } i = 1, 2. \tag{5}$$

Layer 3 is the normalization layer (layer of average nodes). In this layer, the ith node is a circle labeled N, and computes the normalized firing strength as follows:

$$O_i^3 = \overline{w_i} = \frac{w_i}{w_1 + w_2}, \text{ for } i = 1, 2. \tag{6}$$

Layer 4 is the defuzzification layer (layer of consequent nodes). In this layer, every node i marked by a square is an adaptive node with a node function. The output of this layer is calculated by:

$$O_i^4 = \overline{w_i} f_i = \overline{w_i}(p_i x + q_i y + r_i) \text{ for } i = 1, 2, \quad (7)$$

where $\{p_i, q_i, r_i\}$ is the parameter set of this node.

Layer 5 is the output layer. The single node in this layer is a fixed circle node labeled $\sum$, which calculates the final overall output as the summation of all incoming signals. The overall output is computed by this formula:

$$O_i^5 = f_{\text{out}} = \sum_{i=1}^{2} \overline{w_i} f_i = \frac{\sum_{i=1}^{2} w_i f_i}{w_1 + w_2}. \quad (8)$$

Finally, the overall output can be formulated as

$$f_{\text{out}} = \frac{w_1}{w_1 + w_2} f_1 + \frac{w_2}{w_1 + w_2} f_2. \quad (9)$$

Or,

$$f_{\text{out}} = \overline{w_1} f_1 + \overline{w_2} f_2. \quad (10)$$

Substituting Eq. (7) into Eq. (10):

$$f_{\text{out}} = \overline{w_1}(p_1 x + q_1 y + r_1) + \overline{w_2}(p_2 x + q_2 y + r_2). \quad (11)$$

The final output can be written as:

$$f_{\text{out}} = (\overline{w_1} x) p_1 + (\overline{w_1} y) q_1 + (\overline{w_1}) r_1 + (\overline{w_2} x) p_2 + (\overline{w_2} y) q_2 + (\overline{w_2}) r_2. \quad (12)$$

ANFIS Model Development

In this study, the available data obtained from the experimental work were randomly divided into three portions: 70% (112 data points) as the training datasets for the learning process, 20% (32 data points) as the testing datasets to test the precision of the model and, 10% (16 data points) for the validation procedure. Before the training, the data were normalized to be in a range between 0 and +1 in order to decrease their range and increase the precision of the findings. After the normalization process, the data were ready for the training process. The normalization of data was done according to the following equation:

$$X_{\text{n}} = \frac{X_i - X_{\min}}{X_{\max} - X_{\min}}, \quad (13)$$

where X_{n} is the normalized value, X_{i} is the observed value of the variable, $X_{\max}$ is the maximum observed value, and $X_{\min}$ is the minimum observed value.

MATLAB software (MATLAB 8.1.0.604, R2013a, the MathWorks) was used to develop an ANFIS model from the experimental data to forecast SSP. The Sugeno-type fuzzy inference system was used in the modeling of SSP. The grid partition method was employed to classify the input data and in making the rules [36]. In the current study, we used four different types of input membership functions (MFs), including generalized bell (GBELLMF), Gaussian (GAUSSMF), two-sided Gaussian (GASUSS2MF), and the product of two sigmoidal functions (PSIGMF). We identified the best ANFIS structure/model in terms of the input MF (GBELLMF, GAUSSMF, GASUSS2MF, or PSIGMF), where the most appropriate input MF was determined based on statistical indicators. The output MF was selected as a linear function. Furthermore, a hybrid learning algorithm that combines the least squares estimator and the gradient descent method was utilized to estimate the optimum values of the FIS parameters of the Sugeno-type [36]. The number of epochs was selected as 50 owing to their small error.

Multiple Linear Regression (MLR)

The best ANFIS model was compared to multiple linear regression (MLR) to check and confirm the performance and efficiency of the developed ANFIS model. MLR is a commonly used statistical technique for displaying the dependency between two or more parameters by fitting a linear equation to the measured/observed data. The relationship between independent parameters and the dependent parameter can be expressed as follows:

$$\hat{Y} = A_o + \sum_{i=1}^{m} A_j X_j, \quad (14)$$

where $\hat{Y}$ is the dependent parameter, A_o is the intercept, A_j are the partial regression coefficients, and X_j ($j = 1, 2, 3, \ldots, m$) are the independent parameters [29].

By using the regression tool in IBM SPSS statistics 23 and following the same modeling pattern (data points) developed for the ANFIS, 70% of the data points (randomly selected) were used to train/fit the MLR model, 20% of the data points were used to test the model, and the remaining 10% used to validate the model. However, while MLR can be a beneficial predictive technique, because of its dependency on linearly correlated equations it may result in inaccurate findings for complex systems.

Statistical Indicators

In this study, the performance of the developed models was evaluated using different standard statistical performance evaluation criteria. The statistical measures considered were correlation coefficient (CC), root mean square error (RMSE), overall index of model performance (OI), mean absolute error (MAE), and mean absolute relative error (MARE):

$$\text{CC} = \frac{\sum_{i=1}^{n} \left(\text{SSP}_{o,i} - \overline{\text{SSP}}_o\right)\left(\text{SSP}_{p,i} - \overline{\text{SSP}}_p\right)}{\sqrt{\sum_{i=1}^{n} \left(\text{SSP}_{o,i} - \overline{\text{SSP}}_o\right)^2 \times \sum_{i=1}^{n} \left(\text{SSP}_{p,i} - \overline{\text{SSP}}_p\right)^2}}, \quad (15)$$

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{n} \left(\text{SSP}_{o,i} - \text{SSP}_{p,i}\right)^2}{n}}, \quad (16)$$

$$\text{OI} = \frac{1}{2}\left(2 - \frac{\text{RMSE}}{\text{SSP}_{\max} - \text{SSP}_{\min}} - \frac{\sum_{i=1}^{n} \left(\text{SSP}_{o,i} - \text{SSP}_{p,i}\right)^2}{\sum_{i=1}^{n} \left(\text{SSP}_{o,i} - \overline{\text{SSP}}_o\right)^2}\right), \quad (17)$$

$$\text{MAE} = \frac{\sum_{i=1}^{n} |\ \text{SSP}_{o,i} - \text{SSP}_{p,i}|}{n}, \quad (18)$$

$$\text{MARE} = \frac{1}{2}\left(\sum_{i=1}^{n} \left|\frac{\text{SSP}_{o,i} - \text{SSP}_{p,i}}{\text{SSP}_{o,i}}\right| \times 100\right), \quad (19)$$

where $\text{SSP}_{o,i}$ denotes the observed value; $\text{SSP}_{p,i}$ is the predicted value; $\overline{\text{SSP}}_o$ is the mean of the observed values; $\overline{\text{SSP}}_p$ is the mean of the predicted values; $\text{SSP}_{\max}$ is the maximum observed value; $\text{SSP}_{\min}$ is the minimum observed value; and n is the total number of observations. More details about the statistical indicators can be found in Refs. 25–29.

RESULTS AND DISCUSSION

Independent Parameters Selection

In this study, field data obtained from experimental work were used for the training, testing, and validation of ANFIS models. One of the most important steps in the modeling process for a satisfactory prediction of results is the accurate

Table 1. Correlation coefficient matrix for studied parameters.

	T_o	RH	WS	SR	T_F	T_B	M_F	TDS_F	TDS_B	SSP
T_o	1.00	−0.66	−0.14	−0.15	0.91	0.06	0.44	−0.01	−0.15	−0.07
RH	−0.66	1.00	−0.08	0.15	−0.80	0.05	−0.72	0.23	0.45	0.01
WS	−0.14	−0.08	1.00	0.22	−0.01	0.33	−0.34	0.64	0.49	−0.31
SR	−0.15	0.15	0.22	1.00	−0.09	0.82	−0.27	0.22	0.39	0.73
T_F	0.91	−0.80	−0.01	−0.09	1.00	0.13	0.48	0.06	−0.11	−0.06
T_B	0.06	0.05	0.33	0.82	0.13	1.00	−0.40	0.49	0.57	0.40
M_F	0.44	−0.72	−0.34	−0.27	0.48	−0.40	1.00	−0.75	−0.84	0.25
TDS_F	−0.01	0.23	0.64	0.22	0.06	0.49	−0.75	1.00	0.94	−0.40
TDS_B	−0.15	0.45	0.49	0.39	−0.11	0.57	−0.84	0.94	1.00	−0.17
SSP	−0.07	0.01	−0.31	0.73	−0.06	0.40	0.25	−0.40	−0.17	1.00

Abbreviations: T_o, ambient temperature; RH, relative humidity; WS, wind speed; SR, solar radiation; T_F, temperature of feed water; T_B, temperature of brine water; M_F, feed flow rate; TDS_F, total dissolved solids of feed; TDS_B, total dissolved solids of brine; SSP, solar still productivity.

Table 2. Summary statistics for input and output parameters.

Parameter	Type	Unit	Symbol	Min	Max	Avg	SD	CV
Relative humidity	Input	%	RH	12.90	70.00	23.36	12.90	0.55
Solar radiation	Input	W/m^2	SR	75.10	920.69	587.55	181.93	0.31
Feed flow rate	Input	L/min	M_F	0.13	0.25	0.21	0.04	0.20
Total dissolved solids of feed	Input	PPT	TDS_F	41.40	130.00	80.23	29.42	0.37
Total dissolved solids of brine	Input	PPT	TDS_B	46.20	132.80	95.54	29.59	0.31
Solar still productivity	Output	$L/m^2/h$	SSP	0.05	0.97	0.50	0.24	0.48

Abbreviations: Min, minimum value; Max, maximum value; Avg, average value; SD, standard deviation; CV, coefficient of variation.

selection of the input parameters. The significance of this step can be attributed to the fact that these parameters determine the model structures, affect the weighted coefficient and the overall performance of the model. For this purpose, in the current study, a correlation matrix has been performed to evaluate relationships between the dependent parameter (SSP) and independent parameters (T_o, RH, WS, SR, T_F, T_B, TDS_F, and TDS_B), as presented in Table 1. This matrix allows us to recognize the effect of each parameter on SSP and eventually identify the parameter(s), which should be used as input(s) in the ANFIS and MLR models. Moreover, this matrix displays the findings of correlation analysis conducted between each pair of parameters. The strongest correlation has been observed between SSP and SR with a Pearson's correlation coefficient (CC) of +0.734. Furthermore, the study result shows a fine correlation of SSP with TDS_F with CC = –0.402. Herein, the + and – signs refer to the positive correlation and negative correlation, respectively. This is in alignment with the findings of Mashaly *et al.* [4]. Furthermore, this study reveals a significant correlation between SSP and TDS_B with CC = –0.172. Alternatively, an extremely weak correlation is found between SSP and T_o and T_F with CC = –0.072 and –0.061, respectively, and therefore the study analysis does not consider them as input parameters. The correlation analysis also led to the exclusion of WS and T_B due to their high collinearity with other parameters; despite they correlate significantly correlations with the SSP. Although some parameters also appear to be correlated to others, these were included in the modeling process since their inclusion was found to improve its prediction performance, primarily by enhancing the CD. The same argument was also invoked to consider RH as an input parameter with low CC. The input and output parameters used in these models and their range are presented in Table 2.

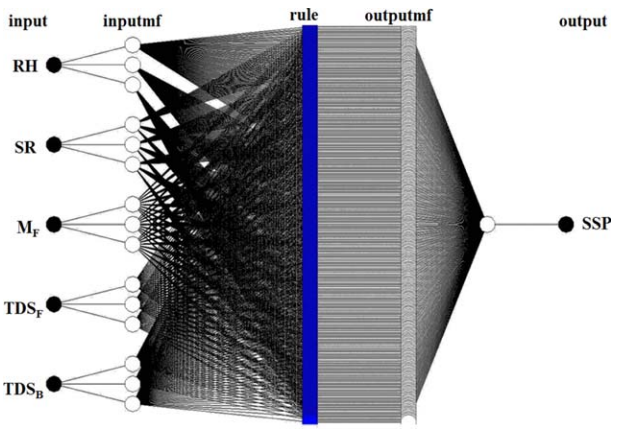

Figure 3. ANFIS model structure used in this study [RH: Relative Humidity; SR: Solar Radiation; M_F: Feed Flow Rate; TDS_F: Total Dissolved Solids of Feed; TDS_B: Total Dissolved Solids of Brine; SSP: Solar Still Productivity]. [Color figure can be viewed at **wileyonlinelibrary.com**]

ANFIS Models Structure

In this study, we developed four ANFIS models with four different types of input membership functions (MFs). The MFs used were GBELLMF, GAUSSMF, GASUSS2MF, and PSIGMF. The ANFIS models developed have five inputs (RH, SR, M_F, TDS_F, and TDS_B) and one output (SSP).The structure of the ANFIS models with the five input parameters are presented in Figure 3. Therefore, in the input layer, five neurons were incorporated. For each neuron, three identical MFs

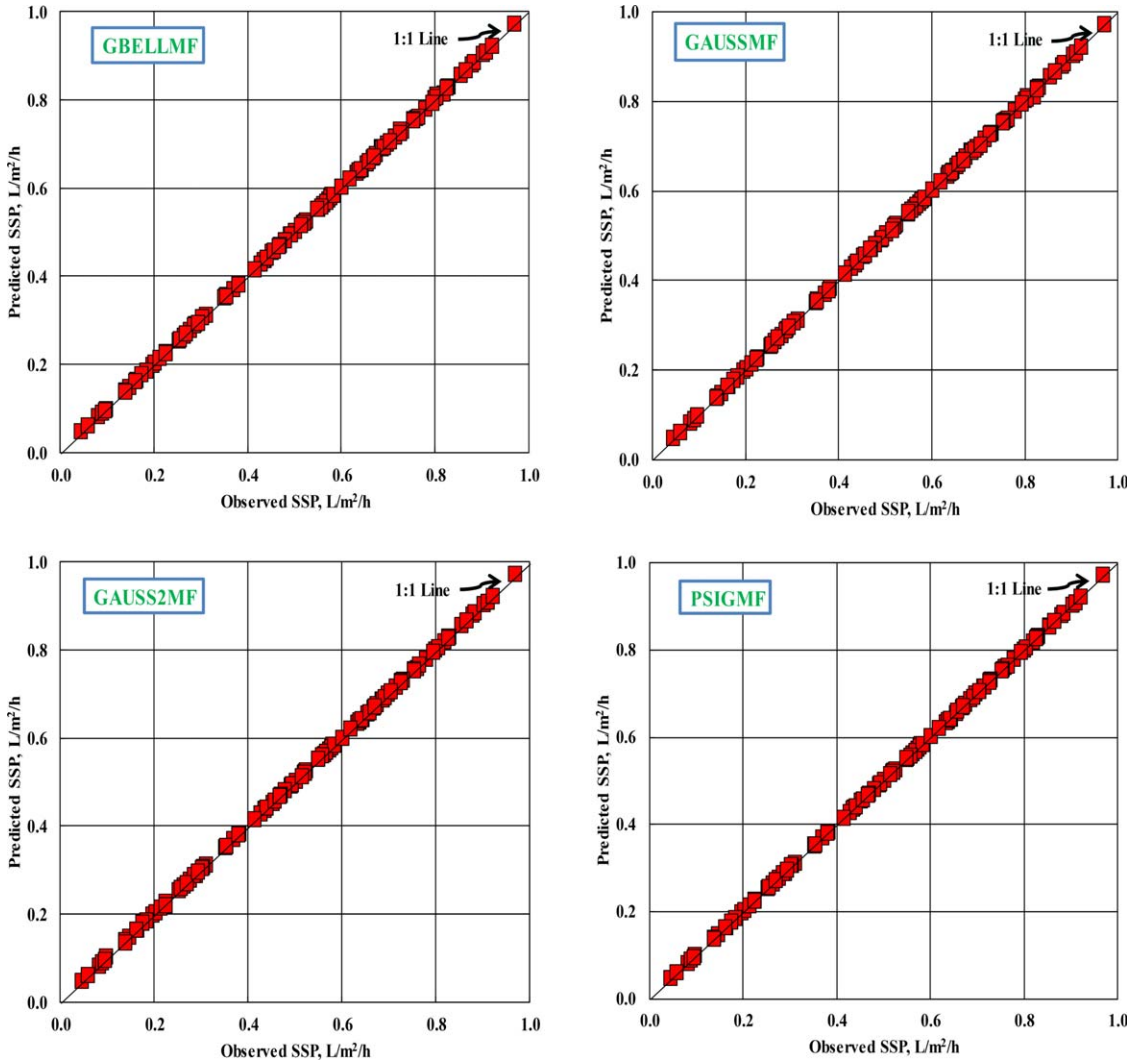

Figure 4. Predicted versus observed solar still production (SSP) during the training process by using different membership functions. [Color figure can be viewed at **wileyonlinelibrary.com**]

were considered with three linguistic terms (low, medium, high), and accordingly, 243 ($3 \times 3 \times 3 \times 3 \times 3$) rules were developed for the implementation of the ANFIS model. The number of nodes was 524, with 1458 linear parameters for all four models. The number of nonlinear parameters was 60, with 1518 parameters in the models with GASUSS2MF and PSIGMF. While the number of nonlinear parameters was 30, and of parameters was 1488 in the model with GAUSSMF. Moreover, the number of nonlinear parameters was 45, and of parameters was 1503 in the model with GASUSS2MF. However, the four ANFIS models were trained, tested, and validated to assess predictability of SSP by ANFIS-based models. The next sections discuss and evaluate the performance of the four developed ANFIS models in the training, testing, and validation phases by the statistical performance indicators.

ANFIS Models Performance

Figure 4 illustrates the relationship between the predicted and observed values of solar still productivity (SSP) by using different MFs during the training process. It comprises four graphs representing the four MFs used to calculate SSP. It is obvious from the figure that the predicted SSP values by using the four MFs, namely GBELLMF, GAUSSMF, GASUSS2MF, and PSIGMF, are in excellent agreement with the observed SSP values.

Table 3 displays the results of the statistical parameters CC, RMSE, OI, MAE, and MARE, which are numerical indicators used to assess the agreement between observed and predicted SSP. The average values of CC, RMSE, OI, MAE, and MARE are 99.99%, 0.0012 L/m^2/h, 99.94%, 0.0007 L/m^2/h, and 0.1698%, respectively, for the developed ANFIS models in the training process. From Table 3, it can be seen that the values of RMSE, MAE, and MARE are very low, and the values of CC and OI are very high for SSP predicted from the GBELLMF, GAUSSMF, GAUSS2MF, and PSIGMF during the training process. Moreover, the CC and OI values are very close to one while RMSE, MAE, and MARE values are close to zero, showing excellent agreement between the observed and predicted results from the four ANFIS models. The tiny deviations between the observed and predicted results in turn highlight the effectiveness of the ANFIS technique in the prediction process of SSP. This agrees with the results of Refs. 14,18, and 24. In the training stage, the performance of all MFs was approximately the same. But in relative terms, the highest performance was obtained with PSIGMF in the training process. The ANFIS model with PSIGMF'statistics were CC = 99.99%, RMSE = 0.0007 L/m^2/h, OI = 99.96%, MAE = 0.0005 L/m^2/h, and MARE = 0.1424%.

Table 3. Statistical performance of the developed ANFIS models with different types of input membership functions and the obtained MLR model during training, testing and validation stages.

Stages	Statistical parameters	Membership functions (ANFIS models)				MLR model
		GBELLMF	GAUSSMF	GAUSS2MF	PSIGMF	
Training	CC	0.9999	0.9999	0.9999	0.9999	0.9654
	RMSE	0.0013	0.0014	0.0012	0.0007	0.1253
	OI	0.9993	0.9992	0.9994	0.9996	0.7991
	MAE	0.0007	0.0007	0.0007	0.0005	0.1094
	MARE	0.1519	0.1498	0.2350	0.1424	29.0776
Testing	CC	0.9809	0.9710	0.9596	0.9427	0.9412
	RMSE	0.0467	0.0601	0.0680	0.0820	0.1337
	OI	0.9507	0.9288	0.9143	0.8855	0.7580
	MAE	0.0333	0.0428	0.0483	0.0562	0.1203
	MARE	9.6451	12.4304	12.8264	15.3335	34.3794
Validation	CC	0.9801	0.9405	0.9133	0.9048	0.9273
	RMSE	0.0441	0.0829	0.0942	0.0895	0.1437
	OI	0.9364	0.8365	0.7998	0.8157	0.6983
	MAE	0.0372	0.0619	0.0648	0.0570	0.1350
	MARE	8.5248	14.2732	14.3701	13.7632	30.3045

Abbreviations: CC, correlation coefficient; RMSE, root mean square error; OI, overall index of model performance; MAE, mean absolute error; MARE, mean absolute relative error.

However, and generally, the statistical values of the observed and predicted values were excellent with all MFs. These findings emphasize the accuracy and efficiency of ANFIS models for estimating SSP by using the four MFs.

Figure 5 shows the 1:1 relationship between the predicted and observed values of SSP by using the developed ANFIS models during the testing process and encompasses four graphs representing the four MFs used to develop the ANFIS models. The figure clearly shows that the SSP values forecasted by using the ANFIS models, coupled with the four MFs are in a good agreement with the observed SSP values. The statistical performance of the ANFIS models with different MFs is shown in Table 3 during the testing process. As indicated in Table 3, the average values of CC, RMSE, OI, MAE, and MARE for all MFs are 96.36%, 0.0642 L/m^2/h, 91.98%, 0.0452 L/m^2/h, and12.5589%, respectively, in the testing process. Furthermore, from Table 3, the comparisons prove that the performance of ANFIS model with GBELLMF is better than that with other MFs during the testing process, where its statistics were CC = 98.09%, RMSE = 0.0467 L/m^2/h, OI = 0.9507%, MAE = 0.0333 L/m^2/h, and MARE = 9.6451%. For GBELLMF, the CC and OI values are close to one while RMSE and MAE values are close to zero, indicating excellent agreement between the observed and predicted values. Also, the MARE value is relatively closer to zero compared to other MFs. On the other hand, the relatively lowest performance has the ANFIS model with PSIGMF with the relatively highest RMSE, MAE, and MARE values of 0.0820 L/m^2/h, 0.0562 L/m^2/h, and 15.3335%, respectively, and the relatively lowest CC and OI values of 94.27 and 88.55%, respectively, in the testing stage. However, the difference between the highest (GBELLMF) and lowest (PSIGMF) performances was small during the testing process.

A good agreement is obtained between the observed and predicted SSP with ANFIS models especially computed by the GAUSSMF as presented by the 1:1 curve in Figure 6 during the validation process. The performance of the ANFIS models with different MFs is presented in Table 3. This good agreement is clearly reflected by the values of the statistical parameters displayed in Table 3 The ANFIS model, using GBELLMF, had a CC value that was about 4.21, 7.31, and 8.32% more accurate than those from the ANFIS models with GAUSSMF, GAUSS2MF, and PSIGMF, respectively. According to Table 3, the RMSE values indicate that the GBELLMF is the most accurate. The RMSE values for GAUSSMF, GAUSS2MF, and PSIGMF are 1.88, 2.14, and 2.03 times, respectively, given by the GBELLMF (0.0441 L/m^2/h). Also, the RMSE value for the GAUSSMF was closer to zero than those from other MFs. However, the best RMSE value was achieved by the ANFIS model with GBELLMF, followed by, respectively, the ANFIS model with GAUSSMF, GAUSS2MF, and PSIGMF.

The OI values of the ANFIS models with different MFs are 93.64, 83.65, 79.98, and 81.57% for GBELLMF, GAUSSMF, GAUSS2MF, and PSIGMF, respectively, during the validation process. The OI value for the ANFIS model with GBELLMF was closer to 100% than its values for the other MFs. The MAE values for the GAUSSMF, GAUSS2MF, and PSIGMF (0.0619, 0.0648, and 0.0570 L/m^2/h, respectively) were almost 1.66, 1.74, and 1.53% times, respectively, that of the value (0.0372 L/m^2/h) for the GBELLMF. The MARE values for GAUSSMF, GAUSS2MF, and PSIGMF were 1.67, 1.69, and 1.62 times, respectively, that of the value for the ANFIS model with GBELLMF. However, these results indicate excellent agreement between the observed results and predicted results from the ANFIS, especially with GBELLMF, and confirm that the ANFIS model with GBELLMF performed better than that with other MFs when using the validation data set.

Overall, in all the modeling stages, the highest prediction capability was the ANFIS model with GBELLMF, followed by GAUSSMF, GAUSS2MF, and PSIGMF. These results confirm with the findings of Refs. 37 and [38]. For the most part, the use of ANFIS models with different MFs gives a high level of accuracy in the modeling process for SSP and this is obviously reflected by the values of the statistical parameters displayed in Table 3. However, as seen from the complete findings in Figures 4–6 and Table 3, the ANFIS models are a reliable and powerful tool for predicting the SSP and provide us with results that are higher in performance and accuracy. The satisfactory nature of these results confirms that the ANFIS technique has good potential for use in SSP forecasting.

MLR Model Performance

The MLR model was developed to confirm the efficiency of the obtained ANFIS model. Here, MLR is selected to

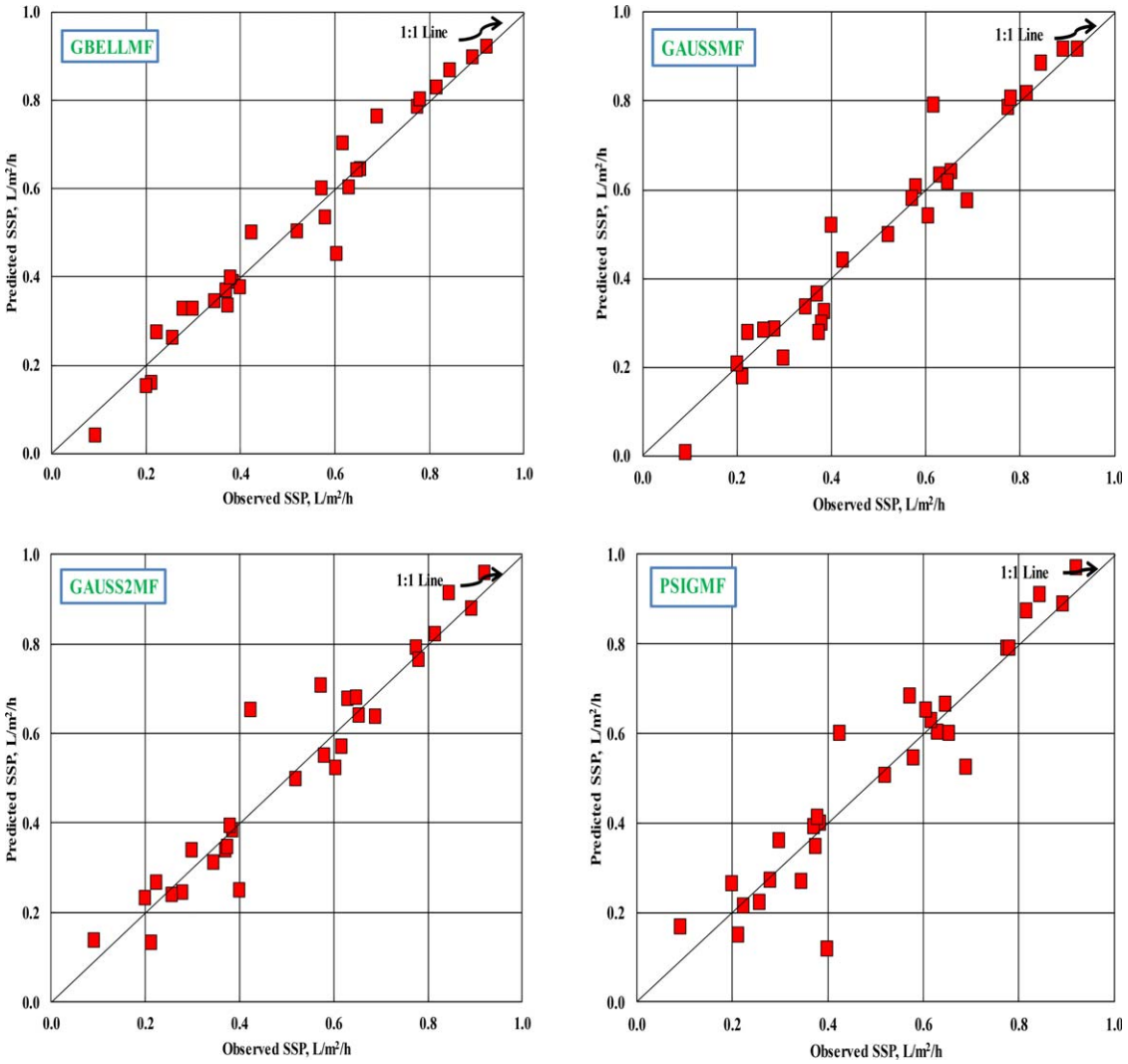

Figure 5. Predicted versus observed solar still production (SSP) during the testing process by using different membership functions. [Color figure can be viewed at **wileyonlinelibrary.com**]

develop a relationship between the SSP and the five variables that affect it (RH, SR, MF, TDSF, and TDSB). The SSP is the dependent variable, while the other variables are independent. The following formula is obtained to estimate SSP based on the MLR model:

$$\mathrm{SSP} = -0.150 - 0.003 \times \mathrm{RH} + 0.001 \times \mathrm{SR} + 1.195 \times M_{\mathrm{F}} - 0.012 \times \mathrm{TDS_F} + 0.010 \times \mathrm{TDS_B}. \quad (20)$$

All variables in this model are significant (P value <0.05). However, Table 4 illustrates the standard error (SE) of the regression coefficients, t-stat, and probability (P value) of the independent parameters. Scatterplots are presented in Figure 7 for the training, testing, and validation data sets using the MLR model. Figure 6 displays the observed SSP on the x axis against the predicted SSP on the y axis. These results show that many points given by the developed MLR model during the training, testing, and validation phases are located above and below the 1:1 line, which indicates that some values predicted by the regression technique were inaccurate. As demonstrated in Table 3, the developed MLR model has CC = 0.9654, RMSE = 0.1253 L/m2/h, OI = 0.7991, MAE = 0.1094 L/m^2/h, and MARE = 29.0776% for the training phase; CC = 0.9412, RMSE = 0.1337 L/m^2/h, OI = 0.7580, MAE = 0.1203 L/m^2/h, and MARE = 34.3794% for the testing phase; and CC = 0.9273, RMSE = 0.1437 L/m^2/h, OI = 0.6983, MAE = 0.1350 L/m^2/h, and MARE = 30.3045% for the validation phase. It can be concluded from the statistical parameters that the MLR model has the relative lowest CC and OI with the highest errors (RMSE, MAE, and MARE). Based on these findings, the MLR model is not accurate enough to use as a forecasting procedure for SSP. Comparing the selected/best ANFIS with MLR model.

To further evaluate the capability of ANFIS and MLR models to compute SSP, their findings are compared with each other in terms of their accuracy, effectiveness, and predictive ability. The results of the statistical parameters CC, RMSE, OI, MAE, and MARE for both the MLR and ANFIS models are given in Table 3 for the SSP estimation. The scatter plots of the predicted (by ANFIS and MLR) and observed SSP values during all the modeling stages for both models (Figures 4–7) clearly indicates the two models' relative proficiency and strength. But the developed MLR model gave the least accurate SSP results, with relatively large scattering, in compared to the best/selected ANFIS model. As can be seen in from Table 3, the statistical indicators of the selected ANFIS model are superior to those of the MLR model.

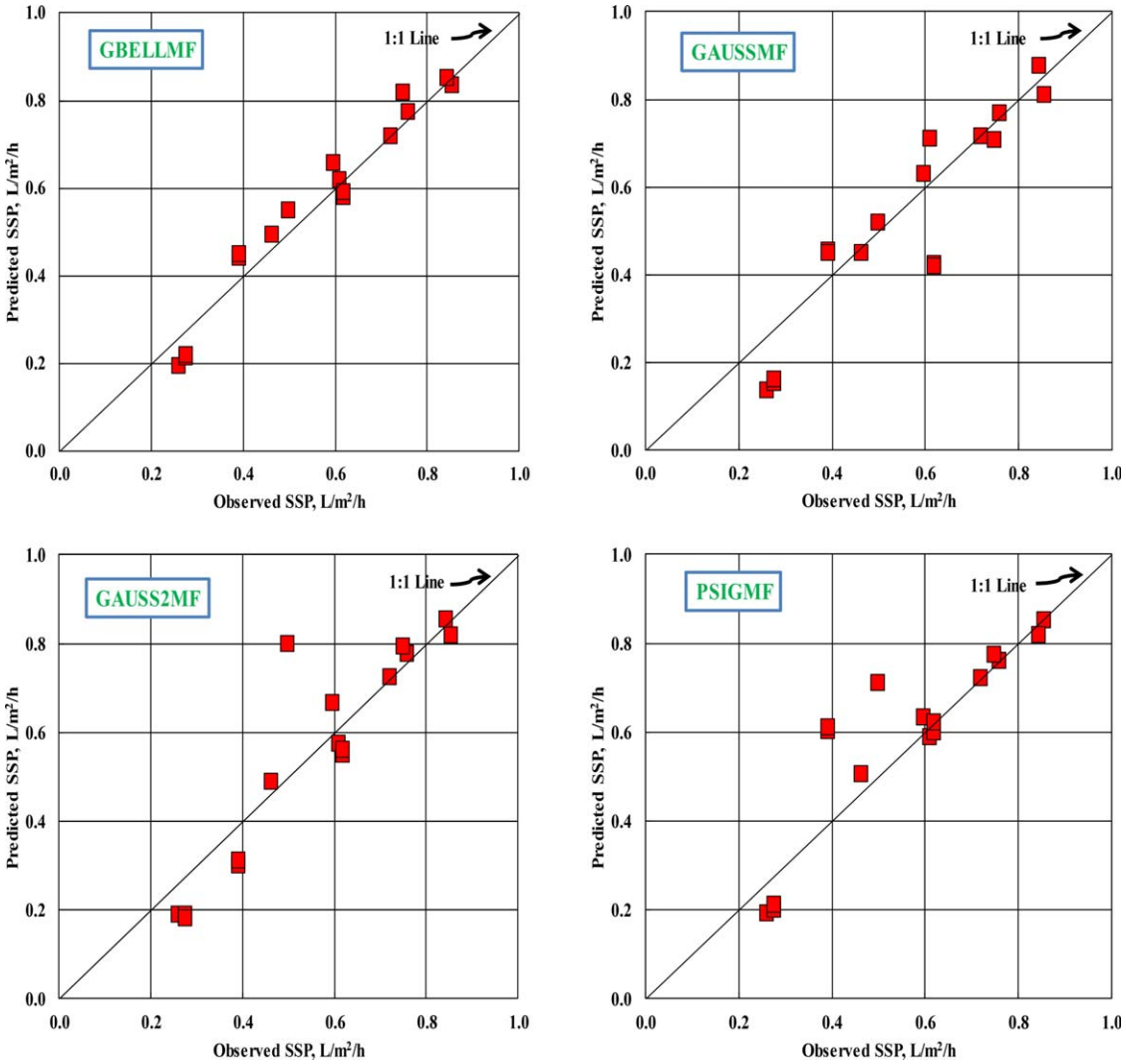

Figure 6. Predicted versus observed solar still production (SSP) during the validation process by using different membership functions. [Color figure can be viewed at **wileyonlinelibrary.com**]

Table 4. Results of multiple linear regression (MLR) analysis.

Model	Values	Std. Error	t-stat	*P* value (Sig)
Constant	−0.150	0.144	−1.041	0.300
RH	−0.003	0.001	−2.489	0.014
SR	0.001	4.3×10^{-5}	20.060	7.1×10^{-38}
M_F	1.195	0.433	2.759	0.007
TDS_F	−0.012	0.001	−12.410	2.2×10^{-22}
TDS_B	0.010	0.001	9.531	6.4×10^{-16}

Abbreviations: RH, relative humidity; SR, solar radiation; M_F, feed flow rate; TDS_F, total dissolved solids of feed; TDS_B, total dissolved solids of brine.

Using the MLR model to estimate SSP increases the RMSE, MAE, and MARE values to 96, 156, and 191 times the ANFIS values, respectively, and reduces the CC and OI values by about 3.45 and 20%, respectively, using the training dataset. Using the ANFIS model, the CC and OI values were reduced by 4.1 and 20.3% compared to the MLR model during the testing process. The RMSE for SSP in the MLR model (0.1337 L/m^2/h) is almost three times that given by the ANFIS model. The MAE value of 0.1203 L/m^2/h in the MLR model was increased by 261.3% over the ANFIS model, while the MARE value in the ANFIS model was decreased by 72% in the MLR model, using the testing dataset. Using the validation data set of the ANFIS model in the comparison demonstrates that the MLR model had a CC value that was about 5.4% less accurate (Table 3). The value of RMSE for the MLR model (0.1437 L/m^2/h) was almost three times that of the value for the ANFIS model, and the OI value for the ANFIS model was 34.1% more accurate than that of the MLR model. The MAE value of 0.0372 L/m^2/h for the ANFIS model was closer to zero than to its value (0.1350 L/m^2/h) for the MLR model. The MARE value for the MLR model was almost four times that of the value for the ANFIS model. The CC, RMSE, OI, MAE, and MARE results for the two approaches confirm that the ANFIS model performs perfectly well.

CONCLUSIONS

In this study, the potential of the ANFIS modeling technique for the prediction SSP was evaluated. The data was obtained through experimental work using solar still to desalinate seawater. Various ANFIS models with different membership functions (MFs) were trained, tested, and validated to discover the best ANFIS model for the modeling of SSP.

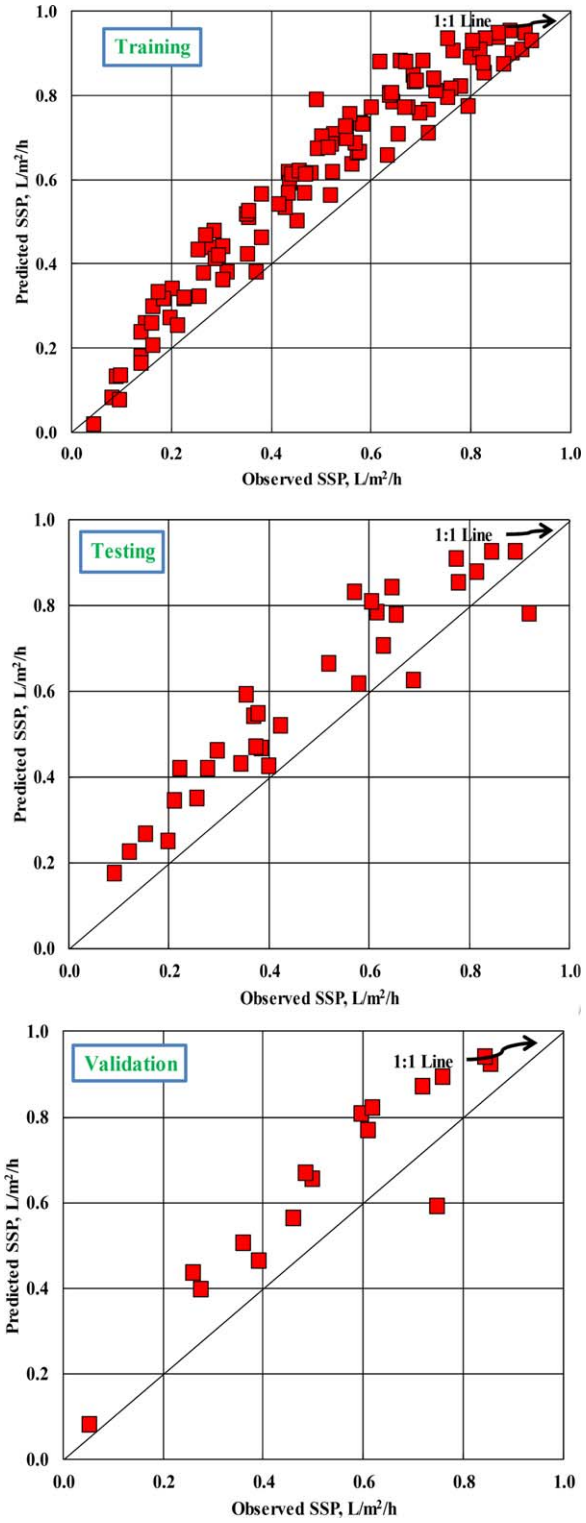

Figure 7. Comparison between observed and predicted SSP values using the MLR model during the training, testing, and validation phases. [Color figure can be viewed at **wileyonlinelibrary.com**]

The ANFIS models were trained using 70% of the available data, tested using 20% of the data, and validated using the remaining 10%. In the modeling process, five input parameters were used, namely, RH, SR, M_F, TDS_F, and TDS_B. We used four different types of MFs in this study, specifically, generalized bell (GBELLMF), Gaussian (GAUSSMF), two-sided Gaussian (GASUSS2MF), and product of two sigmoidal functions (PSIGMF). The performance of the developed ANFIS models was evaluated by using statistical performance indicators, namely, CC, RMSE, OI, MAE, and MARE. Generally, the results reveal the high accuracy, effectiveness and reliability of the ANFIS technique for modeling the SSP. On the basis of statistical performance criteria, it was found that GBELLMF was the best MF in all modeling stages for ANFIS models. Consequently, the highest performance obtained was by the ANFIS model with GBELLMF. Additionally, the results of the best ANFIS were statistically compared with those from MLR. The comparison of the statistical parameters reveals the superiority of the ANFIS model over the MLR model. However, the findings also shed light on some principals regarding the ANFIS-based model application for the prediction of SSP using meteorological and operational parameters. It is hoped that the findings will serve as a reference for future attempts to evaluate the used parameters in this study using other soft computing techniques to model the SSP.

NOMENCLATURE

ANFIS	Adaptive neuro-fuzzy inference system
ANN	Artificial neural network
CC	Correlation coefficient
FL	Fuzzy logic
GASUSS2MF	Two-sided Gaussian membership function
GAUSSMF	Gaussian membership function
GBELLMF	Generalized bell membership function
MAE	Mean absolute error
MARE	Mean absolute relative error
M_F	Feed flow rate
MFs	Membership functions
MLR	Multiple linear regression
OI	Overall index of model performance
PSIGMF	Product of two sigmoidal functions
RH	Relative humidity
RMSE	Root mean square error
SR	Solar radiation
SSP	Solar still productivity
T_B	Temperature of brine water
TDS_B	Total dissolved solids of brine water
TDS_F	Total dissolved solids of feed water
T_F	Temperature of feed water
T_o	Ambient temperature
WS	Wind speed
X_{max}	Maximum measured value
X_{min}	Minimum measured value
$x_{o,i}$	Measured value
$x_{p,i}$	Predicted value
$\overline{x}o$ =	Averaged measured values

ACKNOWLEDGMENTS

The project was financially supported by King Saud University, Vice Deanship of Research Chairs.

LITERATURE CITED

1. Ahmed, F., Mashaly, A.A., Alazba, A.M., Al-Awaadh, Mohamed, A. & Mattar (2015). Area determination of solar desalination system for irrigating crops in greenhouses using different quality feed water, Agricultural Water Management, 154, 1–10.
2. Murugavel, K.K., Chockalingam, K.K., & Srithar, K. (2008). Progresses in improving the effectiveness of the single basin passive solar still, Desalination, 220, 677–686.
3. Prakash, P., & Velmurugan, V. (2015). Parameters influencing the productivity of solar stills—A review, Renewable and Sustainable Energy Reviews, 49, 585–609.

4. Mashaly, A.F., Alazba, A.A., & Al-Awaadh, A.M. (2016). Assessing the performance of solar desalination system to approach near-ZLD under hyper arid environment, Desalination and Water Treatment, 57, 12019–12036.
5. Elango, T., Kannan, A., & Murugavel, K.K. (2015). Performance study on single basin single slope solar still with different water nanofluids, Desalination, 360, 45–51.
6. Köhl, M., Jorgensen, G., Brunold, S., Carlsson, B., Heck, M., & Möller, K. (2005). Durability of polymeric glazing materials for solar applications, Solar Energy, 79, 618–623.
7. Tripathi, R., & Tiwari, G.N. (2006). Thermal modeling of passive and active solar stills for different depths of water by using the concept of solar fraction, Solar Energy, 80, 956–967.
8. Patel, S.G., Bhatnagar, S., Vardia, J., & Ameta, S.C. (2006). Use of photocatalysts in solar desalination, Desalination, 189, 287–291.
9. Badran, A.A., Al-Hallaq, I.A., Salman, I.A.E., & Odat, M.Z. (2005). A solar still augmented with a flat-plate collector, Desalination, 172, 227–234.
10. Abad, H.K.S., Ghiasi, M., Mamouri, S.J., & Shafii, M.B. (2013). A novel integrated solar desalination system with a pulsating heat pipe, Desalination, 311, 206–210.
11. Abdel-Rehim, Z.S., & Lasheen, A. (2007). Experimental and theoretical study of a solar desalination system located in Cairo, Egypt, Desalination, 217, 52–64.
12. Reddy, K.S., Kumar, K.R., O'Donovan, T.S., & Mallick, T.K. (2012). Performance analysis of an evacuated multistage solar water desalination system, Desalination, 288, 80–92.
13. Santos, N.I., Said, A.M., James, D.E., & Venkatesh, N.H. (2012). Modeling solar still production using local weather data and artificial neural networks, Renewable Energy, 40, 71–79.
14. Yaïci, W., & Entchev, E. (2016). Adaptive neuro-fuzzy inference system modelling for performance prediction of solar thermal energy system, Renewable Energy, 86, 302–315.
15. Monika, A.K. (2013). Comparison of fuzzy logic and NEURO fuzzy algorithms for load sensor, International Journal of Soft Computing and Engineering, 3, 219–222.
16. Ghafari, G., & Vafakhah, M. (2012). Rainfall-runoff Modeling Using Artificial Neural Networks and Adaptive Neuro-Fuzzy Inference System Models. In uncertainty Modeling in Knowledge Engineering and Decision Making, Vol. 7. 951–956, World Scientific Publishing Co. Pte. Ltd., Singapore.
17. Atmaca, H., Cetisli, B., & Yavuz, H. S. (2001). The comparison of fuzzy inference systems and neural network approaches with ANFIS method for fuel consumption data. In Second international conference on electrical and electronics engineering papers ELECO, Bursa, Turkey.
18. Kharb, R.K., Shimi, S.L., Chatterji, S., & Ansari, M.F. (2014). Modeling of solar PV module and maximum power point tracking using ANFIS, Renewable and Sustainable Energy Reviews, 33, 602–612.
19. Amirkhani, S., Nasirivatan, S., Kasaeian, A.B., & Hajinezhad, A. (2015). ANN and ANFIS models to predict the performance of solar chimney power plants, Renewable Energy, 83, 597–607.
20. Mohammadi, K., Shamshirband, S., Petković, D., & Khorasanizadeh, H. (2016). Determining the most important variables for diffuse solar radiation prediction using adaptive neuro-fuzzy methodology; case study: City of Kerman, Iran, Renewable and Sustainable Energy Reviews, 53, 1570–1579.
21. Mellit, A., & Kalogirou, S.A. (2011). ANFIS-based modelling for photovoltaic power supply system: A case study, Renewable Energy, 36, 250–258.
22. Shanmugan, S. (2003). Fuzzy logic modeling of floating cum tilted-wick solar still, International Journal of Recent Scientific Research, 4, 579–582.
23. Mamlook, R., & Badran, O. (2007). Fuzzy sets implementation for the evaluation of factors affecting solar still production, Desalination, 203, 394–402.
24. Shanmugan, S., & Krishnamoorthi, G. (2013). Fuzzy logic modeling of single slope single basin solar still, International Journal of Fuzzy Mathematics and Systems, 3, 125–134.
25. Mashaly, A.F., Alazba, A.A., Al-Awaadh, A.M., & Mattar, M.A. (2015). Predictive model for assessing and optimizing solar still performance using artificial neural network under hyper arid environment, Solar Energy, 118, 41–58.
26. Mashaly, A.F., & Alazba, A.A. (2016). Neural network approach for predicting solar still production using agricultural drainage as a feedwater source, Desalination and Water Treatment, 57, 28646–28660.
27. Mashaly, A.F., & Alazba, A.A. (2016). Comparison of ANN, MVR, and SWR models for computing thermal efficiency of a solar still, International Journal of Green Energy, 13, 1016–1025.
28. Mashaly, A.F., & Alazba, A.A. (2015). Comparative investigation of artificial neural network learning algorithms for modeling solar still production, Journal of Water Reuse and Desalination, 5, 480–493.
29. Mashaly, A.F., & Alazba, A.A. (2016). MLP and MLR models for instantaneous thermal efficiency prediction of solar still under hyper-arid environment, Computers and Electronics in Agriculture, 122, 146–155.
30. Mashaly, A.F., & Alazba, A.A. (2017). Artificial intelligence for predicting solar still production and comparison with stepwise regression under arid climate, Journal of Water Supply: Research and Technology-Aqua, 66, 166–177.
31. Jang, R.J.S. (1993). Anfis: Adaptive-network-based fuzzy inference system, IEEE Transactions on Systems, Man, and Cybernetics, 23, 665–685.
32. Svalina, I., Galzina, V., Lujić, R., & Šimunović, G. (2013). An adaptive network-based fuzzy inference system (ANFIS) for the forecasting: The case of close price indices, Expert Systems With Applications, 40, 6055–6063.
33. Inan, G., Göktepe, A.B., Ramyar, K., & Sezer, A. (2007). Prediction of sulfate expansion of PC motor using adaptive neuro-fuzzy methodology, Building and Environment, 42, 1264–1269.
34. Wu, J.-D., Hsu, C.-C., & Chen, H.-C. (2009). An expert system of price forecasting for used cars using adaptive neuro-fuzzy inference, Expert Systems with Applications, 36, 7809–7817.
35. Liu, M., & Ling, Y.Y. (2003). Using Fuzzy neural network approach to estimate contractors' markup, Building and Environment, 38, 1303–1308.
36. Jang, J.S.R., & Sun, C.T. (1995). Neuro-fuzzy modeling and control, Proceedings of IEEE, 83, 378–405.
37. Taghavifar, H., & Mardani, A. (2014). On the modeling of energy efficiency indices of agricultural tractor driving wheels applying adaptive neuro-fuzzy inference system, Journal of Terramechanics, 56, 37–47.
38. Xie, Q., Ni, J.Q., & Su, Z. (2017). A prediction model of ammonia emission from a fattening pig room based on the indoor concentration using adaptive neuro fuzzy inference system, Journal of Hazardous Materials, 325, 301–309.