ANFIS modeling and sensitivity analysis for estimating solar still productivity using measured operational and meteorological parameters

Ahmed F. Mashaly and A. A. Alazba

ABSTRACT

This study investigates a potential application of the adaptive neuro-fuzzy inference system (ANFIS) as a relatively new approach for predicting solar still productivity (SSP). Five variables, relative humidity (RH), solar radiation (SR), feed flow rate (MF), and total dissolved solids of feed (TDSF) and brine (TDSB), were used as input parameters. The data were collected from an experimental solar still system used to desalinate seawater in an arid climate. The data were distributed randomly into training, testing, and validation datasets. A hybrid learning algorithm and eight different membership functions were applied to generate the ANFIS models. Several statistical criteria were used to assess the model performances. The ANFIS model with a generalized bell membership function provided the best prediction accuracy compared with models with other membership functions. The coefficient of correlation values for this model were 0.999, 0.959, and 0.832 for training, testing, and validation datasets, respectively. Sensitivity analysis (SA) was used to show the effectiveness of the considered input parameters for predicting SSP. The SA results indicated that SSP is the most influential parameter on SSP. Generally, the findings indicate the robustness of the ANFIS approach for estimating SSP.

Key words | ANFIS, modeling, sensitivity analysis, solar still

Ahmed F. Mashaly (corresponding author)
A. A. Alazba
Alamoudi Water Research Chair,
King Saud University,
Riyadh,
Saudi Arabia
E-mail: *amashaly@ksu.edu.sa*; *mashaly.ahmed@gmail.com*

A. A. Alazba
Agricultural Engineering Department,
King Saud University,
Riyadh,
Saudi Arabia

INTRODUCTION

Universally, water will be a challenge in the near future with increasing global warming and climate change. Many studies have indicated that by 2025, around 1.8 billion people will suffer from severe water scarcity worldwide (Elango *et al.* 2015). The desalination process, as a solution to the problem of water scarcity, is an increasingly used solution. Among desalination techniques, solar stills require simple maintenance and are readily affordable. However, their productivity is low, i.e., the major obstacle to using solar stills is low productivity.

Many investigations to improve solar still productivity (SSP), which ranges between 2.5 and 5 L/m^2/day, have been conducted. Different materials and solar still designs have been tested, and parameters influencing their performance have been evaluated (Rufuss *et al.* 2016; Sharshir *et al.* 2016). Most researchers have investigated thermal performance and thermal behavior with productivity. However, few studies have focused on estimating and predicting productivity. Among the available approaches for this estimation, and computing parameters, the adaptive neuro-fuzzy inference system (ANFIS) is appropriate. ANFIS is relatively new hybrid technology that simultaneously uses the artificial neural network (ANN) and fuzzy logic (FL).

ANN, FL, and ANFIS systems can be used to model complex nonlinear systems, such as solar thermal systems. ANNs can learn and construct accepted solutions to problems after training with experience and examples (Rafiq *et al.* 2001). Applying ANNs for modeling dynamic systems, such as solar stills, has expanded with the development of high-speed computers. Moreover, ANNs have revealed perfect learning

doi: 10.2166/ws.2017.208

and generalizing capabilities, even on completely ambiguous systems that are only characterized by their input–output parameters. FL represents inference techniques capable of reasoning and patterning similar to human reasoning algorithms, which enable knowledge-based systems (Gorzalczany 2002). FL can use the experience of a human expert and also compensate for inadequate and uncertain system estimates. ANFIS takes advantages from both the ANN and FL, combining them so that the ANN determines and optimizes the membership functions and rules of fuzzy systems and no previous system information is required. By combining these capabilities, more versatile and strong models have been developed. These advantages are the main reasons this technique was chosen for this study.

There has been some published research on the application of ANN and FL techniques for solar still performance modeling (Shanmugan 2003; Mamlook & Badran 2007; Shanmugan & Krishnamoorthi 2013; Mashaly *et al.* 2015; Mashaly & Alazba 2015, 2016a, 2016b, 2016c). However, to the best knowledge of the authors, no studies have investigated SSP modeling using an ANFIS modeling tool. Therefore, this study illustrates the modeling capabilities of the ANFIS approach to modeling SSP. The ANFIS technique was employed using data collected from an experimental study in an arid environment to estimate SSP. Eight different membership functions for ANFIS were investigated and the best membership function was selected. The modeling accuracy of the ANFIS technique was estimated and the functions were compared using a statistical comparison. A sensitivity analysis (SA) was also performed and discussed to evaluate the effects of input parameters on the SSP modeling process.

MATERIALS AND METHODS

Experimental configuration

Experiments were conducted at the Agricultural Research and Experiment Station, Department of Agricultural Engineering, King Saud University, Riyadh, Saudi Arabia (24°44′10.90″N, 46°37′13.77″E), during the February to April 2013 period. Weather data were obtained from a weather station (Vantage Pro2, Davis, USA) located close to the experimental site (24°44′12.15″N, 46°37′14.97″E). The experimental solar still system consisted of one C6000 panel stage (F Cubed Ltd, Carocell Solar Panel, Australia). The area of the panel was 6 m^2. The solar still was manufactured as a panel using modern cost-effective materials, such as coated polycarbonate plastic. The panel heated and distilled a film of water flowing over the panel's absorber mat. The panel was fixed at an angle of 29° to horizontal. The basic construction materials were galvanized steel legs, aluminum frame, and polycarbonate covers. The transparent polycarbonate was coated inside using a special coating material to prevent fogging (patented, F Cubed, Australia). A cross-sectional view of the solar still is presented in Figure 1. The techniques for processing water in the experimental solar still are summarized as follows.

The water was fed to the panel using a centrifugal pump (PKm 60, 0.5 HP, Pedrollo, Italy) with a constant flow rate of 10.74 L/hr. Eight drippers and nozzles fed a water film that flowed over the absorbent mat. Under the absorbent mat there was an aluminum screen that helped distribute the dropped water over the absorber mat. Beneath the aluminum screen, there was a plate also made from aluminum. The aluminum was selected for the manufacturing process because it is hydrophilic and assisted in evenly distributing the sprayed water. The water flowed through and over the absorbent mat. As the solar energy was absorbed and partially collected inside the panel, water was heated and hot air naturally circulated within the panel. The hot air flowed in the upper part toward the top, and then reversed direction toward the bottom. Through this circulation, the humid air contacted the cooled surfaces of the transparent polycarbonate cover and bottom polycarbonate layer, thus water condensed and flowed down the panel and collected as a distilled stream. Seawater was used as feed water input to the system. The solar still system was run from February 23 to April 23, 2013. Raw seawater was obtained from the Persian Gulf, Dammam, Eastern Province of Saudi Arabia (26°26′24.19″N, 50°10′20.38″E). The initial total dissolved solids (TDS), pH, density (ρ), and electrical conductivity (EC) of the raw seawater were 41.4 ppt, 8.02, 1.04 $g{\cdot}cm^{-3}$, and 66.34 mS cm^{-1}, respectively. The productivity or amount of distilled water produced (SSP) during a processing time period was obtained by collecting the accumulated water produced over that time. The

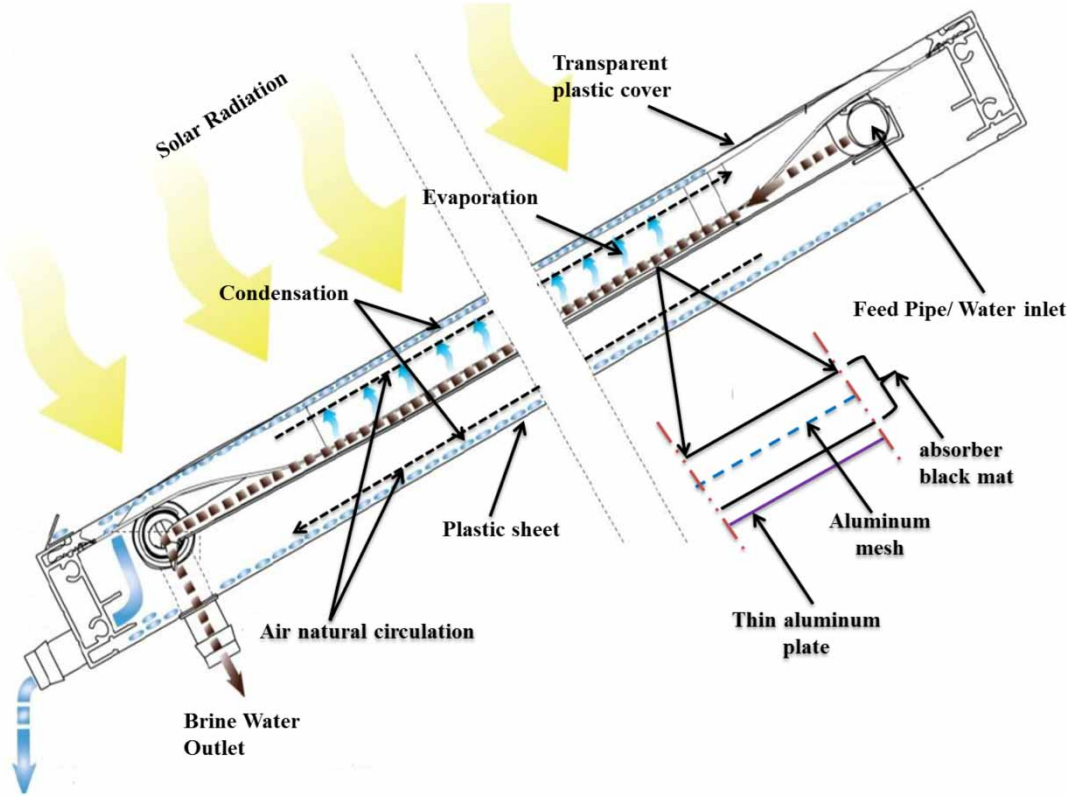

Figure 1 | Cross-sectional view of the solar still used in this study.

temperatures of the feed (TF) and brine (TB) were measured using thermocouples (T-type, UK). Temperature data for feed brine water were recorded on a data logger (177-T4, Testo, Inc., UK) at 1 min intervals. The amount of feed water (MF) was measured using a calibrated digital flow meter mounted on the feed water line (micro-flo, Blue-White, USA). The amounts of brine water and distilled water were measured using a graduated cylinder. TDS and EC were measured using a calibrated (TDS) meter (Cole-Parmer Instrument, Vernon Hills, USA). A pH meter (3510 pH meter, Jenway, UK) was used to determine pH, and ρ was measured with a digital density meter (model: DMA 35_N, Anton Paar, USA). The seawater was fed separately to the panel using a pump, as mentioned previously. The residence time for water passing through the panel was about 20 min. Consequently, the flow rate for feed water, distilled water, and brine water were each measured as 20 min. The total dissolved solids of feed water (TDSF) and brine water (TDSB) were measured every 20 min. Weather data, such as ambient temperature (To), relative humidity (RH), wind speed (WS), and solar radiation (SR), were obtained from the installed weather station, as indicated previously. In summary, there was one dependent variable, SSP, and nine independent variables, To, RH, WS, SR, TF, TB, TDSB, TDSF, and MF. A summary of the experimental data statistics is presented in Table 1.

Adaptive neuro-fuzzy inference system

The ANFIS, which incorporates the best features of the FL and ANN systems, is defined in Jang (1993). In architecture, ANFIS is composed of if–else rules and input–output data couples of FL, and it uses neural network learning algorithms for training. Moreover, ANFIS is an approach to simulate complex nonlinear mappings using 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 to adjust the membership function and associated parameter that approaches the desired datasets.

Table 1 | Summary of experimental data statistics

Statistical index	To	RH	WS	SR	TF	TB	MF	TDSF	TDSB	SSP
Mean	26.64	23.36	2.44	587.55	36.66	49.58	0.21	80.23	95.54	0.50
Minimum	16.87	12.90	0.00	75.10	22.10	27.59	0.13	41.40	46.20	0.05
Maximum	33.23	70.00	12.65	920.69	42.35	68.69	0.25	130.00	132.80	0.97
Standard error	0.29	1.02	0.25	14.38	0.34	0.64	0.00	2.33	2.34	0.02
Median	26.86	18.86	1.30	637.03	37.73	49.19	0.24	73.30	92.15	0.52
Standard deviation	3.68	12.90	3.12	181.93	4.27	8.16	0.04	29.42	29.59	0.24
Kurtosis	−0.56	2.93	3.39	−0.32	−0.02	0.26	−0.73	−1.32	−1.63	−1.06
Skewness	−0.40	1.93	2.03	−0.59	−0.78	0.21	−0.87	0.30	−0.04	−0.08
CV	13.81	55.21	127.53	30.96	11.66	16.45	20.00	36.67	30.98	47.71

To, ambient temperature; RH, relative humidity; WS, wind speed; SR, solar radiation; TF, temperature of feed water; TB, temperature of brine water; MF, feed flow rate; TDSF, total dissolved solids of feed; TDSB, total dissolved solids of brine; SSP, solar still productivity.

The ANFIS learning algorithm is a hybrid learning algorithm comprised of a combination of a back-propagation learning algorithm and least squares method (Liu & Ling 2003; Inan *et al.* 2007; Wu *et al.* 2009; Svalina *et al.* 2013). To understand and simplify the process, a sample having two inputs and an output was considered. Five layers were used to build the ANFIS architecture of the first-order Sugeno-type inference system, which is presented in Figure 2. Two inputs, x and y, and one output, f, along with two fuzzy if–then rules were taken into account as an example. In Figure 2, the circle and square show a fixed node and an adaptive node, respectively. The functions of each layer of the five layers are explained in the following sections. For a first-order Sugeno fuzzy model, two fuzzy if–then rules are as follows:

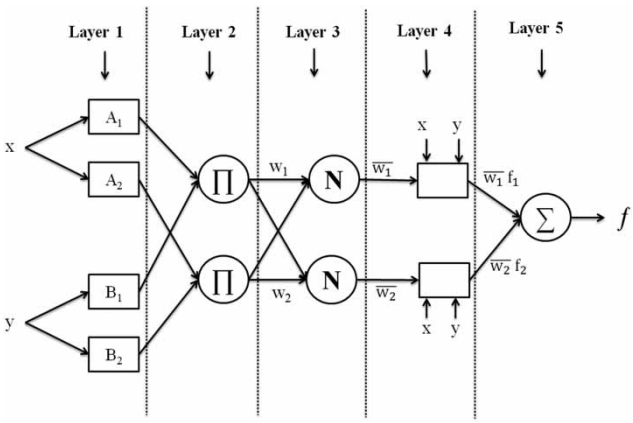

Figure 2 | General ANFIS architecture of the two-input first-order Sugeno-type model with two rules.

Rule 1: If x is A_1 and y is B_1, then $f_1 = p_1x+q_1y+r_1$ (1)

Rule 2: If x is A_2 and y is B_2, then $f_2 = p_2x+q_2y+r_2$ (2)

where, x and y are the inputs, A_1, B_1, A_2, and B_2 are fuzzy sets, and p_1, p_2, q_1, q_2, r_1, and r_2 are the coefficients of the output function 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 as:

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

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

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

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

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

Layer 3 is the normalization layer, i.e., layer of average nodes. In this layer, the ith node is a circle labeled N; the

normalized firing strength is computed as follows:

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

Layer 4 is the defuzzification layer, i.e., 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 as:

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

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 with this formula:

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

Finally, the overall output can be formulated as:

$$f_{\text{out}} = \frac{w_1}{w_1 + w_2} f_1 + \frac{w_2}{w_1 + w_2} f_2 \tag{9}$$

Substituting Equation (7) into Equation (10):

$$f_{\text{out}} = \overline{w_1} f_1 + \overline{w_2} f_2 \tag{10}$$

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

The final output can be written as:

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

ANFIS model development

In this study, the available data obtained from the experiment were randomly divided into three portions: 70% as training data (112 data points) for the learning process, 20% as testing data (32 data points) to test the precision of the model, and 10% for validation (16 data points). Before training, the data were normalized so they fell in the range between 0 and +1 to decrease their range and increase the precision of the findings. After normalization, the data were ready for training. Data normalization was accomplished using the following equation:

$$X_n = \frac{X_i - X_{\min}}{X_{\max} - X_{\min}} \tag{13}$$

where X_n is the normalized value, X_i is the measured value of the variable, $X_{\max}$ is the maximum measured value, and $X_{\min}$ is the minimum measured value.

MATLAB software (MATLAB 8.1.0.604, R2013a, the MathWorks Inc., USA) was used to develop the ANFIS model from the experimental data and forecast the SSP. The Sugeno-type fuzzy inference system was used in the SSP modeling. The grid partition method was employed to classify the input data and make rules (Jang & Sun 1995). In the modeling process, we employed eight different types of input membership functions, including triangular (TRIMF), trapezoidal (TRAPMF), generalized bell (GBELLMF), Gaussian (GAUSSMF), two-sided Gaussian (GAUSS2MF), Pi curve (PIMF), product of two sigmoidal functions (PSIGMF), and difference between two sigmoidal functions (DSIGMF), as presented in Table 2. The output membership function was selected as a linear function. Moreover, a hybrid learning algorithm that combines the least-squares estimator and gradient descent method was used to estimate the optimum values of the FIS parameters of the Sugeno-type (Jang & Sun 1995). Fifty epochs were chosen, owing to their small error.

Evaluation of ANFIS model performance

The performance of the ANFIS model was assessed using the coefficient of correlation (CC), root mean square error (RMSE), efficiency coefficient (ME), overall index of model performance (OI), coefficient of residual mass (CRM), mean absolute error (MAE), and mean absolute relative error (MARE). Detailed descriptions of these statistical parameters for measuring ANFIS model performance are available in Mashaly & Alazba (2016a, 2016b, 2016c). The CC, RMSE, ME, OI, CRM, MAE, and MARE were calculated as follows:

$$CC = \frac{\sum_{i=1}^{n} (SSP_{o,i} - \overline{SSP}_o)(SSP_{p,i} - \overline{SSP}_p)}{\sqrt{\sum_{i=1}^{n} (SSP_{o,i} - \overline{SSP}_o)^2 \times \sum_{i=1}^{n} (SSP_{p,i} - \overline{SSP}_p)^2}} \tag{14}$$

Table 2 | Input membership functions used in the modeling process

Membership functions	Mathematical expression	Description/Purpose
TRIMF	$f(x;a,b,c) = \begin{cases} 0, & x \le a \\ \frac{x-a}{b-a}, & a \le x \le b \\ \frac{c-x}{c-b}, & b \le x \le c \\ 0, & c \le x \end{cases}$	Triangular membership function
TRAPMF	$f(x;a,b,c,d) = \begin{cases} 0, & x \le a \\ \frac{x-a}{b-a}, & a \le x \le b \\ 1, & b \le x \le c \\ \frac{d-x}{d-c}, & c \le x \le d \\ 0, & d \le x \end{cases}$	Trapezoidal membership function
GBELLMF	$f(x;a,b,c) = \frac{1}{1+\left\|\frac{(x-c)}{a}\right\|^{2b}}$	Generalized bell curve membership function
GAUSSMF	$f(x;\sigma,c) = e^{\frac{-(x-c)^2}{2\sigma^2}}$	Gaussian curve membership function
GAUSS2MF	$f(x;\sigma,c) = e^{\frac{-(x-c)^2}{2\sigma^2}}$	Two-sided Gaussian curve membership function
PIMF	$f(x;a,b,c,d) = \begin{cases} 0, & x \le a \\ 2\left(\frac{x-a}{b-a}\right)^2, & a \le x \le \frac{a+b}{2} \\ 1-2\left(\frac{x-a}{b-a}\right)^2, & \frac{a+b}{2} \le x \le b \\ 1, & b \le x \le c \\ 1-2\left(\frac{x-c}{d-c}\right)^2, & c \le x \le \frac{c+d}{2} \\ 2\left(\frac{x-d}{d-c}\right)^2, & \frac{c+d}{2} \le x \le d \\ 0, & x \ge d \end{cases}$	Pi-shaped curve membership function
PSIGMF	$f(x;a,c) = \frac{1}{1+e^{-a(x-c)}}$	Product of two sigmoidal membership functions
DSIGMF	$f(x;a,c) = \frac{1}{1+e^{-a(x-c)}}$	Difference of two sigmoidal membership functions

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}\left(SSP_{o,i} - SSP_{p,i}\right)^2}{n}} \quad (15)$$

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

$$OI = \frac{1}{2}\left(1 - \frac{RMSE}{SSP_{max} - SSP_{min}} + ME\right) \quad (17)$$

$$MAE = \frac{\sum_{i=1}^{n}\left|SSP_{o,i} - SSP_{p,i}\right|}{n} \quad (18)$$

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

where $SSP_{o,i}$ is the measured value, $SSP_{p,i}$ is the predicted value, $\overline{SSP}_{o}$ is the mean of measured values, $\overline{SSP}_{p}$ is the mean of predicted values, SSP_{max} is the maximum measured value, SSP_{min} is the minimum measured value, and n is the number of observations.

RESULTS AND DISCUSSION

Summary statistics and data description

Table 1 provides a statistical summary of the experimental data (To, RH, WS, SR, TF, TB, TDSB, TDSF, MF, and SSP), including maximum, minimum, mean, standard error, median, standard deviation, kurtosis, skewness, and

coefficient of variation (CV) for each parameter. From Table 1, the distribution curves for To, RH, SR, TF, TB, MF, TDSF, TDSB, and SSP are platykurtic, because the kurtosis values were less than 3. However, the distribution curve was leptokurtic with WS, as the kurtosis value was more than 3. The distributions were highly skewed for RH and WS, with skewness values more than +1. The skewness values for SR, TF, and MF were between −1 and −1/2, so the distributions for these variables were moderately skewed. Moreover, the distributions were approximately symmetric for To, TB, TDSF, TDSB, and SSP because the skewness values were between −1/2 and +1/2. From the CV values, it is clear that the To, TF, and TB data were relatively homogeneous ($0.10 \leq CV < 0.20$). In addition, the MF data were relatively heterogeneous ($0.20 \leq CV < 0.30$). Finally, the RH, WS, SR, TDSF, TDSB, and SSP data were heterogeneous ($CV \geq 0.30$).

Table 3 shows the Pearson correlation coefficient (CC) values between parameters. The strongest correlation was found between the SSP and SR, with a CC value of +0.73. Furthermore, SSP was found to be well correlated with TDSF, with a CC of −0.40. The + and − signs refer to positive and negative correlations, respectively. Similar findings have also been reported in the literature (Dev *et al.* 2011; Mashaly *et al.* 2016; Sharshir *et al.* 2016). Furthermore, there was a significant correlation between the SSP and TDSB, with a CC of −0.17. In contrast, only very weak correlations were found between the SSP and To and TF, with CC values of −0.07 and −0.06, respectively; consequently, we did not consider them as input parameters. The correlation analysis also led to the exclusion of WS and TB due to their high collinearity with other parameters, although they had significant correlations with the SSP. Although some parameters also appeared correlated with others, these were included in the modeling process because their inclusion improved the prediction performance, primarily by enhancing the coefficient of determination (CC^2). The same argument was also invoked to consider RH as input parameter with low CC.

The results from the experiments also revealed that the most dominant meteorological parameter affecting the SSP was the SR. It was also found that RH was inversely proportional to the SSP because low RH (drier air) tends to increase the evaporation rate. As the MF increased, the SSP decreased, and conversely, as the TDSF and TDSB decreased, the SSP increased. The reason why the evaporation rate increased can be attributed to the weakness of the ionic bonds for the low TDSF and TDSB. A more complete illustration of these experimental data is provided by Mashaly *et al.* (2016).

ANFIS model structure

In this study, we developed eight ANFIS models with eight different types of input membership functions. The membership functions were TRIMF, TRAPMF, GBELLMF,

Table 3 | Correlation coefficient matrix for the studied parameters

	To	RH	WS	SR	TF	TB	MF	TDSF	TDSB	SSP
To	1.00									
RH	−0.66	1.00								
WS	−0.14	−0.08	1.00							
SR	−0.15	0.15	0.22	1.00						
TF	0.91	−0.80	−0.01	0.09	1.00					
TB	0.06	0.05	0.33	0.82	0.13	1.00				
MF	0.44	−0.72	−0.34	−0.27	0.48	−0.40	1.00			
TDSF	−0.01	0.23	0.64	0.22	0.06	0.49	−0.75	1.00		
TDSB	−0.15	0.45	0.49	0.39	−0.11	0.57	−0.84	0.94	1.00	
SSP	−0.07	0.01	−0.31	0.73	−0.06	0.40	0.25	−0.40	−0.17	1.00

To, ambient temperature; RH, relative humidity; WS, wind speed; SR, solar radiation; TF, temperature of feed water; TB, temperature of brine water; MF, feed flow rate; TDSF, total dissolved solids of feed; TDSB, total dissolved solids of brine; SSP, solar still productivity.

GAUSSMF, GAUSS2MF, PIMF, DSIGMF, and PSIGMF. The developed ANFIS models had five inputs, RH, SR, MF, TDSF, and TDSB, and one output, SSP. The structure of the ANFIS models with five input parameters is presented in Figure 3. Therefore, in the input layer, five neurons were incorporated. For each neuron, three identical membership functions were considered with three linguistic terms {low, medium, high} and accordingly, 243 ($3\times3\times3\times3\times3$) rules were developed to implement the ANFIS model. The structures of the ANFIS models are listed in Table 4. The four ANFIS models were trained, tested, and validated to assess the predictability of the SSP using these models. The next section will discuss and evaluate the performance of the eight developed ANFIS models in the training, testing, and validation stages using statistical performance indicators.

ANFIS model performance and optimal model selection

This section discusses the eight developed ANFIS models and provides additional details oriented on optimizing the ANFIS model. Table 5 provides the statistical parameter results, CC, RMSE, ME, OI, MAE, and MARE, which are numerical indicators that describe the agreement between measured and predicted SSP values during the training, testing, and validation stages. In the training stage, the ANFIS models' CC values ranged from 99.98% to 99.99%, RMSE values from 0.0007 to 0.0048 L/m^2/h, ME values from 99.96% to 99.99%, OI values from 99.72% to 99.96%, MAE values from 0.0005 to 0.0027 L/m^2/h, and MARE values from 0.1424% to 0.8857%. The CC, ME, and OI values were very close to 1 while RMSE, MAE and MARE values were close to 0, indicating excellent agreement between the measured results and predicted results from the ANFIS models during the training stage. The tiny deviations between measured and predicted results in turn highlight the effectiveness of the ANFIS technique in the SSP prediction process. Similar results have also been obtained in the literature (Inan *et al.* 2007; Mohammadi *et al.* 2016; Yaïci & Entchev 2016; Keshavarzi *et al.* 2017).

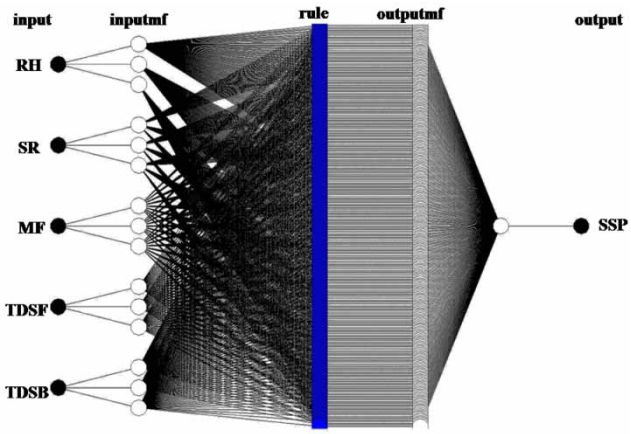

Figure 3 | Structure of the ANFIS model used in this study.

During the testing process, the ANFIS models had CC values ranging from 89.96% to 95.96%, RMSE values from 0.068 to 0.114 L/m^2/h, ME values from 0.7487 to 0.9106, OI values from 0.8056 to 0.9143, MAE values from 0.0448 to 0.0779 L/m^2/h, and MARE values from 12.6466% to 20.4389%. As shown in Table 5, the ANFIS models with high CC, ME, and OI values and low RMSE, MAE, and MARE values predicted the SSP values with an acceptable degree of accuracy during the testing stage. The ANFIS models with GAUSS2MF (CC = 95.96%, RMSE = 0.068 L/m^2/h, ME = 0.9106, OI = 0.9143, MAE = 0.0483 L/m^2/h, and MARE = 12.8264) and with GBELLMF (CC = 95.93%, RMSE = 0.0703 L/m^2/h, ME = 0.9045, OI = 0.9099, MAE = 0.0448 L/m^2/h, and MARE = 12.7663%) produced better results. The ANFIS model with GAUSSMF (CC = 90.17%, RMSE = 0.114 L/m^2/h, ME = 0.7487, OI = 0.8056, MAE = 0.0779 L/m^2/h, and MARE = 20.4389%) performed slightly

Table 4 | ANFIS model structure for SSP modeling

ANFIS parameters	TRIMF	TRAPMF	GBELLMF	GAUSSMF	GAUSS2MF	PIMF	PSIGMF	DSIGMF
Number of nodes	524	524	524	524	524	524	524	524
Number of linear parameters	1,458	1,458	1,458	1,458	1,458	1,458	1,458	1,458
Number of nonlinear parameters	45	60	45	30	60	60	60	60
Total number of parameters	1,503	1,518	1,503	1,488	1,518	1,518	1,518	1,518
Number of fuzzy rules	243	243	243	243	243	243	243	243

Table 5 | Statistical performance of the developed ANFIS models with different types of input membership functions during the training, testing, and validation stages

	Statistical parameters					
Membership functions	**CC**	**RMSE**	**ME**	**OI**	**MAE**	**MARE**
TRAINING						
TRIMF	0.9999	0.0040	0.99970	0.9977	0.0027	0.6376
TRAPMF	0.9998	0.0048	0.99960	0.9972	0.0026	0.8857
GBELLMF	0.9999	0.0013	0.99997	0.9993	0.0007	0.1519
GAUSSMF	0.9999	0.0014	0.99997	0.9992	0.0007	0.1498
GAUSS2MF	0.9999	0.0012	0.99998	0.9994	0.0007	0.2350
PIMF	0.9999	0.0041	0.99970	0.9976	0.0022	0.7844
PSIGMF	0.9999	0.0007	0.99999	0.9996	0.0005	0.1424
DSIGMF	0.9999	0.0007	0.99999	0.9996	0.0005	0.1424
TESTING						
TRIMF	0.8996	0.1109	0.7619	0.8140	0.0686	17.3199
TRAPMF	0.9265	0.0896	0.8446	0.8682	0.0549	12.6466
GBELLMF	0.9593	0.0703	0.9045	0.9099	0.0448	12.7663
GAUSSMF	0.9017	0.1140	0.7487	0.8056	0.0779	20.4389
GAUSS2MF	0.9596	0.0680	0.9106	0.9143	0.0483	12.8264
PIMF	0.9367	0.0871	0.8532	0.8741	0.0529	14.8029
PSIGMF	0.9227	0.0961	0.8211	0.8525	0.0612	16.5259
DSIGMF	0.9227	0.0961	0.8211	0.8525	0.0612	16.5259
VALIDATION						
TRIMF	0.8861	0.1064	0.6920	0.7564	0.0745	16.9053
TRAPMF	0.7950	0.1306	0.5361	0.6581	0.0720	14.5901
GBELLMF	0.8317	0.1263	0.5657	0.6765	0.0798	17.1074
GAUSSMF	0.8980	0.1022	0.7157	0.7718	0.0770	17.2837
GAUSS2MF	0.7413	0.1992	−0.0804	0.2921	0.1116	25.0306
PIMF	0.8458	0.1195	0.6110	0.7049	0.0675	14.5509
PSIGMF	0.7629	0.1541	0.3538	0.5472	0.0792	18.2047
DSIGMF	0.7629	0.1541	0.3538	0.5472	0.0792	18.2047

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

worse. However, the overall performance of the ANFIS model with GAUSS2MF was the best in the testing process.

In the validation process, the average values of ANFIS models with different membership functions were 81.55%, 0.1366 L/m^2/h, 0.4685, 0.6193, 0.080 L/m^2/h, and 17.735% for CC, RMSE, ME, OI, MAE, and MARE, respectively, as shown in Table 5. The performance of the ANFIS model with GAUSSMF (CC = 89.80%, RMSE = 0.1022 L/m^2/h, ME = 0.7157, OI = 0.7718, MAE = 0.077 L/m^2/h, and MARE = 17.2837%) was somewhat better than the other membership functions. Moreover, the TRIMF obtained good results, but slightly worse than the GAUSSMF. Generally, these results indicated good agreement between the measured results and predicted results using the ANFIS model with GAUSSMF during the validation process.

The ANFIS models with different membership functions obtained high accuracy in the SSP modeling process, which is clearly reflected in the statistical parameter values displayed in Table 3. Overall, in all modeling stages, the highest prediction capability was achieved by the ANFIS model with GBELLMF, followed by those with PIMF, GAUSSMF, TRIMF, TRAPMF, PSIGMF, DSIGMF, and GAUSS2MF.

These results are in agreement with the findings of Khoshnevisan *et al.* (2014) and Pourtousi *et al.* (2015). A scatter plot with a line of agreement between the SSP simulated with the ANFIS model and measured SSP values is shown in Figure 4 for the ANFIS model with GBELLMF during the training, testing, and validation processes. As shown, the data were evenly and tightly distributed around the 1:1 line. There was a very close visual agreement between the measured and predicted results with the ANFIS model. This is also shown in the statistical values presented in Table 5.

Sensitivity analysis (SA)

SA is a technique to determine the contribution and relative importance of parameters in the modeling process. It is also used for selecting an input parameter for future experiments, model developments, or field investigations. SA evaluates and describes how model output values are affected by changes in input values. The cosine amplitude method (CAM) is a powerful way to implement an SA (Ross 2010). In this method, CAM determines the significance of each input parameter. Assuming that n data samples are collected from a common data array, X, then the data pairs used to construct a data array X are defined as:

$$X = \{x_1, x_2, x_3, \ldots x_m\} \tag{20}$$

Each of the elements, x_i, in the data array X is a vector of length m, that is:

$$x_i = \{x_{i1}, x_{i2}, x_{i3}, \ldots x_{im}\} \tag{21}$$

Accordingly, each of the data samples can be considered as a point in m-dimensional space, where each point requires m coordinates for a comprehensive description. Each element of a relation, R_{ij}, results from a pairwise comparison. In the CAM technique, the strength of the relationship between the data pairs, x_i and x_j, is estimated and demonstrated using the following equation:

$$R_{ij} = \frac{\sum_{k=1}^{m} x_{ik} x_{jk}}{\sqrt{\left(\sum_{k=1}^{m} x_{ik}^2\right)\left(\sum_{k=1}^{m} x_{jk}^2\right)}} \tag{22}$$

where $i, j = 1, 2, \ldots, n$.

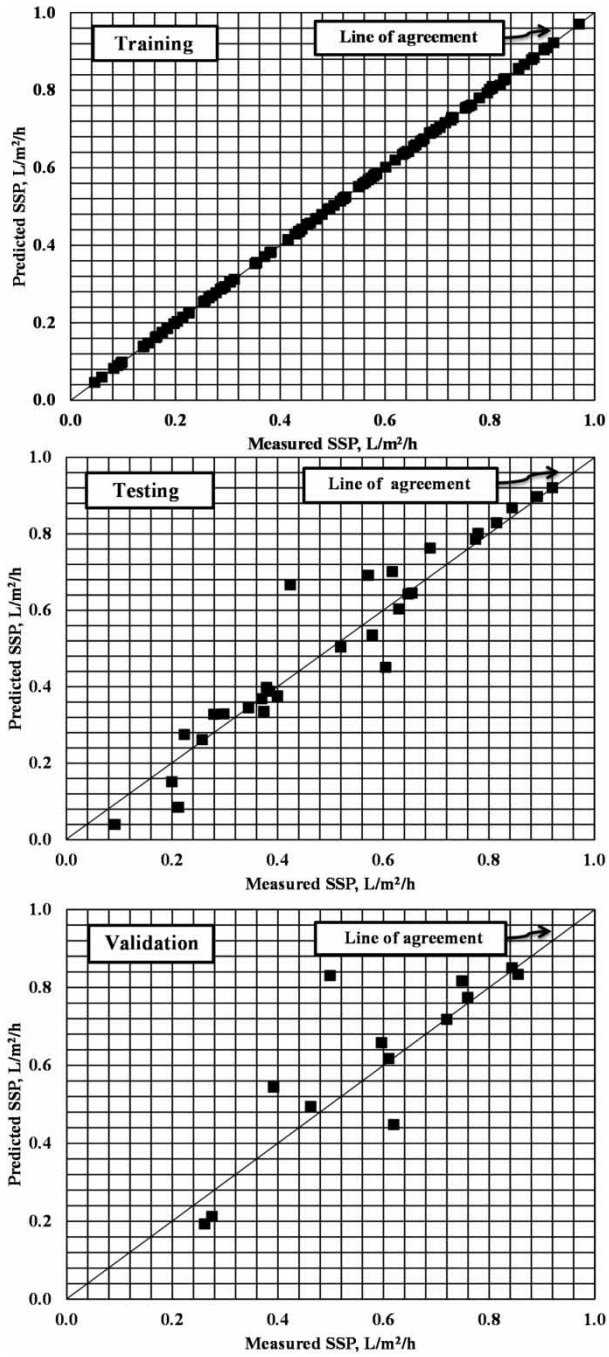

Figure 4 | Predicted versus measured SSP during the training, testing, validation processes using the generalized bell membership function (GBELLMF).

Values of the strength ratio (R_{ij}), which are close to 1, denote the effect and influence of the input parameter on the model output. Increasing values indicate that an input parameter is more effective than other parameters. Using Equation (22), a series of SA was conducted on the input

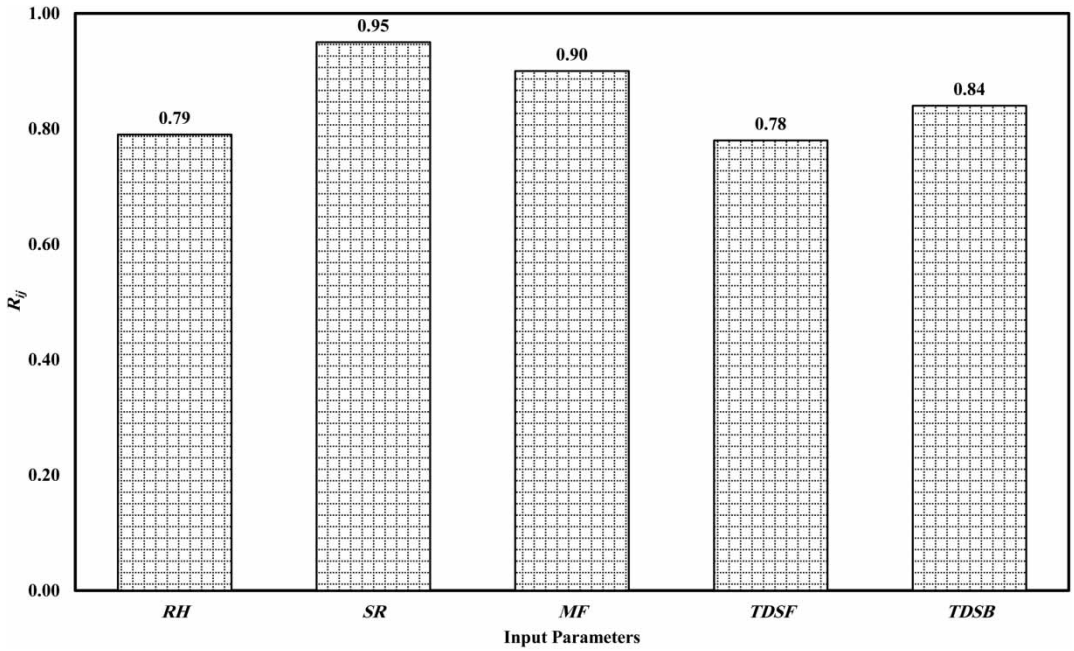

Figure 5 | SA between SSP and each input parameter.

and output parameters. The R_{ij} values between the SSP values predicted with the ANFIS model with GBELLMF and related input parameters using the CAM technique are presented in Figure 5. As shown, SR had the most influence on the SSP, whereas TDSF had the least effect. In addition, the R_{ij} values were almost identical between RH and TDSF. Generally, the obtained R_{ij} values for all parameters were not significantly different and were close to 1, which indicates that all parameters strongly contributed to the developed ANFIS model. Furthermore, this indicates that all input parameters were important in predicting SSP and none should be deleted. The obtained findings are in agreement with other studies (Badran 2001; Dev *et al.* 2011; Mahdi *et al.* 2011; Abujazar *et al.* 2016; Mashaly *et al.* 2016; Sharshir *et al.* 2016; Panchal & Patel 2017).

CONCLUSION

In this study, the structure and potential of ANFIS methodology was assessed for predicting and estimating SSP. The ANFIS technique solves prediction issues with a high degree of precision using human knowledge and experience in modeling and eliciting rules from data without prior assumptions. Five variables were used as input parameters, RH, SR, MF, TDSF, and TDSB. Eight ANFIS membership functions were trained, tested, and validated to identify an optimum supervised ANFIS model for SSP prediction. The performances of the eight ANFIS models were evaluated by comparing the predicted results with experimental results using a variety of standard statistical performance indicators, CC, RMSE, ME, OI, MAE, and MARE. The optimal ANFIS model was developed with the generalized bell-shaped built-in membership function (GBELLMF). An SA method, i.e., CAM, was employed to evaluate the contribution of each parameter in the modeling process. The SA of the ANFIS model results revealed that the most effective parameters on the SSP were SR and MF, whereas TDSF was the least effective parameter. In summary, this study proved that the ANFIS method can be employed as a reliable and powerful method for solving highly nonlinear engineering problems, such as estimating the SSP.

ACKNOWLEDGMENT

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

REFERENCES

Abujazar, M. S. S., Fatihah, S., Rakmi, A. R. & Shahrom, M. Z. 2016 The effects of design parameters on productivity performance of a solar still for seawater desalination: a review. *Desalination* **385**, 178–193.

Badran, A. A. 2001 Inverted trickle solar still: effect of heat recovery. *Desalination* **133** (2), 167–173.

Dev, R., Abdul-Wahab, S. A. & Tiwari, G. N. 2011 Performance study of the inverted absorber solar still with water depth and total dissolved solid. *Applied Energy* **88**, 252–264.

Elango, C., Gunasekaran, N. & Sampathkumar, K. 2015 Thermal models of solar still – a comprehensive review. *Renewable and Sustainable Energy Reviews* **47**, 856–911.

Gorzalczany, M. B. 2002 *Computational Intelligence Systems and Applications*. Springer, Heidelberg, Germany.

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.

Jang, J.-S. R. 1993 ANFIS: adaptive-network-based fuzzy inference system. *IEEE Transactions on Systems, Man, and Cybernetics* **23**, 665–685.

Jang, J.-S. R. & Sun, C. T. 1995 Neuro-fuzzy modeling and control. *Proceedings of the IEEE* **83** (3), 378–406.

Keshavarzi, A., Sarmadian, F., Shiri, J., Iqbal, M., Tirado-Corbalá, R. & Omran, E. S. E. 2017 Application of ANFIS-based subtractive clustering algorithm in soil Cation Exchange Capacity estimation using soil and remotely sensed data. *Measurement* **95**, 173–180.

Khoshnevisan, B., Rafiee, S., Omid, M. & Mousazadeh, H. 2014 Prediction of potato yield based on energy inputs using multi-layer adaptive neuro-fuzzy inference system. *Measurement* **47**, 521–530.

Liu, M. & Ling, Y. Y. 2003 Using fuzzy neural network approach to estimate contractors' markup. *Building and Environment* **38**, 1303–1308.

Mahdi, J. T., Smith, B. E. & Sharif, A. O. 2011 An experimental wick-type solar still system: design and construction. *Desalination* **267** (2–3), 233–238.

Mamlook, R. & Badran, O. 2007 Fuzzy sets implementation for the evaluation of factors affecting solar still production. *Desalination* **203** (1–3), 394–402.

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** (4), 480–493.

Mashaly, A. F. & Alazba, A. A. 2016a Comparison of ANN, MVR, and SWR models for computing thermal efficiency of a solar still. *International Journal of Green Energy* **13** (10), 1016–1025.

Mashaly, A. F. & Alazba, A. A. 2016b MLP and MLR models for instantaneous thermal efficiency prediction of solar still under hyper-arid environment. *Computers and Electronics in Agriculture* **122**, 146–155.

Mashaly, A. F. & Alazba, A. A. 2016c Neural network approach for predicting solar still production using agricultural drainage as a feedwater source. *Desalination and Water Treatment* **57** (59), 28646–28660.

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.

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** (26), 12019–12036.

Mohammadi, K., Shamshirband, S., Kamsin, A., Lai, P. C. & Mansor, Z. 2016 Identifying the most significant input parameters for predicting global solar radiation using an ANFIS selection procedure. *Renewable and Sustainable Energy Reviews* **63**, 423–434.

Panchal, H. N. & Patel, S. 2017 An extensive review on different design and climatic parameters to increase distillate output of solar still. *Renewable and Sustainable Energy Reviews* **69**, 750–758.

Pourtousi, M., Sahu, J. N., Ganesan, P., Shamshirband, S. & Redzwan, G. 2015 A combination of computational fluid dynamics (CFD) and adaptive neuro-fuzzy system (ANFIS) for prediction of the bubble column hydrodynamics. *Powder Technology* **274**, 466–481.

Rafiq, M. Y., Bugmann, G. & Easterbrook, D. J. 2001 Neural network design for engineering applications. *Computers & Structures* **79** (17), 1541–1552.

Ross, T. J. 2010 *Fuzzy Logic with Engineering Applications,* 3rd edn. Wiley, Hoboken, NJ, USA.

Rufuss, D. D. W., Iniyan, S., Suganthi, L. & Davies, P. A. 2016 Solar stills: a comprehensive review of designs, performance and material advances. *Renewable and Sustainable Energy Reviews* **63**, 464–496.

Shanmugan, S. 2003 Fuzzy logic modeling of floating cum tilted-wick solar still. *International Journal of Recent Scientific Research* **4** (5), 579–582.

Shanmugan, S. & Krishnamoorthi, G. 2013 Fuzzy logic modeling of single slope single basin solar still. *International Journal of Fuzzy Mathematics and Systems* **3** (2), 125–134.

Sharshir, S. W., Yang, N., Peng, G. & Kabeel, A. E. 2016 Factors affecting solar stills productivity and improvement techniques: a detailed review. *Applied Thermal Engineering* **100**, 267–284.

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** (15), 6055–6063.

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.

Yaïci, W. & Entchev, E. 2016 Adaptive neuro-fuzzy inference system modelling for performance prediction of solar thermal energy system. *Renewable Energy* **86**, 302–315.

First received 18 July 2017; accepted in revised form 3 October 2017. Available online 24 October 2017