

# Artificial intelligence for predicting solar still production and comparison with stepwise regression under arid climate

Ahmed F. Mashaly and A. A. Alazba

## ABSTRACT

Forecasting the efficiency of solar still production (SSP) can reduce the capital risks involved in a solar desalination project. Solar desalination is an attractive method of water desalination and offers a more reliable water source. In this study, to estimate SSP, we employed the data obtained from experimental fieldwork. SSP is assumed to be a function of ambient temperature, relative humidity, wind speed, solar radiation, feed flow rate, temperature of feed water, and total dissolved solids in feed water. In this study, back-propagation artificial neural network (ANN) models with two transfer functions were adopted for predicting SSP. The best performance was obtained by the ANN model with one hidden layer having eight neurons which employed the hyperbolic transfer function. Results of the ANN model were compared with those of stepwise regression (SWR) model. ANN model produced more accurate results compared to SWR model in all modeling stages. Mean values for the coefficient of determination and root mean square error by ANN model were 0.960 and 0.047 L/m<sup>2</sup>/h, respectively. Relative errors of predicted SSP values by ANN model were about ±10%. In conclusion, the ANN model showed greater potential in accurately predicting SSP, whereas the SWR model showed poor performance.

**Key words** | artificial neural network, modeling, solar still production, stepwise regression

Ahmed F. Mashaly (corresponding author)

A. A. Alazba

Alamoudi Water Research Chair,  
King Saud University,  
Riyadh,  
Saudi Arabia  
E-mail: mashaly.ahmed@gmail.com

A. A. Alazba

Agricultural Engineering Department,  
King Saud University,  
Riyadh,  
Saudi Arabia

## INTRODUCTION

Solar stills are widely used in solar desalination. Nonetheless, solar still production (SSP) is very low. Accordingly, increasing SSP has been the focus of intensive study. Much of the literature has focused on experiments to find a better design for solar stills to improve the SSP (e.g., Tanaka & Nakatake 2009; Kabeel *et al.* 2013; Ayoub *et al.* 2015; Koilraj Gnanadason *et al.* 2015). Usually, these experimental studies are costly, laborious and time-consuming. Consequently, mathematical modeling (MM) may be the best alternative for finding better designs. MM is one of the most effective methods for providing a clear obvious understanding of solar still behavior and enhancing SSP. Moreover, MM by artificial intelligence (AI) gives the most accurate results and is much faster than classical MM.

Artificial neural networks (ANNs), which are one of these AI technologies, are mathematical models that try to mimic the structures and functions of biological neural networks used for solving complex problems (Ali Abdoli *et al.* 2012; Behboudian *et al.* 2014; Nezhad *et al.* 2016). ANNs have been used in various ways to model and predict the performance of various desalination and solar desalination systems ranging from predicting solar still performance (Santos *et al.* 2012; Mashaly & Alazba 2015, 2016a, 2016b, 2016c; Mashaly *et al.* 2015) to optimizing the performance of solar-powered membrane distillation units (Porrizzo *et al.* 2013), to simulating the reverse osmosis desalination process (Khayet *et al.* 2011). These networks have also been able to control multi-stage flash desalination plants (Tayyebi

& Alishiri 2014) as well as analyze seawater desalination systems (Gao *et al.* 2007).

However, SSP is one of the most important elements in the solar desalination process and an essential tool used to analyze solar still performance. Hence, it is necessary to investigate and predict SSP and also consider the effects of meteorological and operational parameters on it. The objective of this research was to develop SSP models using both an ANN model and a stepwise regression (SWR) model. The two approaches were compared in terms of model performance. Also, an assessment of the performance of the ANN and SWR models using a statistical comparison was conducted.

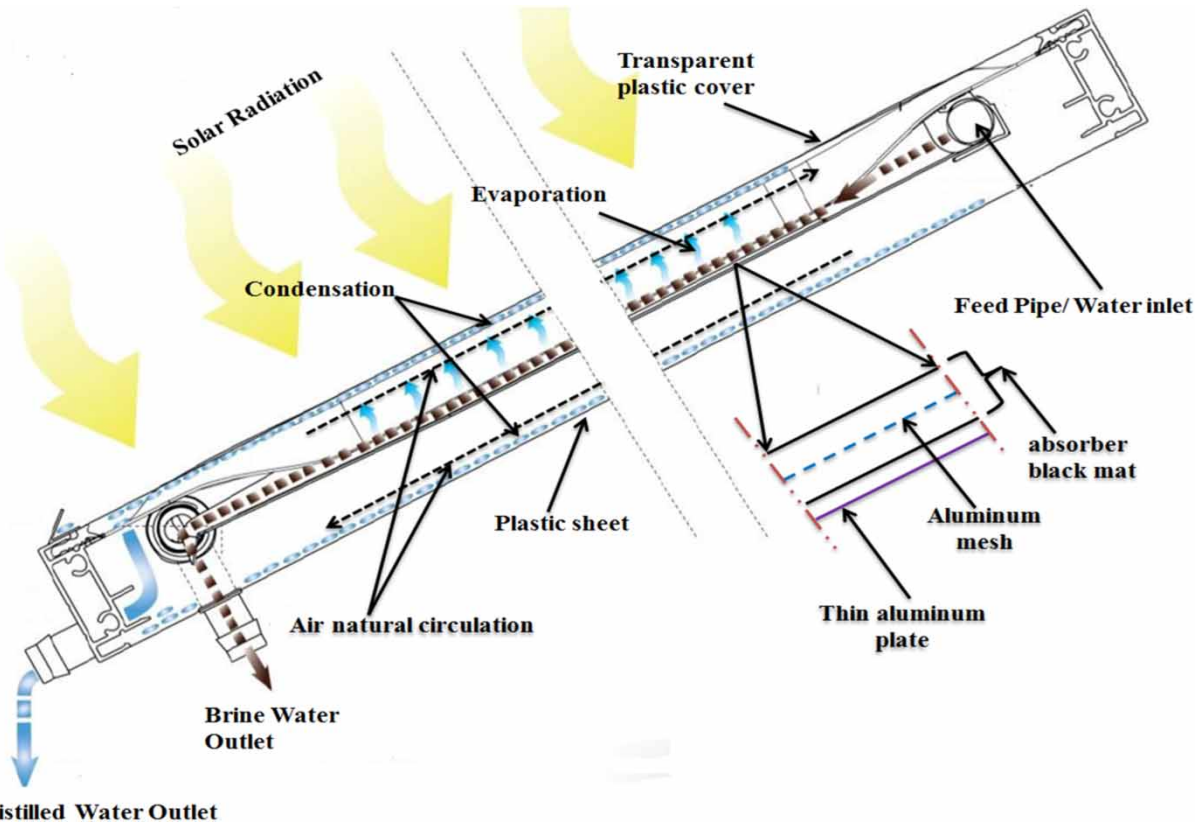
## MATERIALS AND METHODS

### Experimental procedure

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; Davis, USA) 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<sup>2</sup> single stage of C6000 panel (F Cubed, Carocell Solar Panel, Australia). The solar-still panel was 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 an angle of 29° to the 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). A cross-sectional view of the solar still is presented in Figure 1. The working idea of the available system is summarized in the following paragraphs.

The water was fed to the panel using a centrifugal pump (model PKm 60, 0.5 HP, Pedrollo, Italy) with a constant flow rate of 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 flows 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 towards 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 feed water input to the system. The solar still 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 total dissolved solids (TDS) in the three types of water along with their pH, density ( $\rho$ ), and electrical conductivity (EC) were 41.4 ppt, 8.02, 1.04 g.cm<sup>-3</sup>, and 66.34 mS cm<sup>-1</sup>, respectively. The production or the amount of distilled water produced (SSP) during a time period by the system was obtained by collecting the cumulative amount of water produced over time. The temperature of the feed water ( $T_F$ ) was measured by using thermocouples (T-type, UK). Temperature data for feed brine water were recorded on a data logger (model 177-T4, Testo, Inc., UK) at 1 min intervals. The amount of feed water ( $M_F$ ) was measured by calibrated digital flow meter mounted on the feed water line (micro-flo, Blue-White, USA). The amount of brine water and distilled water were measured by graduated cylinder. TDS concentration and EC were measured using a TDS-calibrated meter (Cole-Parmer Instrument, Vernon Hills, USA). A pH meter (model 3510 pH meter, Jenway, UK) was used to measure pH. A digital-density meter (model DMA 35N, Anton Paar, USA) was used to measure  $\rho$ . 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 the panel – was approximately 20 minutes. Therefore, the flow rate of the feed water, the distilled water, and the brine water was measured every 20 minutes. Also, the TDS of feed water ( $TDS_F$ ) were measured every 20 minutes. The weather data,



**Figure 1** | Cross-sectional view of the solar still.

such as ambient temperature ( $T_o$ ), relative humidity ( $RH$ ), wind speed ( $WS$ ), and solar radiation ( $SR$ ), were obtained from the weather station mentioned above. Here, there is one dependent variable, which was the SSP, and seven independent variables, which are  $T_o$ ,  $RH$ ,  $WS$ ,  $SR$ ,  $TDS_F$ ,  $M_F$ , and  $T_F$ .

### Artificial neural network

A particular architecture of ANNs is feed-forward back-propagation neural network (FFBNN), which is popular and commonly applied in many applications, including making predictions. The input parameters are presented to the network through the input layer, which acts as a buffer layer. The number of neurons in this layer is equal to the input/independent parameters. These neurons are connected to the neurons of the adjacent layer and the hidden layer, and all interconnections have a weight ( $W$ ) which represents the strength of the connection. In this respect, in the

hidden layer a summation of weights is passed through a function called the transfer function which decides the relationship between the inputs and output (Eren *et al.* 2012; Valderrama *et al.* 2015; Wadi Abbas Al-Fatlawi *et al.* 2015). The output layer neuron ( $Y_k$ ) is expressed as:

$$Y_k = f \left( \sum_{j=1}^n W_{kj} h_j + B_k \right) \quad (1)$$

where  $f$  is the transfer function;  $W_{kj}$  are the weights from the hidden layer to the output layer;  $B_k$  are the biases in the output layer; and  $h_j$  is the neuron's activation value in the hidden layer mathematically expressed as (Haykin 1999):

$$h_j = f \left( \sum_{i=1}^n W_{ji} X_i + B_j \right) \quad (2)$$

where  $W_{ji}$  are the weights from the input layer to the hidden layer;  $X_i$  are the inputs; and  $B_j$  are the biases in the hidden

layer. The transfer function could be sigmoid (SIG) as shown in Equation (3) or hyperbolic tangent (TANH) as displayed in Equation (4):

$$f(t) = \frac{1}{1 + \exp(-t)} \quad (3)$$

$$f(t) = \frac{1 - \exp(-2t)}{1 + \exp(-2t)} \quad (4)$$

### ANN model development

Commercial Qnet 2000 software was employed in this study to develop the ANN model from the experimental data to predict SSP. The input parameters of the ANN model were  $T_o$ ,  $RH$ ,  $WS$ ,  $SR$ ,  $TDS_F$ ,  $M_F$ , and  $T_F$ . The output parameter was SSP. The obtained data points from the experimental work were randomly used to train (70% of the data points), test (20% of the data points), and validate (10% of the data points) the ANN model. The training, testing, and validation sets have 112, 32, and 16 data points, respectively. In the validation stage, the data used to check the performance of the ANN model are not used in the training stage. Different ANN architectures with one hidden layer were trained. The optimal number of neurons in the hidden layer was determined by trial-and-error procedure, in which the number of neurons in the hidden layer was varied from 1 to 10 to find the best ANN architecture. The training iteration was fixed to 200,000. The learn rate and momentum factor were fixed and were 0.01 and 0.8, respectively. Before the training process, all of the input and output parameters were automatically normalized between 0.15 and 0.85 by the used software. The normalization process accelerates the ANN training, and increases the network's generalization capabilities. This formula was used for the normalization process:

$$X_n = \left( \frac{X_o - X_{\min}}{X_{\max} - X_{\min}} \right) \times (0.85 - 0.15) + 0.15 \quad (5)$$

where  $X_o$  is the original value of input and output parameters;  $X_n$  is normalized value;  $X_{\max}$  and  $X_{\min}$  are maximum and minimum values of input and output parameters, respectively.

### Stepwise regression

This method uses multiple linear regression analysis to determine the relationship between the independent variables and a dependent variable. In this study, SWR analysis was carried out to determine the relations between dependent variable (SSP) and the independent variables ( $T_o$ ,  $RH$ ,  $WS$ ,  $SR$ ,  $TDS_F$ ,  $M_F$ , and  $T_F$ ). The SWR equation in the general form can be written as:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2, \dots, \beta_n X_n \quad (6)$$

where  $Y$  is the output value using the SWR method;  $B_i$  ( $i = 0, 1, 2, \dots, n$ ) are the regression coefficients, and  $X_i$  ( $i = 1, 2, \dots, n$ ) are the input variables.

The SWR includes finding appropriate independent variables and values for the  $\beta$  coefficients. To achieve this aim, SWR analysis was carried out using the Statistical Package for Social Science (IBM SPSS Statistics 22) program (SPSS Inc., Chicago, IL, USA). The basic procedures of SWR were: (1) identifying an initial model; (2) frequently altering the model at the former step by adding or removing independent variable in accordance with the 'stepping criteria: probability-of-F-to-enter  $\leq 0.05$ , probability-of-F-to-remove  $\geq 0.1$ '; (3) ending the search when stepping is no longer potential given the stepping criteria, or when a particular maximum number of steps has been reached. Simply and briefly, SWR technique chooses the most correlated independent variable first, and then chooses the second independent variable which most correlates with the residual variance in the dependent variable. This procedure continues until choice of an additional independent variable does not increase the coefficient of determination (CD) by a significant amount. A detailed description of the SWR can be found in numerous sources (Morrison 1990; Brereton 1995; Hocking 1996).

### Statistical parameters for models' assessment

The performances of both models (ANN and SWR) were evaluated by these statistical error criteria: CD, root mean square error (RMSE), index of agreement (IA), and mean absolute relative error (MARE). The CD illustrates the proportion of the total variance in the observed data

that can be demonstrated by the model. RMSE was used to estimate the sensitivity and extremum effect of the predicted value. IA values closer to one show the better agreement with the model. MARE expresses accuracy as a percentage. The closer MARE is to zero, the better the model accuracy. The mathematical expressions for calculation of these statistical parameters are stated in Equations (7)–(10):

$$CD = \frac{(\sum_{i=1}^n (SSP_{o,i} - \overline{SSP}_o)(SSP_{p,i} - \overline{SSP}_p))^2}{\sum_{i=1}^n (SSP_{o,i} - \overline{SSP}_o)^2 \times \sum_{i=1}^n (SSP_{p,i} - \overline{SSP}_p)^2} \quad (7)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (SSP_{o,i} - SSP_{p,i})^2}{n}} \quad (8)$$

$$IA = 1 - \frac{\sum_{i=1}^n (SSP_{o,i} - SSP_{p,i})^2}{\sum_{i=1}^n (|SSP_{p,i} - \overline{SSP}_o| + |SSP_{o,i} - \overline{SSP}_p|)^2} \quad (9)$$

**Table 1** | Mean, standard deviation and coefficient of variation values of the experimental data used in this study

	Mean ± SD	Ranges	CV (%)
$T_o$ (°C)	26.64 ± 3.68	16.87–33.23	14.00
RH (%)	23.36 ± 12.90	12.90–70.00	55.00
WS (km/h)	2.44 ± 3.12	0.00–12.65	128.00
SR (W/m <sup>2</sup> )	587.55 ± 181.93	75.10–920.69	31.00
$T_F$ (°C)	0.21 ± 0.04	0.13–0.25	20.00
$M_F$ (L/min)	36.66 ± 4.27	22.10–42.35	12.00
TDS <sub>F</sub> (PPT)	80.23 ± 29.42	41.40–130.00	37.00
SSP (L/m <sup>2</sup> /h)	0.50 ± 0.24	0.05–0.97	48.00

SD, standard deviation; CV, coefficient of variation.

**Table 2** | Pearson correlation coefficient values

	$T_o$	RH	WS	SR	$T_F$	$M_F$	TDS <sub>F</sub>
SSP							
Pearson correlation	−0.070	0.014	−0.307*	0.733*	−0.061	0.228*	−0.403*
Sig. (2-tailed)	0.376	0.864	$8 \times 10^{-5}$	$2.82 \times 10^{-28}$	0.445	0.004	$1.28 \times 10^{-7}$

\*Correlation is significant at the 0.01 level (2-tailed).

$$MARE = \frac{1}{n} \left( \sum_{i=1}^n \left| \frac{SSP_{o,i} - SSP_{p,i}}{SSP_{o,i}} \right| \times 100 \right) \quad (10)$$

where  $SSP_{o,i}$  denotes the observed value;  $SSP_{p,i}$  is the predicted value;  $\overline{SSP}_o$  is the mean of observed values;  $\overline{SSP}_p$  is the mean of predicted values; and  $n$  is the whole number of observations.

## RESULTS AND DISCUSSION

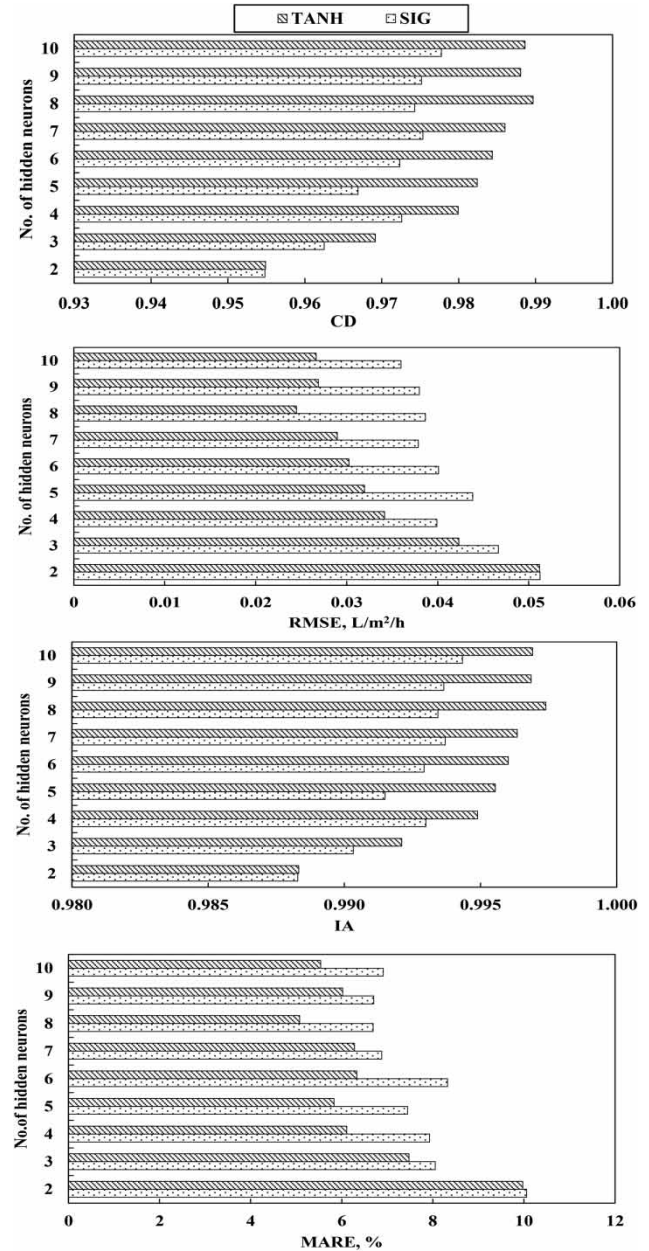
### Data description and general findings

Table 1 shows some of the descriptive statistics for the data obtained from the experimental work used in the ANN and SWR models. Table 2 presents Pearson correlation coefficient values computed using SPSS for all parameters. The mean ± SD (standard deviation) parameter values were  $26.64 \pm 3.68$  °C,  $23.36 \pm 12.90\%$ ,  $2.44 \pm 3.12$  km/h,  $587.55 \pm 181.93$  W/m<sup>2</sup>,  $0.21 \pm 0.04$  °C,  $36.66 \pm 4.27$  L/min,  $80.23 \pm 29.42$  PPT, and  $0.50 \pm 0.24$  L/m<sup>2</sup>/h for  $T_o$ , RH, WS, SR,  $T_F$ ,  $M_F$ , TDS<sub>F</sub>, and SSP, respectively. The coefficient of variation (CV) for all of the parameters was different with the greatest variation being observed in the WS, while the smallest variation was identified in  $M_F$ . Pearson correlation coefficient values for  $T_o$ , WS,  $T_F$ , and TDS<sub>F</sub> were negative, while they were positive for RH, SR, and  $M_F$ . SR shows the highest Pearson correlation coefficient value of 0.733, while RH is the lowest (0.014). A two-tailed significance test was significant for WS, SR,  $M_F$ , and TDS<sub>F</sub> but not significant for  $T_o$ , RH, and  $T_F$ , being greater than 0.001, as indicated in Table 2. The effects of meteorological and operational factors on the SSP were investigated experimentally, and our findings in this regard indicate that the most important factor that affects SSP is the SR. The SSP was observed

to decrease in line with an increase in  $TDS_F$ . A full and complete demonstration of these experimental results can be found in Mashaly *et al.* (2016).

### ANN model selection

To select the best ANN architecture, the number of neurons in the hidden layer and the best transfer function between the layers must be determined through a trial-and-error technique. The number of neurons in the hidden layer of the ANN was changed from 2 to 10 neurons and the transfer function was varied between the SIG and TANH functions in the hidden and output layers. The performance of the different ANN architectures based on the statistical indicators during the training phase is displayed in Figure 2. The ANN architecture was improved with the increase of neurons in the hidden layer, as reflected in the CD, RMSE, IA, and MARE values for the SIG and TANH functions in Figure 2. High values of the CD and IA, and low values of RMSE and MARE, show good model performance. For SIG function in the training phase, the ANN architectures' CD values ranged from 0.955 to 0.978, RMSE values from 0.036 to 0.051 L/m<sup>2</sup>/h, IA values from 0.988 to 0.994, and MARE values from 6.690 to 10.061 L/m<sup>2</sup>/h. For TANH function in the training phase, the ANN architectures' CD values ranged from 0.955 to 0.990, RMSE values from 0.024 to 0.051 L/m<sup>2</sup>/h, IA values from 0.988 to 0.997, and MARE values from 5.074 to 9.973 L/m<sup>2</sup>/h. It can be observed that the TANH function performed better than the SIG function, as seen from the results and Figure 2. However, the most appropriate ANN architecture is achieved when the number of neurons in the hidden layer was 8 and the TANH function was used, as this gave the best SSP predictions with the lowest error. The developed ANN model is 7-8-1 and this means that the developed ANN model contains 7 input neurons, 8 neurons in the hidden layer, and 1 output neuron, as illustrated in Figure 3. The developed ANN model's CD value was 0.990, RMSE value was 0.024 L/m<sup>2</sup>/h, IA value was 0.997, and MARE value was 5.074 L/m<sup>2</sup>/h. The relative contribution of each input to the output (SSP) for the developed ANN model was examined. The results revealed that  $T_o$  and  $WS$  are not significant, contributing only 3.48% and 3.52%, respectively. Also, it has been found that  $TDS_F$  and  $SR$  are



**Figure 2** | Statistical performance of the developed ANN model with various transfer functions and hidden neurons during the training phase.

the dominant inputs, which is in accordance with the results of Tanaka & Nakatake (2006) and Dev *et al.* (2011). The  $RH$ ,  $M_F$ , and  $T_F$  were also found to have a moderate influence on the SSP prediction with a contribution percentage of 16.41, 17.93, and 15.85%, respectively. The developed ANN model can be solved using a spreadsheet (i.e., Microsoft Excel) and expressed by an algebraic system of equations:

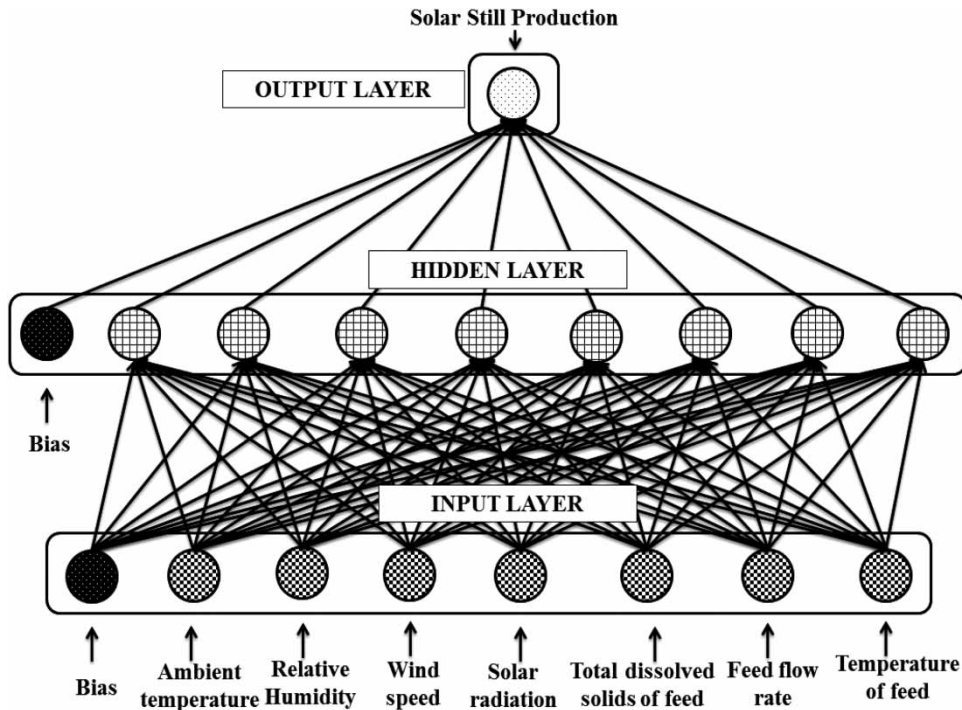


Figure 3 | Architecture of the optimal ANN model.

$$SSP = \frac{1 - \exp[(-7.05 \times F_1 + 4.96 \times F_2 - 3.78 \times F_3 + 7.93 \times F_4 - 5.44 \times F_5 + 5.09 F_6 + 7.58 F_7 - 2.97 F_8) - 2.52]}{1 + \exp[(-7.05 \times F_1 + 4.96 \times F_2 - 3.78 \times F_3 + 7.93 \times F_4 - 5.44 \times F_5 + 5.09 F_6 + 7.58 F_7 - 2.97 F_8) - 2.52]} \quad (11)$$

where  $F_1, F_2, F_3, F_4, F_5, F_6, F_7$  and  $F_8$  are computed as follows:

$$F_1 = \frac{1 - \exp[(-0.78 \times T_o + 2.65 \times RH + 1.55 \times WS - 4.85 \times SR + 1.57 \times T_F - 0.01 \times M_F - 8.42 \times TDS_F) + 3.60]}{1 + \exp[(-0.78 \times T_o + 2.65 \times RH + 1.55 \times WS - 4.85 \times SR + 1.57 \times T_F - 0.01 \times M_F - 8.42 \times TDS_F) + 3.60]} \quad (12)$$

$$F_2 = \frac{1 - \exp[(-1.93 \times T_o + 2.07 \times RH - 0.33 \times WS + 3.80 \times SR + 5.49 \times T_F - 3.62 \times M_F - 1.46 \times TDS_F) - 4.97]}{1 + \exp[(-1.93 \times T_o + 2.07 \times RH - 0.33 \times WS + 3.80 \times SR + 5.49 \times T_F - 3.62 \times M_F - 1.46 \times TDS_F) - 4.97]} \quad (13)$$

$$F_3 = \frac{1 - \exp[(1.87 \times T_o - 4.82 \times RH - 1.95 \times WS - 1.87 \times SR + 3.02 \times T_F + 3.82 \times M_F + 7.94 \times TDS_F) - 6.38]}{1 + \exp[(1.87 \times T_o - 4.82 \times RH - 1.95 \times WS - 1.87 \times SR + 3.02 \times T_F + 3.82 \times M_F + 7.94 \times TDS_F) - 6.38]} \quad (14)$$

$$F_4 = \frac{1 - \exp[(-0.36 \times T_o + 7.28 \times RH - 0.31 \times WS + 4.18 \times SR + 4.55 \times T_F - 11.79 \times M_F + 4.06 \times TDS_F) - 1.62]}{1 + \exp[(-0.36 \times T_o + 7.28 \times RH - 0.31 \times WS + 4.18 \times SR + 4.55 \times T_F - 11.79 \times M_F + 4.06 \times TDS_F) - 1.62]} \quad (15)$$

$$F_5 = \frac{1 - \exp[(-1.14 \times T_o + 6.39 \times RH - 0.30 \times WS - 2.30 \times SR + 0.03 \times T_F - 4.71 \times M_F + 1.40 \times TDS_F) + 0.49]}{1 + \exp[(-1.14 \times T_o + 6.39 \times RH - 0.30 \times WS - 2.30 \times SR + 0.03 \times T_F - 4.71 \times M_F + 1.40 \times TDS_F) + 0.49]} \quad (16)$$

$$F_6 = \frac{1 - \exp[(-0.02 \times T_o + 6.32 \times RH + 1.50 \times WS - 3.85 \times SR - 0.76 \times T_F + 4.29 \times M_F - 4.21 \times TDS_F) - 0.28]}{1 + \exp[(-0.02 \times T_o + 6.32 \times RH + 1.50 \times WS - 3.85 \times SR - 0.76 \times T_F + 4.29 \times M_F - 4.21 \times TDS_F) - 0.28]} \quad (17)$$

$$F_7 = \frac{1 - \exp[(-0.50 \times T_o - 0.22 \times RH + 0.85 \times WS - 3.48 \times SR + 5.86 \times T_F + 1.38 \times M_F - 10.42 \times TDS_F) + 2.08]}{1 + \exp[(-0.50 \times T_o - 0.22 \times RH + 0.85 \times WS - 3.48 \times SR + 5.86 \times T_F + 1.38 \times M_F - 10.42 \times TDS_F) + 2.08]} \quad (18)$$

$$F_8 = \frac{1 - \exp[(-1.59 \times T_o + 0.66 \times RH + 1.40 \times WS + 0.73 \times SR - 0.63 \times T_F + 1.01 \times M_F - 0.62 \times TDS_F) - 0.16]}{1 + \exp[(-1.59 \times T_o + 0.66 \times RH + 1.40 \times WS + 0.73 \times SR - 0.63 \times T_F + 1.01 \times M_F - 0.62 \times TDS_F) - 0.16]} \quad (19)$$

### SWR models

SWR models are developed by iteratively adding and removing the terms from a multi-linear model based on their significance within a regression by using SPSS software. All seven dependent variables ( $T_o$ ,  $RH$ ,  $WS$ ,  $SR$ ,  $T_F$ ,  $M_F$ , and  $TDS_F$ ) were used to create predicting models for SSP. SWR produced five models with 1–5 predictor variables, where  $SR$  was involved in each set of predictor variables, as listed in Table 3. The variables of  $SR$  and  $TDS_F$  have the most effect on SSP modeling. The effect of the variable of  $T_F$  was not considerable, while in the ANN model it was moderately effective. The effect of  $WS$ ,  $M_F$ , and  $RH$  were low. CD values associated with each of the five models ranged from 0.545 to 0.902. Corresponding standard errors of the estimate (SEE) ranged from 0.071 to 0.164 L/m<sup>2</sup>/h. It can be noted from Table 3 that the absence or presence of some of the input variables in the SWR models significantly affects the performances of these models. Model 1 with just the  $SR$  performed worst, with CD = 0.545 and SSE = 0.164 L/m<sup>2</sup>/h. Model 2 performed better than Model 1, owing to the presence of  $TDS_F$ . The CD value of Model 2 was increased by 59.82% than that

for Model 1. Furthermore, the SSE value of Model 2 was decreased by 47.56% than that for Model 1. Following step 2 (Model 2), the accuracy was dramatically unchanged. Models 3, 4, and 5 CD values were increased by 2.30%, 2.87%, and 3.56% than that from Model 2. Also, the SSE values for Models 3, 4, and 5 were decreased by 11.63%, 13.95% and 17.44% than that from Model 2. However, Model 5 showed the best model for estimating SSP (CD = 0.902, SEE = 0.071 L/m<sup>2</sup>/h) which involved  $SR$ ,  $TDS_F$ ,  $WS$ ,  $M_F$ , and  $RH$  as predictor variables.

### Comparing ANN and SWR models

To assess the capability of ANN and SWR models in predicting SSP, their findings in different modeling stages were compared with one other. The performance of the ANN and SWR models was evaluated using the statistical parameters in Table 4. The observed and predicted SSP values from the ANN and SWR models during the training, testing, and validation processes are compared in Figure 4; these are in the form of a scatter (1:1) plot (left panel) and a relative error (RE) plot (right panel). The data were mostly evenly and tightly distributed around the 1:1 line,

**Table 3** | Developed SWR models for estimating SSP

Model	Mathematical expression	CD	SEE (L/m <sup>2</sup> /h)
1	$SSP = -0.074 + 0.001 SR$	0.545	0.164
2	$SSP = 0.232 + 0.001 SR - 0.005 TDS_F$	0.871	0.086
3	$SSP = 0.157 + 0.001 SR - 0.004 TDS_F - 0.016 WS$	0.891	0.076
4	$SSP = -0.066 + 0.001 SR - 0.003 TDS_F - 0.019 WS + 0.685 M_F$	0.896	0.074
5	$SSP = -0.482 + 0.001 SR - 0.002 TDS_F - 0.019 WS + 1.930 M_F + 0.003 RH$	0.902	0.071

CD, coefficient of determination; SEE, standard error of the estimate.



**Table 4** | Statistical performance of the developed ANN and SWR models during training, testing, and validation

	CD	RMSE (L/m <sup>2</sup> /h)	IA	MARE (%)
<b>Training</b>				
ANN	0.990	0.024	0.997	5.074
SWR	0.902	0.146	0.901	36.297
<b>Testing</b>				
ANN	0.918	0.070	0.976	10.593
SWR	0.864	0.152	0.871	26.310
<b>Validation</b>				
ANN	0.972	0.047	0.991	7.152
SWR	0.941	0.166	0.882	25.506

as illustrated in Figure 4, and there was a very close agreement between the observed and predicted values by the ANN model during the training process. The majority of REs of the predicted SSP values (93%) are mostly near  $\pm 10\%$ . The ANN model gave values for CD of 0.990, RMSE of 0.024 L/m<sup>2</sup>/h, IA of 0.997, and MARE of 5.074%, as presented in Table 4 for the training process. The CD and IA values are close to one while RMSE and MARE values are close to zero, demonstrating excellent agreement between the observed and predicted findings from the ANN model. Figure 4 illustrates that the majority of points obtained by the SWR model during the training process are below the 1:1 line. Also, it is seen that the REs are between  $-100$  and  $+24\%$  by using the SWR model and the CD was decreased to 0.902. The RMSE and MARE values of 0.146 and 36.297 in the SWR model, respectively, were increased by 508.33% and 615.35% than those from the developed ANN model. The IA values in the SWR model were decreased by 9.63% than that from the ANN model during the training process.

In the testing process, Figure 4 illustrates how SSP values mostly follow a 1:1 line, showing a strong match between the observed and predicted values. The statistics of the ANN model were CD = 0.918, RMSE = 0.070 L/m<sup>2</sup>/h, IA = 0.976, and MARE = 10.593%, as shown in Table 4. As shown in Figure 4, 71.88% and 18.75% of the REs accumulate between  $-10$  and  $10\%$  for the ANN and SWR models, respectively. The SWR model had a CD value that was about 6% less accurate than the one from the developed ANN model. The value of RMSE for the SWR model

(0.152 L/m<sup>2</sup>/h) was almost double that of the value for the developed ANN model. The SWR model had an IA value that was about 11% less accurate than that from the developed ANN model, while the MARE value for the SWR model was about 2.5 times that of the value for the developed ANN model during the testing process.

By looking at Figure 4 during the validation process, we noted that some values predicted by the SWR model were imprecise, while most of the predicted values were acceptable when using the developed ANN model. The average REs for the ANN and SWR models were  $-2.27\%$  and  $-25.47\%$ , respectively. It is obvious from Table 4 that stronger agreement between the observed and predicted values was obtained using the ANN model. This agreement is reflected in the CD, RMSE, IA, and MARE results. The CD and IA values for the ANN were close to one, while the RMSE value was close to zero. The MARE value for the SWR model was almost 3.57 times that of the value for the ANN model during the validation process. Additionally, the SWR model had CD and IA values that were approximately 3.19% and 11% less accurate, respectively, than those from the ANN model. Additionally, the value of RMSE for the SWR model was almost 3.5 times that of the value for the ANN model. As seen from the findings in Table 4 and Figure 4, the ANN model revealed a better prediction performance than the SWR model, and presented sufficient accuracy in the prediction of SSP. These results are consistent with the studies of Jalali-Heravi & Garkani-Nejad (2002) and Rahimi-Ajdadi & Abbaspour-Gilandeh (2011).

On the basis of the average values in all the modeling stages, the performance of the ANN model (CD = 0.960, RMSE = 0.047 L/m<sup>2</sup>/h, IA = 0.988, and MAE = 7.606 L/m<sup>2</sup>/h) performed better than the SWR model (CD = 0.902, RMSE = 0.155 L/m<sup>2</sup>/h, IA = 0.885, and MAE = 29.371 L/m<sup>2</sup>/h). The average REs were  $-30.55\%$ ,  $-25.96\%$ , and  $-25.47\%$  for the SWR model when using the training, testing, and validation data sets, respectively. The errors indicated that SSP values predicted by the SWR model tended to be underpredicted. This systematic underprediction may be due to inappropriate specification of the relationships between SSP and some variables used in the modeling process. However, the SWR model is easier to use and gives results that are deemed reasonable,

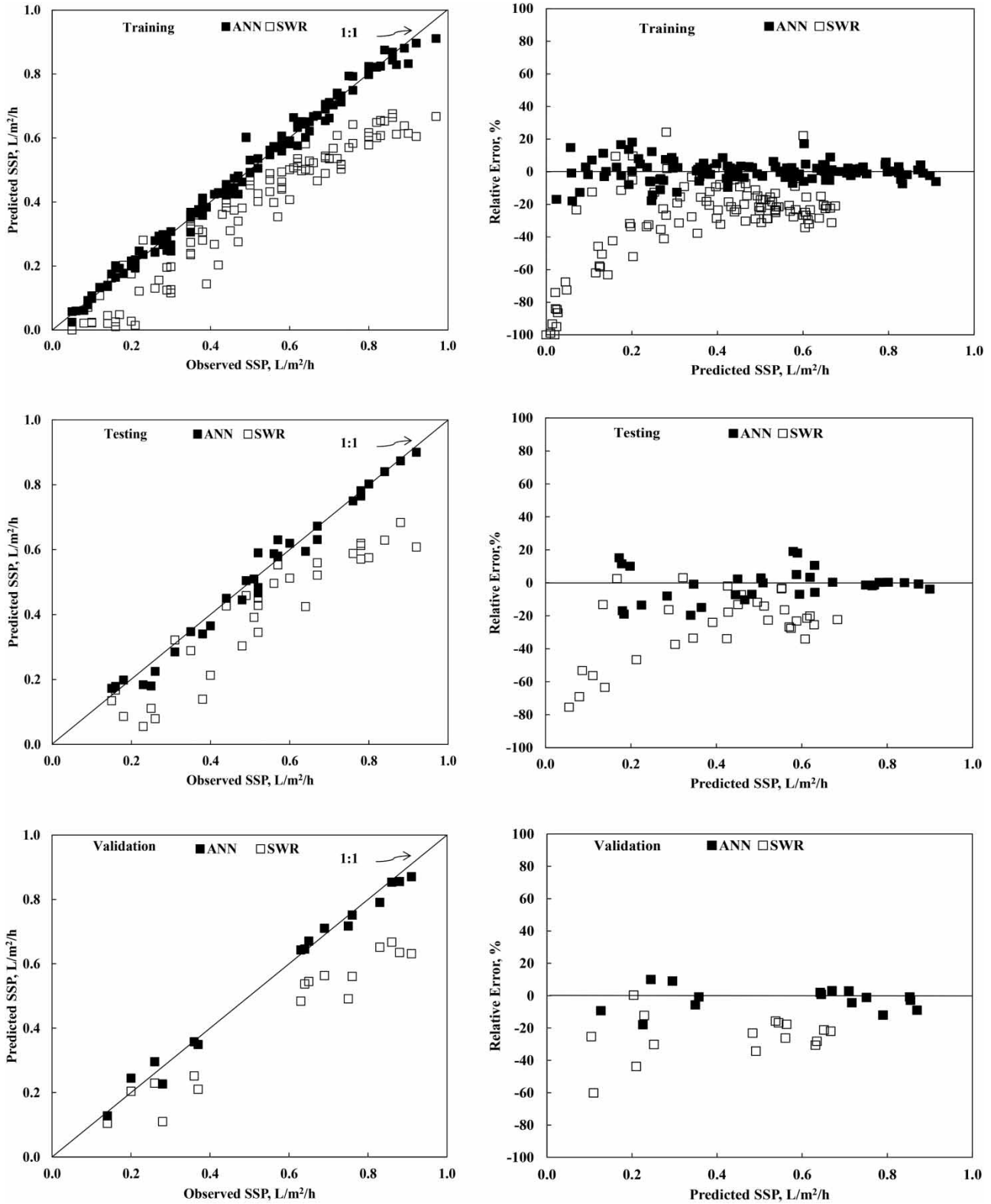


Figure 4 | Performance of the developed ANN and SWR models during training, testing, and validation processes.

but the ANN model gave the most precise estimates. Consequently, the developed ANN model gives a high accuracy in the prediction of SSP and offers flexibility and an ability to model non-linear relationships to get precise results.

## CONCLUSION

ANN and SWR are applied to modeling solar still production (SSP) using ambient temperature ( $T_o$ ), RH, WS, SR, the temperature of feed water ( $T_F$ ), the TDS of feed water ( $TDS_F$ ), and the flow rate of feed water ( $M_F$ ) as explanatory variables. A feed-forward back-propagation algorithm was used in the training process of the ANN model. Several ANN architectures with different numbers of neurons in the hidden layer were trained to determine which architecture gave the best performance and produced minimal errors. The ANN model with 7-8-1 architecture was selected as the best model for modeling the SSP. The performance of the ANN and SWR models was evaluated by the CD, RMSE, IA, and MARE statistical performance indicators. The effectiveness and validity of the ANN model was confirmed owing to the high values of CD and IA, and the low values of RMSE and MARE. Results revealed that ANN can learn the relationships between the input variables and SSP very well; it provided better prediction results than the SWR model. As a final conclusion, the ANN model gives the most precise and reliable estimates, but the SWR model is easier to use though less efficient.

## ACKNOWLEDGEMENT

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

## REFERENCES

- Ali Abdoli, M., Falah Nezhad, M., Salehi Sede, R. & Behboudian, S. 2012 Longterm forecasting of solid waste generation by the artificial neural networks. *Environ. Prog. Sustain. Energy* **31** (4), 628–636.
- Ayoub, G. M., Al-Hindi, M. & Malaeb, L. 2015 A solar still desalination system with enhanced productivity. *Desalin. Water Treat.* **531** (2), 3179–3186.
- Behboudian, S., Tabesh, M., Falahnezhad, M. & Ghavanini, F. A. 2014 A long-term prediction of domestic water demand using preprocessing in artificial neural network. *J. Water Supply. Res. Technol. AQUA* **63** (1), 31–42.
- Brereton, R. G. 1995 *Chemometrics: Data Analysis for the Laboratory and Chemical Plant*. Wiley, New York.
- 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. *Appl. Energy* **88**, 252–264.
- Eren, B., Ileri, R., Dogan, E., Caglar, N. & Koyuncu, I. 2012 Development of artificial neural network for prediction of salt recovery by nanofiltration from textile industry wastewaters. *Desalin. Water Treat.* **50** (1–3), 317–328.
- Gao, P., Zhang, L., Cheng, K. & Zhang, H. 2007 A new approach to performance analysis of a seawater desalination system by an artificial neural network. *Desalination* **205** (1–3), 147–155.
- Haykin, S. 1999 *Neural Networks. A Comprehensive Foundation*. Prentice Hall International, Upper Saddle River, NJ.
- Hocking, R. R. 1996 *Methods and Applications of Linear Models. Regression and the Analysis of Variance*. Wiley, New York.
- Jalali-Heravi, M. & Garkani-Nejad, Z. 2002 Prediction of electrophoretic mobilities of alkyl- and alkenylpyridines in capillary electrophoresis using artificial neural networks. *J. Chromatogr. A* **971** (1), 207–215.
- Kabeel, A. E., Elkenany, E. A. & Hawam, O. M. 2013 Experimental study of the performance of solar still using a reciprocating intermittent spray feeding system. *J. Renew. Sustain. Energy* **5** (1), 013108.
- Khayet, M., Cojocar, C. & Essalhi, M. 2011 Artificial neural network modeling and response surface methodology of desalination by reverse osmosis. *J. Membr. Sci.* **368** (1–2), 202–214.
- Koilraj Gnanadason, M., Senthil Kumar, P., Wilson, V. H. & Kumaravel, A. 2015 Productivity enhancement of a single basin solar still. *Desalin. Water Treat.* **55** (8), 1998–2008.
- Mashaly, A. F. & Alazba, A. A. 2015 Comparative investigation of artificial neural network learning algorithms for modeling solar still production. *J. Water Reuse Desal.* **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. *Int. J. 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. *Comput. Electron. Agric.* **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. *Desalin. Water Treat.* **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. *Sol. 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](#). *Desalin. Water Treat.* **57** (26), 12019–12036.
- Morrison, D. F. 1990 *Multivariate Statistical Methods*, 3rd edn. McGraw-Hill, New York.
- Nezhad, M. F., Mehrdadi, N., Torabian, A. & Behboudian, S. 2016 [Artificial neural network modeling of the effluent quality index for municipal wastewater treatment plants using quality variables: south of Tehran wastewater treatment plant](#). *J. Water Suppl. Res. Technol. AQUA* **65** (1), 18–27.
- Porrizzo, R., Cipollina, A., Galluzzo, M. & Micale, G. 2013 [A neural network-based optimizing control system for a seawater-desalination solar-powered membrane distillation unit](#). *Comput. Chem. Eng.* **54**, 79–96.
- Rahimi-Ajdadi, F. & Abbaspour-Gilandeh, Y. 2011 [Artificial neural network and stepwise multiple range regression methods for prediction of tractor fuel consumption](#). *Measurement* **44** (10), 2104–2111.
- 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](#). *Renew. Energ.* **40** (1), 71–79.
- Tanaka, H. & Nakatake, Y. 2006 [Theoretical analysis of a basin type solar still with internal and external reflectors](#). *Desalination* **186**, 280–299.
- Tanaka, H. & Nakatake, Y. 2009 [Increase in distillate productivity by inclining the flat plate external reflector of a tilted-wick solar still in winter](#). *Sol. Energy* **83**, 785–789.
- Tayyebi, S. & Alishiri, M. 2014 [The control of MSF desalination plants based on inverse model control by neural network](#). *Desalination* **333** (1), 92–100.
- Valderrama, J. O., Campusano, R. A. & Toro, A. S. 2015 [Correlation and prediction of saline solution properties for their use in mineral processing using artificial neural networks](#). *J. Water Reuse Desal.* **5** (4), 454–464.
- Wadi Abbas Al-Fatlawi, A., Rahim, N. A., Saidur, R. & Ward, T. A. 2015 [Improving solar energy prediction in complex topography using artificial neural networks: Case study Peninsular Malaysia](#). *Environ. Prog. Sustain. Energy* **34** (5), 1528–1535.

First received 24 May 2016; accepted in revised form 20 November 2016. Available online 30 January 2017