

Experimental and modeling study to estimate the productivity of inclined passive solar still using ANN methodology in arid conditions

Ahmed F. Mashaly and A. A. Alazba

ABSTRACT

An artificial neural network (ANN) was developed for predicting solar still production (MD) under a hyper-arid environment. A three-layer feed-forward neural network based on back-propagation algorithm was used in the modeling process. The inputs comprise air temperature, relative humidity, wind speed, solar radiation, feed water temperature, feed water total dissolved solids, and feed water flow rate. The output was MD. The ANN model with optimal prediction performance was found by testing several networks. Then, the findings obtained from the ANN model were compared with the findings from the multiple linear regression (MLR) model. The optimal ANN model had a 7-8-1 architecture with a hyperbolic tangent transfer function. Statistical criteria revealed that the ANN model performed better than MLR in predicting MD. The root-mean-square errors during the testing process for MD were 0.070 and 0.128 for the ANN and MLR models, respectively. The coefficient of determination values for the training, testing, and validation data sets in the prediction of MD by ANN were 0.990, 0.918, and 0.945, respectively. The relative errors of the predicted MD values for the ANN model were approximately $\pm 10\%$. Therefore, the ANN model can successfully predict MD.

Key words | ANN, artificial intelligence, desalination, prediction, solar still

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 still is a simple solar device used for converting seawater or brackish water into potable water. It can be fabricated easily using locally available materials and can be maintained economically without the need for skilled labor. Solar still can be a suitable solution to solve the drinking water problem. However, it is not widely used because of its low productivity (Kabeel *et al.* 2015; Mashaly & Alazba 2017a). Consequently, several techniques and researches were introduced to investigate and enhance the design and productivity of solar stills, such as the addition of heat through a flat-plate collector (Riffat *et al.* 2005); the addition of dye in the feed water (Badran 2007); investigation of the effects of climatic, design, and operational variables on the performance of solar stills (Yeh & Chen 1986); and

investigation of the effect of water flowing over the glass cover (Dhiman & Tiwari 1990). Radhwan (2004) investigated the transient performance of a stepped solar still with built-in latent heat thermal energy storage. Badran *et al.* (2004) investigated the simulation and experimentation of an inverted tickle solar still. Aybar *et al.* (2005) investigated the experimentation of an inclined solar water distillation system.

Much research has been reported in the literature focusing on experimental investigations to find better design and improved productivity for solar desalination systems. This experimental research is time-consuming and expensive. Mathematical modeling can be the best method to find better designs and operational variables for solar stills.

Many numerical techniques have been used to model and predict solar still productivity, such as computer simulation (Cooper 1969), thermic circuit and Sankey diagrams (Frick 1970), periodic and transient analysis (Sodha *et al.* 1980; Tiwari & Rao 1984), and iteration methods (Toure & Meukam 1997). These methods are dependent on internal heat and mass transfer processes. Owing to the large amount of data required to validate the heat and mass transfer model, the ability to predict the solar still productivity is restricted by the capability to determine the parameters required to evaluate the model. On the other hand, artificial neural networks (ANNs) have a potential advantage in forecasting the productivity of solar stills because they use fewer parameters, less time, and are more accurate compared to the heat and mass transfer models.

ANNs attempt to mirror the brain functions in a computerized method by restoring the learning mechanism as the basis of human behavior. They are adaptive systems that change their structure based on external or internal information that flows through the network during the learning phase. ANNs can reflect the complex relationship between inputs and outputs successfully. Their predictions can lead to a closer fit with the data than the predictions of the classical modeling techniques, which usually results in more precise predictions (Benli 2013). ANNs have been used in a wide variety of thermal engineering applications, particularly on renewable energy engineering, such as solar energy engineering. ANNs were used to model the layer temperatures in a storage tank of a solar thermal system (Géczy-Víg & Farkas 2010), predict the in-situ daily performance of the solar collectors (Lecoeuche & Lalot 2005), determine the thermal performance of different types of solar collectors (Benli 2013), model solar still distillate production using local weather, analyze the performance of a solar powered membrane distillation system (Porrizzo *et al.* 2013), estimate the thermal performances of solar collectors (Caner *et al.* 2011), evaluate the performance of solar photovoltaic technologies (Velilla *et al.* 2014), and model the thermal performance of solar still (Mashaly & Alazba 2016a, 2016b, 2017b).

However, solar still production has not been clearly elucidated, and no previous research has addressed this topic, particularly in arid conditions. Therefore, the objectives of this study are: (1) to develop mathematical models to

estimate the productivity of the solar still using ANNs; (2) to evaluate the performance of ANNs using a statistical comparison between the productivity obtained from the model and experimental results; and (3) to compare the ANN models with the multiple linear regression (MLR) models in terms of their applicability, suitability, and accuracy in forecasting the productivity of the solar still.

MATERIALS AND METHODS

Experimental set-up

The 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 period from February to April 2013, and the weather data were obtained from a weather station (model: Vantage Pro2, Davis, USA) located close to the experimental site (24°44'12.15"N, 46°37'14.97"E). The solar still system utilized in the experiments consists of one stage of C6000 panel (F cubed, Ltd, Carocell Solar Panel, Australia) with an area of 6 m². The solar still is manufactured as a panel using modern cost-effective materials, such as coated polycarbonate plastic. The panel heats and distills a film of water flowing over the absorber mat of the panel. 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 from the inside by special coating material to prevent fogging (patent for F cubed, Australia). The front and cross-sectional views of the solar still are presented in Figure 1. The operational concept of the available system is summarized in the following paragraphs.

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. Eight drippers/nozzles drip the feed resulting in a film flowing over the absorbent mat. Under the absorbent mat, an aluminum screen helps to distribute the dripping water over the absorbent mat. An aluminum plate is also placed beneath the aluminum screen. Aluminum was selected for the manufacturing process because it is a hydrophilic material, which assists in

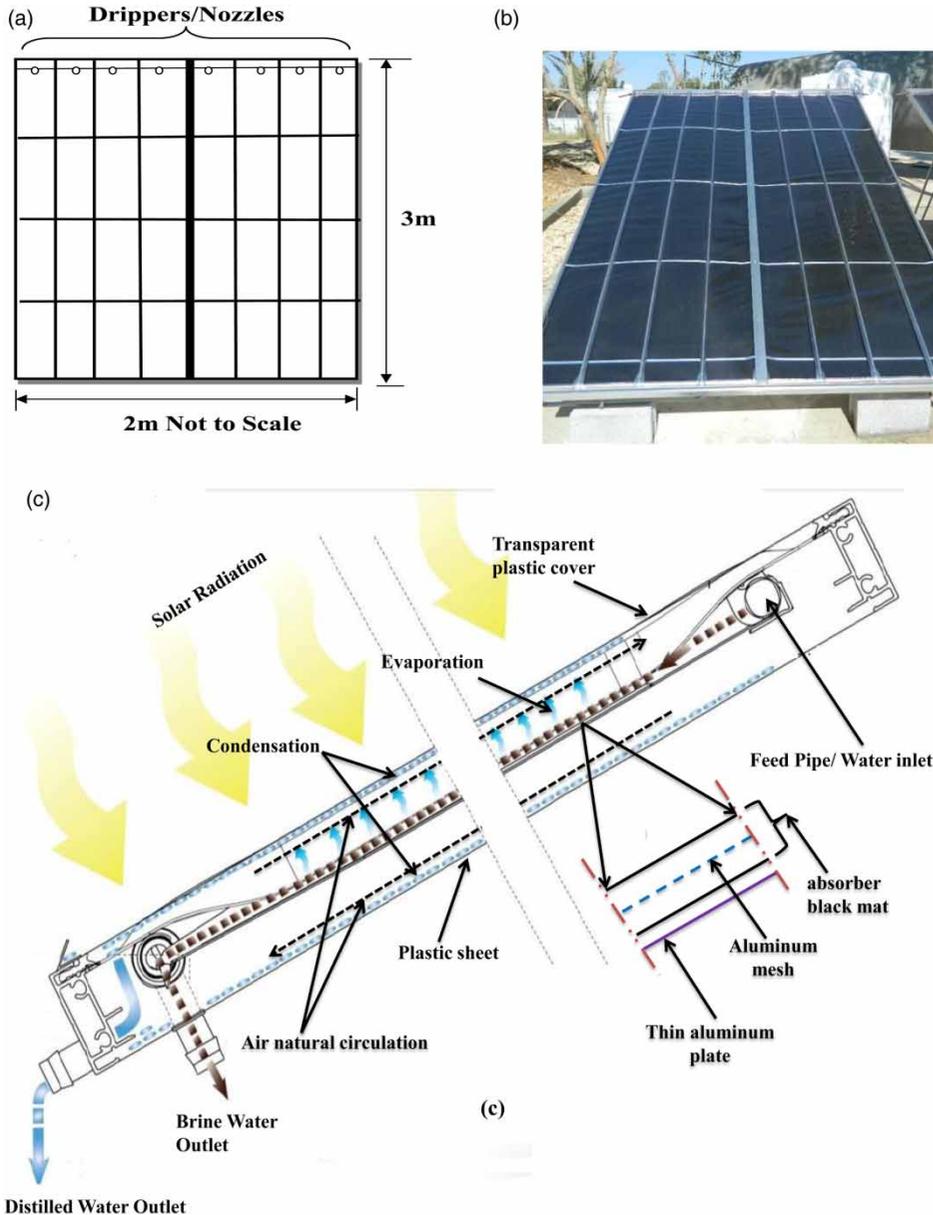


Figure 1 | Solar still panel: front section (a), picture of the panel (b), and cross-sectional view of the solar still panel (c).

the even distribution of the dripping water. The water flows through and over the absorbent mat and the solar energy is absorbed and partially collected inside the panel, which heats the water resulting in hot air that naturally circulates within the panel. The hot air flows in the upper part toward the top, and then reverses direction toward the bottom. With this circulation, the humid air is in contact with the cooled surfaces of the transparent polycarbonate cover and the bottom polycarbonate layer; therefore, the

water condenses and flows down the panel to be collected as distilled steam. Seawater was used as a feed water input to the system. The solar still system was run during the period from 23/02/2013 to 23/04/2013. Raw seawater was obtained from the Gulf, Dammam, East of Saudi Arabia (26°26'24.19" N, 50°10'20.38" E). The initial concentration of the total dissolved solids (TDS), pH, density (ρ), and electrical conductivity (EC) of the raw seawater were 41.4 ppt, 8.02, 1.04 g.cm⁻³, and 66.34 mS.cm⁻¹, respectively. The

productivity or the amount of distilled water produced (MD) 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 using thermocouples (T-type, UK). The temperature data for feed brine water was 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 a 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 a graduated cylinder. TDS and EC were checked using a calibrated (TDS) meter (Cole-Parmer Instrument, Vernon Hills, USA). A pH meter (model: 3,510 pH meter, Jenway, UK) was utilized to determine the acidity. ρ was measured by a digital density meter (model: DMA 35_N, Anton Paar, USA). The seawater was fed separately to the panel using the pump mentioned previously. The residence time for the water to pass through the panel was about 20 min. Consequently, the flow rate for feed water, distilled water, and brine water was measured every 20 min. Furthermore, the total dissolved solids of feed water (TDS_F) were measured every 20 min. The weather data, such as air temperature (T_o), relative humidity (RH), wind speed (U), and solar

radiation (R_s), were obtained from the weather station mentioned previously. Here, the production (MD) of solar desalination/still system is a dependent variable, whereas seven variables are independent, namely, T_o , RH , U , R_s , TDS_F , M_F , and T_F .

Artificial neural networks

In this study, the feed forward back propagation algorithm is used for the ANN model. The feed forward back propagation ANN is the most popular and widely used ANN architecture (Rumelhart *et al.* 1986). It consists of one input layer, one hidden layer, and one output layer. The ANN architecture used in this study is presented in Figure 2. Each of these layers comprises processing units called nodes/neurons of the ANN. Demuth & Beale (2004) stated that each artificial neuron is a unitary computational processor, which has a summing junction operator and a transfer/activation function. The connections among the inputs, neurons, and outputs consist of weights (W) and biases (B). Mathematically, this can be represented as

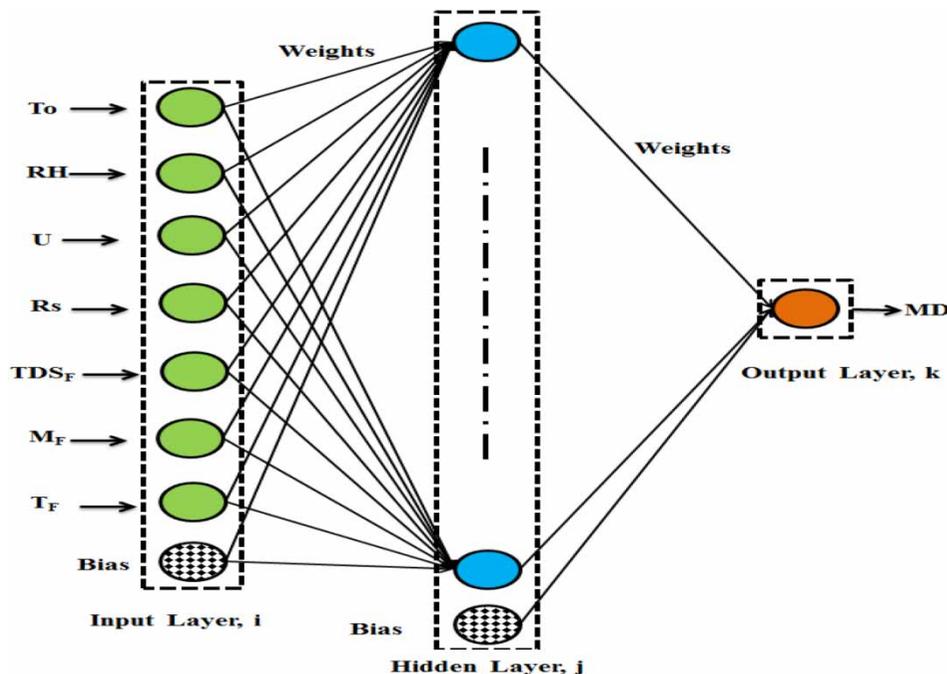


Figure 2 | Architecture of the artificial neural network.

follows (Haykin 1999):

$$Y = F \left(\sum_{j=1}^m W_{kj} F \left(\sum_{i=1}^n W_{ji} X_i + B_j \right) + B_k \right) \tag{1}$$

where Y = the output (MD), W_{kj} = weights between the hidden and output layers; W_{ji} = weights between the input and hidden layers; and X_i = input variables (T_o , RH , U , R_s , TDS_F , M_F , and T_F); m = the number of neurons in the hidden layer; n = the number of neurons in the input layer, B_j and B_k are the bias values of the neurons in the hidden layer and the output layer, respectively, and F is the transfer function. The transfer/activation functions used in the present study were sigmoid and hyperbolic tangent transfer functions. The sigmoid transfer function (SIG) for any variable S is given as follows:

$$F(S) = \frac{1}{1 + \exp(-S)} \tag{2}$$

The hyperbolic tangent transfer function (TANH) for any variable S is given as follows:

$$F(S) = \frac{1 - \exp(-2S)}{1 + \exp(-2S)} \tag{3}$$

The ANN model was developed using Qnet2000 software. The modeling process includes three stages, namely, training, testing, and validation. The available data set, which is composed of 160 data points obtained from the experimental work, was divided randomly into training (70%), testing (20%), and validation (10%) subsets. Therefore, the training, testing, and validation sets have 112, 32, and 16 data points, respectively. Trial and error is the best method to find the optimal number of neurons in the hidden layer (Abutaleb 1991). Consequently, the trial and error method was used to determine the optimum neurons in the hidden layer of the network. Before the modeling process, the data is automatically normalized between 0.15 and 0.85. The normalization accelerates the training process and enhances the network’s generalization capabilities. The iteration was fixed to 200,000. The learning rate and momentum factor were fixed at 0.01 and 0.8, respectively.

Multiple linear regression

MLR is a linear statistical technique that is very useful for predicting the best relationship between a dependent variable and several independent variables (Giacomino *et al.* 2001). The relationship between the input parameter/variable is more than one, and a dependent variable is examined. MLR is based on least squares, which means that the model is fit such that the sum of the squares of differences of the measured and predicted values is minimized. A general MLR model can be expressed by the following equation (Scheaffer *et al.* 2011):

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n \tag{4}$$

where Y is the predicted variable (the output), β_0 is the intercept, $B_1; \dots; \beta_n$ are the regression coefficients, and $X_1; \dots; X_n$ are the predictors (the inputs). In this study, the MLR analysis was carried out using IPM SPSS statistics 22 (Statistical Package for Social Science) software (SPSS Inc., Chicago, IL, USA).

Performance evaluation of the developed models

A large number of statistical criteria are available to examine and test the quality and accuracy of any developed model. The performance evaluation statistics used for the developed models in this study are the determination coefficient (DC), root-mean-square error ($RMSE$), the overall index of model performance (OI), and coefficient of residual mass (CRM). The developed model providing the best prediction outcomes for MD was chosen as the prediction model. The higher DC and OI values present greater similarities between the observed and predicted values. The lower $RMSE$ and CRM values represent more accurate prediction results. The DC , OI , $RMSE$, and CRM values were calculated using Equations (4)–(7), respectively:

$$DC = \frac{(\sum_{i=1}^n (v_{o,i} - \bar{v}_o)(v_{p,i} - \bar{v}_p))^2}{\sum_{i=1}^n (v_{o,i} - \bar{v}_o)^2 \times \sum_{i=1}^n (v_{p,i} - \bar{v}_p)^2} \tag{5}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (v_{p,i} - v_{o,i})^2}{n}} \tag{6}$$

$$OI = \frac{1}{2} \left(1 - \left(\frac{RMSE}{v_{max} - v_{min}} \right) + \left(1 - \frac{\sum_{i=1}^n (v_{o,i} - v_{p,i})^2}{\sum_{i=1}^n (v_{o,i} - \bar{v}_o)^2} \right) \right) \tag{7}$$

$$CRM = \frac{(\sum_{i=1}^n v_{p,i} - \sum_{i=1}^n v_{o,i})}{\sum_{i=1}^n v_{o,i}} \tag{8}$$

where $v_{o,i}$ = observed value; $v_{p,i}$ = correlated value; n = number of observations; v_{max} = maximum observed value; v_{min} = minimum observed value; and \bar{v}_o = averaged observed values.

RESULTS AND DISCUSSION

Experimental field findings and data analysis

The statistical data analysis was carried out using the data analysis tool in Microsoft Excel (MS Excel). Table 1 presents the statistical parameters of the experimental data. The statistical parameters were minimum (MIN), maximum (MAX), mean (AVG), standard deviation (SD), and coefficient of variation (CV). According to the findings of the field experiments, the average MD for the solar still system was 0.50 L/m²/h (approximately 5 L/m²/day), which is consistent with the findings of Radhwan (2004) and Kabeel et al. (2012). The results from the experiments have also revealed that the most dominant meteorological parameter affecting the MD was the R_s . The increase in T_o and U tends to increase the MD. The effect of increasing the U on the MD is more significant than the effect of increasing the T_o because increasing the U causes an increase in the

convective heat transfer coefficient from the cover to the atmosphere. This leads to the decrease in the cover temperature and increase in evaporation and condensation rates inside the solar still, including the MD. RH was inversely proportional to the MD because low RH (drier air) is likely to increase and enhance the evaporation rate. The evaporation rate also increases with the increase of the T_F and thereby higher MD. It was found that with the increase of the M_F , the MD decreases. With the decrease in the TDS_F , the MD increased where the evaporation rate increased, which may be attributed to the weakness of ionic bonds for the low TDS_F . A more complete illustration of these experimental data is given by Mashaly et al. (2016).

Table 2 shows the correlation matrix of the experimental data (all input parameters). The last row in Table 2 lists the correlation coefficient (CC) between the input parameters (T_o , RH , U , R_s , T_F , M_F , and TDS_F) and the output parameter (MD). This table displays that the linear correlation between R_s and MD is 73%. Therefore, any model that employs R_s should be able to estimate the MD satisfactorily. The model's performance can be augmented by considering other parameters that have aerodynamic behaviors on MD, such as RH , U , and T_o . However, the T_o , RH , U , T_F , M_F , and TDS_F are not well correlated with MD. Instead, these parameters are included in the modeling process for better accuracy of MD estimation. Additionally, some of these parameters are correlated to others, and yet these were included in the modeling process because their presence was found to advance the model accuracy. However, the sign of the CC (+, -) is used to denote the correlation, which is either positive or negative. The last row in Table 2 shows that most of these linear correlations were very weak in reflecting the non-linearity of the dominant

Table 1 | Statistical parameters for the experimental data

	To °C	RH %	U km/h	Rs W/m ²	T _F °C	M _F L/min	TDS _F PPT	MD L/m ² /h
MIN	16.87	12.90	0.00	75.10	22.10	0.13	41.40	0.05
MAX	33.23	70.00	12.65	920.69	42.35	0.25	130.00	0.97
AVG	26.64	23.36	2.44	587.55	36.66	0.21	80.23	0.50
SD	3.68	12.90	3.12	181.93	4.27	0.04	29.42	0.24
CV	0.14	0.55	1.28	0.31	0.12	0.20	0.37	0.48

MIN: minimum value; MAX: maximum value; AVG: average value; SD: standard deviation; CV: coefficient of variation; T_o : air temperature; RH : relative humidity; U : wind speed; R_s : solar radiation; T_F : temperature of feed water; M_F : feed flow rate; TDS_F : total dissolved solids of feed; MD: solar still productivity.

Table 2 | Correlation matrix for the experimental data

	To	RH	U	Rs	T _F	M _F	TDS _F	MD
To	1.00							
RH	-0.66	1.00						
U	-0.14	-0.08	1.00					
Rs	-0.15	0.15	0.22	1.00				
T _F	0.91	-0.80	-0.01	-0.09	1.00			
M _F	0.44	-0.72	-0.34	-0.27	0.48	1.00		
TDS _F	-0.01	0.23	0.64	0.22	0.06	-0.75	1.00	
MD	-0.07	0.01	-0.31	0.73	-0.06	0.25	-0.40	1.00

To: air temperature; RH: relative humidity; U: wind speed; Rs: solar radiation; T_F: temperature of feed water; M_F: feed flow rate; TDS_F: total dissolved solids of feed; MD: solar still productivity.

processes, which supports the use of ANNs. Moreover, the solar distillation process is considered to be highly nonlinear.

Optimal ANN architecture selection

Table 3 presents the statistical performance of the ANN model with various node numbers in the hidden layer and

transfer/activation function. The number of nodes in the hidden layer with identifying activation functions between the layers was determined through a trial and error procedure to select the best ANN model architecture. Figure 3 illustrates the statistical performance of the ANN model with various hidden nodes and activation functions during the training process. In the beginning, as shown in Table 3, with two nodes and SIG function in the hidden

Table 3 | Statistical performance of the ANN model with various node numbers in the hidden layer and transfer functions (bold values refer to the optimum architecture)

ANN	TF	Network statistics			Average contribution of the input node on output, %						
		SD	MXE	CC	To	RH	U	Rs	T _F	M _F	TDS _F
7-2-1	SIG	0.051	0.288	0.977	5.70	12.28	8.97	44.01	5.06	16.21	7.77
	TANH	0.051	0.289	0.977	5.55	12.37	9.27	43.77	5.31	16.35	7.38
7-3-1	SIG	0.047	0.274	0.981	7.54	11.55	4.80	40.26	11.04	4.21	20.60
	TANH	0.042	0.248	0.984	9.20	19.97	6.74	35.27	8.00	2.62	18.19
7-4-1	SIG	0.040	0.256	0.986	9.65	10.83	5.95	28.07	13.64	9.24	22.63
	TANH	0.034	0.209	0.990	10.81	12.82	3.93	25.75	11.02	17.23	18.43
7-5-1	SIG	0.044	0.268	0.983	4.43	18.35	5.44	36.62	11.30	4.49	19.36
	TANH	0.032	0.197	0.991	10.22	21.53	4.60	19.04	12.61	12.54	19.47
7-6-1	SIG	0.040	0.247	0.986	5.77	11.63	7.40	34.22	15.15	7.72	18.11
	TANH	0.030	0.181	0.992	9.19	13.84	9.79	22.10	13.55	10.64	20.88
7-7-1	SIG	0.038	0.250	0.988	5.80	12.70	6.90	31.17	11.22	10.97	21.24
	TANH	0.029	0.176	0.993	7.65	20.01	4.56	19.48	14.94	13.95	19.40
7-8-1	SIG	0.039	0.263	0.987	4.70	14.90	4.27	26.33	16.13	10.69	22.97
	TANH	0.024	0.113	0.995	3.48	16.41	3.52	20.04	17.93	15.85	22.78
7-9-1	SIG	0.038	0.262	0.988	4.06	17.94	5.02	27.48	14.49	8.52	22.49
	TANH	0.027	0.145	0.994	8.27	17.80	8.83	28.95	13.96	8.97	13.23
7-10-1	SIG	0.036	0.248	0.989	8.37	11.96	4.49	30.4	11.91	11.24	21.63
	TANH	0.027	0.131	0.994	10.47	11.03	6.02	25.87	16.83	12.07	17.71

TF: transfer function; SD: standard deviation, CC: correlation coefficient; MXE: maximum error; To: ambient temperature; RH: relative humidity; U: wind speed; Rs: solar radiation; T_F: temperature of feed water; M_F: feed flow rate; TDS_F: total dissolved solids of feed.

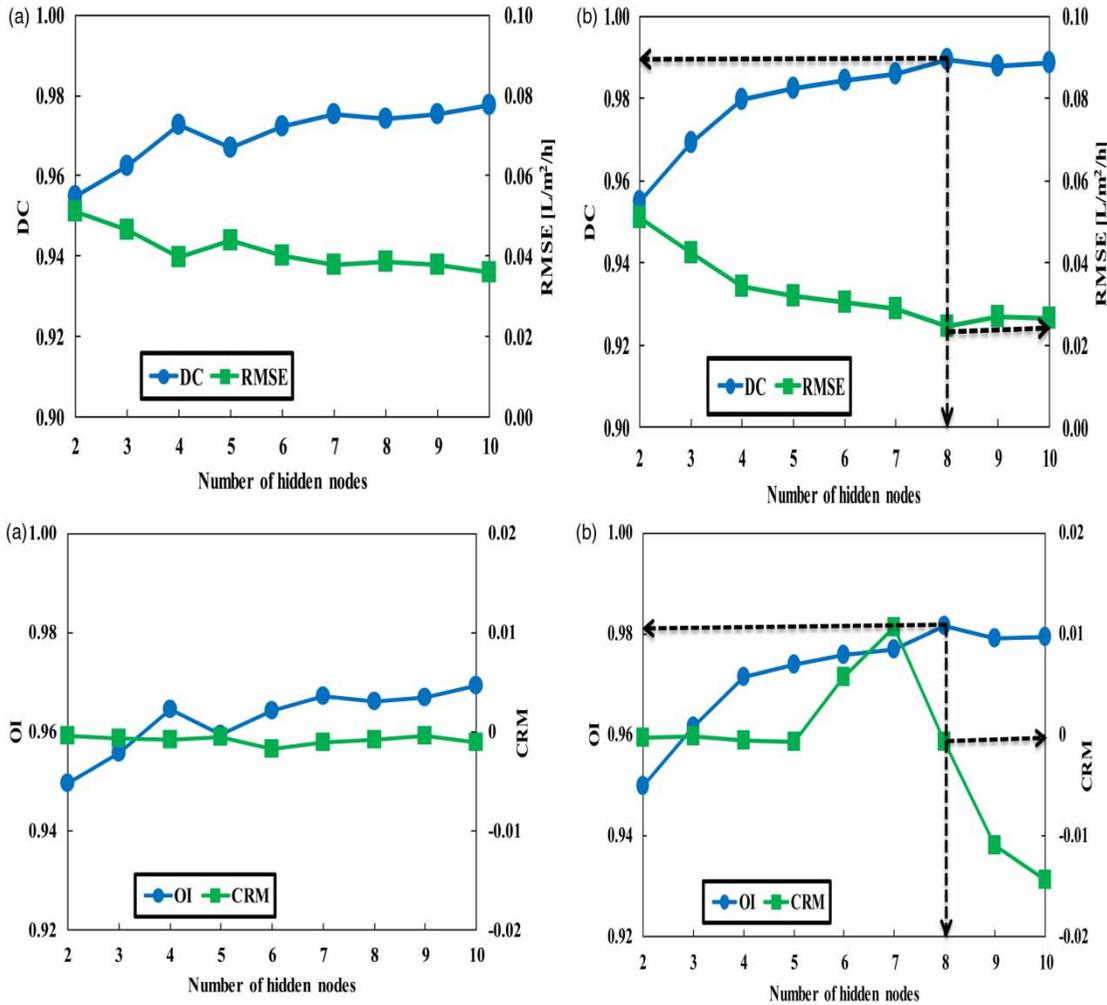


Figure 3 | Statistical performance of the ANN model for MD with various hidden nodes and transfer functions during the training process. (a) Sigmoid (SIG), (b) hyperbolic tangent (TANH).

layer, the statistical parameter values for the *SD*, maximum error (*MXE*), and *CC* were 0.051 L/m²/h, 0.288 L/m²/h, and 0.977, respectively. Using the TANH function, the *SD*, *MXE*, and *CC* values did not differ significantly at this node. It was found that increasing the number of nodes in the hidden layer led to the improvement in the values of statistical parameters. When the number of nodes was five, the *SD*, *MXE*, and *CC* values were 0.044 L/m²/h, 0.268 L/m²/h, and 0.983, respectively for the SIG function. Moreover, the *DC*, *RMSE*, *OI*, and *CRM* values in Figure 3 at this node (five) for the SIG function were 0.967, 0.044 L/m²/h, 0.960, -0.001, respectively. The results were improved using the TANH function where the *SD*, *ME*, and *CC* values were 0.032 L/m²/h, 0.197 L/m²/h, and 0.991, respectively. Furthermore, the *DC*, *RMSE*, *OI*, and *CRM*

values were 0.982, 0.032 L/m²/h, 0.974, and -0.001, respectively, as depicted in Figure 3. Additionally, increasing the number of nodes to eight gave a marked improvement in the ANN model, particularly by the TANH function. The ANN model then tended to become stable and somewhat weak. At eight nodes, the *SD*, *MXE*, and *CC* for this architecture were 0.024 L/m²/h, 0.113 L/m²/h, and 0.995, respectively, for the TANH function. Additionally, the *DC*, *RMSE*, *OI*, and *CRM* values for the TANH function were 0.990, 0.024 L/m²/h, 0.982, and -0.001, respectively, as shown in Figure 3. For the SIG function, the *CC*, *ME*, and *SD* values were 0.987, 0.263 L/m²/h, and 0.039 L/m²/h, respectively. Moreover, the *DC*, *RMSE*, *OI*, and *CRM* values for the SIG function as presented in Figure 3 were 0.974, 0.039 L/m²/h, 0.966, -0.001, respectively. The

TANH function is more accurate than the SIG function. Thus, as demonstrated in Table 3 and Figure 3, the TANH function performed better than the SIG function, and there was an obvious improvement in the model when the number of hidden nodes was increased and the TANH function was used. Consequently, the best architecture was 7-8-1 as presented by the dashed line in Figure 3 and bold values in Table 3. This architecture was obtained using the TANH function and provided the best prediction of MD with the lowest error. The average contribution of each input node (variable) on the output is shown in Table 3 (bold). This factor gives the relative importance of each input variable to the training of the ANN model and is usually used to select the input variables in problems with many inputs. It can be realized that the variable with the smallest contribution is *U*. The variable with the highest contribution is *TDS_F*. The TANH function of the developed ANN model is given with the connection weightings and bias values as shown in Table 4 for forecasting the MD values as follows:

$$MD = \frac{1 - \exp(-2S_k)}{1 + \exp(-2S_k)} \tag{9}$$

where *S_k* is the sum of the TANH functions multiplied by their weights (the sum of hidden signals). It is represented as follows:

$$S_k = (3.53 \times F_1 - 2.48 \times F_2 + 1.89 \times F_3 - 3.96 \times F_4 + 2.72 \times F_5 - 2.55 \times F_6 - 3.79 \times F_7 + 1.48 \times F_8) + 1.26 \tag{10}$$

where the TANH function (*F_j*) used for this model is expressed as follows:

$$F_j = \frac{1 - \exp(-2S_j)}{1 + \exp(-2S_j)} \tag{11}$$

where *S_j* is the sum of the input variables multiplied by their weights. It can be determined as:

$$S_k = (W_{j1} \times T_o + W_{j2} \times RH + W_{j3} \times U + W_{j4} \times Rs + W_{j5} \times T_F + W_{j6} \times M_F + W_{j7} \times TDS_F) + B_j \tag{12}$$

where the connection weights *W_{ji}* and hidden biases *B_j* are presented in Table 4. This algebraic system of equations can be easily programmed in a spreadsheet (i.e. Microsoft Excel) to forecast the MD of the solar still.

Performance analysis of ANN and MLR models

The MLR model was used to compare the effectiveness and accuracy of the developed ANN model. The MLR model was developed from the MD data used in the ANN training process. The model was expressed as follows:

$$MD = -0.521 - 0.005 T_o + 0.004 RH - 0.018 U + 0.001 Rs + 0.005 T_F + 1.898 M_F - 0.002 TDS_F \tag{13}$$

Table 4 | Connection weights and biases for the developed ANN model

#HN	<i>W_{ji}</i>							<i>B_j</i>
	<i>T_o</i>	<i>RH</i>	<i>U</i>	<i>Rs</i>	<i>T_F</i>	<i>M_F</i>	<i>TDS_F</i>	
1	0.39	0.96	-0.93	0.18	0.57	0.01	0.25	0.80
2	-1.32	-1.04	2.41	-3.64	-3.19	-3.16	0.11	-0.33
3	-0.77	0.17	0.98	0.16	0.15	-0.75	-0.42	-0.70
4	2.42	-1.90	0.94	-2.09	1.15	1.93	1.74	-0.36
5	-0.78	-2.74	-1.51	-2.27	-0.02	0.38	-2.93	0.32
6	0.00	1.81	-1.91	5.89	2.36	-2.15	-0.69	-0.51
7	4.21	0.73	-3.97	-2.03	-0.70	2.10	5.21	0.31
8	-1.80	2.48	3.19	0.81	-0.25	0.14	-1.04	0.08

HN: no. of hidden neurons; *W_{ji}*: connection weights between input and hidden layer; *B_j*: hidden biases; *T_o*: ambient temperature; *RH*: relative humidity; *U*: wind speed; *Rs*: solar radiation; *T_F*: temperature of feed water; *M_F*: feed flow rate; *TDS_F*: total dissolved solids of feed.

Equation (12) shows that the T_o , U , and TDS_F were inversely proportional to MD. Furthermore, the RH , Rs , T_F , and M_F were directly proportional to MD. Table 5 illustrates the standard error (SE) of the regression, probability (p -value), and t statistic (t -stat) of the MLR model parameters. The significance of each coefficient in Equation (12) was determined by t -stat and p -value, which are presented in Table 5. Larger t -stat and smaller p -value indicate greater significance of the corresponding coefficient. Table 5 also shows the meaningfulness degrees of the input variables. This degree of meaningfulness is determined via the p -value less than 0.05. By reviewing the p -values from Table 5, a significant relationship was found between independent variables (RH , U , Rs , M_F , and TDS_F) and dependent variable (MD) at a statistical significance level of 0.05. This finding can be attributed to the p -value of these variables, which is less than 0.05. The T_o and T_F were not statistically significant as their p -value is greater than 0.05. Thus, the significance ranking of the input variables is determined as Rs , U , M_F , TDS_F , and RH .

Figure 4 indicates the comparison between the predicted versus observed MD using ANN and MLR models during the training process. For the ANN model, the data points were mostly evenly and tightly distributed around the 1:1 line. There was a very close visual agreement between the observed MD and the results obtained by the ANN model. Furthermore, the figure indicates that many points given by the MLR model during the training process are located above and below the 1:1 line for the output.

Table 5 | Standard error of regression coefficients, t statistic, and probability of MLR model parameters

Model parameters	SE	t-Stat	p-Value
Intercept	0.189	-2.751	0.007
T_o	0.005	-1.026	0.307
RH	0.001	2.562	0.012
U	0.004	-4.789	6×10^{-6}
Rs	4×10^{-5}	29.297	4.68×10^{-52}
T_F	0.006	0.874	0.384
M_F	0.532	3.570	0.001
TDS_F	0.001	-2.826	0.006

T_o : ambient temperature; RH : relative humidity; U : wind speed; Rs : solar radiation; T_F : temperature of feed water; M_F : feed flow rate; TDS_F : total dissolved solids of feed.

Thus, the figure shows that the ANN model gives an excellent match between the observed and predicted values. The overall performance of the ANN model and the MLR model was assessed using the statistical analyses shown in Table 6, which supports the better performance of the ANN model compared to the MLR model. From Table 6 and using the training data set, the MLR model had a DC value that was about 8% less accurate than that from the ANN model. The $RMSE$ value for the MLR model (0.119 L/m²/h) was almost five times the value for the ANN model (0.024 L/m²/h). Meanwhile, the ANN model had an OI value that was about 17% more accurate than that from the MLR model. The CRM value for the ANN model was closer to zero than its value for the MLR model. Figure 5 shows the relative errors of the predicted MD values for the ANN and MLR models during the training phase. The relative errors of the predicted MD values for the ANN model were mostly around +10 to -10%, except for a few data points. For the MLR model, the figure shows more relative errors than the ANN model.

Figure 4 presents a comparison of the observed and predicted values for both the ANN and MLR models using the testing data set. The figure shows the fit of the trained ANN between the predicted and observed MD values. The tight banding around the 1:1 line demonstrates the remarkable agreement between the predicted and observed data. From Table 6 and using the testing data set, the MLR model had a DC value that was about 5% less accurate than from the ANN model. The $RMSE$ value for the MLR model (0.128 L/m²/h) was almost double that of the value for the ANN model, and the OI value for the ANN model was approximately 16% more accurate than the MLR model. The CRM value for the ANN model was closer to zero than that of the MLR model. Furthermore, the CRM value for the MLR was nearly 7.5 times the value of the ANN model. Figure 5 shows the relative errors for the ANN and MLR models during the testing process. Figure 5 indicated that the relative errors of the predicted MD values are not considerable and most of it falls in the domain of +10 to -10% for the ANN model.

Figure 4 describes the relationship between the observed and predicted values of MD using the ANN and MLR models during the validation process. Similar to the training and testing processes, the ANN model provides better

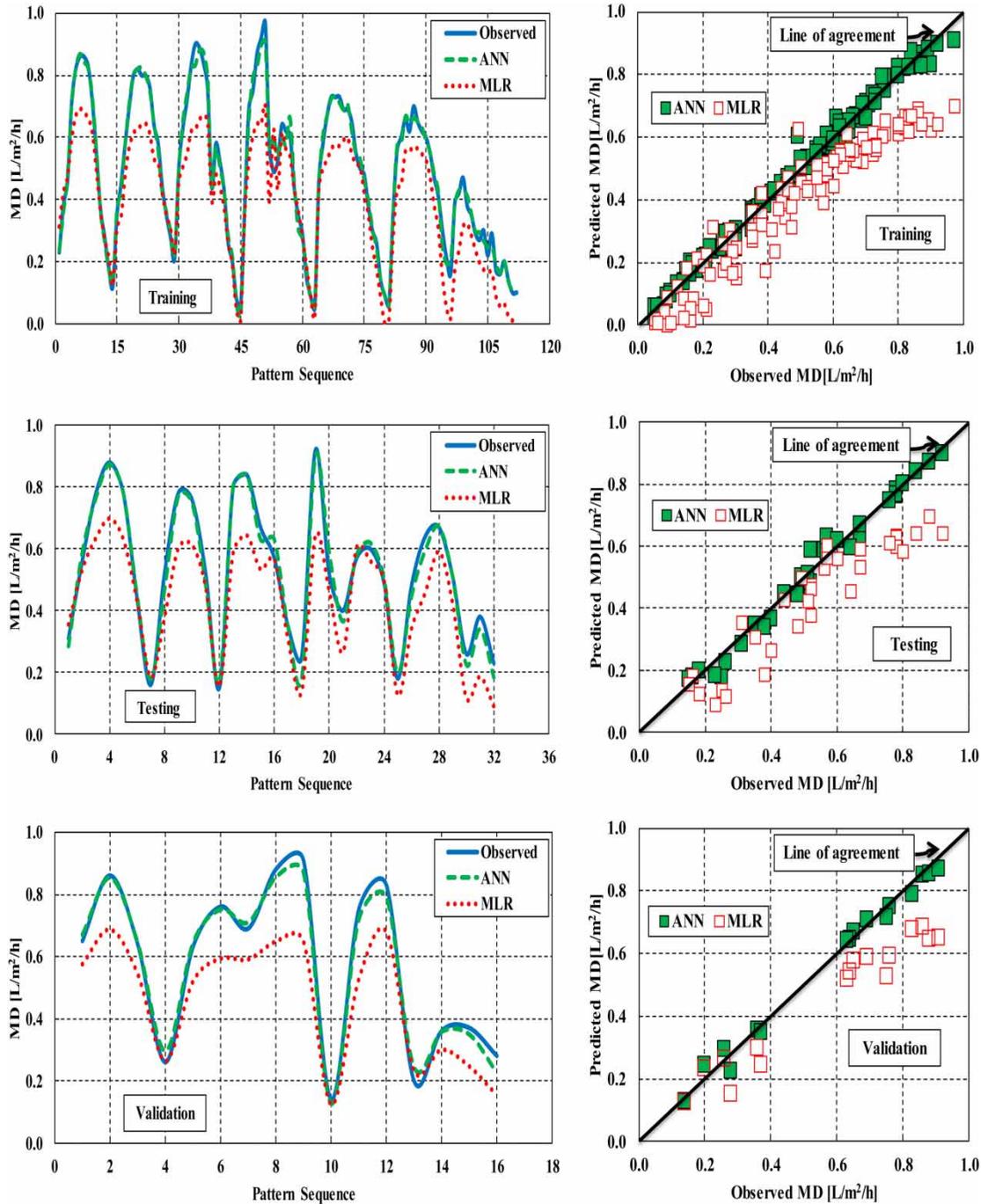


Figure 4 | Comparison between the observed and predicted values of MD using ANN and MLR models during the training, testing, and validation processes.

agreement between the observed and predicted values than the MLR model. The figure illustrates that the ANN model provides an excellent match between the observed and predicted values. From Table 6 and using the validation data

set, the MLR model had a DC value that was approximately 3% less accurate than from the ANN model. The RMSE value for the MLR model (0.142 L/m²/h) was almost three times the value of the ANN model. The OI value for the

Table 6 | Statistical parameters for assessing the performance of the ANN and MLR models during training, testing, and validation processes

Statistical parameters	Training		Testing		Validation	
	ANN	MLR	ANN	MLR	ANN	MLR
DC	0.990	0.910	0.918	0.868	0.972	0.945
RMSE	0.024	0.119	0.070	0.128	0.047	0.142
OI	0.982	0.813	0.903	0.745	0.953	0.752
CRM	-0.001	-0.190	-0.025	-0.186	-0.027	-0.204

DC: determination coefficient; RMSE: root mean-square error; OI: overall index of model performance; CRM: coefficient of residual mass; ANN: artificial neural network; MLR: multiple linear regression.

ANN model was closer to one than for the MLR model. Moreover, the OI value for the ANN model was approximately 20.1% more accurate than the MLR model. The CRM value for the ANN model was closer to zero than for the MLR model. Additionally, the CRM value for the MLR model was almost 7.5 times the value of the ANN model. Moreover, Figure 5 displays the relative errors of the predicted MD values using the validation data set for the ANN and MLR models. The relative errors of the predicted MD for the ANN model were mostly in the vicinity of $\pm 10\%$. The low relative errors demonstrate the strength of the ANN model.

Figures 4 and 5 demonstrate the inaccuracy of some values predicted and obtained from the MLR model, while most of the predictions were highly precise when using the ANN model. This shows that the MLR model is not an accurate predicting technique for MD. Thus, the ANN model produced a better fit with the observed data during the training, testing, and validation processes. Table 6 shows that better agreement between the observed and predicted MD values is obtained using the ANN model. This agreement is reflected in the DC, RMSE, OI, and CRM results, as mentioned previously. This result agrees with the findings of Şahin *et al.* (2013), El Badaoui *et al.* (2013), and Mashaly & Alazba (2016c, 2017c).

CONCLUSIONS

The capability of solar stills to yield water is highly beneficial for small communities under hyper-arid environments. Consequently, solar stills should be optimally designed and operated, and the prediction of solar still production or

water being distilled (MD) is one of the important parameters that should be precisely determined. The prediction of MD helped in determining the amount of potential distilled water attainable by the solar still and to ensure the adequacy of productivity; that is, that sufficient water quantities are achieved. This will help in decision making that will support various development plans. One technique of predicting the MD is the use of the ANN model. Seven variables were used as inputs to the ANN model in the input layer, namely, T_o , RH, U , R_s , T_F , TDS_F , and M_F . One neuron in the output layer represents the output (MD). A feed-forward back-propagation algorithm was used to train the ANN model. Several neural network architectures with different numbers of neurons in the hidden layer were trained and tested to determine the architecture that gave the minimal error and best performance. Eight neurons were the best number of neurons in the hidden layer. The 7-8-1 architecture was the optimal ANN architecture. TANH was used as the activation function in the hidden and output layers and was better than the SIG function. The findings from the developed ANN model were compared with those from the MLR. The performance of the models was evaluated by DC, RMSE, OI, and CRM. From the results, the ANN model demonstrated better prediction performance than the MLR model and revealed adequate precision in the forecasting of MD. The results revealed that the developed ANN model has a very high DC and OI between the predicted and the observed values of MD. Furthermore, the developed ANN model has a very low RMSE and CRM between the predicted and the observed values of MD. These results support the applicability of the developed ANN model. The MLR model results were also satisfactory for predicting MD, but it was

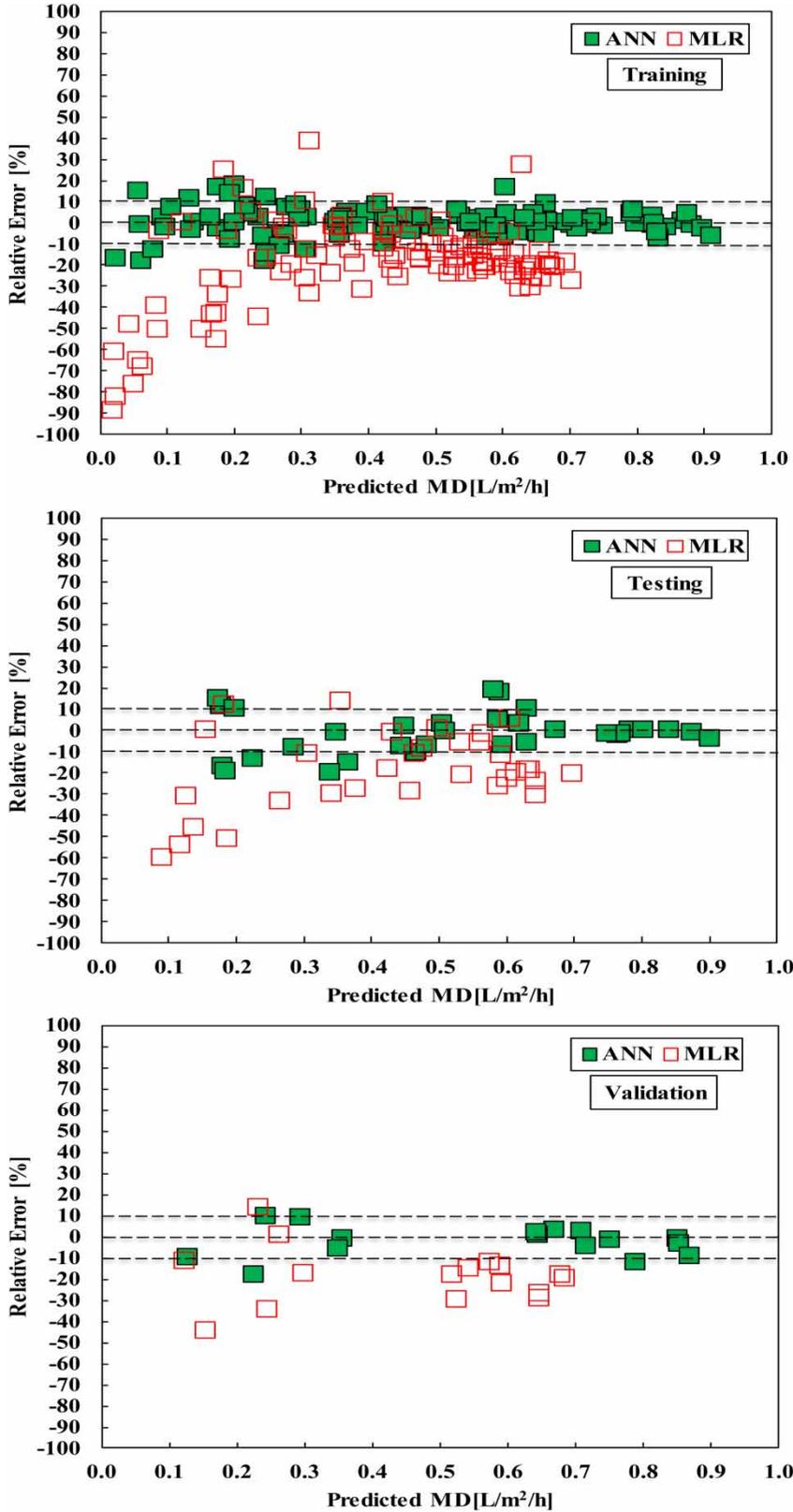


Figure 5 | Relative errors for the ANN and MLR models using the training, testing, and validation data sets.

less accurate compared to the ANN model. In this study, the ANN model was proven to be a sufficient, accurate, and successful tool for modeling the MD without the need for comprehensive experimental investigations. Therefore, this investigation allows a preliminary decision on the usability under conditions in which the solar still is required.

ACKNOWLEDGEMENT

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

REFERENCES

- Abutaleb, A. S. 1991 A neural network for the estimation of forces acting on radar targets. *Neural Netw.* **4**, 667–678.
- Aybar, H. Ş., Egelioglu, F. & Atikol, U. 2005 An experimental study on an inclined solar water distillation system. *Desalination* **180**, 285–289.
- Badran, O. O. 2007 Experimental study of the enhancement parameters on a single slope solar still productivity. *Desalination* **209**, 136–143.
- Badran, A. A., Assaf, L. M., Kayed, K. S., Ghaith, F. A. & Hammash, M. I. 2004 Simulation and experimental study for an inverted trickle solar still. *Desalination* **164**, 77–85.
- Benli, H. 2013 Determination of thermal performance calculation of two different types solar air collectors with the use of artificial neural networks. *Int. J. Heat Mass Transf.* **60**, 1–7.
- Caner, M., Gedik, E. & Kecebas, A. 2011 Investigation on thermal performance calculation of two type solar air collectors using artificial neural network. *Expert Syst. Appl.* **38**, 1668–1674.
- Cooper, P. I. 1969 Digital simulation of transient solar still processes. *Solar Energy* **12**, 313–331.
- Demuth, H. & Beale, M. 2004 *Neural Network Toolbox: For Use with MATLAB (Version 4.0)*. The Math Works, Inc., Natick, MA, USA.
- Dhiman, N. K. & Tiwari, G. N. 1990 Effect of water flowing over the glass cover of a multi-wick solar still. *Energy Convers. Manag.* **30**, 245–250.
- El Badaoui, H., Abdallaoui, A. & Chabaa, S. 2013 Using MLP neural networks for predicting global solar radiation. *Int. J. Eng. Sci.* **2**, 48–56.
- Frick, B. 1970 Some new considerations about solar stills. In: *Proceedings of International Solar Energy Congress*. International Solar Energy Society, Melbourne, Australia, p. 395.
- Géczy-Víg, P. & Farkas, I. 2010 Neural network modelling of thermal stratification in a solar DHW storage. *Solar Energy* **84**, 801–806.
- Giacomino, A., Abollino, O., Malandrino, M. & Mentasti, E. 2011 The role of chemometrics in single and sequential extraction assays: a review. Part II. Cluster analysis, multiple linear regression, mixture resolution, experimental design and other techniques. *Anal. Chim. Acta* **688**, 122–139.
- Haykin, S. 1999 *Neural Networks: A Comprehensive Foundation*, 2nd edn. Prentice Hall, Upper Saddle River, NJ, USA, p. 842.
- Kabeel, A., Hamed, M. H. & Omara, Z. 2012 Augmentation of the basin type solar still using photovoltaic powered turbulence system. *Desa. Water Treat.* **48**, 182–190.
- Kabeel, A. E., Omara, Z. M. & Younes, M. M. 2015 Techniques used to improve the performance of the stepped solar still – A review. *Renew. Sustain. Energy Rev.* **46**, 178–188.
- Lecoeuche, S. & Lalot, S. 2005 Prediction of the daily performance of solar collectors. *Int. Commun. Heat and Mass Transf.* **32**, 603–611.
- Mashaly, A. F. & Alazba, A. A. 2016a 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. 2016b 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. 2016c Neural network approach for predicting solar still production using agricultural drainage as a feedwater source. *Desal. Water Treat.* **57** (59), 28646–28660.
- 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. *Desal. Water Treat.* **57**, 12019–12036.
- Mashaly, A. F. & Alazba, A. A. 2017a Application of adaptive neuro-fuzzy inference system (ANFIS) for modeling solar still productivity. *J. Water Suppl. Res. Technol. Aqua* **66** (6), 367–380.
- Mashaly, A. F. & Alazba, A. A. 2017b Thermal performance analysis of an inclined passive solar still using agricultural drainage water and artificial neural network in arid climate. *Solar Energy* **153**, 383–395.
- Mashaly, A. F. & Alazba, A. A. 2017c Artificial intelligence for predicting solar still production and comparison with stepwise regression under arid climate. *J. Water Suppl. Res. Technol. Aqua* **66** (3), 166–177.
- 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.
- Radhwan, A. M. 2004 Transient performance of a steeped solar still with built-in latent heat thermal energy storage. *Desalination* **171**, 61–71.
- Riffat, S. B., Zhao, X., Boukhanouf, R. & Doherty, P. S. 2005 Theoretical and Experimental investigation of a novel hybrid heat-pipe solar collector. *Int. J. Green Energy* **1**, 515–542.
- Rumelhart, D. E., Hinton, G. E. & Williams, R. J. 1986 Learning representations by back-propagating errors. *Nature* **323**, 533–536.

- Şahin, M., Kaya, Y. & Uyar, M. 2013 Comparison of ANN and MLR models for estimating solar radiation in Turkey using NOAA/AVHRR data. *Adv. Space Res.* **51**, 891–904.
- Scheaffer, R., Mulekar, M. & McClav, J. 2011 *Probability and Statistics for Engineers*, 5th edn. Brooks/Cole, Boston, USA, p. 599.
- Sodha, M. S., Nayak, J. K., Tiwari, G. N. & Kumar, A. 1980 Double basin solar still. *Energy Convers. Manage.* **20**, 23–32.
- Tiwari, G. N. & Rao, V. S. V. B. 1984 Transient performance of single basin solar still with water flowing over the glass cover. *Desalination* **49**, 231–241.
- Toure, S. & Meukam, P. 1997 A numerical model and experimental investigation for a solar still in climatic conditions in Abidjan (Côte d'Ivoire). *J. Renew. Energy* **11**, 319–330.
- Velilla, E., Valencia, J. & Jaramillo, F. 2014 Performance evaluation of two solar photovoltaic technologies under atmospheric exposure using artificial neural network models. *Solar Energy* **107**, 260–271.
- Yeh, H. & Chen, L. 1986 The effects of climatic, design and operational parameters on the performance of wick-type solar distillers. *Energy Convers. Manage.* **26**, 175–180.

First received 21 June 2017; accepted in revised form 10 April 2018. Available online 3 May 2018