Estimation of daily dew point temperature using genetic programming and neural networks approaches

Jalal Shiri, Sungwon Kim and Ozgur Kisi

ABSTRACT

The present study investigates the ability of two different artificial neural network (ANN) models and gene expression programming (GEP) technique for estimating daily dew point temperature by using recorded weather data. The weather data used consist of 8 years of daily records of air temperature, wind speed, relative humidity, atmospheric pressure, incoming solar radiation and dew point temperature from two weather stations (Seoul and Incheon, in the Republic of Korea). Two different data management scenarios are applied in this paper. In the first scenario, weather data obtained from each station are used to estimate T_{dew} at the same station (at-station approach). In the second scenario, the ANN and GEP models are used for estimating dew point temperature of each station by using the data of the other station (cross-station application), through the optimal input combinations of the first scenario. Comparison of the results reveals that the GEP model surpasses ANN in estimating daily dew point temperature values.

Key words | agro-climatology, dew point temperature, modeling, soft computing

Jalal Shiri (corresponding author)
Water Engineering Department,
Faculty of Agriculture,
University of Tabriz,
Tabriz,
Iran
E-mail: *j_shiri2005@yahoo.com*

Sungwon Kim
Department of Railroad and Civil Engineering,
Dongyang University,
Yeongju,
Republic of Korea

Ozgur Kisi
Canik Basari University,
Architecture and Engineering Faculty,
Civil Engineering Department,
Samsun,
Turkey

INTRODUCTION

Dew is a condensation of atmospheric moisture on objects that are colder than the dew point temperature of the surrounding air. From an agricultural point of view, dew phenomenon may decrease the vapor pressure deficit in the vicinity of the dew drops leading to better photosynthesis (Slatyer 1967) and enhancing water content recovery after extreme water losses (Went 1955). Factors affecting the formation of dew phenomenon in natural ecosystems are radiation exchange between the Earth's surface and atmosphere, turbulent heat, and water vapor pressure (Atzema *et al.* 1990).

Dew point temperature corresponds to the circumstances in which the gas is cooled down at a constant pressure so that the relative humidity (R_H) rises above 100% (Berning 2012). This is the temperature at which the moisture (water vapor) in the air begins to condense into dew or water droplets. Dew point temperature is termed a conservative property because changes in temperature do not convert an air's dew point temperature. However, the addition or extraction of moisture, to or from an air, will increase or decrease the dew point temperature, respectively. CP-165/UM psychrometric computer is used for the calculation of dew point temperature by using the dry-bulb temperature and the wet-bulb depression. The wet-bulb depression is equal to the difference between the wet-bulb temperature and dry-bulb temperature.

The accurate estimation/prediction of the dew point temperature is very important as it determines whether it will rain or snow as well as how high the danger is for a grass or brush fire during a dry spell. It also can be used for determining the amount of available moisture in the air (Shank *et al.* 2008) as well as for estimating the near surface humidity which is crucial from an agricultural viewpoint. Irrigated agriculture also needs the values of this variable for irrigation scheduling (Mahmood *et al.* 2008). Many hydro-climatologic models require dew point values as one of the important input variables for estimating reference evapotranspiration (Mahmood & Hubbard 2005). So far, a number of studies have proposed various approaches for estimating dew point temperature.

doi: 10.2166/nh.2013.229

In recent years, the application of artificial intelligence (e.g., artificial neural networks (ANNs) and genetic programming (GP)) techniques in hydrology, agro-meteorology and climatology has become viable and notable papers have been published. Recent experiments have reported that ANN may offer some promising results in hydro-meteorology and water resources engineering (see ASCE Task Committee 2000a, b). Abdel-Aal (2004) applied an abductive neural network approach to forecast hourly air temperature. Smith *et al.* (2005) developed an enhanced ANN for air temperature prediction by including information on seasonality and modifying parameters of an existing ANN model. Shank *et al.* (2008) applied ANN for predicting dew point temperature. Shiri *et al.* (2011) compared neuro-fuzzy and neural network models for estimating pan evaporation values in the state of Illinois, USA. Kim *et al.* (2012) compared three kinds of ANN models for estimating daily pan evaporation values in different climatic zones and demonstrated the superiority of ANN models over other applied traditional methods. Kisi & Shiri (2012a) introduced a new wavelet-neuro-fuzzy model for predicting short-term groundwater level fluctuations. Pour Ali Baba *et al.* (2013) applied neuro-fuzzy and neural network techniques for estimating daily reference evapotranspiration using available and estimated climatic data. Abdellatif *et al.* (2013) applied a hybrid generalized linear and Levenberg–Marquardt artificial neural network model for downscaling future rainfall.

The application of GP (i.e., gene expression programming, GEP) to modeling issues in hydrology and agro-meteorology also includes, for example, modeling risks in water supply (Babovic *et al.* 2001), modeling rainfall–runoff (e.g., Savic *et al.* 1999; Kisi *et al.* 2013), modeling suspended sediment load (e.g., Kisi & Shiri 2012b), predicting stream flow (Kisi & Shiri 2010), predicting groundwater table depth fluctuations (Shiri & Kisi 2011a; Shiri *et al.* 2013), modeling evaporation and evapotranspiration (Shiri & Kisi 2011b; Shiri *et al.* 2012), seawater level forecasting (Kisi *et al.* 2012), wind speed forecasting (Kisi *et al.* 2011), rainfall prediction (Kisi & Shiri 2011), and estimating daily incoming solar radiation (Landeras *et al.* 2012).

The main purpose of this study is to investigate the ability of GEP and two different ANN methods in estimating daily dew point temperature. Daily weather data from two weather stations, Seoul and Incheon, in the Republic of Korea are used in the present study. GEP models are compared with two different ANN methods, Elman discrete recurrent ANN models trained with (1) conjugate gradient learning algorithm and (2) quick prop learning algorithm.

MATERIALS AND METHODS

Study area and data descriptions

In the present study, daily weather data from two weather stations in the Republic of Korea, namely, Seoul and Incheon stations (operated by the Korea Meteorological Administration), were used for estimating dew point temperature values. Figure 1 shows the locations of the studied weather stations. Table 1 presents the coordinates of the studied stations along with the continentality indexes of

Figure 1 | Geographical positions of the studied weather stations.

Table 1 | Summary of the geographical positions of the studied weather stations

Station	Latitude (°N)	Longitude (°W)	Altitude (m)	CI^{S}	CI^{CU}
Seoul	37° 57′	126° 96′	85.5	27.62	2.04
Incheon	37° 47′	126° 62′	68.9	26.64	1.56

each weather station. The selected indexes were Supan (CI^{S}) and Currey (CI^{CU}) indexes. These indicators were selected for their simplicity, as they only demand temperature and latitude records.

$$CI^{S} = M_i - m_i$$

$$CI^{CU} = \frac{M_i - m_i}{1 + \theta/3}$$

where M_i is the maximum monthly average temperature (°C); m_i is the minimum monthly average temperature (°C); θ is the latitude (degrees). From the table it can be followed that there are small differences between the stations from the continentality viewpoint.

The data sample consists of daily records of average air temperature (T_{mean}), relative humidity (R_H), atmospheric pressure (P), solar radiation (R_S), and wind speed (W_S) as well as dew point temperature (T_{dew}). The existing data cover a period of 8 years (1999–2006), of which the first 6 years (75% of whole data) were used for training the applied models and the last 2 years were reserved for independent testing of the models. Figure 2 shows the time series plot of the observed dew point temperature values during the study period. Table 2 represents the statistic parameters of the used climatic variables in the studied stations. In the table, the terms X_{mean}, X_{max}, X_{min}, S_X, C_V and C_{SX} denote the mean, maximum, minimum, standard deviation, coefficient of variation and skewness coefficient, respectively. In both stations, W_S data show skewed distribution (see the C_{SX} values in Table 2).

Artificial neural networks

Artificial neural networks are massively parallel-distributed data processing systems consisting of a large number of highly interconnected artificial nodes with performance characteristics resembling biological neural networks of the human brain (Haykin 1999). Due to the advantages of

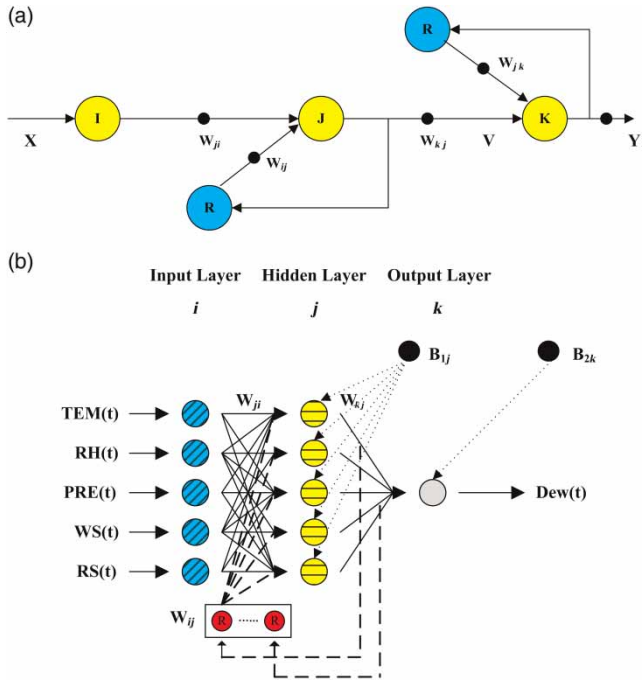

Figure 2 | (a) DRNNM architecture with recurrent feedback; (b) schematic diagram of EDRNNM.

Table 2 | Statistical parameters of the used data set during the study period

Parameter	Unit	X_{mean}	X_{max}	X_{min}	S_X	C_V	C_{SX}
Seoul weather station							
T_{mean}	°C	12.7	31	−14.6	9.7	0.77	−0.26
R_H	%	67.2	97.5	28	14.2	0.21	−0.14
P	hPa	1016	1039	991	8.1	0.01	0.01
R_S	MJ/m^2	12.9	36.2	0.02	7.2	0.55	0.29
W_S	$m.\ s^{-1}$	2.6	9.7	0	1.3	0.49	1.19
T_{dew}	°C	6.4	25.3	−24.1	11.4	1.8	−0.25
Incheon weather station							
T_{mean}	°C	12.9	30.4	−15.5	10.2	0.79	−0.28
R_H	%	41	95	8	16.3	0.39	0.57
P	hPa	1016	1039	991	8.1	0.01	0.02
R_S	MJ/m^2	11.8	31.1	0.01	6.5	0.54	0.35
W_S	$m.\ s^{-1}$	2.1	6.7	0.1	0.87	0.41	0.75
T_{dew}	°C	5.3	24.3	−25.4	11.6	2.1	−0.24

ANNs in modeling, they have become extremely popular for the forecasting of water resources variables. Feedback in the ANN-based model is advantageous for certain situations. When the output of a node is fed back into a node

in an earlier layer, the output of that node is a function of both inputs from the previous layer at time t and its own output that existed at an earlier time, $t-\Delta t$, where Δt is the time for one cycle of calculation. Such networks exhibit characteristics similar to short-term memory, because the output of the networks depends on current and prior inputs (Li *et al.* 1989; Ryeu & Chung 1996; Giles *et al.* 1997). The ANN-based models that contain such feedback and discrete time interval are called the discrete recurrent neural networks model (DRNNM). Recurrent networks are made by adding feedback from the output node to the hidden layer and from the hidden node to the input layer through the recurrent nodes, which are buffer nodes, labeled R, and the corresponding connection weights W_{ij} and W_{jk}, as shown in Figure 2(a). The output to nodes J and K must exist before any feedback occurs. From Figure 2(a), when input time $t=0$ is applied to the input, the outputs of nodes J and K at time $t=0$ are given by Equations (1) and (2).

$$V(0) = \Phi_1[W_{ji} \cdot X(0)] \tag{1}$$

$$Y(0) = \Phi_2[W_{kj} \cdot V(0)] = \Phi_2[W_{kj}[\Phi_1[W_{ji} \cdot X(0)]]] \tag{2}$$

where Φ = transfer function; $X(0)$ = the input value of input node at time $t=0$; $V(0)$ = the output value of hidden node at time $t=0$, and $Y(0)$ = output value of output node at time $t=0$. As the process proceeds step by step, the feedback terms in the nodes of J and K are not included until time $t=1$. The feedback terms, however, must be added for all subsequent times. The output of the networks for nodes J and K at time $t=1$ is given by Equations (3) and (4).

$$\begin{aligned} V(1) &= \Phi_1[[W_{ji} \cdot X(1)] + [W_{ij} \cdot V(0)]] \\ &= \Phi_1[[W_{ji} \cdot X(1)] + [W_{ij} \cdot [\Phi_1[W_{ji} \cdot X(0)]]]] \end{aligned} \tag{3}$$

$$\begin{aligned} Y(1) &= \Phi_2[W_{kj} \cdot V(1)] + [W_{jk} \cdot Y(0)] \\ &= \Phi_2[[W_{kj} \cdot \Phi_1[[W_{ji} \cdot X(1)] + [W_{ij} \cdot [\Phi_1[W_{ji} \cdot X(0)]]]] \\ &\quad + [W_{jk} \cdot \Phi_2[W_{kj} \cdot [\Phi_1[W_{ji} \cdot X(0)]]]]] \end{aligned} \tag{4}$$

The outputs of the networks for nodes J and K at time $t=2$ can be calculated by using Equations (3) and (4). As the most recent inputs are introduced, the influence of the earlier term becomes negligible in DRNNM architecture (Tsoukalas & Uhrig 1997). DRNNM may be more powerful than feed forward neural networks model (FFNNM), since it can recognize and recall temporal and spatial patterns. The behavior of DRNNM, however, is much more complex than that of FFNNM. For the architecture of FFNNM, the output is constant for a fixed input and is a function only of the networks input. For the architecture of DRNNM, however, the output of the networks is a function of time. For a given input and initial networks output, the response of DRNNM may converge to a stable output (Hagan *et al.* 1995).

Construction of Elman discrete recurrent neural network model

Elman discrete recurrent neural network model (EDRNNM), which was suggested by Elman (1990), is constructed for dew point temperature estimation in this study. All calculation processes for the training and testing of EDRNNM are carried out by the NeuroSolution 5.0 software (Neuro Dimension Inc. 2005). Kim & Kim (2008) developed EDRNNM for flood stage forecasting in South Korea. They used sensitivity analysis to reduce the uncertainty of input data information of EDRNNM. With the results of sensitivity analysis, they could avoid unnecessary data collection and operate the flood stage forecasting system economically. In this study, EDRNNM is trained by the conjugate gradient and QuickProp back-propagation algorithm (BPA). The results of output node with five input combinations, Dew(t), is given by Equation (5), and Figure 2(b) represents the developed EDRNNM architecture with five hidden nodes.

$$\mathrm{Dew}(t) = \Phi_2\left[\left[\sum_{k=1}^{1} W_{kj} \cdot \Phi_1\left(\left(\sum_{j=1}^{5} W_{ji} \cdot X(t)\right) . + \left(\sum_{j=1}^{5} W_{ji} \cdot \left(\Phi_1\left(\sum_{j=1}^{5} W_{ji} \cdot X(t-\alpha)\right)\right)\right)\right) + B_1\right] + B_2\right] \tag{5}$$

where i, j, k = input, hidden and output layers of EDRNNM; Dew(t) = the dew point temperature; $\Phi_1(\cdot)$ = the transfer function in hidden layer; $\Phi_2(\cdot)$ = the transfer function in output layer; W_{ji} = the connection weights between input

and hidden layers; W_{kj} = the connection weights between hidden and output layers; W_{ij} = the connection weights between recurrent nodes and hidden layer; B_1 = the biases in hidden layer; B_2 = the bias in output layer; $X(t)$ = the climatic variables in input layer; and α = the lead-time of the climatic variables in this study.

The main difference between the EDRNNM and DRNNM structures is as follows. The EDRNNM has only feedback from the outputs of the hidden layer to the inputs of the same hidden layer, whereas the DRNNM has feedbacks from the outputs of the output layer to the inputs of the same output layer as well as from the outputs of hidden layer to the inputs of the same hidden layer at the same time. The complicated internal procedure of the EDRNNM can be reduced by the use of only one-feedback connection. Therefore, the EDRNNM consists of a three-layer network with feedback only from the outputs of the hidden layer to the inputs of the same hidden layer. This recurrent connection allows the EDRNNM to detect and generate time-varying patterns (Kim & Cho 2003; Kim & Kim 2008).

Gene expression programming

The procedure starts by random generation of chromosomes of the certain program (initial population), then the generated chromosomes are expressed and the fitness of each individual program is evaluated against a set of fitness cases (Ferreira 2006). The programs are then selected according to their own fitnesses (their performance in that particular environment). The mentioned process is repeated until a good solution can be found for the studied phenomenon. In the present work, the GeneXpro program was applied for modeling daily dew point temperature.

The application of GEP involves the following general steps:

1. Selection of fitness function.
2. Choosing the set of terminals T and the set of functions F to create the chromosomes.
3. Choosing the chromosomal architecture.
4. Choosing the linking function.
5. Choosing the genetic operators.

The first step consists of selecting the fitness function. For the present problem, the root relative square error was selected based on literature review (e.g., Shiri & Kisi 2011a; Kisi *et al.* 2011, 2012). The second step consists of choosing the set of terminals T and the set of functions F, to create the chromosomes. In the current problem, the terminal set includes recorded weather parameters: $\{T_{mean}, R_H, P, R_S$, and $W_S\}$. The choice of the appropriate function depends on the viewpoint of the user. In this study, different mathematical functions were utilized ($\{+, -, *, /\}$,$\{\sqrt{}, \sqrt[3]{}, \ln(x), e^x, x^2, x^3, \sin(x), \cos(x), \text{arctg}(x)\}$). The third step is to choose the chromosomal architecture. Length of head, $h = 8$, and three genes per chromosomes were employed. The fourth step is to choose the linking function. The linking function must be chosen as 'addition' or 'multiplication' for algebraic subtrees (Ferreira 2001). Here, the subtrees were linked by addition. The fifth and final step is to choose the genetic operators. The parameters used per run are summarized as follows: number of chromosomes: 30, head size: 8, number of genes: 3, linking function: addition, fitness function error type: root relative squared error, mutation rate: 0.044, inversion rate: 0.1, one point recombination rate: 0.3, two point recombination rate: 0.3, gene recombination rate: 0.1, gene transposition rate: 0.1, insertion sequence transposition rate: 0.1, root insertion sequence transposition: 0.1. These are default parameters of the GeneXpro model and can be applied for modeling issues (e.g., Shiri & Kisi 2011a, b; Landeras *et al.* 2012). The parsimony pressure tool was applied to penalize the parse trees of each GEP model for condensing the models' expressions and avoiding producing the nested functions.

Statistical parameters

Four statistical evaluation parameters were used to assess the models' performances.

1. Coefficient of determination (R^2): provides information for linear dependence between observations and corresponding estimates and can be calculated using Equation (6):

$$R^2 = \left(\frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n} (x_i - \bar{x})^2 \sum_{i=1}^{n} (y_i - \bar{y})^2}} \right) \quad (6)$$

2. Root mean square error (RMSE): describes the average magnitude of the errors by giving more weight on large errors and is calculated through Equation (7):

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{i=1}^{n}(x_i - y_i)^2}{n}} \quad (7)$$

3. Mean absolute error (MAE): is a linear scoring rule and describes only the average magnitude of the errors, ignoring their direction:

$$\mathrm{MAE} = \frac{\sum_{i=1}^{n}|x_i - y_i|}{n} \quad (8)$$

4. Nash–Sutcliffe coefficient (NS): the higher the NS value the better the model performance, and vice versa:

$$\mathrm{NS} = 1 - \frac{\sum_{i=1}^{n}(x_i - y_i)^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2} \quad (9)$$

A perfect match between simulated and observed pan evaporation values would yield NS = 1.0 where, x_i and y_i denote the observed and corresponding simulated values at ith time step, respectively. Also, $\bar{x}$ and $\bar{y}$ represent the mean observed and simulated values, respectively. Finally, n represents the number of time steps.

Legates & McCabe (1999) argued that R^2 should not be applied alone as performance indicator, since it only represents the linear dependency between the parameters and is sensitive to outliers. Therefore, other statistical measures, such as MAE and RMSE, should be applied to evaluate the models' performance. The combined use of these parameters can provide enough insight about the applied models.

RESULTS AND DISCUSSION

At-station application scenario

Based on the physical factors affecting the dew formation, input variables are selected to be T_{mean}, W_S (corresponding to turbulent temperature), R_S (corresponding to radiation exchange), P and R_H (corresponding to water vapor pressure). Table 3 sums up the introduced input combinations for estimating T_{dew}. In the present study the input variables were introduced step by step, so the final combinations structure consisted of single-, double-, triple-, quadruple-, and quintuple-input combinations. The testing performance of the applied models is listed in Tables 4 and 5 for the Seoul and Incheon weather stations, respectively. In the tables, the ANN 1 and ANN2 models correspond to the Elman discrete recurrent ANN models trained with (1) conjugate gradient learning algorithm and (2) quick prop learning algorithm, respectively. From the tables it can be seen that the GEP model surpasses the ANN in all introduced input combinations at both stations. Nonetheless, the intercomparison of the ANN models reveals that the ANN1 model gives higher accuracy than the ANN2 for both stations and all input combinations. The comparison of the input combinations used for estimating T_{dew} indicates that the quintuple-input models offer more accurate results than the other applied combinations, whereas of the remainder, the single-input models are of the lowest accuracy. It can be clearly

Table 3 | The selected input combinations

Models	T_{mean}	R_H	P	R_S	W_S
Single-input models					
1	*				
2		*			
3			*		
4				*	
5					*
Double-input models					
1	*	*			
2	*		*		
3	*			*	
4	*				*
Triple-input models					
1	*	*	*		
2	*	*		*	
3	*	*			*
Quadruple-input models					
1	*	*		*	*
2	*	*	*		*
Quintuple-input model					
1	*	*	*	*	*

Table 4 | RMSE and MAE values of the applied models in the Seoul station during the test period (2005–2006)

Input combinations	RMSE (°C)			MAE (°C)		
	GEP	ANN-1	ANN-2	GEP	ANN-1	ANN-2
Single-input models						
$\boldsymbol{T_{mean}}$	**3.55**	**3.54**	**3.61**	**2.76**	**2.74**	**2.79**
R_H	8.71	8.66	8.71	7.53	7.46	7.51
P	7.97	7.99	8.09	6.41	6.49	6.58
R_S	11.1	12	12.3	9.54	10.3	10.5
W_S	12.19	12.2	12.3	10.4	10.4	10.4
Double-input models						
$\boldsymbol{T_{mean}, R_H}$	**0.53**	**0.81**	**2.41**	**0.39**	**0.67**	**2.01**
T_{mean}, P	3.45	3.54	3.49	2.68	2.75	2.71
T_{mean}, R_S	2.72	2.73	2.76	2.02	2.11	2.11
T_{mean}, W_S	3.57	3.57	3.62	2.76	2.79	2.81
Triple-input models						
T_{mean}, R_H, P	0.62	0.97	2.93	0.44	0.76	2.22
T_{mean}, R_H, R_S	0.89	0.95	2.96	0.51	0.73	2.22
$\boldsymbol{T_{mean}, R_H, W_S}$	**0.57**	**0.94**	**2.93**	**0.50**	**0.69**	**2.01**
Quadruple-input models						
T_{mean}, R_H, W_S, R_S	0.61	2.88	2.78	0.37	1.54	2.14
$\boldsymbol{T_{mean}, R_H, W_S, P}$	**0.45**	**2.14**	**2.73**	**0.34**	**1.57**	**2.14**
Quintuple-input model						
$\boldsymbol{T_{mean}, R_H, W_S, R_S, P}$	**0.39**	**2.49**	**2.72**	**0.31**	**1.97**	**1.97**

Note: ANN1, Elman discrete recurrent neural networks model with conjugate gradient learning algorithm; ANN2, Elman discrete recurrent neural networks model with quick prop learning algorithm.

Table 5 | RMSE and MAE values of the applied models in the Incheon station during the test period (2005–2006)

Input combinations	RMSE (°C)			MAE (°C)		
	GEP	ANN-1	ANN-2	GEP	ANN-1	ANN-2
Single-input models						
$\boldsymbol{T_{mean}}$	**3.03**	**3.07**	**3.07**	**2.38**	**2.44**	**2.45**
R_H	8.05	7.63	7.75	6.73	6.74	6.51
P	7.66	7.71	7.77	6.21	6.29	6.35
R_S	11.1	11.2	11.6	9.36	9.57	9.95
W_S	11.2	11.2	11.2	9.33	9.39	9.35
Double-input models						
$\boldsymbol{T_{mean}, R_H}$	**0.65**	**2.66**	**2.75**	**0.51**	**2.06**	**2.15**
T_{mean}, P	2.99	2.96	2.96	2.35	2.38	2.36
T_{mean}, R_S	2.73	2.74	2.79	2.17	2.18	2.17
T_{mean}, W_S	3.04	3.04	3.04	2.39	2.39	2.39
Triple-input models						
T_{mean}, R_H, P	0.43	1.41	2.7	0.31	1.08	2.06
T_{mean}, R_H, R_S	0.52	0.87	2.5	0.31	1.59	1.96
$\boldsymbol{T_{mean}, R_H, W_S}$	**0.43**	**0.78**	**2.5**	**0.31**	**1.02**	**1.02**
Quadruple-input models						
T_{mean}, R_H, W_S, R_S	0.66	1.7	2.31	0.45	1.21	1.82
$\boldsymbol{T_{mean}, R_H, W_S, P}$	**0.56**	**1.6**	**2.15**	**0.41**	**1.15**	**1.71**
Quintuple-input model						
$\boldsymbol{T_{mean}, R_H, W_S, R_S, P}$	**0.55**	**1.92**	**2.55**	**0.37**	**1.54**	**2.02**

Note: ANN1, Elman discrete recurrent neural networks model with conjugate gradient learning algorithm; ANN2, Elman discrete recurrent neural networks model with quick prop learning algorithm.

seen from the single-input models in Tables 4 and 5 that T_{mean} is the most important variable on T_{dew}. Double-input combinations were obtained by adding other variables to the most effective variable T_{mean} step by step. From double-input models, the T_{mean} and R_H input combination seems to be more effective on T_{dew} than the other double-input combinations. Among the triple-input combinations, T_{mean}, R_H, and W_S combination has the best accuracy in estimating T_{dew}. Comparison of quadruple-input models reveals that the T_{mean}, R_H, W_S, and P combination is more effective on T_{dew} than the other combination. According to the test results given in Tables 4 and 5, the most effective variables on T_{dew} are ranked as the T_{mean}, R_H, W_S, P, and R_S. It is clear from Tables 4 and 5 that the triple-, quadruple-, and quintuple-input models for GEP produced better results compared with those of ANN1 and ANN2, respectively. Comparison shows that GEP is more powerful than ANN1 and ANN2 to generalize the daily dew point temperature.

Observed and estimated T_{dew} values by using the optimal single-input models during the test period are shown in Figure 3 in the form of scatterplots. It is clearly seen from the graphs that a slight difference exists among the GEP, ANN1 and ANN2 models for both stations. This confirms the RMSE and MAE values given in Table 4. Figure 4 illustrates the estimates of the optimal double-input models for the Seoul and Incheon stations during the test period. It can be obviously seen from the fit line equations and R^2 values that the GEP model comprising T_{mean} and R_H inputs performs better than the ANN1 and ANN2 models for both stations. The ANN2 model shows the worst accuracy. Comparison of the two stations' results indicates that the models are more successful in Seoul station than

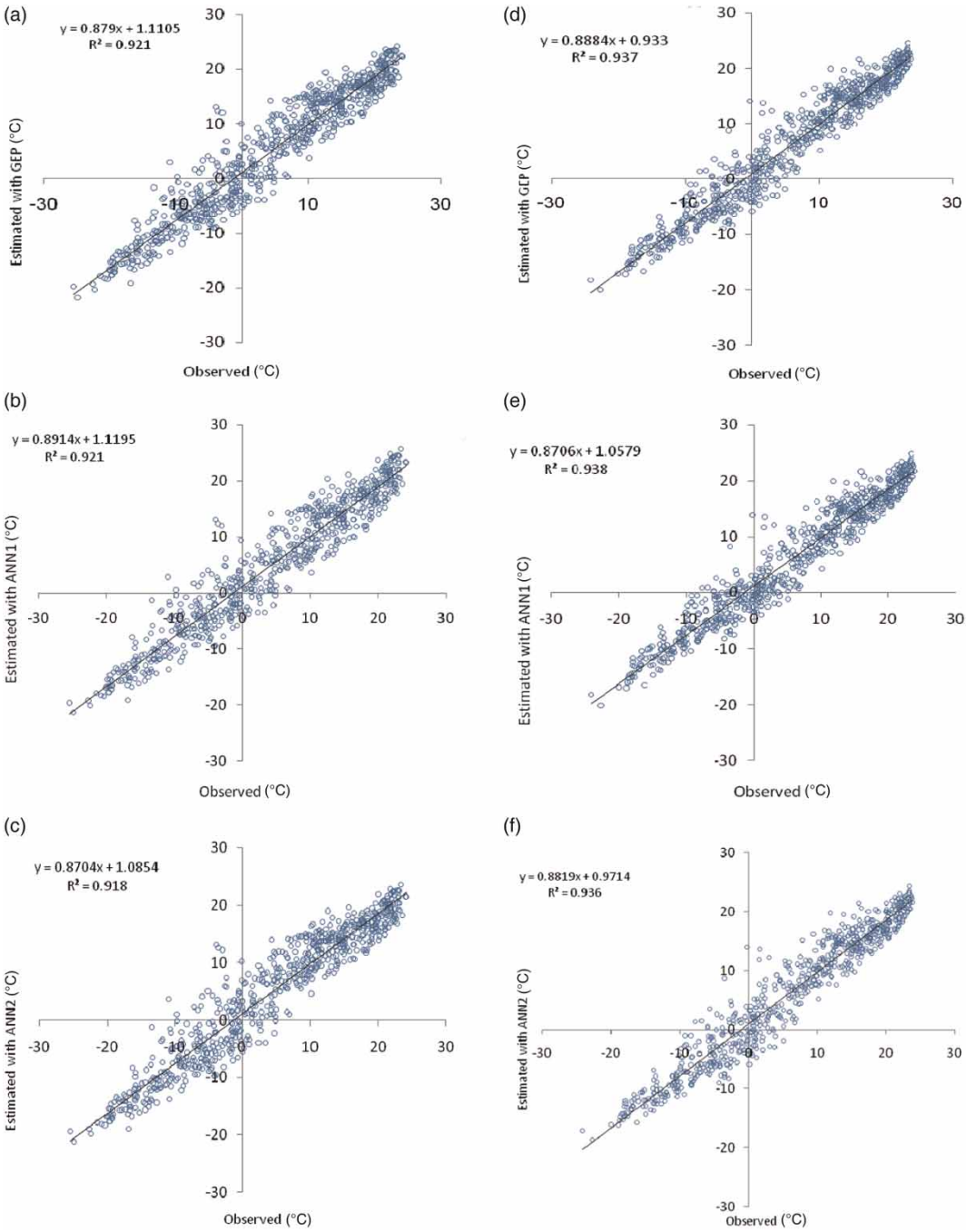

Figure 3 | Observed and estimated values of T_{dew} by using the optimal single-input models in Seoul ((a), (b), (c)) and Incheon ((d), (e), (f)) stations (testing period: 2005–2006).

Incheon. The reason behind this may be the fact that the Incheon station has higher skewed distributed R_H data than the Seoul station. The estimates of the optimal triple-input models during the test period are shown in Figure 5. Here, also, the GEP model provides less scattered estimates than the ANN models for both stations. Observed and estimated T_{dew} values by using the optimal quadruple- and quintuple-input models during the test period are illustrated in Figures 6 and 7. It is clear from these figures that there is a significant difference between the GEP and ANN models in

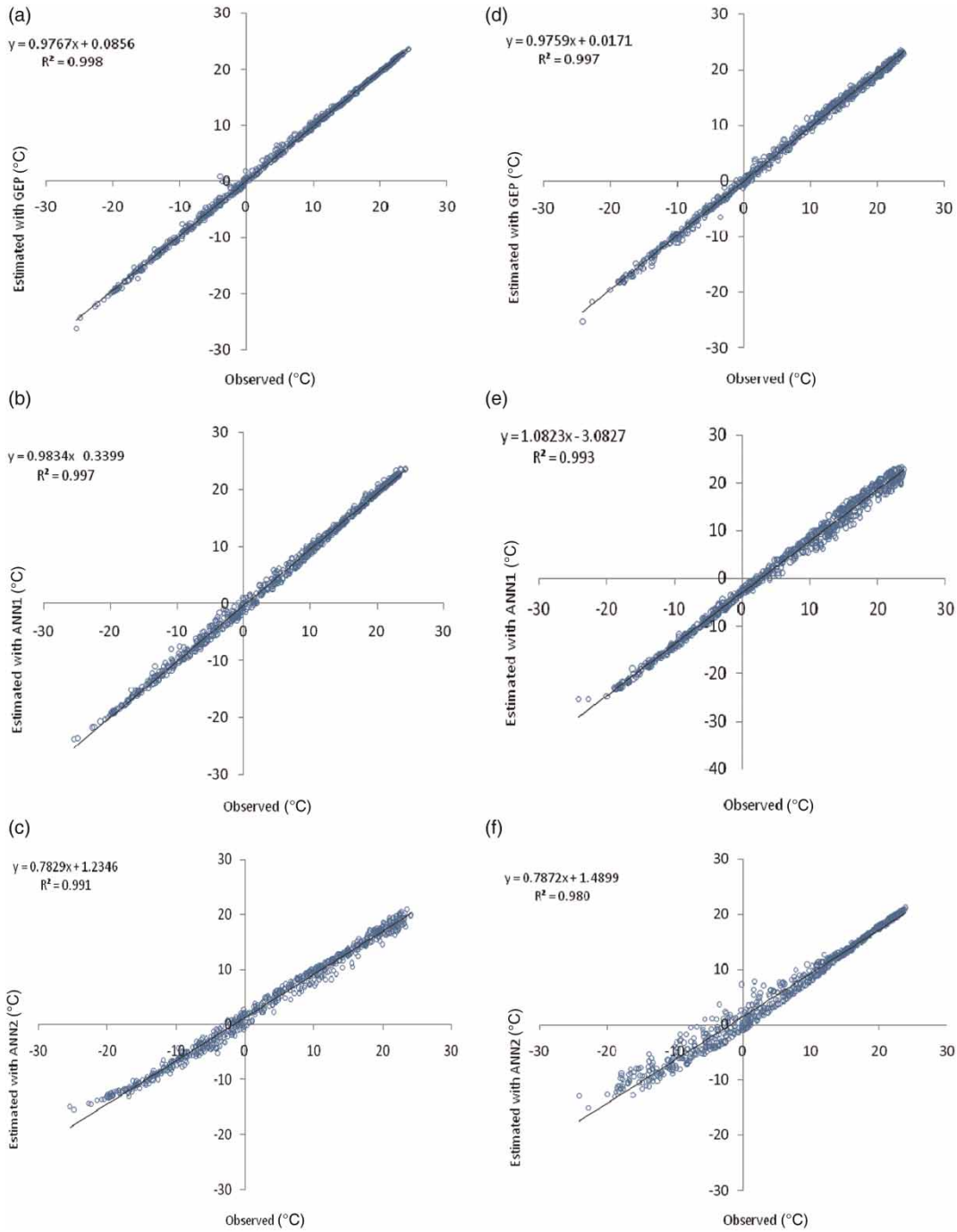

Figure 4 | Observed and estimated values of T_{dew} by using the optimal double-input models in Seoul ((a), (b), (c)) and Incheon ((d), (e), (f)) stations (testing period: 2005–2006).

estimating T_{dew}. Figure 8(a) displays the *NS* values of optimal input combinations of each category. It is clear from the figure that the GEP models also perform (having higher NS values) better than the ANN models for both stations with respect to the NS criteria.

Cross-station application scenario

The estimation of dew point temperature by using nearby station weather data is most important because in some circumstances the weather data of one station may be missing.

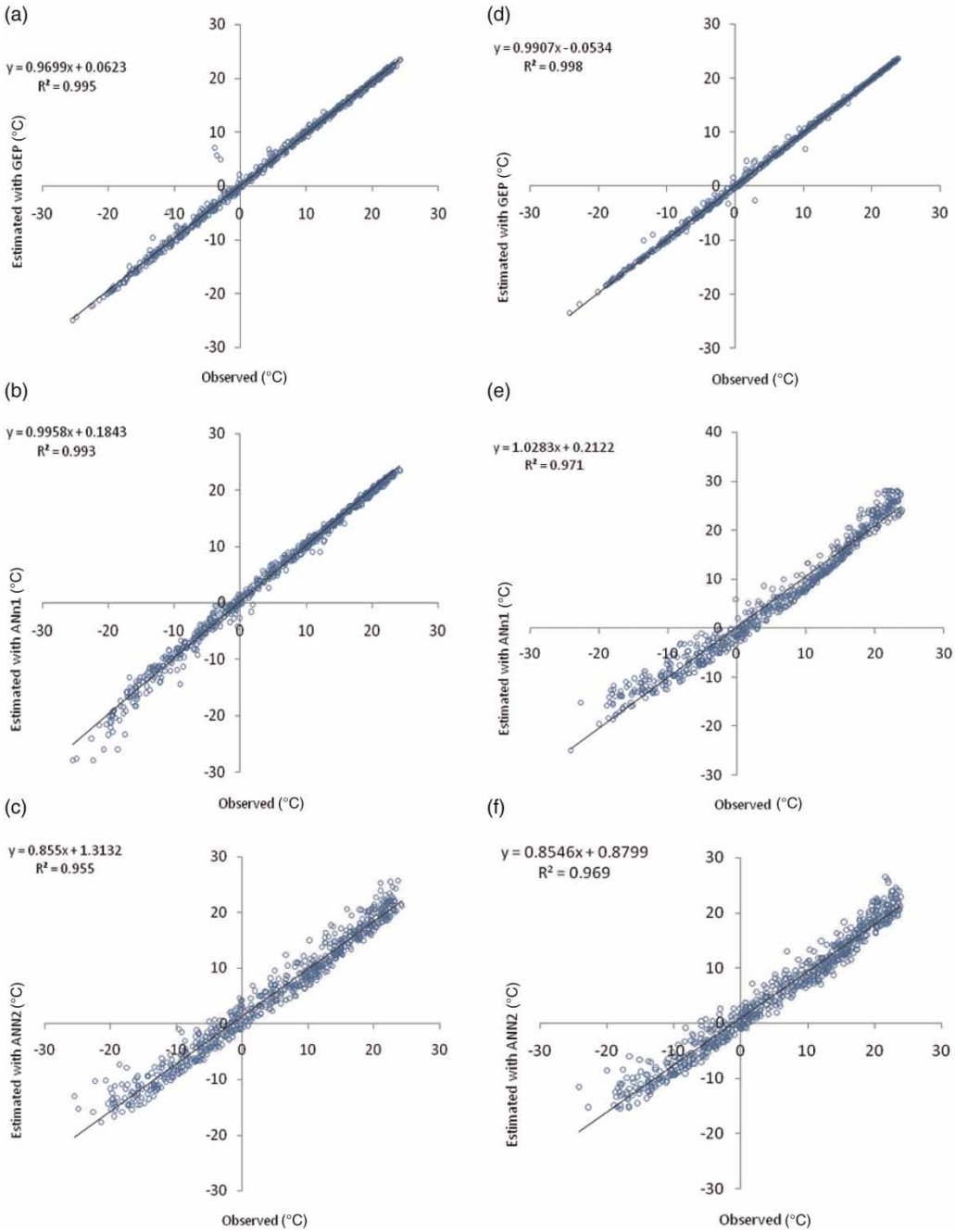

Figure 5 | Observed and estimated values of T_{dew} by using the optimal triple-input models in Seoul ((a), (b), (c)) and Incheon ((d), (e), (f)) stations (testing period: 2005–2006).

Regression techniques are commonly used for solution of this problem. This section of the paper will focus on the cross-station application of GEP and ANN techniques also. In this way, the weather data from Incheon weather station are used to estimate T_{dew} values of Seoul station (Case I) and vice versa (Case II), using the optimal input combinations of each category. The multi-linear regression (MLR) model is also considered for comparison in this part of the study. The correlation matrix of the studied stations' data is given in Table 6. According to the Pearson correlation values in the table, it is clear that the T_{mean} is the most effective parameter on T_{dew} for both stations. The

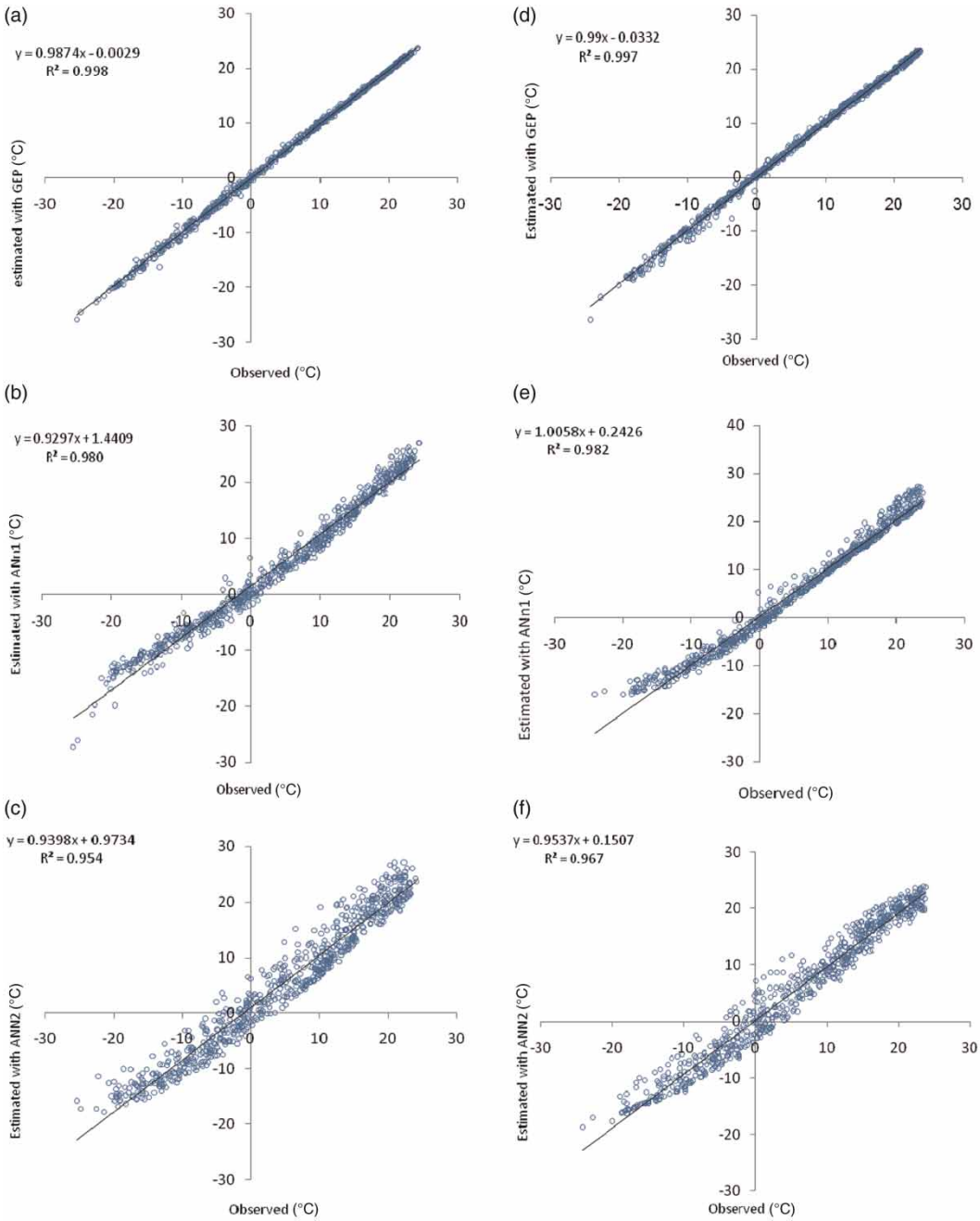

Figure 6 | Observed and estimated values of T_{dew} by using the optimal quadruple-input models in Seoul ((a), (b), (c)) and Incheon ((d), (e), (f)) stations (testing period: 2005–2006).

effectiveness ranks of the parameters on T_{dew} are: T_{mean}, R_S, P, R_H and W_S with respect to linear correlation values given in Table 6. There are small differences between the two stations from a continentality viewpoint as mentioned in the section Study area and data descriptions (see Table 1). It can be clearly seen from the correlation matrix that there is a strong correlation (0.990) between T_{dew} data of the Incheon and Seoul stations. Each station also has similar correlations between their meteorological parameters and corresponding T_{dew} parameter. This indicates why the data for the two stations are strongly correlated.

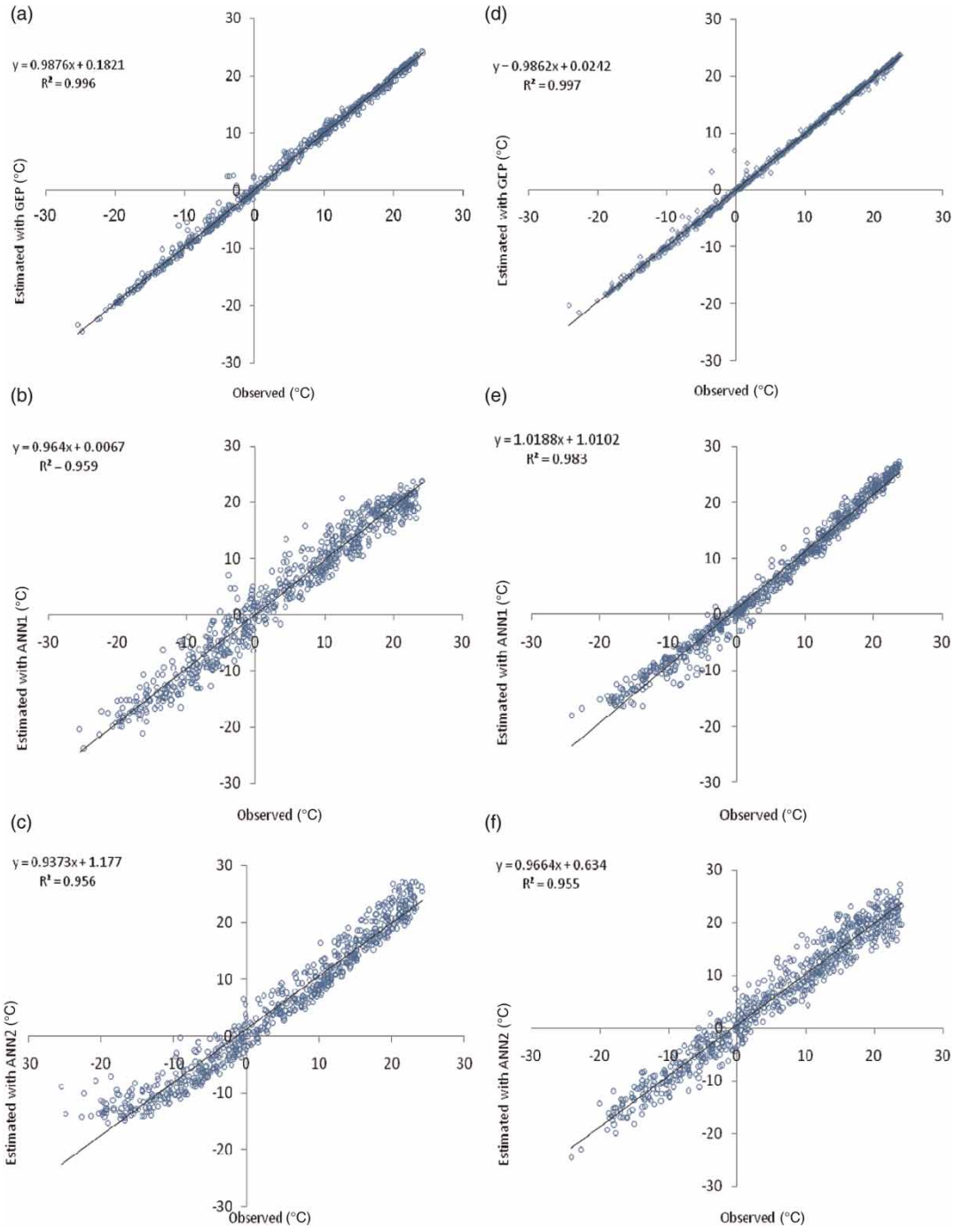

Figure 7 | Observed and estimated values of T_{dew} by using the quintuple-input models in Seoul ((a), (b), (c)) and Incheon ((d), (e), (f)) stations (testing period: 2005–2006).

Table 7 summarizes the statistical indices of the cross-station application scenario. It is clear from the table that using only the T_{mean} data of Incheon station is not sufficient for estimating T_{dew} values of Seoul station. In Case I, the GEP and ANN models whose inputs are the T_{mean} and R_H perform better than the other models. The MLR models give inferior results when compared to the GEP and ANN models. In Case II, however, the MLR models

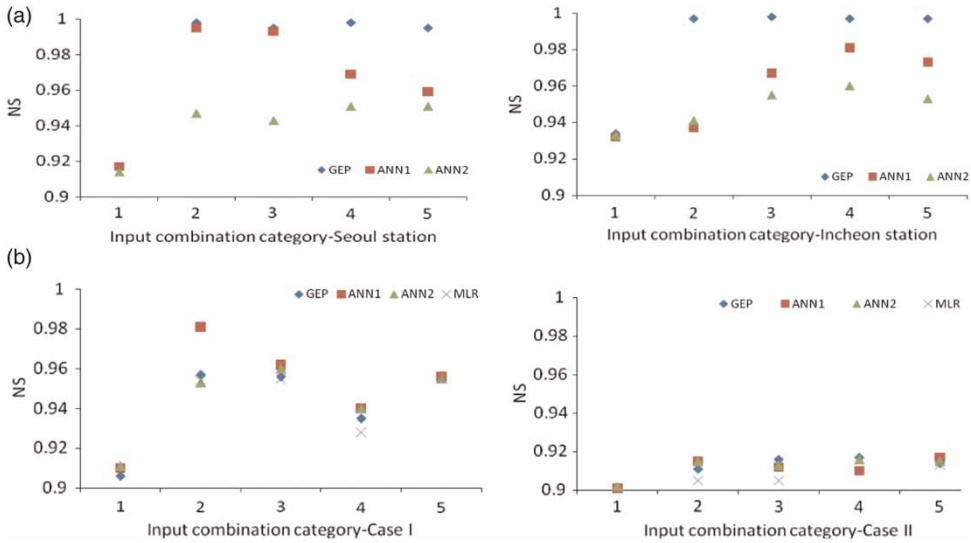

Figure 8 | NS values of optimal input combinations in both scenarios.

Table 6 | Correlation matrix of the two stations' data

		Seoul						Incheon					
		T_{mean}	R_H	P	R_S	W_S	T_{dew}	T_{mean}	R_H	P	R_S	W_S	T_{dew}
Seoul	T_{mean}	1											
	R_H	−0.277	1										
	P	0.411	−0.167	1									
	R_S	−0.765	−0.056	−0.488	1								
	W_S	0.334	0.053	−0.473	−0.174	1							
	T_{dew}	**0.957**	**−0.284**	**0.655**	**−0.782**	**0.126**	1						
Incheon	T_{mean}	0.993	−0.284	0.425	−0.757	0.304	**0.956**	1					
	R_H	−0.351	0.765	−0.147	0.000	−0.034	**−0.340**	−0.353	1				
	P	0.489	−0.167	0.895	−0.548	−0.300	**0.686**	0.483	−0.138	1			
	R_S	−0.764	−0.046	−0.487	0.990	−0.173	**−0.781**	−0.758	−0.001	−0.561	1		
	W_S	0.337	0.057	−0.368	−0.211	0.853	**0.163**	0.312	0.009	−0.269	−0.195	1	
	T_{dew}	**0.960**	**−0.285**	**0.617**	**−0.781**	**0.157**	**0.990**	**0.965**	**−0.334**	**0.695**	**−0.785**	**0.175**	1

seem to be more successful. The optimal GEP and ANN models whose inputs are T_{mean} and R_H perform slightly better than the optimal MLR model whose inputs are T_{mean}, R_H, and W_S. The difference between the double-input GEP and ANN models and triple/quintuple-input models are not significant. GEP models perform better than both the ANN models in estimating T_{dew} values of Seoul station using the nearby station's input data. Figure 8(b) represents the *NS* values of the two cases. The ANN1 model having its highest *NS* value for the second input combination of Case I has better accuracy than the GEP and ANN2 models with respect to NS criteria. In Case II, however, GEP, ANN1 and ANN2 models seem to have similar accuracy from the NS viewpoint. In some combinations, the better accuracy of the MLR models with respect to NS is clearly seen from the figure.

Nevertheless, one of the strong points of the GP (i.e., GEP) model is in giving explicit mathematical expressions for the studied phenomenon. Mathematical expressions of the GEP model for the first and second scenarios are respectively given in Tables 8 and 9.

Table 7 | Statistical parameters of cross-station application of GEP and ANN models during the test period (2005–2006)

Input combinations	RMSE (°C)				MAE (°C)			
	GEP	ANN-1	ANN-2	MLR	GEP	ANN-1	ANN-2	MLR
Case I								
T_{mean}	3.78	3.86	3.87	3.93	2.90	2.94	2.99	2.94
T_{mean}, R_H	2.55	2.58	3.24	4.11	1.65	1.72	2.57	3.06
T_{mean}, R_H, W_S	2.58	2.72	3.25	4.10	1.66	1.79	2.59	3.06
T_{mean}, R_H, W_S, P	3.14	3.34	3.36	4.01	2.33	2.61	2.58	2.91
T_{mean}, R_H, W_S, R_S, P	2.6	3.27	3.61	4.04	1.71	2.35	2.53	2.96
Case II								
T_{mean}	3.71	3.71	3.71	3.85	2.81	2.78	2.80	2.84
T_{mean}, R_H	3.41	3.44	3.42	3.44	2.20	2.24	2.24	2.24
T_{mean}, R_H, W_S	3.42	3.49	3.44	3.43	2.24	2.28	2.55	2.25
T_{mean}, R_H, W_S, P	3.42	3.55	3.42	3.43	2.35	2.34	2.35	2.19
T_{mean}, R_H, W_S, R_S, P	3.65	3.50	3.50	3.49	2.33	2.24	2.29	2.39

Note: Case I, T_{dew} values of Seoul station are estimated through observed data from Incheon station; Case II, T_{dew} values of Incheon station are estimated through observed data from Seoul station.

Table 8 | GEP expressions of at-station application scenario

Models	GEP expressions
Seoul station	
1	$T_{dew} = T_{mean} + \sqrt[3]{\exp(-0.11T_{mean})} + \frac{1.14\,(T_{mean} - 3.31)}{\cos(T_{mean}) + 10.9} - 8.67$
2	$T_{dew} = T_{mean} + \sqrt[3]{2T_{mean} + 7.01R_H - 49.14} + \sqrt[3]{T_{mean} + [(T_{mean} - R_H)(26.7 - R_H)]} - 26.8$
3	$T_{dew} = T_{mean} + 14.7(R_H - W_S) + \left[1.85\cos\left(\sqrt[3]{R_H + 5.39}\right)\right]^3 + \left[\cos\sqrt[3]{R_H - W_S + 6.18} - 0.61\right]^3 - 7.34$
4	$T_{dew} = T_{mean} + \sqrt{R_H - T_{mean} - 3.51} + \text{arctg}\left[W_S \frac{[0.392(T_{mean} - R_H)]^3}{P}\right] - [\exp\{\sin[\ln(0.24R_H)]\}]*8.36$
5	$T_{dew} = T_{mean} + 0.15R_H + \text{arctg}(W_S + 0.74) - \sqrt{\cos(0.98R_S) + R_S} + \left[\cos\sqrt{R_H + R_S}.\text{arctg}(0.44R_S)\right] - 15.86$
Incheon station	
1	$T_{dew} = 8.07T_{mean} - 7.02$
2	$T_{dew} = T_{mean} - \frac{590}{R_H} - \frac{T_{mean} - \sqrt{R_H} + 8.44}{0.391R_H} + \sqrt{0.03R_H^{1.5} + 1.25R_H^{0.33}} - 6.27$
3	$T_{dew} = T_{mean} + \ln[(R_H - T_{mean})^2] + \sqrt[3]{R_H - 8.26} + \sqrt{R_H^2 + 6.88}$
4	$T_{dew} = T_{mean} + \sqrt{R_H - T_{mean}} + \frac{T_{mean}}{R_H} + \frac{([W_S/R_H] - 9.93)*(84.37 - R_H)}{R_H} - 13.31$
5	$T_{dew} = T_{mean} - 5.06\cos\sqrt{R_H + 0.99} + \frac{\sin(R_H^{0.66})}{\text{arctg}(R_S + 1.38)} - \frac{T_{mean}\sqrt{RS + 3.25}}{T_{mean} + R_H} + 1.72$

Note: Numbers 1–5 at each station, respectively, denote the optimal input combination of each single-, double-, triple-, quadruple-, and quintuple-input GEP model category (Tables 4 and 5).

Table 9 | GEP expressions of cross-station application scenario

Input combinations	GEP expression
Case I	
T_{mean}	$T_{dew} = 7.59T_{mean} - 8.69$
T_{mean}, R_H	$T_{dew} = \text{arctg}\left(\frac{-4.18}{R_H}\right) - \text{arctg}(\exp[42.5 - R_H]) - \exp\left(\frac{T_{mean}}{R_H}\right) + T_{mean} + 0.0015R_H^2 - 13.34$
T_{mean}, R_H, W_S	$T_{dew} = T_{mean} - \ln\left[\exp\left(\frac{T_{mean}}{R_H}\right)\right] - R_H^{0.66} + \sqrt{2R_H(R_H - 8.62)} - 9.41$
T_{mean}, R_H, W_S, P	$T_{dew} = 2\cos\left[\sin\sqrt[3]{W_s} - 3.02\sqrt[3]{R_H - 3.02}\right] + \cos\left[\sqrt[3]{T_{mean} + R_H - 2.74} + T_{mean}\right] - 6.96$
T_{mean}, R_H, W_S, R_S, P	$T_{dew} = T_{mean} + \cos\sqrt[3]{R_S} + \sqrt[3]{4.63R_H - R_S} - 0.04\left[\frac{2.27T_{mean} + P + 2.38}{R_H}\right] - 2.14$
Case II	
T_{mean}	$T_{dew} = \cos[0.2T_{mean} - 0.2 + \text{arctg}(T_{mean} - 2.87)] + T_{mean} - 5.41$
T_{mean}, R_H	$T_{dew} = T_{mean} + \ln[36.6 - T_{mean}] - \frac{23.95}{\ln(R_H)} + 0.99[R_H - 17.84]^{0.66} - 16.07$
T_{mean}, R_H, W_S	$T_{dew} = T_{mean} + \frac{\ln(R_H)}{T_{mean} + R_H} - \frac{0.32T_{mean}(W_S + 8.1)}{R_H} + 0.23(2R_H - 13.91) - 17.56$
T_{mean}, R_H, W_S, P	$T_{dew} = T_{mean} + 2\left[R_H^2 - R_H + P\right] - 6.07$
T_{mean}, R_H, W_S, R_S, P	$T_{dew} = T_{mean} + \frac{RS}{0.6\sqrt{\exp(RS)}} + \frac{54}{[(T_{mean} * RS)/(P + RS)] - 0.1RH} + \frac{RS}{\sin(-2.16WS) - 44.5} + 2.86$

Note: Case I, T_{dew} values of Seoul station are estimated through observed data from Incheon station; Case II, T_{dew} values of Incheon station are estimated through observed data from Seoul station.

The goal in the present paper is a generalized function which will capture the process physics or at least perform statistically equivalent on the training and blind test data. The valid performance metric is the result on totally blind data (data omitted; not used at all in training). The present results, derived from a single chronological data set assignment, should be confirmed in further studies through data set scanning approaches to discuss the minimum data length for similar studies. Further research should also deal with models fed with more climatic inputs as well as under different climatic conditions.

CONCLUSIONS

The accuracy of GEP and two different ANN methods was investigated for estimating daily dew point temperature in the present study. Various input combinations of daily air temperature, wind speed, relative humidity, atmospheric pressure, and incoming solar radiation data from two weather stations, Seoul and Incheon, in the Republic of Korea were used. Comparison of the models' results revealed that the GEP model surpassed both ANN models in estimating daily dew point temperature values of both stations. Comparison of the two stations' results indicated that the GEP and ANN models are more successful in Seoul station than Incheon station. First cross-station application indicated that the ANN and GEP models performed better than the MLR model in estimating dew point temperature of Seoul station by using the data of Incheon station. Second cross-station application revealed that the optimal MLR model had slightly worse accuracy than the optimal GEP and ANN models in estimating dew point temperature of Incheon station by using the data of Seoul station. GEP models were found to be better than the ANN models in both cross-station applications. The study showed that the T_{dew} values of Seoul station can be successfully estimated using the T_{mean} and R_H data of Incheon station.

REFERENCES

Abdel-Aal, R. E. 2004 Hourly temperature forecasting using abductive networks. *Eng. Appl. Artif. Intel.* **17**, 543–556.

Abdellatif, M., Atherton, W. & Alkhaddar, R. 2013 A hybrid generalized linear and Levenberg–Marquardt artificial neural network approach for downscaling future rainfall in North Western England. *Hydrol. Res.* doi:10.2166/nh.2013.045.

ASCE Task Committee 2000a Artificial neural networks in hydrology. I: preliminary concepts. *J. Hydrol. Eng.* **5** (2), 115–123.

ASCE Task Committee 2000b Artificial neural networks in hydrology. II: hydrological applications. *J. Hydrol. Eng.* **5** (2), 124–137.

Atzema, A. J., Jacobs, A. F. G. & Wartena, L. 1990 Moisture distribution within a maize crop due to dew. *Neth. J. Agric. Sci.* **38** (2), 117–129.

Babovic, V., Kanizares, R., Jenson, H. R. & Klinting, A. 2001 Neural networks as routine for error updating of numerical models. *J. Hydrol. Eng.* **127** (3), 181–193.

Berning, T. 2012 The dewpoint temperature as a criterion for optimizing the operating conditions of proton exchange membrane fuel cells. *Int. J. Hydrogen Energy* **37**, 10265–10275.

Elman, J. L. 1990 Finding structure in time. *Cognitive Sci.* **14**, 179–211.

Ferreira, C. 2001 Gene expression programming: a new adaptive algorithm for solving problems. *Complex Syst.* **13** (2), 87–129.

Ferreira, C. 2006 *Gene Expression Programming: Mathematical Modeling by an Artificial Intelligence*. Springer, Berlin, 478 pp.

Giles, C. L., Lawrence, S. & Tsoi, A. C. 1997 Rule inference for financial prediction using recurrent neural networks. In *Proceedings of 1997 IEEE/IAFE Conference on Computational Intelligence for Financial Engineering*, IEEE Press, Piscataway, NJ, pp. 253–259.

Hagan, M. T., Demuth, H. B. & Beale, M. 1995 *Neural Network Design*. PWS Publishing Company, Boston, MA.

Haykin, S. 1999 *Neural Networks: a Comprehensive Foundation*. Prentice-Hall, Upper Saddle River, NJ.

Kim, S. & Cho, J. S. 2003 Uncertainty analysis of flood stage forecasting using time-delayed patterns in the small catchment. In *International Symposium on Disaster Mitigation and Basin-Wide Water Management Niigata 2003*, IAHR/AIRH, Niigata, Japan, pp. 465–474.

Kim, S. & Kim, H. S. 2008 Uncertainty reduction of the flood stage forecasting using neural networks model. *J. Am. Water Resour. Ass.* **44** (1), 148–165.

Kim, S., Shiri, J. & Kisi, O. 2012 Pan evaporation modeling using neural computing approach for different climatic zones. *Water Resour. Manage.* **26** (11), 3231–3249.

Kisi, O. & Shiri, J. 2010 A comparison of genetic programming and ANFIS in forecasting daily, monthly and daily streamflows. In *Proceedings of the International Symposium on Innovations in Intelligent Systems and Applications*, 21–24 June 2010, Kayseri and Cappadocia, Turkey, pp. 118–122.

Kisi, O. & Shiri, J. 2011 Precipitation forecasting using wavelet-genetic programming and wavelet-neuro-fuzzy conjunction models. *Water Resour. Manage.* **25** (13), 3135–3152.

Kisi, O. & Shiri, J. 2012a Wavelet and neuro-fuzzy conjunction model for predicting water table depth fluctuations. *Hydrol. Res.* **43** (3), 286–300.

Kisi, O. & Shiri, J. 2012b River suspended sediment estimation by climatic variables implication: comparative study among soft computing techniques. *Comput. Geosci.* **43**, 73–82.

Kisi, O., Shiri, J. & Makarynskyy, O. 2011 Wind speed prediction by using different wavelet conjunction models. *Int. J. Ocean Climate Syst.* **2** (3), 189–208.

Kisi, O., Shiri, J. & Nikoofar, B. 2012 Forecasting daily lake levels using artificial intelligence approaches. *Comput. Geosci.* **41**, 169–180.

Kisi, O., Shiri, J. & Tombul, M. 2013 Modeling rainfall-runoff process using soft computing techniques. *Comput. Geosci.* **51**, 108–117.

Landeras, G., Lopez, J. J., Kisi, O. & Shiri, J. 2012 Comparison of gene expression programming with neuro-fuzzy and neural network computing techniques in estimating daily incoming solar radiation in the Basque Country (Northern Spain). *Energ. Convers. Manage.* **62**, 1–13.

Legates, D. R. & McCabe, G. J. 1999 Evaluating the use of goodness-of-fit measures in hydrologic and hydroclimatic validation. *Water Resour. Res.* **35** (1), 233–241.

Li, J., Michel, A. N. & Porod, W. 1989 Analysis and synthesis of a class of neural networks: linear systems operating on a closed hypercube. *IEEE Trans. Circuits Syst.* **36** (11), 1405–1422.

Mahmood, R. & Hubbard, K. G. 2005 Assessing bias in evapotranspiration and soil moisture estimates due to the use of modeled solar radiation and dew point temperature data. *Agric. Forest Meteorol.* **130**, 71–84.

Mahmood, R., Hubbard, K. G., Leeper, R. D. & Foster, S. A. 2008 Increase in near-surface atmospheric moisture content due to land use changes: evidence from the observed dewpoint temperature data. *Mon. Weather Rev.* **136**, 1554–1561.

Neuro Dimension Inc. 2005 *Developers of NeuroSolutions V5.01: Neural Network Simulator*. Gainesville, FL.

Pour Ali Baba, A., Shiri, J., Kisi, O., Fakheri Fard, A., Kim, S. & Amini, R. 2013 Estimating daily reference evapotranspiration using available and estimated climatic data by adaptive neuro-fuzzy inference system (ANFIS) and artificial neural networks (ANN). *Hydrol. Res.* **44** (1), 131–146.

Ryeu, J. K. & Chung, H. S. 1996 Chaotic recurrent neural networks and their application to speech recognition. *J. Neurocomput.* **13**, 281–294.

Savic, A. D., Walters, A. G. & Davidson, J. W. 1999 A genetic programming approach to rainfall-runoff modeling. *Water Resour. Manage.* **13**, 219–231.

Shank, D. B., Hoogenboom, G. & McClendon, R. W. 2008 Dew point temperature prediction using artificial neural networks. *J. Appl. Meteorol. Climatol.* **47**, 1757–1769.

Shiri, J. & Kisi, O. 2011a Comparison of genetic programming with neuro-fuzzy systems for predicting short-term water table depth fluctuations. *Comput. Geosci.* **37** (10), 1692–1701.

Shiri, J. & Kisi, O. 2011b Application of artificial intelligence to estimate daily pan evaporation using available and estimated climatic data in the Khozestan Province (South Western Iran). *J. Irrig. Drain. Eng.* **137** (7), 412–425.

Shiri, J., Dierickx, W., Pour-Ali Baba, A., Nemati, S. & Ghorbani, M. A. 2011 Estimating daily pan evaporation from climatic data of the state of Illinois, USA using adaptive neuro-fuzzy inference system and artificial neural network. *Hydrol. Res.* **42** (6), 491–502.

Shiri, J., Kisi, O., Landeras, G., Lopez, J. J., Nazemi, A. H. & Stuyt, L. C. P. M. 2012 Daily reference evapotranspiration modeling by using genetic programming approach in the Basque Country (Northwestern Spain). *J. Hydrol.* **414–415**, 302–316.

Shiri, J., Kisi, O., Yoon, H., Lee, K. K. & Nazemi, A. H. 2013 Predicting groundwater level fluctuations with meteorological effect implications – a comparative study among soft computing techniques. *Comput. Geosci.* **56**, 32–44.

Slatyer, R. O. 1967 *Plant-Water Relationships*. Academic Press, London.

Smith, B. A., McClendon, R. W. & Hoogenboom, G. 2005 An enhanced artificial neural network for air temperature prediction. *Proc. World Acad. Sci. Eng. Tech.* **7**, 7–12.

Tsoukalas, L. H. & Uhrig, R. E. 1997 *Fuzzy and Neural Approaches in Engineering*. John Wiley & Sons, New York.

Went, F. W. 1955 Fog, mist dew and other sources of water. Yearbook Agriculture. US Department of Agriculture, pp. 103–109.

First received 15 December 2012; accepted in revised form 13 July 2013. Available online 17 August 2013