

A hybrid computational intelligence method for predicting dew point temperature

Mohsen Amirmojahedi¹ · Kasra Mohammadi² · Shahaboddin Shamshirband³ · Amir Seyed Danesh⁴ · Ali Mostafaipoor⁵ · Amirrudin Kamsin³

Received: 27 May 2015 / Accepted: 5 October 2015 / Published online: 24 February 2016
© Springer-Verlag Berlin Heidelberg 2016

Abstract Recently, utilization of hybrid models has gained remarkable attention as they take the advantage of specific nature of each technique to enhance the precision and reliability of the predictions. In this research work, a new hybrid approach combined the extreme learning machine (ELM) with wavelet transform (WT) algorithm is proposed to predict daily dew point temperature. To test the validity of the proposed approach, the daily weather data sets for port of Bandar Abass situated in the south costal part of Iran are used. The merit of the proposed ELM-WT method is verified against the ELM, support vector machines and artificial neural network techniques based upon several well-known statistical indicators. The achieved results demonstrate that the hybrid ELM-WT method presents absolute superiority over other powerful techniques applied. It is found that among four considered

sets of parameters with 1, 2 and 3 inputs, further accuracy can be achieved using combination of average ambient temperature (T_{avg}) and relative humidity (R_h). For the best ELM-WT model using T_{avg} and R_h as inputs, the statistical indicators of mean absolute percentage error, mean absolute bias error, root mean square error and coefficient of determination are 6.1664 %, 0.5495, 0.7621 and 0.9953 °C, respectively. Based on relative percentage error (RPE), for the best ELM-WT model 91 % of the predictions fall within the RPE acceptable range of -10 and $+10$ %. In a nutshell, the results of this study convincingly advocate that coupling the ELM with WT would be particularly appealing to offer accurate predictions and favorable enhancement in the precision of ELM.

Keywords Daily dew point temperature · Extreme learning machine (ELM) · Wavelet transform (WT) algorithm · Prediction · Statistical analysis · Hybrid approach

✉ Kasra Mohammadi
kmohammadi@umass.edu

✉ Shahaboddin Shamshirband
shamshirband@um.edu.my

Amir Seyed Danesh
a.seyeddaneh@gmail.com

¹ Department of Civil Engineering, Faculty of Engineering, University of Malaya, 50603 Kuala Lumpur, Malaysia

² Department of Mechanical and Industrial Engineering, University of Massachusetts, Amherst 01003, Massachusetts, USA

³ Department of Computer System and Technology, Faculty of Computer Science and Information Technology, University of Malaya, Kuala Lumpur 50603, Malaysia

⁴ Payame Noor University, Rasht branch, Iran

⁵ Industrial Engineering Department, Yazd University, Yazd, Iran

Introduction

Dew is a condensation of atmospheric moisture on the objects which their temperatures are lower than the dew point temperature of the surrounding ambient. In fact, the dew formation occurs once the surface air temperature decline to the dew point temperature. Therefore, the dew point temperature can be defined as the temperature at which moisture or water vapor in the air starts into dew or water droplets. The radiation exchange between the surface of Earth and atmosphere, water vapor pressure as well as turbulent heat are among the major elements influencing the dew formation (Atzema et al. 1990; Shiri et al. 2014). Based upon an agricultural vantage point, dew

phenomenon can reduce the vapor pressure deficit in the proximity of the dew drops eventuating in a more favorable photosynthesis (Slatyer 1967) and improving the recovery of water content after extreme water losses (Went 1955).

The availability of accurate and reliable dew point temperature data plays a notable role in various hydrological, climatological and agronomical related researches. Dew point temperature is typically utilized along with relative humidity to identify the moisture's level in the air. It can also be used in conjunction with wet bulb-temperature for computing the ambient temperature which provides the possibility for being prepared against the potential frosts which may harm crops (Snyder and Melo-Abreu 2005; Shank 2006). Dew point temperature can be used to provide a favorable estimate of the near-surface humidity that influences the stomatal closure in plants where a low level humidity may result in declining the plants productivity (Kimball et al. 1997; Shank 2006). The dew would be really significant for plant survival particularly in the arid areas with rare rainfall (Agam and Berliner 2006). The dew point temperature is as an especially significant element in various hydrological and climatological models for the purpose of reference evapotranspiration estimation (Hubbard et al. 2003; Shank 2006). It is also significant to mention that the dew point temperature data may be useful for the purpose of dam project flood study. Furthermore, dew point depression, defined as the difference between the temperature and dew point temperature, influence the water balance by changing the amount of water evaporated or transpired from a basin. Therefore, the knowledge of dew point temperature can also be useful for this aim.

Over the previous years, the artificial intelligence and computational intelligence techniques have been successfully applied in the estimation of dew point temperature.

Shank et al. (2008) utilized ANN technique for prediction of dew point temperature from 1 to 12 h ahead based upon the previous weather data sets. They used measured data of 20 stations in Georgia State in USA for developing general models to predict dew point temperature in the whole Georgia State. Zounemat-Kermani (2012) evaluated the capability of multi linear regression (MLR) and Levenberg–Marquardt (LM) feed-forward neural network for estimation of hourly dew point temperature in a location in Ontario, Canada. It was found that LM-NN model provide further accuracy compared to the MLR model. Nadig et al. (2013) developed combined air and dew point temperatures models using ANN technique to provide an enhancement in the predictions of both temperatures. Their results demonstrated that the combined method decline the predictions error. Shiri et al. (2014) assessed the capability of two ANN models and gene expression programming (GEP) technique to estimate daily dew point temperature in two stations of Korea. Their results indicated that the GEP

model outperforms the ANN models. Kim et al. (2014) utilized two soft computing techniques for estimation of daily dew point temperature in California, USA. By providing comparisons with a conventional regression model, they found that developed soft computing models are more precise in estimating daily dew point temperature.

Mohammadi et al. (2016) applied adaptive neuro fuzzy inference system (ANFIS) to select the most influential parameters for prediction of daily dew point temperature. They analyzed the influence of eight different parameters on dew point temperature prediction in two cities of Iran. Their results showed that, despite climate difference between the selected studied areas, for both cities water vapor pressure is the most relevant parameter while relative humidity is the least relevant parameter. They concluded that using more than two input parameters cannot be proper and advisable.

As a key artificial intelligence method, the neural network (NN) is capable of solving the complicated nonlinear problems which cannot be solved easily via the classic parametric methods. The main shortcoming of NNs is their learning time requirement. Huang et al. (2004) introduced extreme learning machine (ELM) as an algorithm for a single-layer feedforward neural network (SLFN). The algorithm can reduce the required time for training a NN (Huang et al. 2004). ELM is able to produce good generalization performance and speed up the learning process (Huang et al. 2004; Liang et al. 2006) due to its capability in simplifying the training. The ELM also enjoys a faster speed in the learning process than other traditional algorithms such as the back-propagation (BP) (Huang et al. 2015). As a consequence, several researches have been performed to apply the ELM algorithm for problems' solving in different realms of science (Ghouti et al. 2013; Zhao et al. 2013; Nian et al. 2014; Wang and Han 2014; Wang et al. 2014; Wong et al. 2015; Yu et al. 2014).

Lately, coupling different approaches to build a hybrid model has received considerable attention since it is possible to take the advantage of specific nature of each technique for enhancing the precision. In fact, the particular features of each technique are able to capture different patterns in the data series. Based upon the theoretical and empirical findings it has been proved that hybrid approaches would be particularly effective and promising for different scientific fields to enhance the prediction accuracy and reliability (Partal and Kisi 2007; Kisi and Cimen 2012; Liang et al. 2012; Kavousi-Fard et al. 2014; Xiong et al. 2014; Sudheer et al. 2014; Shrivastava and Panigrahi 2014; Shamshirband et al. 2015, 2016; Mohammadi et al. 2015; Olatomiwa et al. 2015).

Therefore, in this study, the ELM is coupled with wavelet transform (WT) algorithm to propose a hybrid approach for prediction of daily dew point temperature. To

test the validity of the proposed method named ELM-WT, daily weather data sets for port of Bandar Abass located in the south costal part of Iran have been utilized as a case study. Wavelet analysis is utilized for preprocessing and decomposing the time series of weather data into its various components, after which the decomposed components are utilized as inputs for the ELM model. The merit of the proposed ELM-WT approach is validated against ELM, support vector machines (SVM) and artificial neural network (ANN). The performances evaluations are conducted on the basis of several widely utilized statistical parameters to provide an appropriate comparative study.

Methodology

In this research work, a new method by hybridizing the extreme learning machine (ELM) with wavelet transform (WT) algorithm is proposed to predict daily dew point temperature. The precision of the hybrid ELM-WT method is compared with support vector machine (SVM) and artificial neural network (ANN) approaches. In this section, the description of the employed methods is presented.

Extreme learning machine (ELM)

Huang at el. (2004) introduced ELM as a tool of learning algorithm for single-layer feedforward neural network (SLFN) architecture (Huang et al. 2004, 2006a). The ELM determines random input weight parameters and analytically defines the output weights parameters (Huang et al. 2015).

Single hidden layer feed-forward neural network (SLFN)

Mathematical representation of SLFN function incorporates additives and radial basis function (RBF) intermediate nodes (hidden nodes) in an integrated manner as (Huang et al. 2006b; Liang et al. 2006):

$$f_L(x) = \sum_{i=1}^L \beta_i G(a_i, b_i, x), \quad x \in R^n, \quad a_i \in R^n \quad (1)$$

where $f_L(x)$ is the output function of ELM for generalized SLFNs, x is input, the learning parameters of hidden nodes are represented by a_i and b_i , L is number of hidden nodes and β_i is the connecting weight between the i th hidden and output node. $G(a_i, b_i, x)$ shows the output of the i th hidden node with regard to the input x . The additive hidden node with the activation function of $g(x) : R \rightarrow R$ (e.g., sigmoid and threshold) is (Huang et al. 2006a):

$$G(a_i, b_i, x) = g(a_i \cdot x + b_i), \quad b_i \in R \quad (2)$$

where a_i is weight vector which connects input layer to i th hidden-node and b_i is the bias of i th node a_i . Internal product of vector a_i and x in R^n is represented as x . Activation function $g(x) : R \rightarrow R$ finds $G(a_i, b_i, x)$ for RBF hidden node as (Huang et al. 2006a):

$$G(a_i, b_i, x) = g(b_i \|x - a_i\|), \quad b_i \in R^+ \quad (3)$$

The center and impact factors for RBF i th node is represented by a_i and b_i , respectively. Entire positive real value set is represented by R^+ . A particular SLFN case containing RBF node in their hidden layers forms the RBF network. Arbitrary N distinct samples are represented as $(x_i, t_i) \in R^n \times R^m$. The x_i and t_i are input vector and target vector, respectively. The size of x_i and t_i are $n \times 1$ and $m \times 1$, respectively. Provided that a SLFN is capable of approximating these N samples, it means that there exists β_i , a_i and b_i as (Huang et al. 2006a):

$$f_L(x) = \sum_{i=1}^L \beta_i G(a_i, b_i, x), \quad j = 1, \dots, N. \quad (4)$$

Equation (4) may be expressed neatly as:

$$H\beta = T \quad (5)$$

where

$$H(\tilde{a}, \tilde{b}, \tilde{x}) = \begin{bmatrix} G(a_1, b_1, x_1) & \cdots & G(a_L, b_L, x_1) \\ \cdots & \cdots & \cdots \\ G(a_1, b_1, x_N) & \cdots & G(a_L, b_L, x_N) \end{bmatrix}_{N \times L} \quad (6)$$

With $\tilde{a} = a_1, \dots, a_L$; $\tilde{b} = b_1, \dots, b_L$; $\tilde{x} = x_1, \dots, x_L$

$$\beta = \begin{bmatrix} \beta_1^T \\ \vdots \\ \beta_L^T \end{bmatrix}_{L \times m} \quad \text{and} \quad T = \begin{bmatrix} t_1^T \\ \vdots \\ t_L^T \end{bmatrix}_{N \times m} \quad (7)$$

where H is a matrix of SLFN's hidden layer output. In matrix H , i th output of hidden mode forms the i th column of H for the corresponding inputs x_1, \dots, x_N . In this study, 13 hidden nodes are used.

Principle of ELM

ELM designed as a SLFN with L hidden neurons can learn L distinct samples with zero error (Huang et al. 2004, 2006a). Even if the number of hidden neurons (L) is less than the number of distinct samples (N), ELM is still capable of assigning random parameters to the hidden nodes and computing the output weights by pseudo inverse of H giving only a small error $\varepsilon > 0$. The hidden node parameters of ELM should not be tuned throughout training and can easily be assigned with random values.

Discrete wavelet transform

Wavelet transform (WT) represents the basis of the mathematical expression to decompose time series frequency signal into various components (Mallat 1989; Mallat 2009; Peng and Chu 2004), and is also known as the signal processing algorithm from Fourier transforms. The key superiority over Fourier transform is its ability to perform accurate analysis based on resulting decomposed components with scaled-fit resolution that aids to enhance the size of the study model due to its capability to obtain the required information at various levels (Adamowski and Chan 2011). This is ideal for data analysis applications with a time domain and frequency owing to its ability to extract information from transient and non-periodic signals (Jawerth and Sweldens 1994). Recently, these methods have generated enormous interest for engineering applications (Burrus et al. 1997; Kalteh 2013).

On the other hand, continuous wavelet transform (CWT) with signal $f(t)$, is defined as the time-scale technique of signal processing as

$$W_f(a, b, \psi) = \frac{1}{\sqrt{|a|}} \int_{-\infty}^{\infty} f(t) \psi^* \left(\frac{t-b}{a} \right) dt, \quad b \in R, a \in R, a \neq 0 \tag{8}$$

where ψ is the mother wavelet function, $\psi^*(t)$ denotes the complex conjugate of ψ , t is the time, b describes the time shifting parameter. By discretizing Eq. (8), the discrete wavelet transform (DWT) can be found whereby the parameters a and b can be found as:

$$a = a_0^m, \quad b = na_0^m b_0, \quad a_0 > 1, \quad b_0 \in R, \tag{9}$$

where n and m are integer numbers that control the scale and translation, respectively, a_0 is a fixed dilation step, and b_0 is a translation factor that depends on the aforementioned dilation step.

DWT of $f(t)$ can be written as:

$$W_f(m, n, \psi) = a_0^{-m/2} \int_{-\infty}^{\infty} f(t) \psi^* (a_0^{-m}t - nb_0) dt \tag{10}$$

According to (Mallat 1989), when $a_0 = 2$ and $b_0 = 1$ Eq. (10) becomes binary wavelet transform:

$$W_f(m, n, \psi) = a_0^{-m/2} \int_{-\infty}^{\infty} f(t) \psi^* (a_0^{-m}t - nb_0) dt \tag{11}$$

$W_f(a, b, \psi)$ or $W_f(m, n, \psi)$ represents the features of original time series in frequency (a or m) and time domain (b or n) at the same time (Wang and Ding 2003). When a or m gain small values, the frequency resolution of wavelet transform becomes low, but the time domain resolution would then be high. When a or m gain large values, the frequency resolution of wavelet transform would then be

high, but the time domain resolution would have low values.

Support vector machine (SVM)

SVM is a soft computing method that has been applied in a large number of scientific fields (Lee and Verri 2003; Lu and Wang 2005; Asefa et al. 2006; Ji and Sun 2013; Sun 2013). The details of theory and evolution of SVM developed by Vapnik can be found in Vapnik (2000) and Vapnik and Vapnik (1998).

On the basis of Vapnik’s theory, the SVM functions are offered by Eqs. (12–15), where $R = \{x_i, d_i\}_i^n$ is utilized for assuming a set of data points, the input space vector is shown by x_i , and the desired value and data size are defined as d_i and n , respectively. The SVM approximates the function as represented by Eqs. (12) and (13):

$$f(x) = w\varphi(x) + b \tag{12}$$

$$R_{SVMs}(C) = \frac{1}{2} \|w\|^2 + C \frac{1}{n} \sum_{i=1}^n L(x_i, d_i) \tag{13}$$

where $\varphi(x)$ shows high dimensional space characteristic mapped from x . Also, w and b are normal vector and scalar, respectively. Furthermore, $C \frac{1}{n} \sum_{i=1}^n L(x_i, d_i)$ represents error or risk. Factors b and w are measured by minimizing the regularized risk equation by introducing the positive slack variables ξ_i and ξ_i^* as (Vapnik and Vapnik 1998):

$$\text{Minimize } R_{SVMs}(w, \xi^{(*)}) = \frac{1}{2} \|w\|^2 + C \sum_{i=1}^n (\xi_i + \xi_i^*) \tag{14}$$

$$\text{Subject to } \begin{cases} d_i - w\varphi(x_i) + b_i \leq \varepsilon + \xi_i \\ w\varphi(x_i) + b_i - d_i \leq \varepsilon + \xi_i^* \\ \xi_i, \xi_i^* \geq 0, \quad i = 1, \dots, l \end{cases}$$

where $\frac{1}{2} \|w\|^2$ is the regularization term, C characterizes the error penalty feature utilized to regulate the trade-off between the regularization term and empirical error, ε denotes the loss function associated with approximation accuracy of the trained data point and the factors number in the training data is l .

Lagrange multiplier and optimality constraints utilized to solve Eq. (12) are consequently attained by a generic function as follows:

$$f(x, a_i a_i^*) = \sum_{i=1}^n (a_i - a_i^*) K(x, x_i) + b \tag{15}$$

where $K(x, x_i) = \varphi(x_i)\varphi(x_j)$ and the term K is the kernel function, which is dependent on the two vectors x_i and x_j in the feature space $\varphi(x_i)$ and $\varphi(x_j)$, respectively.

The kernel function, K , creates a non-linear mapping; thus, it can be used for operation in a higher dimensional space. This is used for calculating the inner product that can be served as a function for the original input points. The SVM's flexibility for using the kernel functions is significant where it discreetly changes the information into a higher-dimensional feature space. The achieved results in this space characterize the outcomes of the lower-dimensional, original input space.

Four kernel functions including sigmoid, linear, polynomial, and radial basis functions are used for SVM models. Nevertheless, radial basis function (RBF) is usually considered as an ideal function because of its capability for adaptable, simple, reliable and effective computation for the sake of optimization particularly for compatibility in handling the complicated parameters (Rajasekaran et al. 2008; Yang et al. 2009; Wu and Wang 2009). The non-linear radial basis kernel function is defined as:

$$K(x_i, x_j) = \exp\left(-\gamma \|x_i - x_j\|^2\right) \tag{16}$$

where x_i and x_j are vectors in the input space, such as the vectors of features computed from training and testing. γ is defined by $\gamma = -\frac{1}{2\sigma^2}$ for which σ is the Gaussian noise level of standard deviation.

RBF was adopted in this study for prediction of daily dew point temperature. The accuracy of the SVM models largely depends on the appropriate selection of user-defined parameters of C , γ and ε associated with kernels. In this study, the optimal values of user-defined parameters for the SVM model are $C = 2.47$, $\gamma = 0.67$, and $\varepsilon = 0.62$, obtained by trial and error procedure.

In Fig. 1, the hidden nodes' centres are the SVM's support vectors and weights, named as the Lagrange multipliers ($\bar{\alpha}_i = \alpha_i - \alpha_i^*$) that determine the relative significance of the training data sets for the final output.

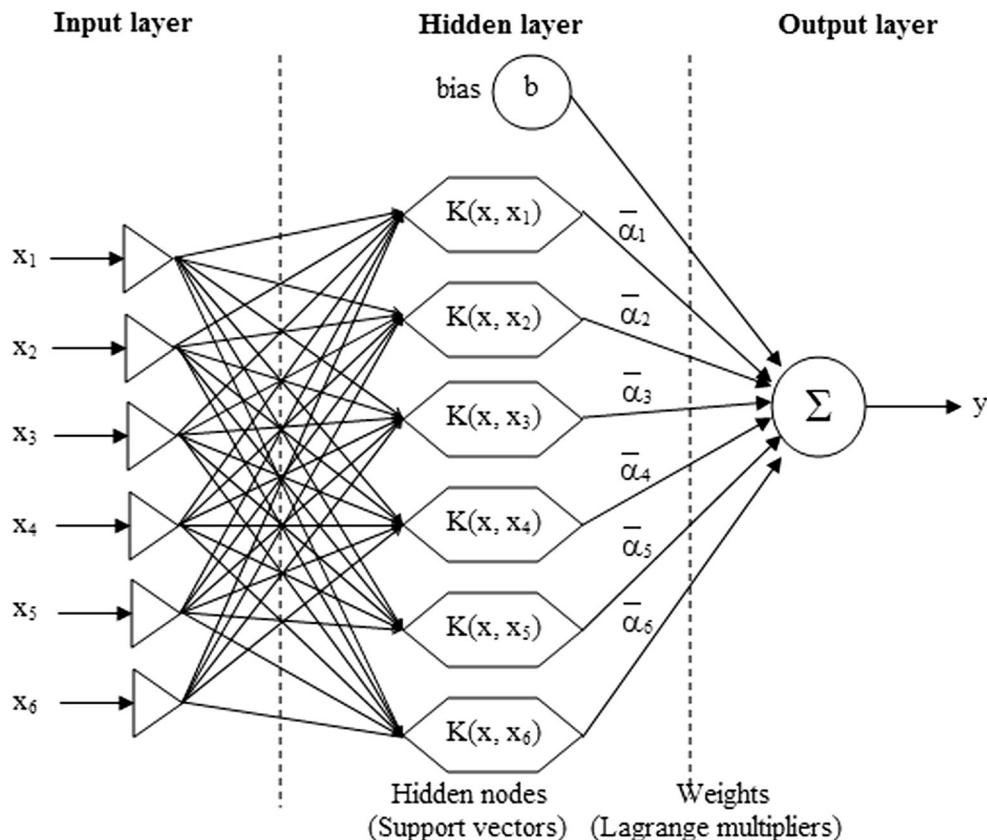
Artificial neural network (ANN)

An usual ANN has 3 layers namely input, output and hidden layers (Aziz and Wong 1992; Balkhair 2002; Chau 2007; Schalkoff 1997).

The input vectors are $D \in \mathbb{R}^n$ and $D = (X_1, X_2, \dots, X_n)^T$, the outputs of q neurons in the hidden layer are $Z = (Z_1, Z_2, \dots, Z_n)^T$, the outputs of the output layer are $Y \in \mathbb{R}^m$, $Y = (Y_1, Y_2, \dots, Y_n)^T$, and the weight and the threshold between the input layer and the hidden layer are w_{ij} and y_j , respectively.

The following equations represent the neuron outputs in hidden layer and output layer (Schalkoff 1997):

Fig. 1 The network architecture of SVM



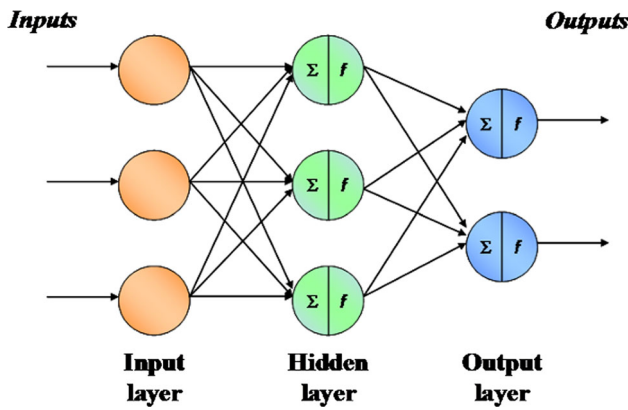


Fig. 2 ANN structure

$$Z_j = f \left(\sum_{i=1}^n w_{ij} X_i \right) \tag{17}$$

$$Y_k = f \left(\sum_{j=1}^q w_{kj} Z_j \right) \tag{18}$$

where a transfer function f is used to offer the rule for mapping the neuron’s total input to its output. A proper selection means introducing a non-linearity into the network’s design. Input and output layers are defined through training data. Number of inputs and number of outputs define the number of input nodes and output nodes, respectively. In this study, the key parameters of the developed ANN model were chosen manually by trial and error procedure. One of the main ANN parameter is learning rate. If the learning rate value be high, then the NN may learn more rapidly; however, in case there is a high variability in the input set then the network cannot learn favorably or at all. Usually, it is better to set the factor to a small value. In this article learning rate of ANN is chosen 0.2 by trial and error procedure. Momentum is also one of the most important ANN parameter. Momentum allows a change to weight of ANN which persists for a number of adjustment cycles. Momentum factor controls the magnitude of the persistence which could improve rate of learning in some situations by helping to smooth out

unusual conditions in the training set. The chosen momentum factor by trial and error procedure in this study is 0.1. The structure of the ANN model is shown in Fig. 2.

Case study and data collection

To appraise the merit of the developed hybrid ELM–WT model, measured daily weather data sets for city of Bandar Abass were utilized in the present study. Bandar Abass, the capital of the Hormozgan province, is located in the southern part of Iran at 27°13’N and 56°22’E, and its elevation is 9.8 m above the sea level. The low level of precipitation, short cool season and long warm season are among the climatic features of this area. According to the Köppen classification the climate condition of Bandar Abass is categorized as BWh, which relates to arid desert hot (Kottek et al. 2006). It is worthwhile to mention that Bandar Abass which is the biggest port of Iran is considered as a significant location in the country in terms of agricultural and hydrological view points. In fact, Hormozgan province and Bandar Abass have an active agriculture sector, ranking first in Iran terms of production of some crops. Also, there are several dams in this region. Therefore, due to importance of accurate dew point temperature in agricultural, hydrological and crop modeling, Bandar Abass was nominated in this research work as a case study.

For this research work, 10 years measured data provided by Iranian Meteorological Organization (IMO) for the period of 2000–2009 have been utilized. Based upon the physical factors influencing the formation of dew, three widely available parameters have been nominated as inputs to predict daily dew point temperature. The used data sets consist of measured daily dew point temperature (T_{dew}), average ambient temperature (T_{avg}), relative humidity (R_h) and atmospheric pressure (P). It should be mentioned that daily average ambient temperature is the average of maximum and minimum air temperatures throughout the day.

The accessible data for this study were divided into two data sets for training and testing. Generally, there is no rule

Table 1 Descriptive statistics for daily weather data utilized for training and testing

Period	Variable	Min	Max	Mean	SD	Range	Correlation with T_{dew}
Training data	T_{dew} (°C)	−9.70	32.30	19.11	7.24	42.00	1.00
	T_{avg} (°C)	11.00	38.40	27.05	6.28	27.40	0.81
	R_h (%)	13.90	99.50	66.46	12.64	85.60	0.73
	P (mb)	990.90	1023.60	1008.50	7.71	32.70	−0.80
Testing data	T_{dew} (°C)	−8.00	29.60	19.31	7.02	37.60	1.00
	T_{avg} (°C)	12.30	37.70	27.64	6.03	25.40	0.80
	R_h (%)	16.10	87.90	64.35	11.52	71.80	0.64
	P (mb)	993.30	1023.10	1007.88	7.66	29.80	−0.78

to choose the size of these data sets. In this study, nevertheless, 7 years data sets for the period of 2000–2006 were utilized to train the models and the remaining 3 years for the period of 2007–2009 were used to test the models. In another word, the models were trained by 2555 days and tested using 1095 days.

Some descriptive statistics including mean values, standard deviation, minimum and maximum values and the range of the data utilized as well as the correlation coefficient between the dew point temperature and the considered input variables for both training and testing data sets are listed in Table 1. The variables of T_{avg} , R_h and P were selected as inputs owing to their broad accessibility in the long-term format as well as their favorable correlation with T_{dew} ; thus, convenient predictions can be performed with favorable accuracy using a limited number of input parameters. It is noticed that there are positive correlations between T_{dew} and two variables of T_{avg} and R_h whereas negative correlations exist between T_{dew} and P . This means that T_{dew} increase with increasing T_{avg} and R_h while T_{dew} decrease with increasing P .

Statistical performance assessment

The proficiency of the proposed hybrid ELM-WT model to predict daily dew point temperature is appraised using different statistical indicators which their descriptions are offered briefly in the following.

The bias error (BE) represents the deviation of predicted data from measured data and it is used to recognize that whether the predicted data are either larger or smaller than the measured values. BE is calculated by:

$$BE = (X_{i,pred} - X_{i,meas}) \tag{19}$$

where $X_{i,pred}$ and $X_{i,meas}$ are the i th predicted and measured values, respectively.

The relative percentage error (RPE) presents the percentage deviation between the predicted and measured data and its values falling within the interval of -10 to $+10$ %

are typically considered acceptable (Duzen and Aydin 2012). RPE is calculated by:

$$RPE = \left(\frac{X_{i,pred} - X_{i,meas}}{X_{i,meas}} \right) \times 100 \tag{20}$$

The mean absolute percentage error (MAPE) shows the mean absolute percentage deviation between the predicted and measured data. Also, the mean absolute bias error (MABE) indicates the average quantity of total absolute bias errors between the predicted and measured values. The MAPE and MABE are obtained respectively, by:

$$MAPE = \frac{1}{x} \sum_{i=1}^x \left| \frac{X_{i,pred} - X_{i,meas}}{X_{i,meas}} \right| \times 100 \tag{21}$$

$$MABE = \frac{1}{x} \sum_{i=1}^x |X_{i,pred} - X_{i,meas}| \tag{22}$$

where x is the total number of observations.

The root mean square error (RMSE) identifies the models' accuracy by providing comparison between the predicted and real data. The RMSE has always a positive value and is calculated by:

$$RMSE = \sqrt{\frac{1}{x} \sum_{i=1}^x (X_{i,pred} - X_{i,meas})^2} \tag{23}$$

The correlation coefficient (R) provides a measure of the linear relationship between the predicted and measured values, obtained by:

$$R = \frac{\sum_{i=1}^x (X_{i,pred} - \bar{X}_{pred}) \cdot (X_{i,meas} - \bar{X}_{meas})}{\sqrt{\left[\sum_{i=1}^x (X_{i,pred} - \bar{X}_{pred})^2 \right] \left[\sum_{i=1}^x (X_{i,meas} - \bar{X}_{meas})^2 \right]}} \tag{24}$$

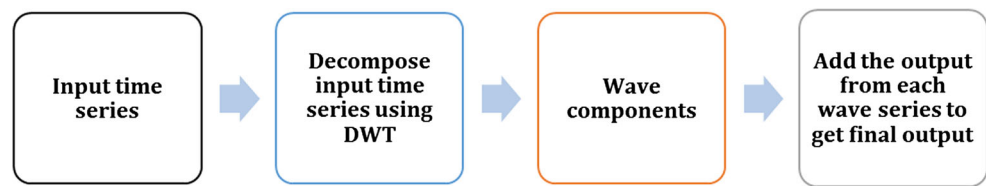
where \bar{X}_{pred} and \bar{X}_{meas} are the average of predicted and measured values, respectively.

The smaller values of BE, RPE, MABE, MAPE and RMSE indicate higher accuracy of the predictions. While greater values of R show higher linear relationship between the predicted and measured values so that its value of -1 or $+1$ shows a perfect linear relationship.

Table 2 User-defined parameters for the developed ELM, SVM and ANN models

ELM		SVM		ANN	
Number of layers	3	–	–	Number of layers	3
Neurons	Input: 6 Hidden: 3, 6, 10 Output: 1	Neurons	–	Neurons	Input: 6 Hidden: 3, 6, 10 Output: 1
–	–	Cost	$C = 2.47$	Number of iteration	1000
–	–	Learning rate	$\varepsilon = 0.62$	Activation function	Sigmoid Function
Learning rule	ELM for SLFNs	Regularization parameter	$\gamma = 0.67$	Learning rule	Back propagation

Fig. 3 Flow chart of proposed Wavelet-based parameter determination approach for the ELM



Results and discussion

In this research work, the ELM was coupled with WT algorithm to propose a new hybrid approach named ELM-WT for prediction of daily dew point temperature. The merit of the proposed ELM-WT method was verified against ELM, SVM and ANN.

Table 2 presents the parameters of the ELM, SVM and ANN modeling frameworks employed in this study.

The following steps are considered for the ELM modeling:

Step 1: Initiate the population based on the ELM function.

Step 2: Evaluate the fitness function of each parameter

Step 3: Step 2 and Step 3 are iterated repetitively until the optimal iteration time is satisfied.

Step 4: The optimal parameters of ELM function can be determined. Then based on the optimized parameters, the hidden layer function matrix is computed.

Step 5: Determine the final output weights.

In this study, wavelet analysis is employed to decay series of the dew point temperature data to individual components, where the decomposed components can be considered for data inputs into the ELM model expressed. Figure 3 depicts the flow chart for obtaining the optimal ELM parameters using the wavelet algorithm.

Basically, the ability of each model and technique to provide precise predictions is dependent on appropriate selection of input parameters. For this study, three highly accessible parameters of average ambient temperature (T_{avg}), relative humidity (R_h) and atmospheric pressure (P) were selected. These parameters have important influences on formation of dew so that T_{avg} corresponds to the turbulent temperature. Also, R_h and P correspond to the water vapor pressure (Shiri et al. 2014). The primary evaluation revealed that T_{avg} is more effective compared to R_h and P to obtain acceptable prediction of daily dew point

Table 3 The considered models with different input elements

Model	Input element
(1)	T_{avg}
(2)	T_{avg}, R_h
(3)	T_{avg}, P
(4)	T_{avg}, R_h, P

Table 4 Performance assessment of the all developed ELM-WT, ELM, SVM and ANN models on the basis of the statistical parameters

Model	Training data				Testing data			
	MAPE (%)	MABE (°C)	RMSE (°C)	R	MAPE (%)	MABE (°C)	RMSE (°C)	R
ELM-WT (1)	40.7595	2.8637	4.3812	0.8230	42.1637	2.9394	4.4333	0.8210
ELM-WT (2)	36.6282	2.7581	4.0819	0.8477	38.5362	2.8678	4.1610	0.8430
ELM-WT (3)	5.7261	0.5153	0.7038	0.9956	6.1664	0.5495	0.7621	0.9953
ELM-WT (4)	5.9359	0.5185	0.7059	0.9952	6.2047	0.5503	0.7653	0.9951
ELM (1)	48.3972	3.2295	4.6868	0.8132	50.5339	3.2990	4.7552	0.8091
ELM (2)	46.1257	3.1654	4.4413	0.8189	47.9237	3.2172	4.5250	0.8148
ELM (3)	9.5487	0.8272	1.0158	0.9922	9.9590	0.8517	1.0529	0.9917
ELM (4)	9.5983	0.8365	1.0067	0.9918	10.0576	0.8642	1.0625	0.9905
SVM (1)	50.6613	3.4443	4.8396	0.7979	52.9299	3.4671	4.9073	0.7910
SVM (2)	48.6803	3.3132	4.5784	0.8075	50.4545	3.3539	4.3913	0.8019
SVM (3)	11.1067	1.0153	1.2456	0.9911	11.6036	1.0547	1.3178	0.9889
SVM (4)	11.1482	1.0546	1.2984	0.9903	11.7503	1.1012	1.3436	0.9851
ANN (1)	53.8352	3.5782	5.0805	0.7851	56.7015	3.6986	5.1582	0.7769
ANN (2)	52.6402	3.4931	4.9820	0.7937	54.4701	3.5481	4.9123	0.7832
ANN (3)	13.8815	1.1951	1.4441	0.9863	14.4457	1.2318	1.5188	0.9835
ANN (4)	13.9271	1.2279	1.4856	0.9855	14.4929	1.2901	1.5481	0.9805

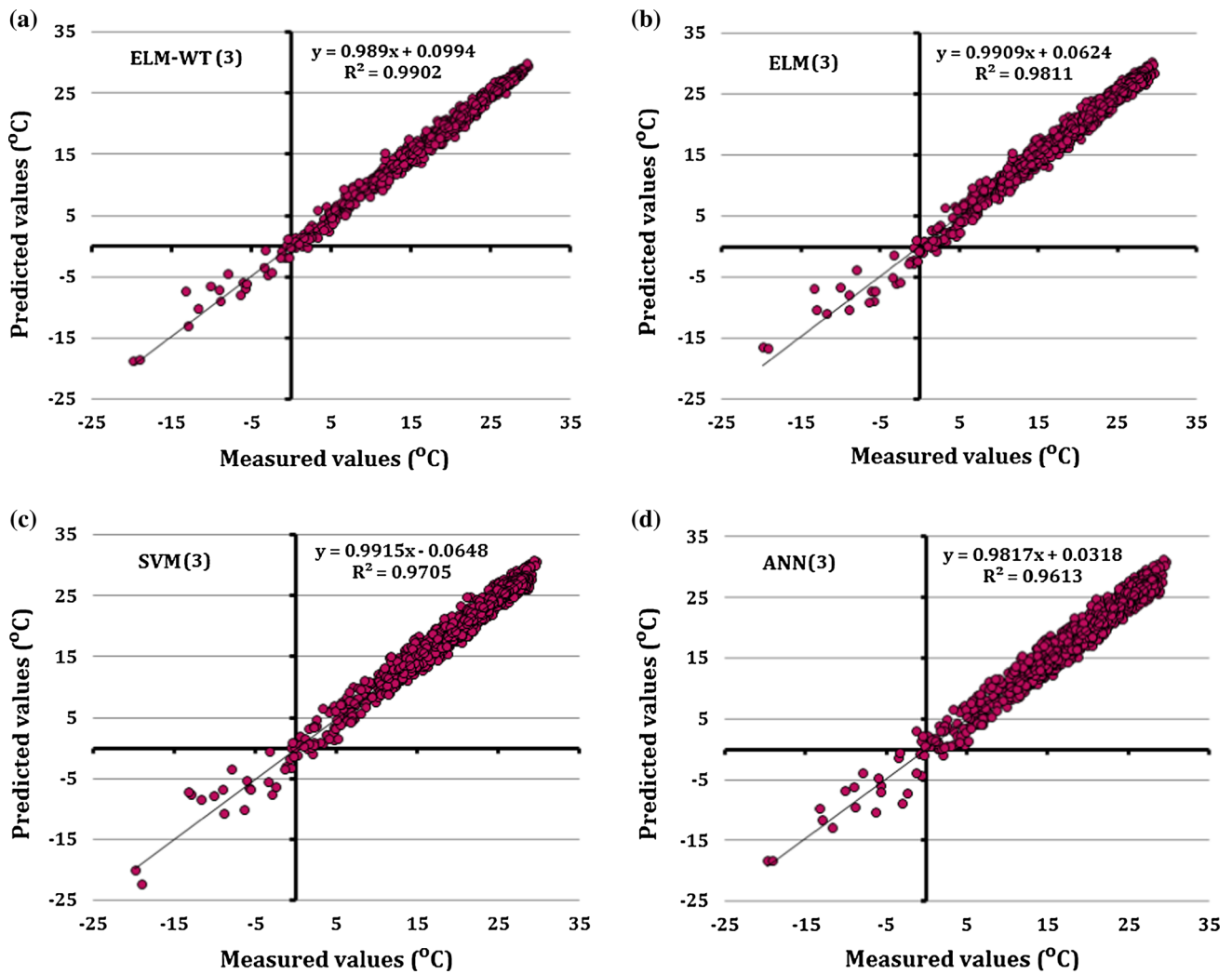


Fig. 4 Scatter plots of the predicted daily dew point temperatures by: **a** ELM-WT (3), **b** ELM (3), **c** SVM (3) and **d** ANN (3) versus the measured values

temperature. Thus, further analysis is only based on the combination of T_{avg} with R_h and P . To achieve this, four models with 1, 2 and 3 input combinations are established using each approach and later explored. These four models with the used input elements are presented in Table 3.

The performance evaluations of the models are conducted via several reliable and widely utilized statistical indicators. For this purpose, a comparison between the predicted values of daily dew point temperature via ELM-WT, ELM, SVM and ANN models and the measured ones was performed by the statistical indicators of MAPE, MABE, RMSE and R. Table 4 displays the performances of all developed models in terms of calculated statistical indicators. The results clearly indicate that for all considered combinations of inputs, the proposed ELM-WT approach presents absolute superiority over other employed techniques. This conclusion is drawn due to the lower values of MAPE, MABE and RMSE as well as the higher

values of R. It is also noticed that the ELM, SVM and ANN approaches, respectively placed in the next ranks in terms of offering higher accuracy. The presented results in Table 4 convincingly advocate that coupling the ELM with WT would be highly promising as it leads to favorable improvement in the precision of ELM for prediction of dew point temperature. From the attained results it is also found that selection of proper input combinations has a notable influence on the prediction accuracy obtained by each method. It is observed that model (3) and (4) of each approach utilizing T_{avg} and R_h as well as T_{avg} , R_h and P as input elements presents much greater precision compared to models (1) and (2). Nevertheless, the model (3) has slightly better performance than model (4) and less complexity in terms of required inputs. Thus, utilization of model (3) with two inputs of T_{avg} and R_h is undoubtedly more appropriate. For the developed ELM-WT (3) model, statistical parameters of MAPE, MABE, RMSE and R are

achieved as 6.1664 %, 0.5495, 0.7621 and 0.9953 °C, respectively for the testing phase. While for the ELM (3) placed in the next rank the obtained values are 10.0576 %, 0.8642, 1.0625 and 0.9905 °C, respectively.

In the following the merit of ELM-WT (3) is further assessed compared to ELM (3), SVM (3) and ANN (3).

The scatter plots of the predicted daily dew point temperature values by ELM-WT (3), ELM (3), SVM (3) and ANN (3) models against measured data are presented in Fig. 4a–d, respectively for the testing phase. The values of coefficient of determination (R^2), as an indicator of the linear relationship between the predicted and observed data, were calculated and presented for all figures. It is highly clear that the highest R^2 between the predicted dew point temperature values and the measured values are attained by the ELM-WT (3) model. In fact, it is observed that the dispersion degree of the scatter plots of the ELM-WT models is highly lower than other plots presented for the ELM, SVM and ANN models. This demonstrates the higher degree of linear relationship between the predicted values by ELM-WT (3) and the measured values.

To evaluate the performance of the developed ELM-WT (3), ELM (3), SVM (3) and ANN (3) models in different days of the year, a statistical analysis was performed by computing the daily bias error (BE) and daily relative percentage error (RPE). These evaluations would definitely be useful to determine the capability of the models for dew point temperature prediction in different days throughout the years. Figure 5a–d illustrates the daily BE between the measured dew point temperature data and the predicted values by ELM-WT (3), ELM (3), SVM (3) and ANN models, respectively for the testing phase. Clearly, the daily BE achieved for the ELM-WT (3) model has the lowest values compared to the other models. In fact, for the ELM-WT (3) model the attained daily BE are highly favorable so that for most of the days the BE falls within the range of -1 to 1 °C.

RPE analysis was also utilized to evaluate the day by day performance of the developed models. On this account, if a higher number of days fall in the lower ranges of RPE, the precision of the model will be higher. As mentioned, based on the RPE definition, the RPE values which fall in

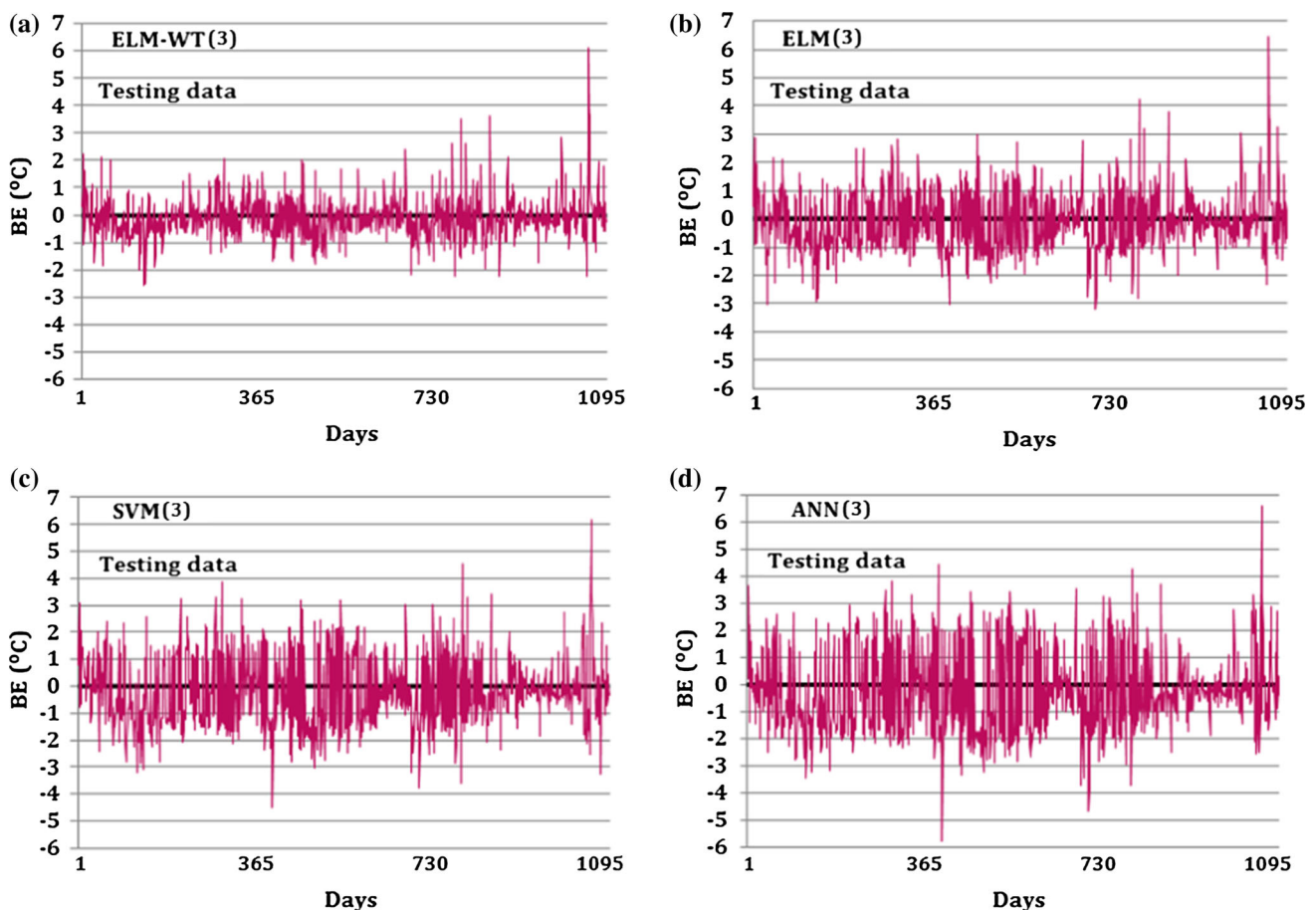


Fig. 5 Daily bias error (BE) between the predicted daily dew point temperatures by: **a** ELM-WT, **b** ELM, **c** SVM and **d** ANN and the measured values

the interval of -10 to 10 % are typically considered acceptable (Duzen and Aydin 2012).

The attained results show that for the ELM-WT (3), ELM (3), SVM (3) and ANN (3) models, of the 1095 days considered for the testing phase, the predictions fall within the acceptable range, respectively in the 991, 925, 855, 797 days. This means that the predicted values by the ELM-WT (3), ELM (3), SVM (3) and ANN (3) models, respectively for 91, 85, 78, and 73 % of the testing data set fall within the RPE acceptable range of -10 and $+10$ %. Thus, it is found that the ELM-WT (3) model offers much greater performance in terms of RPE analysis since in a higher number of days its predictions fall in the acceptable range of -10 and $+10$ %.

The combination of BE and RPE analysis clearly demonstrates that the ELM-WT (3) model enjoys high capability to predict daily dew point temperature in different days throughout the years.

Conclusion

Accessibility to the accurate and reliable dew point temperature data is of indispensable importance in many scientific areas such as hydrology, climatology and agronomy. Hybrid models which enjoy the particular features of different techniques would be effective to attain further accuracy and reliability in the predictions. In this research work, a new hybrid method based on integrating the ELM with WT algorithm was proposed for prediction of daily dew point temperature. For this aim, daily climate data of an Iranian station placed in the south costal of the country were utilized as a case study. According to the physical factors influencing the formation of dew, three highly available meteorological parameters of T_{avg} , R_h and P were considered as input elements. Wavelet analysis was utilized for preprocessing and decomposing the time series of weather data into its various components, after which the decomposed components were utilized as inputs for the ELM model. To certify the suitability of the ELM-WT method, its predictions were compared against powerful techniques of ELM, SVM and ANN. To offer thorough comparisons and performance evaluations, the widely-used statistical indicators of BE, RPE, MAPE, MABE, RMSE and R were utilized. Based upon the achieved results it was conclusively proved that the hybrid ELM-WT approach favorably outperforms other examined techniques. In fact, the results demonstrated that integrating the WT into ELM boosts the effectiveness of ELM for prediction.

The primary evaluation showed that T_{avg} is more effective than R_h and P to attain acceptable predictions. Among four considered sets of parameters with 1, 2 and 3 inputs the combination of T_{avg} and R_h were determined as

more effective inputs. For the best ELM-WT model using T_{avg} and R_h as inputs, the statistical indicators of MAPE, MABE, RMSE and R were obtained equal to 6.1664 %, 0.5495, 0.7621 and 0.9953 °C, respectively. The BE and RPE analysis showed that the best ELM-WT model, using T_{avg} and R_h as inputs, enjoys high performance to predict daily dew point temperature in different days throughout the years. For the best ELM-WT model, 91 % of the predictions fell within the RPE acceptable range of -10 and $+10$ %.

To sum up, the proposed hybrid ELM-WT approach would be highly efficient to offer precise predictions of daily dew point temperature and also higher accuracy than the single ELM technique.

References

- Adamowski J, Chan HF (2011) A wavelet neural network conjunction model for groundwater level forecasting. *J Hydrol* 407(1):28–40
- Agam N, Berliner PR (2006) Dew formation and water vapor absorption in semi-arid environments—a review. *J Arid Environ* 65:572–590
- Asefa T, Kemblowski M, McKee M, Khalil A (2006) Multi-time scale stream flow predictions: the support vector machines approach. *J Hydrol* 318:7–16
- Atzema AJ, Jacobs AFG, Wartena L (1990) Moisture distribution within a maize crop due to dew. *Neth J Agric Sci* 38:117–129
- Aziz A, Wong K (1992) Neural-network approach to the determination of aquifer parameters. *Ground Water GRWAAAP* 30:164–166
- Balkhair K (2002) Aquifer parameters determination for large diameter wells using neural network approach. *J Hydrol* 265:118–128
- Burrus CS, Gopinath RA, Guo H (1997) Introduction to wavelets and wavelet transforms: a primer. Prentice Hall, Englewood Cliffs
- Chau K (2007) Reliability and performance-based design by artificial neural network. *Adv Eng Softw* 38:145–149
- Duzen H, Aydin H (2012) Sunshine-based estimation of global solar radiation on horizontal surface at Lake Van region (Turkey). *Energy Convers Manag* 58:35–46
- Ghouthi L, Sheltami TR, Alutaibi KS (2013) Mobility prediction in mobile ad hoc networks using extreme learning machines. *Proc Comput Sci* 19:305–312
- Huang GB, Zhu QY, Siew CK (2004) Extreme learning machine: a new learning scheme of feedforward neural networks. *Int Jt Conf Neural Netw* 2:985–990
- Huang GB, Zhu QY, Siew CK (2006a) Extreme learning machine: theory and applications. *Neurocomputing* 70:489–501
- Huang GB, Chen L, Siew CK (2006b) Universal approximation using incremental constructive feedforward networks with random hidden nodes. *IEEE Trans Neural Netw* 17:879–892
- Huang G, Huang GB, Song S, You K (2015) Trends in extreme learning machines: a review. *Neural Netw* 61:32–48
- Hubbard KG, Mahmood R, Carlson C (2003) Estimating daily dew point temperature for the northern Great Plains using maximum and minimum temperature. *Agron J* 95(2):323–328
- Jawerth B, Sweldens W (1994) An overview of wavelet based multiresolution analyses. *SIAM Rev* 36(3):377–412
- Ji Y, Sun S (2013) Multitask multiclass support vector machines: model and experiments. *Pattern Recogn* 46(3):914–924

- Kalteh AM (2013) Monthly river flow forecasting using artificial neural network and support vector regression models coupled with wavelet transform. *Comput Geosci* 54:1–8
- Kavousi-Fard A, Samet H, Marzbani F (2014) A new hybrid modified firefly algorithm and support vector regression model for accurate short term load forecasting. *Expert Syst Appl* 41:6047–6056
- Kim S, Singh VP, Lee CJ, Seo Y (2014) Modeling the physical dynamics of daily dew point temperature using soft computing techniques. *KSCE J Civ Eng*. doi:10.1007/s12205-014-1197-4
- Kimball JS, Running SW, Nemani R (1997) An improved method for estimating surface humidity from daily minimum temperature. *Agric For Meteorol* 85:87–98
- Kisi O, Cimen M (2012) Precipitation forecasting by using wavelet-support vector machine conjunction model. *Eng Appl Artif Intell* 25:783–792
- Kottek M, Grieser J, Beck C, Rudolf B, Rubel F (2006) World map of the Koppen–Geiger climate classification updated. *Meteorol Z* 15(3):259–263
- Lee S-W, Verri A (2003) Support vector machines for computer vision and pattern recognition. World Scientific, Singapore
- Liang NY, Huang GB, Rong HJ, Saratchandran P, Sundararajan N (2006) A fast and accurate on-line sequential learning algorithm for feedforward networks. *IEEE Trans Neural Netw* 17:1411–1423
- Liang R-P, Huang S-Y, Shi S-P, Sun X-Y, Suo S-B, Qiu J-D (2012) A novel algorithm combining support vector machine with the discrete wavelet transform for the prediction of protein sub cellular localization. *Comput Biol Med* 42:180–187
- Lu WZ, Wang WJ (2005) Potential assessment of the “support vector machine” method in forecasting ambient air pollutant trends. *Chemosphere* 59:693–701
- Mallat SGA (1989) Theory for multiresolution signal decomposition: the wavelet representation. *IEEE Trans Pattern Anal Mach Intell* 11(7):674–693
- Mallat SGA (2009) A wavelet tour of signal processing: the sparse way, 3rd edn. Academic Press, Burlington
- Mohammadi K, Shamsheerband S, Seyed Danesh A, Zamani M, Sudheer C (2015) Horizontal global solar radiation estimation using hybrid SVM-firefly and SVM-wavelet algorithms: a case study. *Nat Hazards*. doi:10.1007/s11069-015-2047-5
- Mohammadi K, Shamsheerband S, Petkovic D, Yee PL, Mansor Z (2016) Using ANFIS for selection of more relevant parameters to predict dew point temperature. *Appl Therm Eng* 96:311–319
- Nadig K, Potter W, Hoogenboom G, McClendon R (2013) Comparison of individual and combined ANN models for prediction of air and dew point temperature. *Appl Intell* 39:354–366. doi:10.1007/s10489-012-0417-1
- Nian R, He B, Zheng B, Heeswijk MV, Yu Q, Miche Y et al (2014) Extreme learning machine towards dynamic model hypothesis in fish ethology research. *Neurocomputing* 128:273–284
- Olatomiwa L, Mekhilef S, Shamsheerband S, Mohammadi K, Petkovic D, Sudheer C (2015) A support vector machine-firefly algorithm-based model for global solar radiation prediction. *Sol Energy* 115:632–644
- Partal T, Kisi O (2007) Wavelet and neuro-fuzzy conjunction model for precipitation forecasting. *J Hydrol* 342:199–212
- Peng Z, Chu F (2004) Application of the wavelet transform in machine condition monitoring and fault diagnostics: a review with bibliography. *Mech Syst Signal Process* 18(2):199–221
- Rajasekaran S, Gayathri S, Lee TL (2008) Support vector regression methodology for storm surge predictions. *Ocean Eng* 35(16):1578–1587
- Schalkoff RJ (1997) Artificial neural networks. McGraw-Hill Higher Education, New York
- Shamsheerband S, Petković D, Pavlović NT, Sudheer C, Torki A, Altameem TA, Gani A (2015) Support vector machine firefly algorithm based optimization of lens system. *Appl Opt* 54:37–45
- Shamsheerband S, Mohammadi K, Khorasanizadeh H, Yee PL, Lee M, Petković D, Zalnezhad E (2016) Estimating the diffuse solar radiation using a coupled support vector machine-wavelet transform model. *Renew Sustain Energy Rev* 56:428–435
- Shank DB (2006) Dew point temperature prediction using artificial neural networks, MS thesis. Harding University
- Shank DB, Hoogenboom G, McClendon RW (2008) Dew point temperature prediction using artificial neural networks. *Appl Meteorol Climatol* 47:1757–1769
- Shiri J, Kim S, Kisi O (2014) Estimation of daily dew point temperature using genetic programming and neural networks approaches. *Hydrol Res* 45(2):165–181
- Shrivastava NA, Panigrahi BK (2014) A hybrid wavelet-ELM based short term price forecasting for electricity Markets. *Electr Power Energy Syst* 55:41–50
- Slatyer RO (1967) Plant–water relationships. Academic Press, London
- Snyder RL, Melo-Abreu JPD (2005) Frost protection: fundamentals, practice and economics, vol 1. Food and Agricultural Organization of the United Nations, Rome
- Sudheer C, Sohani SK, Kumar D, Malik A, Chahar BR, Nema AK et al (2014) A support vector machine-firefly algorithm based forecasting model to determine malaria transmission. *Neurocomputing* 129:279–288
- Sun S (2013) A survey of multi-view machine learning. *Neural Comput Appl* 23(7–8):2031–2038
- Vapnik V (2000) The nature of statistical learning theory. Springer, New York
- Vapnik VN, Vapnik V (1998) Statistical learning theory, vol 2. Wiley, New York
- Wang W, Ding J (2003) Wavelet network model and its application to the prediction of hydrology. *Nat Sci* 1(1):67–71
- Wang X, Han M (2014) Online sequential extreme learning machine with kernels for nonstationary time series prediction. *Neurocomputing* 145:90–97
- Wang DD, Wang R, Yan H (2014) Fast prediction of protein–protein interaction sites based on Extreme Learning Machines. *Neurocomputing* 128:258–266
- Went FW (1955) Fog, mist dew and other sources of water. Year book agriculture. US Department of Agriculture, Washington, DC, pp 103–109
- Wong PK, Wong KI, Vong CM, Cheung CS (2015) Modeling and optimization of biodiesel engine performance using kernel-based extreme learning machine and cuckoo search. *Renew Energy* 74:640–647
- Wu KP, Wang SD (2009) Choosing the kernel parameters for support vector machines by the inter-cluster distance in the feature space. *Pattern Recogn* 42(5):710–717
- Xiong T, Bao Y, Hu Z (2014) Multiple-output support vector regression with a firefly algorithm for interval-valued stock price index forecasting. *Knowl Based Syst* 55:87–100
- Yang H, Huang K, King I, Lyu MR (2009) Localized support vector regression for time series prediction. *Neurocomputing* 72(10):2659–2669
- Yu Q, Miche Y, Séverin E, Lendasse A (2014) Bankruptcy prediction using extreme learning machine and financial expertise. *Neurocomputing* 128:296–302
- Zhao Z, Li P, Xu X (2013) Forecasting model of coal mine water inrush based on extreme learning machine. *Appl Math Inf Sci* 7:1243–1250
- Zounemat-Kermani M (2012) Hourly predictive Levenberg–Marquardt ANN and multi linear regression models for predicting of dew point temperature. *Meteorol Atmos Phys* 117:181–192