© 2011, Scienceline Publication

Journal of Civil Engineering and Urbanism

Volume 1, Issue 1: 25-38 (2011)

A Wavelet-Genetic Programming Model for Predicting Short-Term and Long-Term Air Temperatures

Ozgur Kisi[1], Jalal Shiri[2*], Amir-Hossein Nazemi[2]

[1.] *Engineering Faculty, Civil Engineering Department, Hydraulics Divisions, Erciyes University, Kayseri, Turkey.*
[2.] *Faculty of Agriculture, Water Engineering Department, University of Tabriz, Tabriz, Iran.*

*Corresponding author's Email: j_shiri2005@yahoo.com

ORIGINAL ARTICLE

ABSTRACT

A new conjunction wavelet-gene expression programming (WGEP) method for predicting air temperature values is proposed in this paper. The conjunction method combines the discrete wavelet and genetic programming methods. The daily and monthly air temperature data from two weather stations of Mahabad and Urmieh in Iran were used as case studies and the accuracy of the single gene expression programming (GEP) and wavelet-gene expression programming (WGEP) models were compared with each other. First, the daily air temperatures were used as inputs to the GEP and WGEP models to forecast one-, two- and three day as well as thirty-day ahead air temperatures. Then, the monthly air temperatures were used as inputs to the GEP and WGEP models to forecast one-month ahead air temperatures. The comparison results indicated that the WGEP model significantly increased the accuracy of single GEP model especially in forecasting long-term (thirty-day and one-month ahead) air temperatures. The thirty-day and one-month ahead air temperatures of the Mahabad Station were also estimated using the data of nearby Urmieh Station. It was found that the WGEP model performed much better than the single GEP model in cross-station application.

Keywords: Air temperature, discrete wavelet, genetic programming, cross application

INTRODUCTION

Air temperature is an important climatic variable used in determining site suitability for agricultural and forest crops [1], parameterizing the habitat of plant species [2], determining the pattern of vegetational zonation [3], predicting the energy consumption of a passive solar buildings [4], predicting the soil surface temperature [5] and modeling hourly diffuse solar radiation [6]. A number of attempts have been carried out to model air temperature variations [7-12] which have emphasized the need to accurate estimation of air temperature in various aspects of meteorology, hydrology and agro-hydrology.

Artificial Intelligence (AI) approaches have been successfully used in a wide range of scientific applications including water resources engineering, agro-hydrology and agro-meteorology [e.g., 13-22]. Also a fuzzy modeling approach for predicting air temperature has been suggested [23]. Nonetheless, adductive neural network approach as well as enhanced artificial neural network (ANN), has been applied for air temperature prediction [24, 25] along with the ANN application for dew point forecast [26].

The methodology of Genetic Programming (GP) was first proposed by Koza [27], as a generalization of Genetic Algorithms (GA) [28]. The fundamental difference between GP and GAs lies in the nature of individuals, where in GAs individuals are linear strings of fixed length (as chromosomes), while in GP individuals are nonlinear entities of different sizes and shapes (as parse trees). Major advantages of GP are that it can be applied to areas where (i) the interrelationships among the relevant variables are poorly understood (or where it is suspected that the current understanding may well be less than satisfactory), (ii) finding the ultimate solution is hard, (iii) conventional mathematical analysis does not, or cannot, provide analytical solutions, (iv) an approximate solution is acceptable (or is the only result that is ever likely to be obtained), (v) small improvements in the performance are routinely measured (or easily measurable) and highly valued, and (vi) there is a large amount of data, in computer readable form, that requires examination, classification, and integration (such as satellite observations) [29]. One of the strong points of using GP over other data driven techniques is that it can produce explicit formulations (model expression) of the relationship that rules the physical phenomenon. Such expressions may be subject to some physical interpretations. Actually, the comprehensibility of GP models is also a way to reduce the risk of over-fitting to training data and improve generalization of resulting models. In this way, one may perform knowledge discovery using GP, finding some confirmation of well-known physical relationships and evolving interesting new formulae, as an up-gradation of particular cases of study. GP has been applied for modeling risks in water supply [30], rainfall-runoff modeling [31, 32, 33], suspended sediment transport modeling in streams [34], predicting of compressive and tensile strength of limestone [35], forecasting sea water level [36], estimating short-term and long-term river flow [37], soil liquefaction modeling [38], predicting groundwater table depth fluctuations [39] and estimating daily pan evaporation values using recorded and estimated weather variables [40].

In the last decade, wavelet transform has become a useful technique for analyzing variations, periodicities, trends in time series. It has been used for quantifying stream flow variability [41], decomposition of interdecadal and interannual components of rainfall data in rainy season [42], studying the rainfall spectrum and its evolution of North China in rainy season with summer monsoon decaying interdecadal time scale [43], identifying and describing variability in annual Canadian stream flows and to gain insights into the dynamic link between the streamflows and the dominant modes of climate variability in the Northern Hemisphere [44], illustrating new wavelet analysis methods in the field of hydrology [45], demonstrating the application of new wavelet indicators to several improvements in the analysis of global hydrological signal fluctuations and of their mutual time varying relationships [46], determining the possible trends in annual total precipitation series [47], precipitation forecast [48], simulation and prediction of monthly discharge time series [49], predicting short term and long term stream flows [50], predicting daily precipitation values (using WGEP model) [51], predicting monthly stream flows [52] and predicting hourly as well as daily wind speed values [53].

In the present paper, a conjunction model (wavelet-GEP) was applied to predict daily and monthly air temperatures. Air temperature data considered are decomposed into wavelet sub-series by discrete wavelet transform. Then, GEP model is constructed with appropriate wavelet sub-series as input, and original air temperature time series as output. To the best authors' knowledge, the presented study is the first application for air temperature forecasting using wavelet and GEP in the literature.

Data Used

Daily recorded air temperature data from two stations located in the West-Azarbayjan Province in North-West Iran which covers a time period of 10 years (from March 2000 to April 2009) are used in this study. Table1 represents some of the statistical properties of the applied data. For each station, the first six years data (60% of whole data) are applied for training the models, two years for testing and the remaining two years are applied for models validation. Such a manner (data division in three parts) is much better than the data division in two parts. First, one can obtain models' parameters by using training data and then, choose the optimal model according to their testing performances. Finally, the evaluation and comparison the optimal models can be achieved by using different data (validation) sets which are not used for model development stages (training and testing).

Modeling Procedure

The modeling procedure of time series analysis consists of three major phases as follows [54]:

Phase 1: reviewing the data for any possible discontinuity in both dependent and independent data set and choosing the appropriate software; dividing the data into training, validation and application blocks.

Phase 2: implementing the time series analysis as per selected modeling application; setting the parameters of selected software and producing the results. This phase depends on the time-series analysis technique, which for GEP the primary object is to identify the relationship between independent and dependent variables.

Phase 3: Post-processing the results in relation to training, validation and application and if applicable, carrying out some sensitivity analysis.

MATERIALS and METHODS

Overview of Genetic Programming

GEP (Gene Expression Programming) is comparable to GP yet evolves computer programs of different sizes and shapes encoded in linear chromosomes of fixed lengths. The chromosomes are composed of multiple genes, each gene encoding a smaller subprogram. Furthermore, the structural and functional organization of the linear chromosomes allows the unconstrained operation of important genetic operators such as mutation, transposition and recombination. The advantages of a system like GEP are clear from nature, but the most important are [55]: (i) the chromosomes are simple entities: linear, compact, relatively small, easy to manipulate genetically (replicate, mutate, recombine, etc.); (ii) the expression trees are exclusively the expression of their respective chromosomes; they are entities upon which selection acts, and according to fitness, they are selected to reproduce with modification. In the present work the GeneXpro program was used for modeling air temperature [55]. The procedure to forecast air temperature is as follows. The first step is the fitness function. For this problem, the fitness function, f_i, of an individual program, i, is expressed as [55]:

$f_i = \sum_{j=1}^{n}\left(M - \left|C_{i,j} - T_j\right|\right)$; in which M is the range of selection, $C_{i,j}$ is the value predicted by individual program i for fitness case j, and T_j is the target value for fitness case j. For a perfect fit, $C_{i,j}=T_j$. 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 air temperature values: {T_i, T_{i-1}, T_{i-2} and T_{i-3} where T_j denotes the air temperature at time j}. The study examined the various combinations of these parameters as inputs to the GEP models to evaluate the degree of effect of each of these variables on air temperature at specified time step. These variables were added into input combinations several times with one different variable added into the input combination. The choice of the appropriate function is not so obvious and depends on the viewpoint and guess of user. In this study, different mathematical functions were utilized, including basic arithmetic operators ({+, -, *, /}) as well as some of the other basic mathematical functions ({$\sqrt{\ }$, $\sqrt[3]{\ }$, ln(x), e^x, x^2, x^3}). The preliminary investigation of parse tree (and choosing the appropriate function set) shows that this function set has more accuracy. However, the full study about the effect of function set and parse tree on the models' performance is beyond the scope of this paper. The third step is to choose the chromosomal architecture. Length of head, h=8, and three genes per chromosomes are employed. The fourth step is to choose the linking function. The linking function must be chosen as "addition" or "multiplication" for algebraic sub trees [55]. Here, the sub trees are 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. It is noted that these parameters are default values of GeneXpro program.

Discrete Wavelet Transform (DWT)

Wavelet function $\psi(t)$, or so-called the mother wavelet, can be defined as $\int_{-\infty}^{+\infty}\psi(t)dt = 0$. The function $\psi_{a,b}(t)$ can be obtained through compressing and expanding $\psi(t)$ as follows

$$\psi_{a,b}(t) = |a|^{-1/2}\psi\left(\frac{t-b}{a}\right) \quad b \in R,\ a \in R,\ a \neq 0 \tag{1}$$

Where $\psi_{a,b}(t)$ = the successive wavelet, a = the scale or frequency factor, b = a time factor; R = the domain of real numbers.

If $\psi_{a,b}(t)$ satisfies Equation (1), for the time series f(t) $\in L^2(R)$ or finite energy signal, successive wavelet transform of f(t) is defined as

$$W_\psi f(a,b) = |a|^{-1/2}\int_R f(t)\bar{\psi}\left(\frac{t-b}{a}\right)dt \tag{2}$$

Where $\bar{\psi}(t)$ = complex conjugate functions of $\psi(t)$. It can be seen from Equation (2) that the wavelet transform is the decomposition of f(t) under different resolution level (scale). In other words, to filter wave for f(t) with different filter is the essence of wavelet transform.

The successive wavelet is often discrete in real applications. Let $a = a_0^j$, $b = kb_0a_0^j$, $a_0 > 1$, $b_0 \in R$, k, j are integer numbers. Discrete wavelet transform of f(t) can be written as:

$$W_\psi f(j,k) = a_0^{-j/2}\int_R f(t)\bar{\psi}\left(a_0^{-j}t - kb_0\right)dt \tag{3}$$

The most common (and simplest) choice for the parameters a_0 and b_0 is 2 and 1 time steps, respectively. This power of two logarithmic scaling of the time and scale is known as dyadic grid arrangement and is the simplest and most efficient case for practical purposes [56]. Equation (3) becomes binary wavelet transform when a_0 = 2, b_0 = 1:

$$W_\psi f(j,k) = 2^{-j/2}\int_R f(t)\bar{\psi}\left(2^{-j}t - k\right)dt \tag{4}$$

The characteristics of the original time series in frequency (a or j) and time domain (b or k) at the same time are reflected by $W_\psi f(a,b)$ or $W_\psi f(j,k)$. When the frequency resolution of wavelet transform is low, but the time domain resolution is high a or j becomes small. When the frequency resolution of wavelet transform is high, but the time domain resolution is low a or j becomes large [57].

For a discrete time series f(t), where occurs at different time t (i.e., here integer time steps are used), the DWT can be defined as

$$W_\psi f(j,k) = 2^{-j/2}\sum_{t=0}^{N-1} f(t)\bar{\psi}\left(2^{-j}t - k\right) \tag{5}$$

Where $W_\psi f(j,k)$ is wavelet coefficient for the discrete wavelet of scale a = 2^j, b = 2^jk.

DWT operates two sets of function viewed as high-pass and low-pass filters. The original time series are passed through high-pass and low-pass filters and separated at different scales. The time series is decomposed into one comprising its trend (the approximation) and one comprising the high frequencies and the fast events (the detail) [58]. In the present study, the detail coefficients and approximation (A) sub-time series are obtained using the Equation (5).

Main structure of wavelet-GEP model

The aim of wavelet-GEP (WGEP) model is to predict the 1-, 2- and 3-day ahead air temperatures employing sub-series components (DWs) obtained using DWT on original data. For this purpose, firstly the original time series are decomposed into a certain number of DWs Mallat DWT algorithm [56].The WGEP is constructed in which the DWs of original input time series are input of the GEP and the original output time series are output of the GEP. In the study, the current and previous air temperature time series are decomposed into various DWs at different resolution levels by using DWT to forecast i-day/month ahead air temperature values. Ten and six resolution levels were employed for the daily and monthly air temperatures in this study. Also, ten resolution levels of DWs indicating DW_1 (2^1–day mode), DW_2 (2^2–day mode), DW_3 (2^3–day mode which is nearly weekly mode), DW_4 (2^4–day mode), DW_5 (2^5–day mode which is nearly monthly mode), DW_6 (2^6–day mode), DW_7 (2^7–day mode), DW_8 (2^8–day mode), DW_9 (2^9–day mode) and DW_{10} (2^{10}–day mode) and six resolution levels of DWs indicating DW_1 (2^1–month mode), DW_2 (2^2–month mode), DW_3 (2^3–month mode), DW_4 (2^4–month mode), DW_5 (2^5–month mode) and DW_6 (2^6–month mode) and one approximation (A) signal are respectively employed for the daily and monthly air temperatures in this study. The approximate signal indicates the trend (low frequency) of the original air temperature time series. The correlation coefficients were calculated between each DWs sub-time series and original air temperature time series and the effective DWs components are selected for the daily and monthly air temperatures. For the WGEP models, the new series obtained by adding the effective DWs components (DW7, DW8 and DW9 for daily air temperatures, and DW2, DW3, DW4, DW5 and DW6 for monthly air temperatures) and approximation component are used as inputs to the GEP model to forecast i-day/month ahead air temperatures.

APPLICATIONS

Performance evaluation

Three statistical evaluation criteria were used to assess the model performance:

(1) Coefficient of determination (R^2):

$$R^2 = \frac{\sum_{i=1}^{N}\left(T_{io}-\bar{T}_{io}\right)\left(T_{if}-\bar{T}_{if}\right)}{\sqrt{\sum_{i=1}^{N}\left(T_{io}-\bar{T}_{io}\right)^2\sum_{i=1}^{N}\left(T_{if}-\bar{T}_{if}\right)^2}} \tag{6}$$

(2) Root mean square error (RMSE):

$$\text{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(T_{io}-T_{if}\right)^2} \tag{7}$$

(3) Scatter index (SI):

$$SI = \frac{RMSE}{\bar{T}_{io}} \tag{8}$$

Where T_{io} is the value observed at the i^{th} time step, T_{if} is the corresponding forecasted value, N is number of time steps, $\bar{T}_{io}$ is the mean of observational values and $\bar{T}_{if}$ is the mean value of the simulations. According to literature [59], the coefficient of determination (R^2) standardizes for differences between observed and corresponded simulated means and variances. Large values of R^2 can be obtained even when the model accuracy is low. Therefore, it is appropriate to apply a dimensional parameter (such as RMSE) which describes the average differences between the recorded and predicted values of air temperature in units of the temperature. Also a dimensionless RMSE expressed as a percentage of mean observed value (here called as SI) provides a relative measure with respect to mean observed air temperature. The combined use of these criteria provides a sufficient evaluation of each model's performance.

RESULTS AND DISCUSSION

This paper aims at representation of air temperature prediction by GEP and hybrid Wavelet-GEP models. At the first step, several input combinations were tried using GEP to forecast air temperature in two stations. At the second step, GEP model was evaluated for air temperature forecasting using wavelet subseries. It is relevant to note that each of these steps includes two parts as follows: Part1 in which various combinations of daily air temperature values (i.e., T_i,…, T_{i-3}, where T_i denotes the air temperature value at time i) were applied as inputs for GEP and wavelet-GEP (WGEP) to predict daily temperature values at times i+1, i+2, i+3 and i+30; and Part2 in which monthly air temperature values were predicted using GEP and WGEP. The latter is given with detailed information in the next steps.

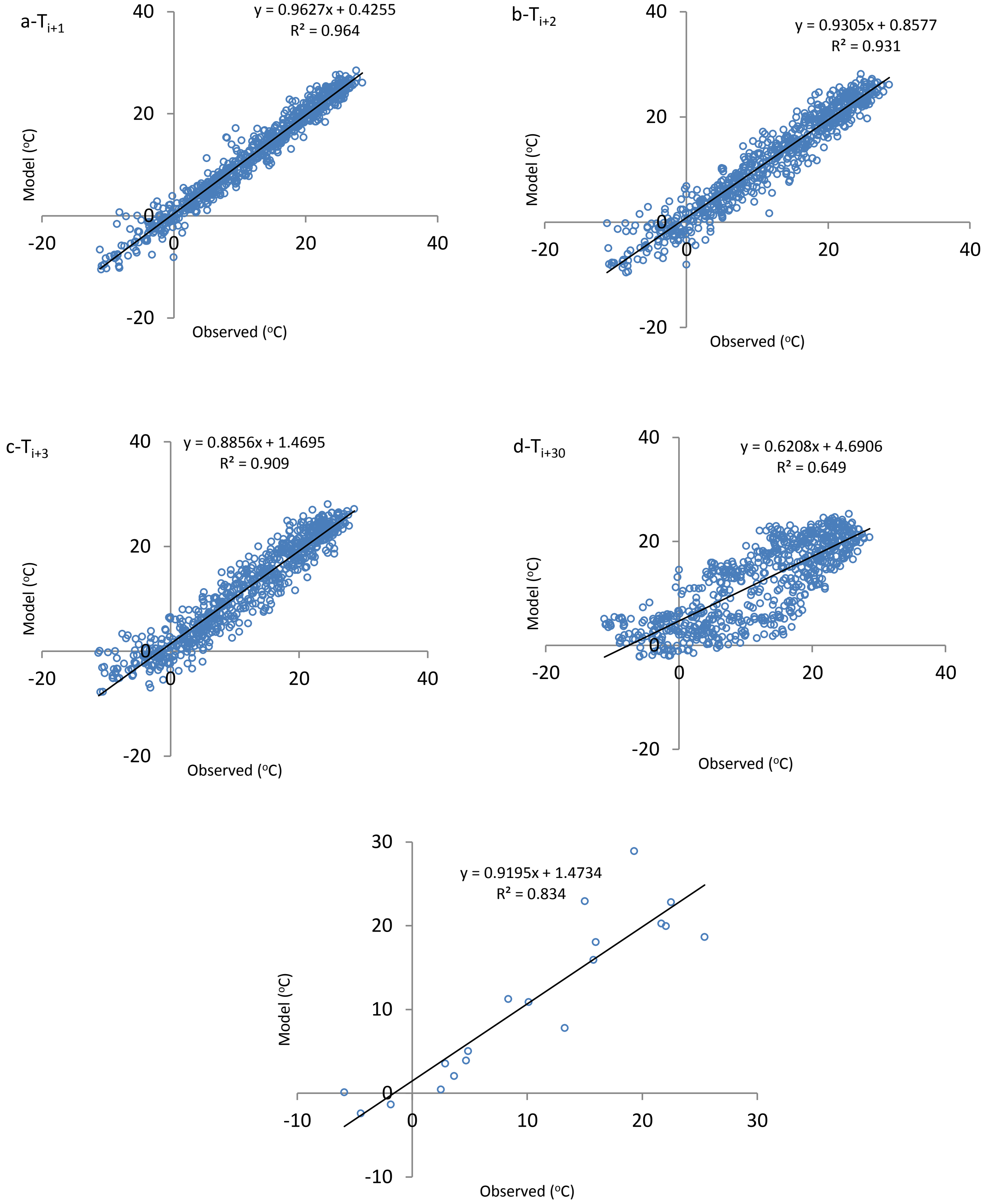
GEP models

The following combinations of input data of daily air temperature values were evaluated:

(i) T_i
(ii) T_{i-1}, T_i
(iii) T_{i-2}, T_{i-1}, T_i
(iv) T_{i-3}, T_{i-2}, T_{i-1}, T_i

Table 2 gives the coefficient of determination (R^2), root mean square error (RMSE) and scatter index (SI) of the GEP models for the both stations during the test period. From this table it can be seen that introducing the air temperature of the current day as well as temperature of the one, two and three previous days as model inputs, produces the best results for both Urmieh and Mahabad stations. However, the results for Urmieh Station seem to be better than those of Mahabad Station. A reason behind this may be the high skewness coefficient of Mahabad data (see Table1). The performance evaluation measures of the GEP models during the validation period are summarized in Table3. Similar to the testing period, this table shows that forecasting air temperature based on the current day air temperature values as well as one-, two- and three- previous days temperature values provided the best performance for every three prediction intervals at the both stations. Generally, the GEP model performances in Urmieh Station are better than Mahabad.

From the meteorological viewpoint, a model would be of more applicability and reliability if it could produce long-term (such as monthly) air temperature values. Therefore, a part of this study will discuss on the producing one-month ahead predictions of air temperature. In order to make such predictions, firstly the optimal daily GEP model was applied. In this way, the models whose inputs are the air temperature of the current day and one-, two-, and three previous days (T_{i-3}, T_{i-2}, T_{i-1}, T_i) was applied to forecast air temperature value at time i+30 (one-month ahead). Table 4 represents the performance evaluation measures of the monthly predictions models. It is noted that there are two sets of simulation results in this table as follows: first set of results produced by using daily air temperature values (e.g., T_i, T_{i-1},…) as input parameters to predict one-month ahead air temperature value as output (i.e., T_{i+30} is the model output); and the second set of the results which applied monthly air temperature values as input parameters to produce one-month ahead air temperature value (as output parameter). The parameters $MT_i,\ldots, MT_{i-3}$ represent the monthly air temperature values in the current month as well as one-, two- and three-previous months. In the latter application used the same input combination order as the optimal daily predictions model for sake of consistency and comparability of the obtained outcomes. It can be easily seen from this table that forecasting one-month ahead air temperature by using daily temperature values as model inputs gives better results in Urmieh Station during the test period which was not resulted for Mahabad. However, the validation results reveals that application of monthly temperature values as model inputs give much better results than the case in which daily values are introduced as input parameters. Figure1 displays the observed and forecasted air temperature values by the GEP model for Urmieh Station during the validation period.



Fgiure 1: Observed and forecasted air temperature values using GEP model for Urmieh station by using daily (a, b, c, d) and monthly (e) temperature values as input parameters

From the scatter plots of this figure it can be seen that increasing the prediction interval from 1-day to 3-day head interval, detracts from the GEP model accuracy both from the correlation viewpoint and data scattering. For monthly predictions it can be easily observed from this figure that the application of monthly air temperature values as GEP input parameters produces better results than those produced with application of daily temperature values as GEP inputs, with relatively high R^2 values and low scattering. The same conclusions as Urmieh Station can be resulted for Mahabad Station (Figure2).

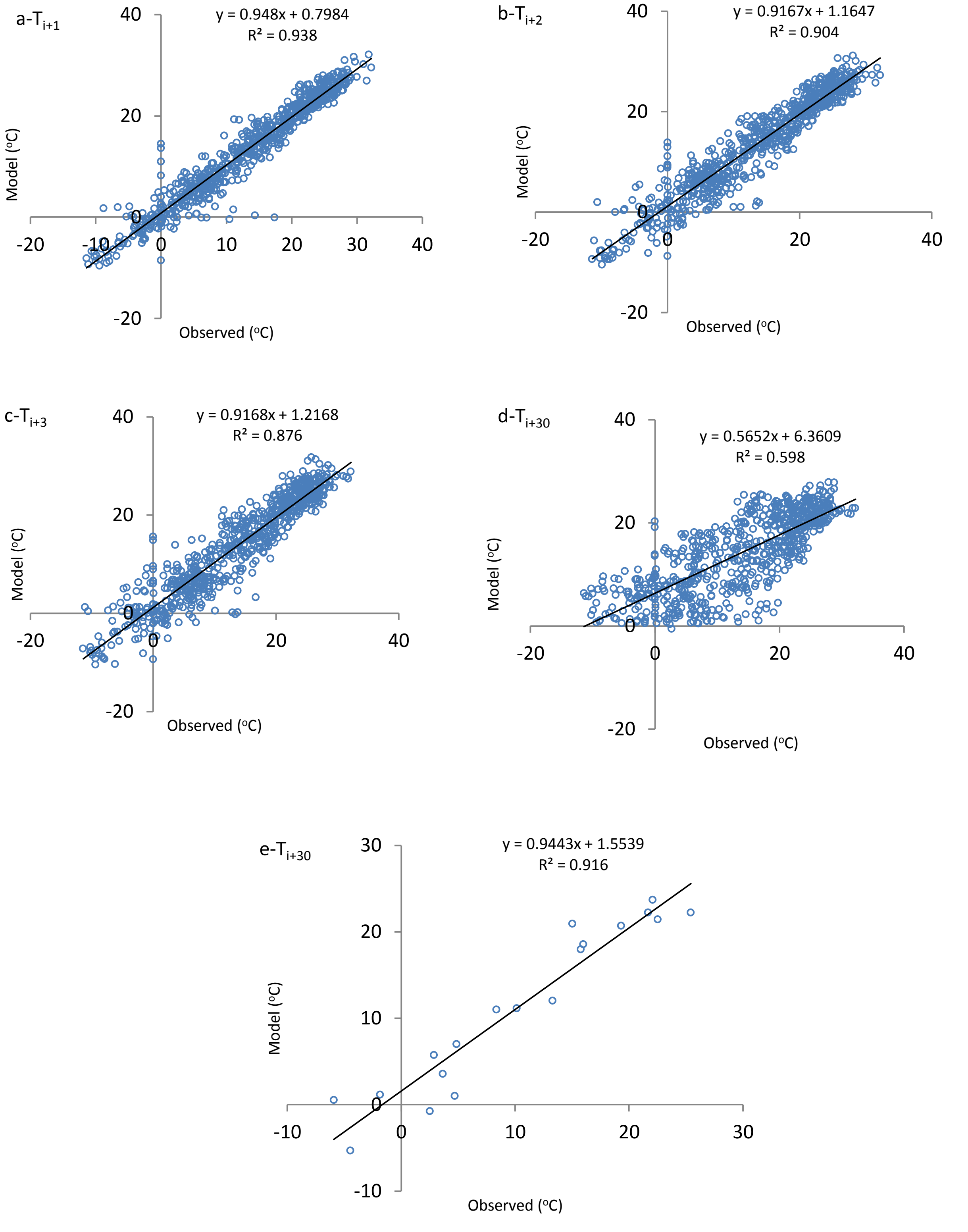


Figure 2: Observed and forecasted air temperature values using GEP model for Mahabad station by using daily (a, b, c, d) and monthly (e) temperature values as input parameters

To cite this paper: Kisi O, Shiri J and Nazemi AH. 2011. A wavelet-genetic programming model for predicting short-term and long-term air temperatures. *J. Civil Eng. Urban.* 1(1): 25-38.
Journal homepage: http://www.ojceu.ir/main

As mentioned earlier, one of the strong points of the GP (i.e., GEP) is that it can produce some mathematical expressions that rule the phenomena. Such mathematical expressions for predicting t+30 days air temperature values are given here. Based on the validation results of the single GEP models (Table4), application of monthly air temperature values gives better results for predicting one-month ahead air temperature values. Figure3 displays the GEP formulation of this model for Urmieh Station which is actually as follows:

exp(((G1C0^2)+(((G1C1-d(2))+G1C1)/exp(d(0)))))+d(2)+((((G3C1-G3C0)*d(2))/(d(3)+G3C0))-(d(0)-(G3C0+G3C1)))

where the actual parameters are

$d(0)=MT_{i-3}$, $d(2)=MT_{i-1}$, $d(3)=MT_i$;

and the constant coefficients are

G1C0 = -0.704529; G1C1 = -7.46106; G3C0 = 8.163604; G3C1 = 4.860138;

After putting the corresponding values in the general expression, the final equation becomes

$$MT_{i+30} = MT_{i-1} - MT_{i-3} - \frac{3.3MT_{i-1}}{MT_i + 8.16} + Exp\left\{\left[-\frac{MT_{i-1}+15.28}{ExpMT_{i-3}}\right] + 0.495\right\} - 13.02 \qquad (9)$$

It can be seen from the Eqn. (9) that the model is not much sensitive to MT_{i-2}.

For Mahabad Station the GEP formulation for the same input combination is:

$$MT_{i+30} = MT_i + \frac{1.82MT_{i-3}+MT_{i-2}-MT_i}{(MTi-2)^2+2.58} + 3.67\left[MT_{i-2} + MT_{i-3} + 3.25\right] + 5.67 \qquad (10)$$

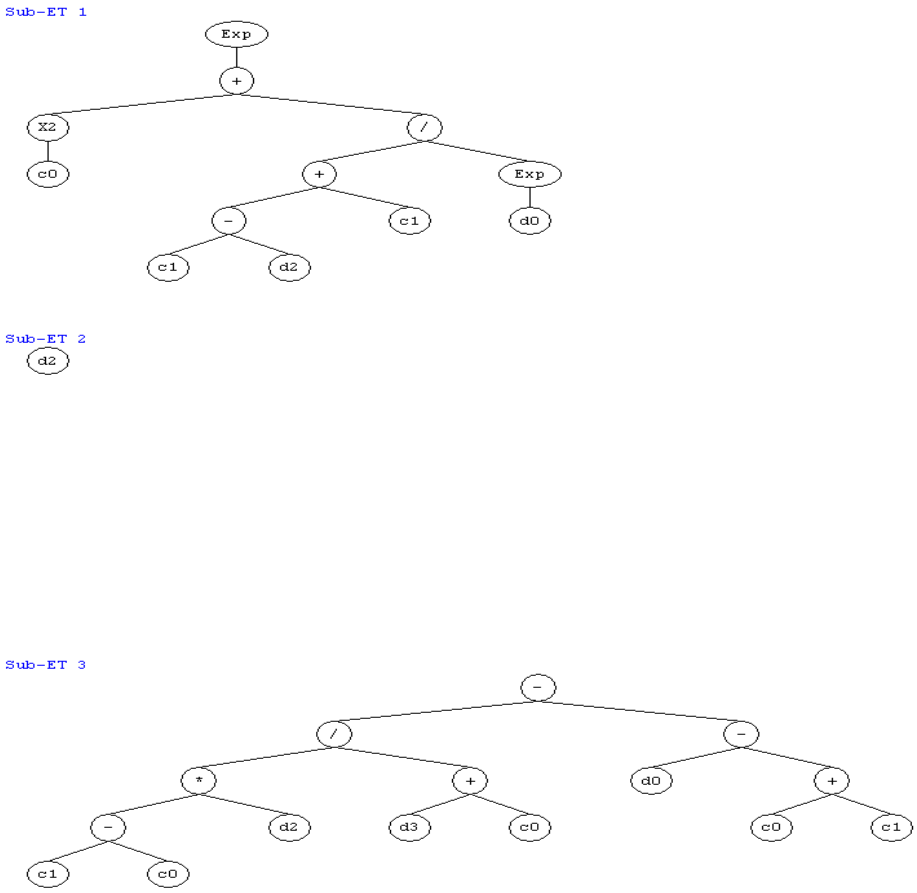


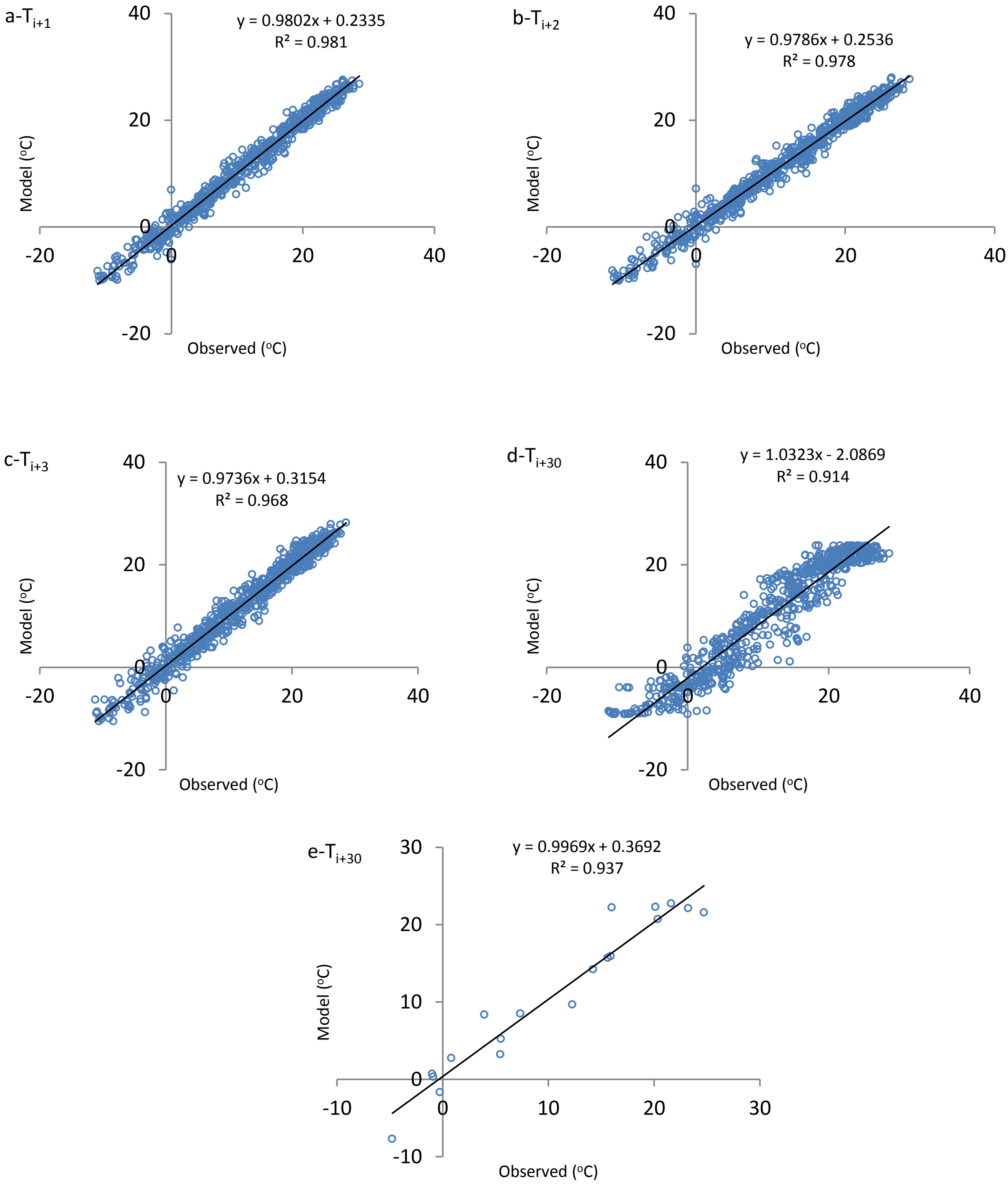
Figure 3. GEP expression for predicting *t+30* days air temperature

Wavelet-Genetic Programming (WGEP) models

In the hybrid wavelet-genetic programming (WGEP) models the discrete wavelet components were employed (instead of original time series data) as input parameters of GEP model for the air temperature prediction. Similar to the previous section, four different combinations of input data (the new series) were evaluated for forecasting as in the single GEP models. The error statistics of the WGEP models are represented in Table5. From this table it is clear that, similar to the single GEP models, the input combination (v) whose inputs are the air temperature value of the current day as well as the values of the one-, two- and three previous days, produces the best results for every three prediction intervals in both the stations. Comparison of the results of Tables2 and 5 reveals that application of hybrid wavelet-GEP model improves the predictions to some extent. The statistical performance of WGEP models during the validation period are presented in Table 6. Table shows that WGEP model has a significant positive effect on daily air temperature forecast. As seen from the table the WGEP model (iv) (with input combination iv) has the lowest RMSE and SI values as well as the highest R^2

To cite this paper: Kisi O, Shiri J and Nazemi AH. 2011. A wavelet-genetic programming model for predicting short-term and long-term air temperatures. *Online J. Civil Eng. Urban.* 1(1): 25-37.
Journal homepage: http://www.ojceu.ir/main

values among other WGEP models. Table7 illustrates the statistical performance measures for one-month ahead forecast of air temperature for wavelet-GEP model for both the stations. From this table it is clear that application of the wavelet coefficients of monthly air temperature values as GEP input parameters produces better results than the state in which the wavelet coefficients of daily temperature values are introduced as GEP input parameters. A comparison between the results presented in Table4 and 7 shows that the hybrid WGEP model improves the accuracy of monthly air temperature forecast to great extent. Figure4 gives the observed and predicted temperature values produced by wave-GEP model during the validation period in Urmieh Station.



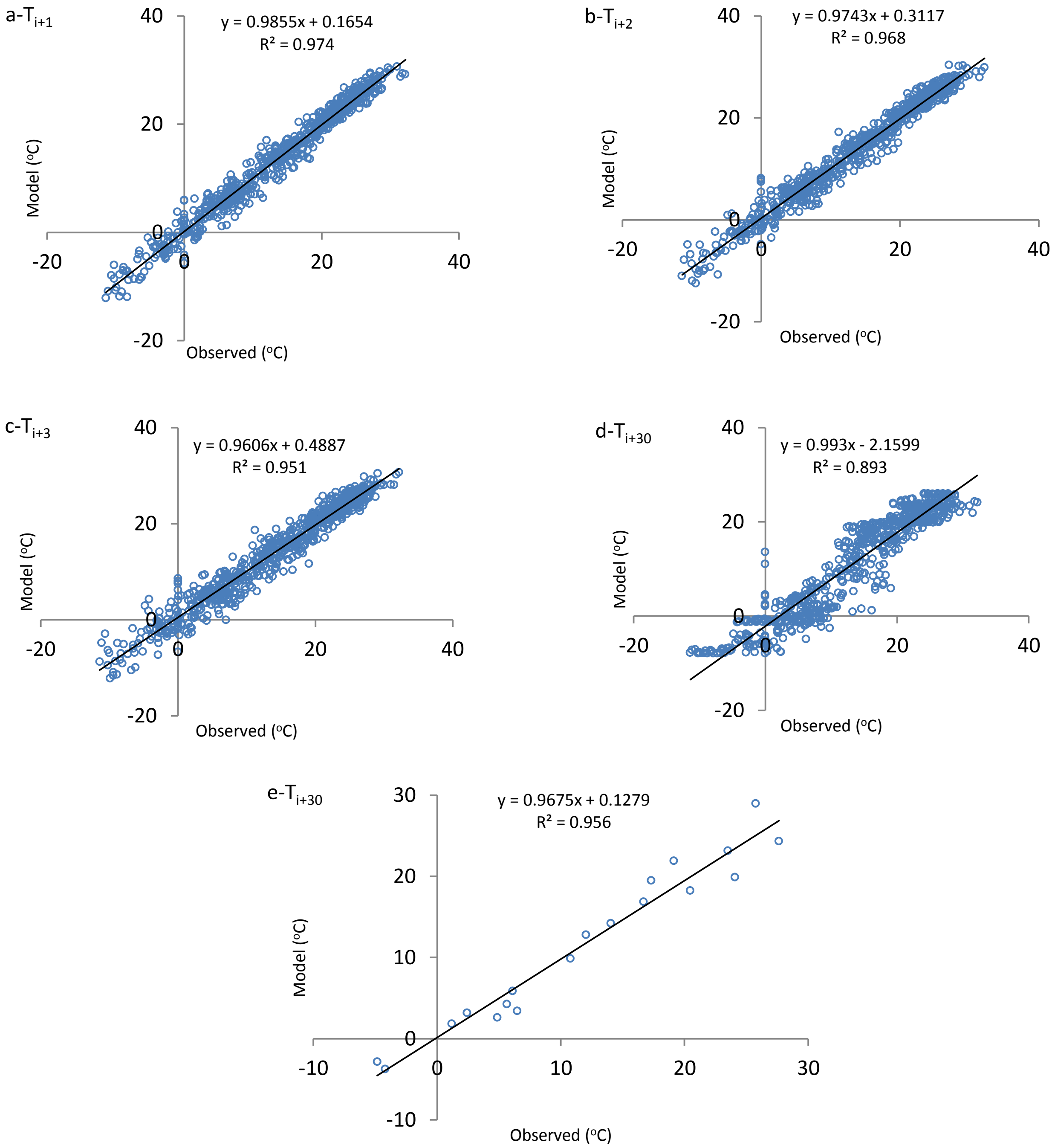
Fgiure 4: Observed and forecasted air temperature values using WGEP model for Urmieh station by using daily (a, b, c, d) and monthly (e) temperature values as input parameters

Similar to the single GEP model, increasing the prediction interval from 1- to 3- day ahead decreases from the models accuracy. Also application of wavelet coefficients of the monthly temperature values as input parameters gives much better results than WGEP models implemented with daily coefficients for monthly forecast. Comparison of Figures1 and 4 reveals that the hybrid WGEP model gives better results than GEP model with relatively high correlation and low scattering. From the straight fit line equation (assuming that the equation is y=ax+b) it can be observed that a, and b coefficients are closer to 1 and 0, respectively, for WGEP model, which exhibits the high generalization capacity of hybrid models. Also for monthly forecast these coefficients are respectively, closer to 1 and 0 for the state in which the monthly coefficients are applied as GEP inputs. Figure5 displays the observed and modeled air temperature values of Mahabad Station by using the WGEP model during the validation period. The same conclusion as Urmieh Station is resulted for Mahabad Station in the case of applying WGEP model. Therefore, the WGEP model seems to be more adequate than the GEP in predicting air temperature values. The GEP formulation for predicting monthly air temperature values (by using the monthly values as input variables) in Urmieh station is:

$$T_{i+30} = 0.792(-0.890WT_{i-1}) + 1.75WT_i - WT_{i-1} + Ln\sqrt[3]{WT_{i-3}} + \frac{WT_{i-3}^3}{Exp[WT_{i-3}]} \tag{11}$$

where the WT_i, $WT_{i-1}$and WT_{i-3} denote the wavelet coefficients of the monthly air temperature values at times i, i-1 and i-3, respectively. From Eq. (11) it is clear that, similar to the single GEP model, the WGEP model is not sensitive to the air temperature value at time i-2 in this case. The GEP mathematical expression for Mahabad Station is:

$$T_{i+30} = \frac{WT_i + 2.73}{WT_{i-1} + 0.265}WT_{i-1} + 0.293(WT_{i-1} + LnWT_{i-3}) + WT_i + 0.15WT_{i-2} - 0.2 \tag{12}$$



Fgiure 5. Observed and forecasted air temperature values using WGEP model for Mahabad station by using daily (a, b, c, d) and monthly (e) temperature values as input parameters

To cite this paper: Kisi O, Shiri J and Nazemi AH. 2011. A wavelet-genetic programming model for predicting short-term and long-term air temperatures. *Online J. Civil Eng. Urban.* 1(1): 25-37.
Journal homepage: http://www.ojceu.ir/main

Prediction of air temperature values of Mahabad Station using the data of Urmieh Station

Prediction of air temperature values using nearby station data is an important issue, since the data of some stations are sometimes missing. To solve such a problem regression techniques are usually applied. This part of the study is focused on the investigation of GEP and WGEP model performances to solve this problem. In this application the Urmieh Station air temperature data were used as inputs to the GEP and WGEP models to predict air temperature data of Mahabad Station at time t+30. In the first application (application I) the daily air temperature values and corresponded wavelet coefficients were applied as model inputs and in the second application (application II) the monthly air temperature values and corresponded wavelet coefficients were applied as inputs for GEP and WGEP models. Table8 represents the error statistics of the models for both applications during the test period. The WGEP model seems to be better than GEP for both the applications however, the performance of GEP in application II (by using the monthly data as the input parameters of the model) is much better than those resulted for application I. From the results of this table it can be clearly observed that the hybrid WGEP model can be applied for predicting air temperature data by using the data of another station (cross-application). A comparison between the results of WGEP model in both applications reveals that introducing the wavelet coefficients of monthly air temperature data (application II) gives better results than the application I. A review of the validation results of the both WGEP models dictate that the WGEP model for application II produces better results (with R^2 value of 0.933, RMSE value of 2.647 ^oC and SI value of 0.198) than those of application I (with R^2 value of 0.863, RMSE value of 5.111 ^oC and SI value of 0.383). From these results it can be concluded that the WGEP model is a successful approach for predicting air temperature data at one station by using the data of the other nearby station. The observed and predicted air temperature values of cross-application for daily and monthly data based models in validation period are shown in Figure6. It is clear from the scatter plots that the WGEP models perform much better than the single GEP models for the both applications.

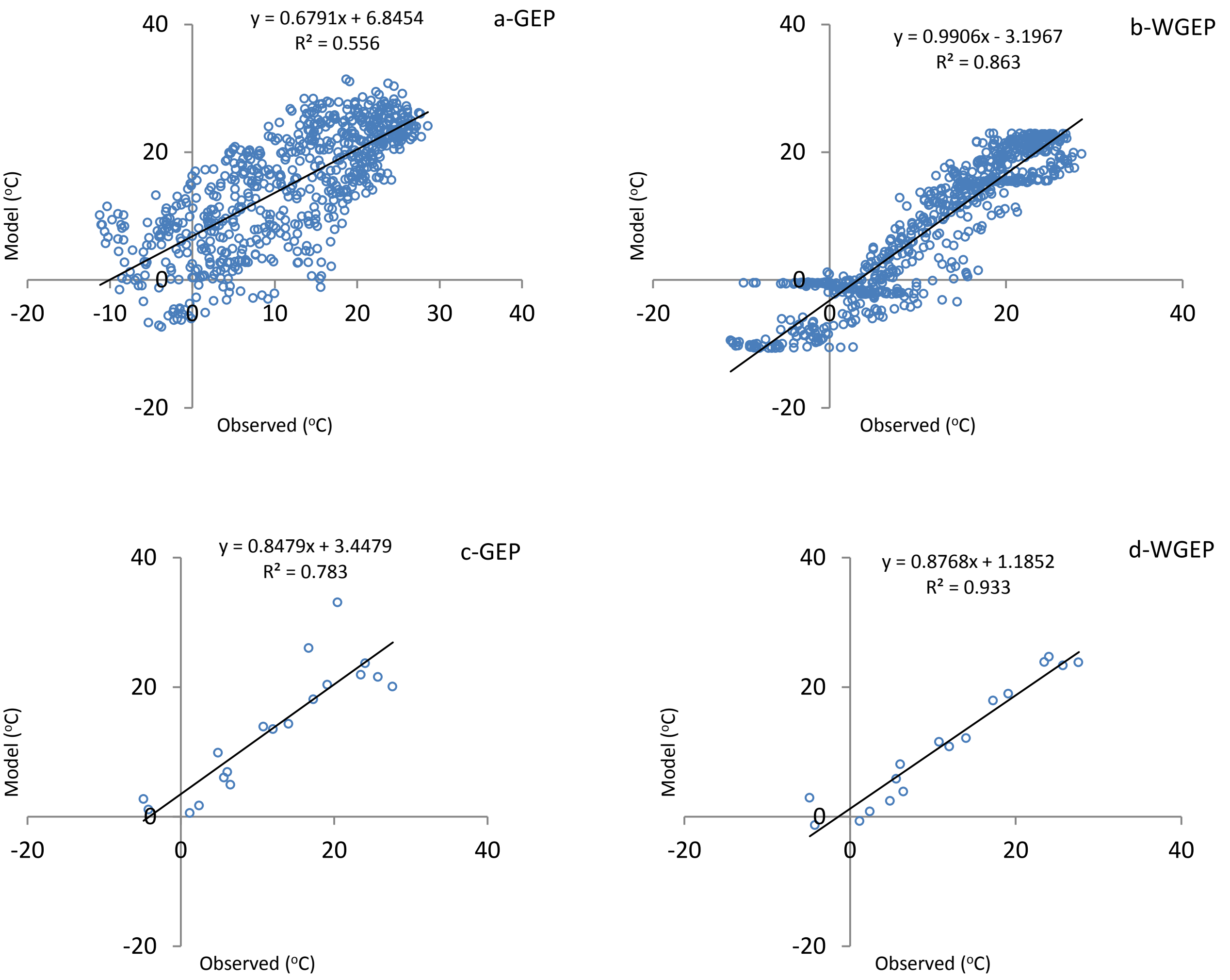


Figure 6. Observed and estimated air temperature values of cross-application for daily data based models (a, b) and monthly data based models (c, d).

CONCLUSIONS

The accuracy of wavelet-genetic programming conjunction model for modeling short- and long-term air temperatures was studied in this paper. In the first part of the study, the WGEP models were compared with those of the single GEP models using the previously recorded daily and monthly air temperature inputs. At the first step, the previous daily air temperature values and their wavelet coefficients were respectively used as inputs to the GEP and WGEP models to forecast one, two, three and thirty days ahead air temperatures. At the second step, the monthly previous air temperatures and their wavelet coefficients were respectively used as inputs to the GEP and WGEP models to forecast one month ahead air temperatures. The comparison results indicated that the WGEP model considerably increased the accuracy of single GEP model especially in forecasting long-term (thirty days and one month ahead) air temperatures. In the second part of the study, the WGEP models were compared with those of the single GEP models in estimating long-term (thirty days and one month ahead) air temperatures of Mahabad Station using the daily and monthly data of nearby Urmieh Station. Based on the comparison of these results, it was found that the WGEP model performed much better than the single GEP model in cross-station application.

REFERENCES

1. Hudson G, Wackernagel H. 1994. Mapping temperature using kriging with external drift: theory and example from Scotland. Int. J. Climatology, 14: 77-91.
2. Rubio A, Sanchez O, Gomez V, Grana D, Elena R, Blanco A. 2002. Auto-ecology of chestnut tree forest in Castilla, Spain. Investigation Agraria: sistemas de RecursosForestales, 11(2): 373-393.
3. Richardson A.D, Lee X, Friedland AJ. 2004. Microclimatology of treeline spruce-fir forests in mountains of the northeastern United States. Agric. Forest Meteorol., 125: 53-66.
4. Kalogirou, S.A., and Bojic, M. 1999. Artificial neural networks for the prediction of the energy consumption of a passive solar building. Energy, 25(5), 479-491.
5. George R.K. 2001. Prediction of soil temperature by using artificial neural networks algorithms. Non linear analysis, 47(3): 1737-1748.
6. Soarse J, Oliveria A.P, Boznar M.Z, Mlakar P, Escobedo J.F, Machado A.J. 2003. Modeling hourly diffuse solar radiation in the city of Sao Paulo using a neural network technique.Applied Energy, 79(2): 201-214.
7. Bartos I, Janosi IM. 2006. Nonlinear correlations of daily temperature records over land. Nonlinear Processes in Geophysics, 13: 571-576.
8. Kiraly A, Janosi I.M. 2002. Stochastic modeling of daily temperature fluctuations. Physical Review E, 65: 051102 (1-6).
9. Benth F.E, Saltyte-Benth J. 2004. Stochastic modeling of temperature variations with a view towards weather derivatives. Statistical Research Report, No.1, Depart. Of. Math., University of OSLO.28 pp.
10. Gyure B, Bartos I, Janosi I.M. 2007. Nonlinear statistics of daily temperature fluctuations reproduced in a laboratory experiment. Physical Review E, 76, 037301.
11. Alfaro E, Gershunov A, Cayan D, Steinemann A, Pierce D, Barnett T. 2004. A method for prediction of California summer air surface temperature. EOS Transactions, American Geophysical Union, 85(51): 553-564.
12. Guan B.T, Hsu H.W, Wey T.H, Tsao L.S. 2008. Modeling monthly mean temperatures for the mountain regions of Taiwan by generalized additive models. Agric. Forest Meteorol. Doi:10.1016/2008.08.10
13. Schoof J.T, Pryor S.C. 2001. Downscaling temperature and precipitation: a comparison of regression-based methods and artificial neural networks. International Journal of Climatology, 21(7): 773-790.
14. Chang F.J, Chen Y.C, Liang J.M. 2002. Fuzzy Clustering Neural Network as Flood Forecasting Model. Nordic Hydrology, 33(4): 275-290.
15. Kisi O. 2004a. River flow modeling using artificial neural networks. ASCE Journal of Hydrologic Engineering, 9(1): 60-63.
16. Kisi O. 2004b. Multi layer perceptions with Levenberg-marquardt training algorithm for suspended sediment concentration prediction and estimation. Hydrological Sciences Journal, 49(6): 1025-1040.
17. Kisi O. 2005. Suspended sediment estimation using neuro-fuzzy and neural network approaches. Hydrological Scinces Journal, 50(4): 683-696.
18. Kisi O. 2006a. Generalized regression neural networks for evapotranspiration modeling. Hydrological Scinces Journal, 51(6): 1092-1105.
19. Kisi O. 2006b. Evapotranspiration estimation using feed forward neural networks. Nordic hydrology, 37(3): 247-260.
20. Kisi O. 2007. Evapotranspiration modeling from climate data using a neural computing technique. Hydrological Processes, 21(6): 1925-1934.
21. Morid S, Smakhtin V, Bagherzadeh K. 2007. Drought forecasting using artificial neural networks and time series of drought indices. International Journal of Climatology, 27(15): 2103-2111.
22. Landeras G, Ortiz-Barredo A, Lopez J.J. 2008. Comparison of artificial neural network models and empirical and semi-empirical equations for daily reference evapotranspiration estimation in the Basque Country (Northern Spain). Agricultural Water Management, 95: 553-565.
23. Tatli H, Sen Z. 1999. A new fuzzy modeling approach for predicting the maximum daily temperature from a time series. Turkish Journal of Engineering and environmental science, 23: 173-180.

24. Abdel-Aal RE. 2004. Hourly temperature forecasting using abductive networks. Engineering Applications of Artificial Intelligences, 17: 543-556.
25. Smith B.A, McClendon R.W, Hoogenboom G. 2005. An enhanced artificial neural network for air temperature prediction. Proceedings of World Academy of Science, Engineering and Technology (PWASET), **7**: 7-12.
26. Shank D.B, Hoogenboom G, McClendon R.W. 2008. Dewpoint temperature prediction using artificial neural networks.Journal of applied meteorology and climatology, 47: 1757- 1769.
27. Koza J.R. 1992. Genetic Programming: On the programming of computers by means of Natural Selection. The MIT Press, Cambridge, MA 840 pp.
28. Goldberg D.E. 1989. Genetic algorithms in search, optimization, and machine learning. Addison – Wesley, Reading MA 432 pp.
29. Banzhaf W, Nordin P, Keller PE, Francone FD. 1998. Genetic Programming. Morgan Kaufmann. San Francisco, CA 512 pp.
30. Babovic V, Kanizares R, Jenson HR, Klinting A. 2001. Neural networks as routine for error updating of numerical models. ASCE Journal of Hydrologic Engineering, 127(3): 181-193.
31. Savic A.D, Walters A.G, Davidson J.W. 1999. A genetic programming approach to rainfall-runoff modeling. Water Resources Management, 13: 219-231.
32. Liong S.Y, Gautam T.R, Khu S.T, Babovic, V, Keijzer M, Muttil N. 2002. Genetic programming: A new paradigm in rainfall runoff modeling. Journal of American Water Resources Association, 38(3): 705-718.
33. Aytek A, Alp M. 2008. An application of artificial intelligence for rainfall runoff modeling. Journal of Earth Systems Science, 117(2): 145-155.
34. Aytek A, Kisi O. 2008. A genetic programming approach to suspended sediment modeling. Journal of Hydrology, 351: 288-298.
35. Baykasoglu A, Gullu H, Canakci H, Ozbakir L. 2008. Prediction of compressive and tensile strength of limestone via genetic programming. Expert Systems with Applications, 35: 111-123.
36. Ghorbani M.A, Khatibi R, Aytek A, Makarynskyy O, Shiri J. 2010. Sea water level forecasting using genetic programming and comparing the performance with artificial neural networks. Computers & Geosciences, 36: 620-627.
37. Kisi O, Shiri J. 2010. A comparison of genetic programming and ANFIS in forecasting daily, monthly and yearly stream flows. In: Proceedings of the International Symposium on Innovations in Intelligent Systems and Applications (INISTA). PP 118-122, 21-24 June 2010, Kayseri and Cappadocia, Turkey.
38. Kayadelen C. 2010. Soilliquefaction modeling by geneexpression modeling and neurofuzzy. Expert Systems with Application 2010. doi: 10.1016/j.eswa. 2010. 09. 071
39. Shiri J, Kisi O. 2011a. Comparison of genetic programming with neuro-fuzzy systems for predicting short-term water table depth fluctuations. Computers & Geosciences 2011. doi: 10.1016/j.cageo. 2010.11.010.
40. Shiri J, Kisi, O. 2011b. Application of artificial intelligence to estimate daily pan evaporation using available and estimated climatic data in the Khozestan Province (Southwestern Iran). ASCE Journal of Irrigation and Drainage Engineering 137(7): 412-425.
41. Smith L.C, Turcotte D.L, Isacks B. 1998. Stream flow characterization and feature detection using a discrete wavelet transform. Hydrological Processes 12: 233-249.
42. Lu R.Y. 2002. Decomposition of interdecadal and interannual components for North China rainfall in rainy season. Chinese Journal of Atmosphere (in Chinese), 26: 611-624.
43. Xingang D, Ping W, Jifan C. 2003. Multiscale characteristics of the rainy season rainfall and interdecadal decaying of summer monsoon in North China. Chinese Science Bulletin, 48: 2730-2734.
44. Coulibaly P, Burn H.D. 2004. Wavelet analysis of variability in annual Canadian stream flows. Water Resources Research, 40, W03105.
45. Labat D. 2005. Recent advances in wavelet analyses: Part 1. A review of concepts. Journal of Hydrology, 314(1-4): 275-288.
46. Labat D, Ronchail J, Guyot J.L. 2005. Recent advances in wavelet analyses: Part 2—Amazon, Parana, Orinoco and Congo discharges time scale variability. Journal of Hydrology, 314(1-4): 289-311.
47. Partal T, Küçük M. 2006. Long-term trend analysis using discrete wavelet components of annual precipitations measurements in Marmara region (Turkey). Physics and Chemistry of the Earth, 31: 1189-1200.
48. Partal T, Kisi O. 2007. Wavelet and neuro-fuzzy conjunction model for precipitation forecasting. Journal of Hydrology, 342: 199-212.
49. Zhou H.C, Peng Y, Liang G.H. 2008. The Research of Monthly Discharge Predictor-corrector Model Based on Wavelet Decomposition. Water Resources Management, 22(2): 217-227.
50. Shiri J, Kisi O. 2010. Short term and long term stream flow forecasting using a wavelet and neuro-fuzzy conjunction model. Journal of Hydrology 394: 486-493.
51. Kisi O, Shiri J. 2011. Precipitation forecasting using wavelet-genetic programming and wavelet-neuro-fuzzy conjunction models. Water Resources Management, DOI: 10.1007/s 1126396-011-9849-3.
52. Kisi O, Partal T. 2011. Wavelet and neuro-fuzzy conjunction model for stream flow forecasting. Hydrology Research, 42(6): 447-456.
53. Kisi O, Shiri J, Makarynskyy O. 2011. Wind speed prediction by using different wavelet conjunction models. International Journal of Ocean and Climate Systems (IJOCS), 2(3): 189-208.
54. Khatibi R. 2004. Treating uncertainty in flood forecasting – development, uptake and barriers. In: Proceedings 2nd ACTIF Workshop, Delft, Netherlands.

55. Ferreira C. 2001. Gene expression programming: a new adaptive algorithm for solving problems. Complex Systems, 13(2): 87-129.
56. Mallat S.G. 1989. A theory for multi resolution signal decomposition: the wavelet representation. IEEE Transactions on Pattern Analysis and Machine Intelligence, 11 (7): 674–693.
57. Wang W, Ding J. 2003. Wavelet network model and its application to the prediction of the hydrology. Nature and Science, 1(1): 67-71.
58. Kisi O. 2009. Neural networks and wavelet conjunction model for intermittent stream flow forecasting. ASCE Journal of Hydrologic Engineering, 14(8): 773-782.
59. Legates D.R, McCabe G.J. 1999. Evaluating the use of goodness-of-fit measures in hydrologic and hydro climatic model validation. Water Resources Research, 35(1): 233-241.