

Estimation of soil temperature using gene expression programming and artificial neural networks in a semiarid region

Javad Behmanesh¹ · Saeid Mehdizadeh¹

Received: 18 October 2015 / Accepted: 3 January 2017 / Published online: 11 January 2017
© Springer-Verlag Berlin Heidelberg 2017

Abstract Soil temperature (T_s) is one of the most important parameters which affect physical and chemical properties of soil. In the present study, two biologically inspired approaches for artificial intelligence including gene expression programming (GEP) and artificial neural networks (ANN), as well as multiple linear regression (MLR) were used to estimate the soil temperature at six different depths (5, 10, 20, 30, 50 and 100 cm) for the Sanandaj synoptic station in a semiarid region in western Iran. Twelve combinations of meteorological parameters, such as minimum and maximum air temperatures, relative humidity, wind speed, sunshine hours and extraterrestrial radiation, were used as input variables. The full data set containing soil temperature and atmospheric parameters, which spans the time period from 1997 to 2008, was divided into training (1997–2004) and testing (2005–2008) data sets. To evaluate the accuracy of the models, determination coefficient (R^2) and root mean square error (RMSE) were calculated. The results showed that the GEP, ANN and MLR were able to model T_s at different depths. However, the performance of the ANN approach was the best.

Keywords Artificial neural networks · Gene expression programming · Soil temperature

List of symbols

T_{\min}	Daily minimum air temperature (°C)
T_{\max}	Daily maximum air temperature (°C)
RH	Daily relative humidity (%)
U_2	Daily wind speed at 2 m height (m s^{-1})
n	Daily sunshine hours (hr)
R_a	Daily extraterrestrial radiation ($\text{MJ m}^{-2} \text{day}^{-1}$)
T_s	Daily soil temperature (°C)
w	Weight of each neuron
x	Input to each neuron
s	Summation of inputs multiplication in corresponding weights
f	Transfer function
E	Error function of network in the back-propagation algorithm
η	Learning constant in the back-propagation algorithm
y_j	The output value of j th neuron
R^2	Determination coefficient
RMSE	Root mean square error (°C)
P_i	i th estimated T_s (°C)
O_i	i th observed T_s (°C)
P_{av}	Average of the estimated T_s values (°C)
O_{av}	Average of the observed T_s values (°C)
N	Number of observations
Y	Dependent variable in MLR (i.e., T_s)
X_1, \dots, X_n	Independent variables in MLR
a_0, \dots, a_n	Constant coefficients of MLR approach

✉ Javad Behmanesh
j.behmanesh@urmia.ac.ir

Saeid Mehdizadeh
saeid.mehdizadeh@gmail.com

¹ Water Engineering Department, Urmia University, Urmia, Iran

Introduction

Soil temperature (T_s) and its spatial and temporal variations affect physical and chemical processes in soil. Soil temperature is an important meteorological parameter for transferring heat energy from the atmosphere to soil and vice versa, solar energy applications such as the passive heating and cooling of buildings, frost depth prediction (construction depth of drainage structures and urban water supply networks), agricultural applications (crop growth, root development and potential evapotranspiration), hydrology (effect on climate change), geology, agronomy (determining suitable depth and time for seeds planting) and environmental studies (Mihalakakou 2002; Kocak et al. 2004; Jackson et al. 2008; Yilmaz et al. 2009). Soil temperature depends on a variety of environmental factors, including meteorological conditions such as surface global radiation and air temperature; soil physical parameters such as surface albedo, water content and texture; topographical variables such as elevation, slope and aspect; and other surface characteristics such as leaf area index and ground litter stores (Kang et al. 2000; Garcia-Suarez and Butler 2006; Paul et al. 2004).

Soil temperature is measured by using sensors. For example, THERM200 is a soil temperature probe, which has a temperature span from -40 to 85 °C. It outputs a voltage proportional to the temperature, so no complex equations are required to calculate the temperature from voltage. It is highly accurate with 0.125 °C of resolution. Empirical equations and models can be suitable to estimate T_s because of high costs of direct measurements. Soil temperature can be estimated by two different approaches based upon: (1) soil heat flow and energy balance and (2) empirical correlations with easily acquirable variables (Kang et al. 2000). The former approach can give accurate predictions for a well-evaluated site. However, in many sites, there is not sufficient data for calculating heat transfer equations such as surface global radiation (Kang et al. 2000). In addition, several studies have been conducted to estimate soil temperature by using various models, such as analytical, semi-analytical, numerical and experimental models (Hanks et al. 1971; Ghuman and Lal 1982; Paul et al. 2004; Prangnell and McGowan 2009; Droulia et al. 2009). Ghuman and Lal (1982) predicted soil temperature by Fourier analysis in a tropical area. The results demonstrated the high accuracy of the Fourier analysis. Usowicz and Walczak (1994) presented a mathematical model of heat flow to predict soil temperature. The results showed that the estimated values of soil temperature had acceptable accuracy. Droulia et al. (2009) estimated subsurface ground temperature profiles of a bare soil. They used an experimental plot located in the Agricultural University of

Athens campus and analytical and semiempirical models. It was concluded that the proposed models may serve as useful tools for predicting T_s .

Recently, artificial intelligence techniques have been increasingly used to estimate meteorological and environmental parameters such as soil temperature. Artificial intelligence methods can be used as alternative techniques. These methods do not require the internal relationship between variables of any investigated system. Simple solutions for multivariable problems and factual calculation (for variables with nonlinear relationships) are other advantages of artificial intelligence methods (Zadeh 1992; Chaturvedi 2008; Huang et al. 2010). Gene expression programming (GEP) and artificial neural networks (ANN) are two examples of biologically inspired approaches to artificial intelligence.

Multiple linear regression (MLR) is one of the most simple methods to model variables which have a linear relationship. Nevertheless, the GEP and ANN are biologically inspired methods which have the ability to model variables with nonlinear relationship. The relationships between soil temperature and meteorological parameters are generally nonlinear (Jungqvist et al. 2014).

The GEP was used by many researchers in a wide range of sciences. For example, rainfall–runoff modeling (Aytek et al. 2008), modeling of daily river discharge (Güven 2009), modeling of groundwater table fluctuations (Shiri et al. 2013), estimation of potential evapotranspiration (i.e., is a measure of the ability of the atmosphere to remove water from the surface through the processes of evaporation and transpiration assuming no control on water supply) (Traore and Güven 2012; Shiri et al. 2014a), solar radiation modeling (Landeras et al. 2012; Mehdizadeh et al. 2016), estimation of dew point temperature (Shiri et al. 2014b), function finding in component thermodynamical selection (Guo et al. 2014) and estimating the peak flood (Zorn and Shamseldin 2015) were accomplished by using GEP. To our knowledge, no study has been yet conducted about soil temperature estimation using GEP. However, the ANN was extensively used to predict T_s in many studies. George (2001) predicted T_s in Gujarat, India, by using ANN with relative humidity, wind speed and air temperature serving as input parameters. The results showed that the ANN was able to estimate T_s . Mihalakakou (2002) modeled daily and annual variations of soil surface temperatures in Athens and Dublin by a deterministic equation (based on the transient heat conduction differential equation and using as boundary condition the energy balance equation at the ground surface) and neural network approaches. It was found that the proposed neural network had the capability to estimate the soil surface temperature distribution. Bilgili (2010) estimated the monthly T_s using linear and nonlinear

regression and artificial neural networks in Adana, Turkey. He concluded that the ANN showed a better performance than both regression methods. Tabari et al. (2011) compared the ANN and MLR methods to estimate T_s in Isfahan, an arid region of Iran. The results showed that the ANN predictions were superior to the MLR. Bilgili et al. (2013) estimated T_s by multi-nonlinear regression and ANN at eight stations in Turkey. They concluded that the ANN model provides a simple and accurate prediction of T_s . Also, the ANN had a better performance than the nonlinear regression method. Tabari et al. (2014) predicted short-term soil temperature using ANN for two weather stations located in humid (Sari) and arid (Zahedan) regions of Iran. They concluded that the ANN can be successfully applied to provide accurate and reliable short-term soil temperature forecasts. Kisi et al. (2015) modeled monthly T_s in Mersin, Turkey, by using three different neural techniques which are multilayer perceptron, radial basis neural networks and generalized regression neural networks. Radial basis neural networks were found to be better than the generalized regression neural networks and multilayer perceptron in estimating monthly T_s at 5 and 10 cm depths, while the multiple linear regression and generalized regression neural networks gave the best accuracy at 50 and 100 cm depths, respectively. Further studies have been conducted to estimate soil temperature by Kisi et al. (2016), Hosseinzadeh Talaei (2014), Kim and Singh (2014), Shaker et al. (2014), Napagoda and Tilakaratne (2012).

As mentioned, accurate estimation of soil temperature is one of the most important problems in agricultural and environmental fields. Since strawberry is planted in Sanandaj and the time of planting this crop is a function of soil temperature, the investigation of soil temperature and its prediction have an important role in obtaining considerable yield. Literature review revealed that the

GEP has not been used to estimate T_s . Furthermore, the GEP can give an algebraic equation which can be easily used in the future. Moreover, the ANN is commonly used in modeling of environmental variables with the nature of nonlinear dynamic. Beside the GEP and ANN models, the MLR method was employed to predict T_s . Twelve combinations of meteorological parameters, such as minimum air temperature, maximum air temperature, extraterrestrial radiation, relative humidity, wind speed and sunshine hours, were used as inputs in GEP, ANN and MLR.

Materials and methods

Study area and meteorological data

The considered site is Sanandaj which is located at latitude $35^\circ 20'N$ and longitude $47^\circ 00'E$ in western Iran (Fig. 1). Sanandaj has an area of 2906 km^2 and is located 1373.4 m above free water level. According to the de Martonne index (1925), Sanandaj is located in a semiarid region. The meteorological conditions in Sanandaj during the study period (1997–2008) are summarized in Table 1.

The data for the present study were included minimum and maximum air temperatures (T_{\min} , T_{\max}), relative humidity (RH), wind speed (U_2), sunshine hours (n) and soil temperature (T_s) at six different depths (5, 10, 20, 30, 50 and 100 cm). The used data were collected from the Islamic Republic of Iran Meteorological Organization (IRIMO) for the period 1997–2008. It should be noted that the average measured soil temperature in three hours (03:00, 09:00 and 15:00) was considered as mean daily T_s . Other meteorological variables are measured at 00:00, 03:00, 06:00, 09:00, 12:00, 15:00, 18:00 and 21:00. The data between 1997 and 2004 as well as from 2005 to 2008 were used in training and testing stages, respectively.

Fig. 1 Location of studied area in the west of Iran

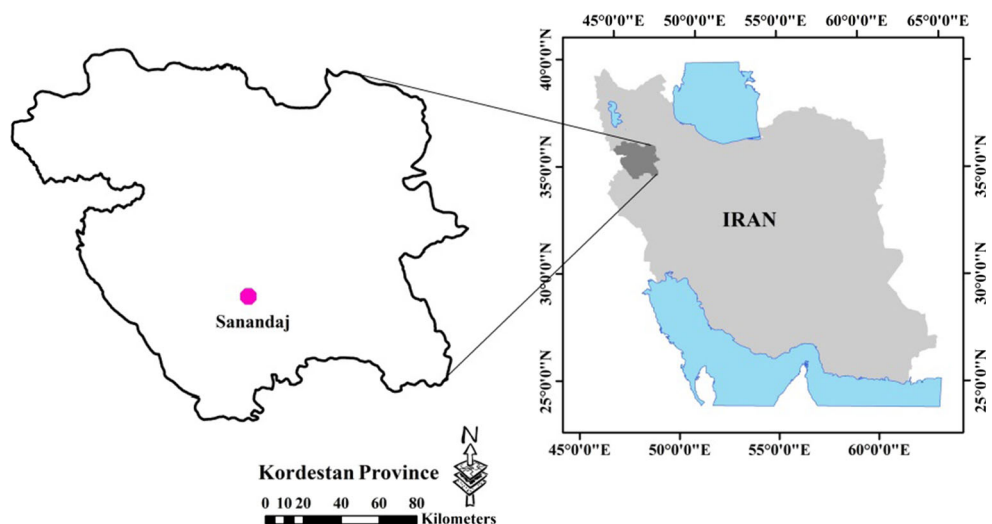


Table 1 Statistical characteristics of daily meteorological parameters for Sanandaj during 1997–2008

Parameter	X_{\min}	X_{\max}	X_{mean}	X_{stdev}
Minimum air temperature, T_{\min} (°C)	−22.4	28.2	6.4	8.3
Maximum air temperature, T_{\max} (°C)	−8.2	42.2	22.5	11.5
Extraterrestrial radiation, R_a ($\text{MJ m}^{-2} \text{day}^{-1}$)	16.4	41.7	29.9	9.0
Relative humidity, RH (%)	13.0	96.0	48.3	17.3
Wind speed at 2 m height, U_2 (m s^{-1})	0.0	8.2	1.8	1.2
Sunshine hours, n (hrs)	0.0	13.2	8.3	3.7

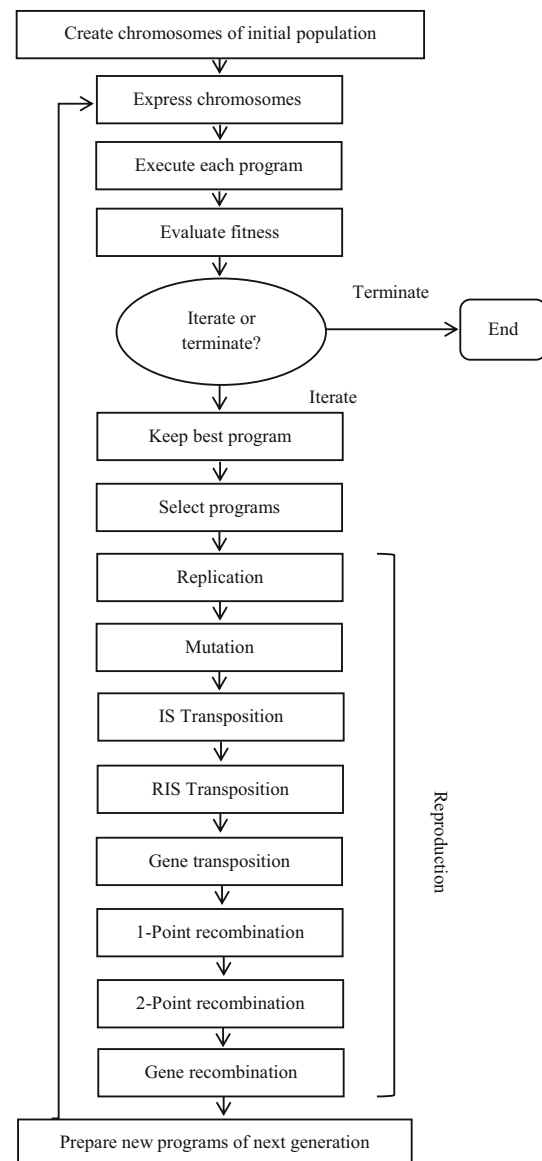
Gene expression programming (GEP)

GEP was presented by Ferreira (2001). GEP is one of the several machine learning techniques which are based on the concept of Darwinian evolution. GEP is a member of a broad family of techniques called evolutionary algorithms. All these techniques are based on the Darwinian principle of reproduction and survival of the fitness. In fact, in this method, the best population is selected. Otherwise, the new population is reproduced to obtain the best population. One important point that distinguishes the GEP from many other artificial intelligence techniques is the representation of the solutions. GEP returns algebraic equations. The most important advantages of the GEP to other intelligent models are: (1) The chromosomes are simple entities: linear, compact, relatively small, easy to manipulate genetically (replicate, mutate, recombine, transpose, etc.) and (2) the expression trees are exclusively the expression of their respective chromosome (Ferreira 2001).

The flowchart of a GEP algorithm is shown in Fig. 2. The generation of an initial population is the first stage in GEP. This can be done randomly or incorporate prior knowledge on the problem to be solved. Then, the chromosomes are expressed in the form of an expression tree. The results are evaluated using a fitness function to determine the suitability of a solution. If a solution of satisfactory quality is found, the evolution process will be stopped and the best obtained solution to this stage will be reported. If the stop condition is not satisfied, the best solution from the present generation will be copied to the next generation. The chromosomes are selected to reproduce with modification. During reproduction, it is the chromosomes of the individuals, not the expression trees, which are reproduced with modification and transmitted to the new generation. The process is repeated for a certain number of generations or until a solution has been found.

In this study, GeneXpro Tools 4.0 program was employed to develop planning models based on gene expression programming. To apply GEP to a given problem (e.g., T_s), the user needs to provide the following,

1. In the first step, a terminal set is selected which includes the independent variables of an investigated

**Fig. 2** Flowchart of a GEP algorithm (Ferreira 2001)

phenomenon. In this study, the input variables were different combinations of meteorological variables, e.g., minimum and maximum air temperatures, relative humidity, wind speed at 2 m height, sunshine hours and extraterrestrial radiation (Table 2).

Table 2 Different GEP and ANN scenarios and the respective input variables

Scenario	Input combinations
GEP1, ANN1	T_{min}, T_{max}
GEP2, ANN2	T_{min}, T_{max}, RH
GEP3, ANN3	T_{min}, T_{max}, n
GEP4, ANN4	T_{min}, T_{max}, U_2
GEP5, ANN5	T_{min}, T_{max}, R_a
GEP6, ANN6	$T_{min}, T_{max}, R_a, RH$
GEP7, ANN7	$T_{min}, T_{max}, R_a, U_2$
GEP8, ANN8	T_{min}, T_{max}, R_a, n
GEP9, ANN9	$T_{min}, T_{max}, R_a, RH, U_2$
GEP10, ANN10	$T_{min}, T_{max}, R_a, RH, n$
GEP11, ANN11	$T_{min}, T_{max}, R_a, U_2, n$
GEP12, ANN12	$T_{min}, T_{max}, R_a, RH, U_2, n$

Table 3 Setup of the genetic operators in GEP

Mutation rate	0.044	Two-point recombination	0.3
Inversion rate	0.1	Gene recombination rate	0.1
One-point recombination	0.3	Gene transposition rate	0.1

- The second stage is to select the function set. In the present study, the function set includes the four basic arithmetic operators $\{+, -, \times, \div\}$ and a set of algebraic and transcendental functions $\{x^2, x^3, \sqrt{x}, \sqrt[3]{x}, Lnx, e^x, Sinx, Cosx, Atanx\}$.
- The third step is to choose an index for evaluating model's accuracy. In this research, the root mean square error (RMSE) was used as fitness function.
- In the fourth step, the structure and architecture of the chromosomes are selected. In this process, head size (head contains special functions that activates function set) and the number of genes (these genes code for expression trees of different sizes and shapes) and chromosomes (composed of one or more genes of equal size) were selected 8, 3 and 30, respectively. Also, addition function was used as a linking function between expression trees. Finally, the genetic operators (to modify chromosomes) and corresponding rates were selected (Table 3).
- In fifth step, a stop condition must be defined. In this study, the generation number equal to 1000 was employed as stop criterion.

It should be noted that the GEP is not a deterministic model and it is likely to achieve different results from repeating GEP runs. In the present study, all set of scenarios (see Table 2) were used to run the GEP. Then, the best runs for each scenario were selected.

Artificial neural networks (ANN)

Artificial neural networks are an artificial intelligence technique inspired by biologically neural networks. The ANN establishes a model between a set of input and output variables. Artificial neural networks typically consist of three layers: input layer, hidden layer and output layer (Fig. 3a). Each layer consists of one or several nodes (neurons). The input layer contains the input variables. The output layer contains the output variable(s) of the ANN. The hidden layer(s) processes the information between input and output layers.

The structure of a neuron is shown in Fig. 3b. The input to a node is given by a vector $x = (x_1, x_2, \dots, x_n)$ with n components. The inputs are multiplied with corresponding weights $w = (w_1, w_2, \dots, w_n)$ and added up to calculate transfer function variable, s ,

$$s = \sum_{n=1}^n w_n \cdot x_n = w^T \cdot x \tag{1}$$

These weights are optimized by the model. To obtain the output of the respective neuron, s is transferred using a nonlinear transfer function f ,

$$y = f(s) \tag{2}$$

The nonlinear transfer function usually is defined as a sigmoid function,

$$f(s) = \frac{1}{(1 + e^{-s})} \tag{3}$$

Further detailed information about the ANN can be found in Haykin (1998).

In this study, a three-layer feed-forward neural networks (Fig. 3a) with a back-propagation learning algorithm was employed to estimate soil temperature. To develop the ANN models, the ANN toolbox in MATLAB R2014a was used.

In the present study, the input layer contains several nodes defined by one of the various combinations of meteorological parameters (Table 2). The output layer contains one node representing the soil temperature at a certain depth. The number of neurons in the hidden layer was optimized by trial and error procedure.

Back-propagation algorithm

More than 70% of the studies implementing ANN for environmental and hydrological applications employed the back-propagation learning algorithm because of its simplicity and robustness (Kumar et al. 2011). The back-propagation algorithm is divided into two stages: forward and backward stages. In the forward stage, all samples

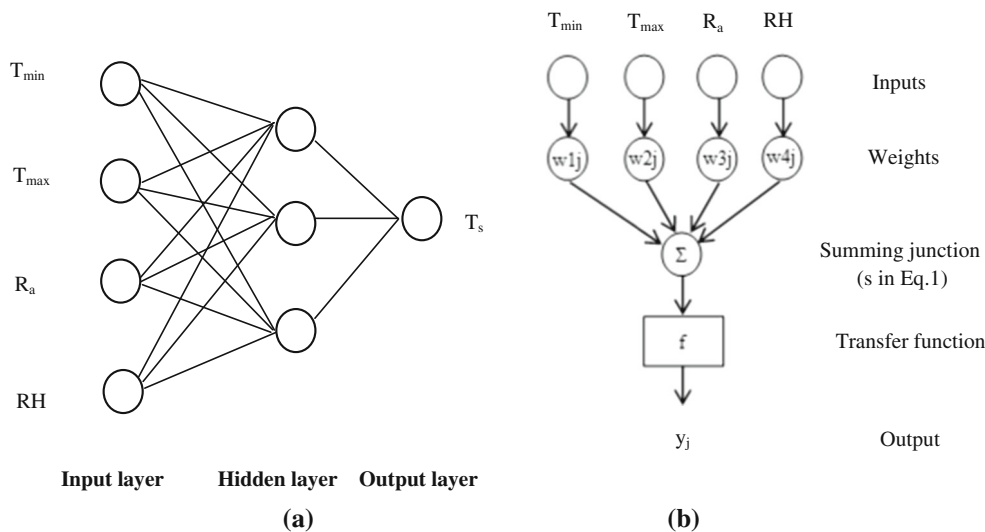


Fig. 3 Architecture of a three-layer feed-forward neural network with four input nodes (ANN6) (a) and structure of a neuron (b)

enter the network and the neurons are continuously updated from input to output layer. In this stage, all inputs into a neuron are calculated and each input is multiplied to its weight. The weights are randomly applied by the ANN, and the optimum values are obtained. The input value of j th neuron is calculated as,

$$s_j = \sum_i x_i w_{ji} \tag{4}$$

where x_i is the i th neuron value and w_{ji} is the i th neuron connection weight to j th neuron. Then, s_j is converted to the neuron output by an activation function (f),

$$f(s_j) = y_j. \tag{5}$$

The tangent–sigmoid function is often used as an activation function. In the forward stage, weights are not adjusted and remain constant. The outputs obtained are compared with the actual values, and the error function of network is calculated. In the backward stage, the errors (Eq. 6) are used to update the weights from output to input layer. This stage aims to minimize the errors between network output and reference data. The forward and backward stages are repeated several times. The iteration process is stopped until minimizing the error function (Eq. 6) is obtained. A popular method for this is the steepest descent method. In this method, the weights are adjusted as follows (Eqs. 6 and 7),

$$E(w) = \frac{1}{2} \sum_{j,k} (y_{j,k} - o_{j,k})^2 \tag{6}$$

where w is the weight, $y_{j,k}$ is the output value of j th neuron in the last layer which is obtained from k th learning sample and $o_{j,k}$ is the actual value of j th neuron in k th learning sample. There are different methods to minimize the value of $E(w)$.

$$w_{j,i}^{n+1} = w_{j,i}^n - \eta \frac{\partial E}{\partial w_{j,i}^n} \tag{7}$$

where $w_{j,i}^n$ is the weight of i th neuron to j th neuron at time n , η is the learning constant (between 0 and 1) which is optimized by model and $\frac{\partial E}{\partial w_{j,i}^n}$ is the derivation at $w_{j,i}^n$.

Further detailed information about the back-propagation algorithm can be found in Rojas (1996).

Multiple linear regression (MLR)

Besides applied artificial intelligence models, i.e., GEP and ANN, the MLR method was employed to estimate T_s . The general form of the MLR approach is as follows:

$$Y = a_0 + a_1 X_1 + a_2 X_2 + \dots + a_n X_n \tag{8}$$

where a_0, \dots, a_n are constants which are obtained by linear regression between dependent variable (Y , i.e., T_s) and independent variables used in the present study (X_1, \dots, X_n , i.e., different combinations of predictors).

Evaluation criteria

To evaluate the performance of GEP, ANN and MLR for soil temperature prediction, two statistical indices, the determination coefficient (R^2) and the root mean square error (RMSE) were used (Eqs. 9 and 10):

$$R^2 = \left(\frac{\sum_{i=1}^N (P_i - \bar{P}_i)(O_i - \bar{O}_i)}{\sqrt{\sum_{i=1}^N (P_i - \bar{P}_i)^2 \sum_{i=1}^N (O_i - \bar{O}_i)^2}} \right)^2 \tag{9}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^N (P_i - O_i)^2}{N}} \tag{10}$$

where P_i is the i th estimated T_s using GEP, ANN and MLR approaches; O_i is the i th observed T_s ; P_{av} is the average of the estimated T_s values (the average of the P_i); O_{av} is the average of the observed T_s values (the average of the O_i); and N is the number of observations.

Results and discussion

The variations of the annual mean daily soil temperature at different soil depths are presented in Fig. 4. In the studied period, maximum and minimum soil temperatures occurred at 5 and 20 cm depths, respectively. The maximum and minimum soil temperatures at 5 and 20 cm depths were 19.9 °C (in 2008) and 16.6 °C (in 1997), respectively.

The statistical characteristics of daily soil temperature at various depths in the studied period are shown in Table 4. Maximum and minimum soil temperatures were observed at 5 cm depth (equal to 44.4 °C and -3.9 °C, respectively). From surface to the deeper depths, minimum soil temperature increased and maximum soil temperature decreased. The standard deviation of the soil temperature is highest in the surface layer. This is most likely caused the meteorological conditions which have a stronger effect on the surface layer than on deeper soil layers.

Results of the GEP models

To develop the GEP models, various function sets were tested. The results showed that a function set consisting of basic arithmetic operators $\{+, -, \times, \div\}$, as well as algebraic and transcendental functions $\{\sin x, \cos x, A \tan x, x^2, x^3, \sqrt{x}, \sqrt[3]{x}, \ln x, e^x\}$, showed the best performance compared to all other tested function sets. The values of R^2 and RMSE for the GEP in all twelve scenarios are presented in Table 5. The values of RMSE range from 2.027 °C at the depth of 10 cm in GEP5 to 4.812 °C at 100 cm depth (in GEP12). Moreover, R^2 varies from 0.686 (100 cm) to 0.974 (10 cm) in GEP12 and GEP5, respectively. The performance of GEP1 is acceptable, especially for 10, 20, 30 and 100 cm depths. At 30 cm depth, the GEP1 was the best scenario ($R^2 = 0.941$ and $RMSE = 2.593$ °C). Of all scenarios with three inputs (GEP2 to GEP5), the GEP2 (with T_{min} , T_{max} and RH predictors) showed the best performance at 20, 30, 50 and 100 cm depths. At 5 and 10 cm depths, the GEP4 and GEP5 were the best scenarios. At all depths, by increasing the number of input variables to four (GEP6 to GEP8), the GEP7 yielded the best results in estimating T_s , except at 30 cm depth. Therefore, with regard to the input variables in GEP6 to GEP8, it can be concluded that the use of wind speed with accompaniment of T_{min} , T_{max} and R_a showed better performance in comparison with adding RH and n to

Fig. 4 Annual mean daily soil temperature at different depths in the studied period

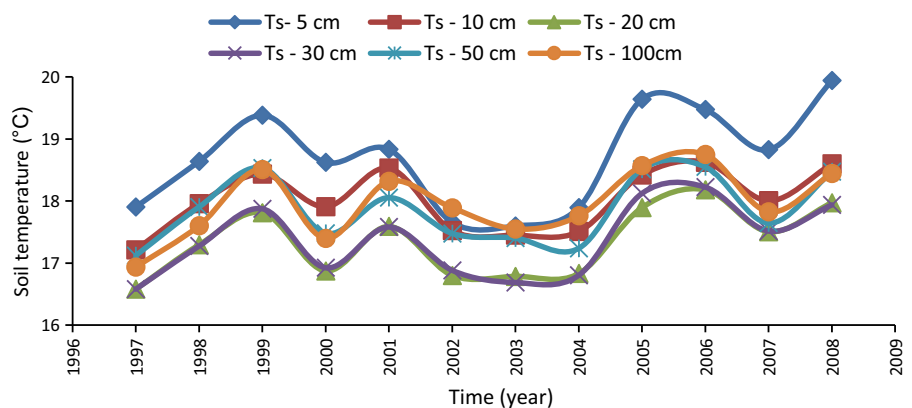


Table 4 Statistical characteristics of daily soil temperature for Sanandaj during 1997–2008

Characteristic	Soil depth (cm)					
	5	10	20	30	50	100
Minimum soil temperature, $T_{s,min}$ (°C)	-3.9	-3.8	-1.3	0.0	1.5	5.2
Maximum soil temperature, $T_{s,max}$ (°C)	44.4	40.7	37.0	33.7	33.6	33.4
Average soil temperature, $T_{s,mean}$ (°C)	18.7	18.0	17.3	17.4	17.9	18.0
Standard deviation, $T_{s,stdev}$ (°C)	12.5	11.7	10.4	9.8	9.1	7.7

Table 5 Statistical indices for the GEP models at different depths for 12 different scenarios (see Table 2)

Depth (cm)	Indices	Scenario											
		GEP1	GEP2	GEP3	GEP4	GEP5	GEP6	GEP7	GEP8	GEP9	GEP10	GEP11	GEP12
5	R^2	0.960	0.965	0.964	0.966	0.962	0.961	0.964	0.956	0.960	0.966	0.960	0.965
	RMSE*	2.804	2.833	2.611	2.607	2.789	2.768	2.615	3.031	2.836	2.575	2.713	2.592
10	R^2	0.970	0.966	0.944	0.971	0.974	0.958	0.964	0.960	0.966	0.940	0.960	0.962
	RMSE	2.149	2.244	2.892	2.129	2.027	2.536	2.316	2.440	2.242	3.021	2.427	2.370
20	R^2	0.950	0.948	0.950	0.942	0.948	0.948	0.959	0.942	0.954	0.913	0.923	0.949
	RMSE	2.505	2.566	2.704	2.694	2.684	2.569	2.270	2.931	2.419	3.473	3.059	2.571
30	R^2	0.941	0.939	0.930	0.938	0.928	0.941	0.923	0.923	0.925	0.881	0.932	0.924
	RMSE	2.593	2.632	2.853	2.681	2.901	2.657	3.099	2.974	2.913	3.692	2.843	2.930
50	R^2	0.900	0.908	0.907	0.904	0.913	0.899	0.914	0.859	0.908	0.857	0.868	0.901
	RMSE	3.371	3.144	3.242	3.398	3.293	3.375	3.022	3.786	3.135	3.901	3.674	3.222
100	R^2	0.776	0.770	0.723	0.746	0.749	0.760	0.833	0.766	0.803	0.711	0.809	0.686
	RMSE	3.847	4.000	4.341	4.122	4.160	4.296	3.304	3.961	3.725	4.468	3.570	4.812

* RMSE is in °C

Bold values indicate statistical indices of the best scenario in each depth

GEP5. In the scenarios with five inputs (GEP9 to GEP11), different results were obtained. The GEP10 at surface layer (5 cm depth), the GEP9 at 10, 20 and 50 cm depths and the GEP11 at 30 and 100 cm depths were the best scenarios to estimate T_s . In general, by considering six parameters as inputs (GEP12), the estimation accuracy of T_s is not necessarily increased.

According to the statistical indices in all scenarios which have been presented in Table 5, it can be concluded that the GEP10 at 5 cm depth ($R^2 = 0.966$ and $RMSE = 2.575$ °C), the GEP5 at 10 cm depth ($R^2 = 0.974$ and $RMSE = 2.027$ °C), the GEP1 at 30 cm depth ($R^2 = 0.941$ and $RMSE = 2.593$ °C), the GEP7 at the depths of 20 cm ($R^2 = 0.959$ and $RMSE = 2.270$ °C), 50 cm ($R^2 = 0.914$ and $RMSE = 3.022$ °C) and 100 cm ($R^2 = 0.833$ and $RMSE = 3.304$ °C) achieved the highest accuracy in estimating T_s .

Figures 5 and 6 show scatter and time series plots comparing observed and estimated soil temperature in testing stage (2005–2008) for the best scenarios at each depth, respectively. It is obvious from the given R^2 of the fitted lines in the scatter plots (Fig. 5) that the estimated values from GEP5 at 10 cm depth are closer to the observed soil temperature values than those of the other different scenarios and depths. Temporal variations of predicted soil temperature and its comparison with observed data (Fig. 6) reveal that the observed and predicted values of soil temperature match closely. It can be concluded from Fig. 6 that in warmer days, the values of T_s have been underestimated by GEP. It is clear that the model is adapted with all data and in the extreme points,

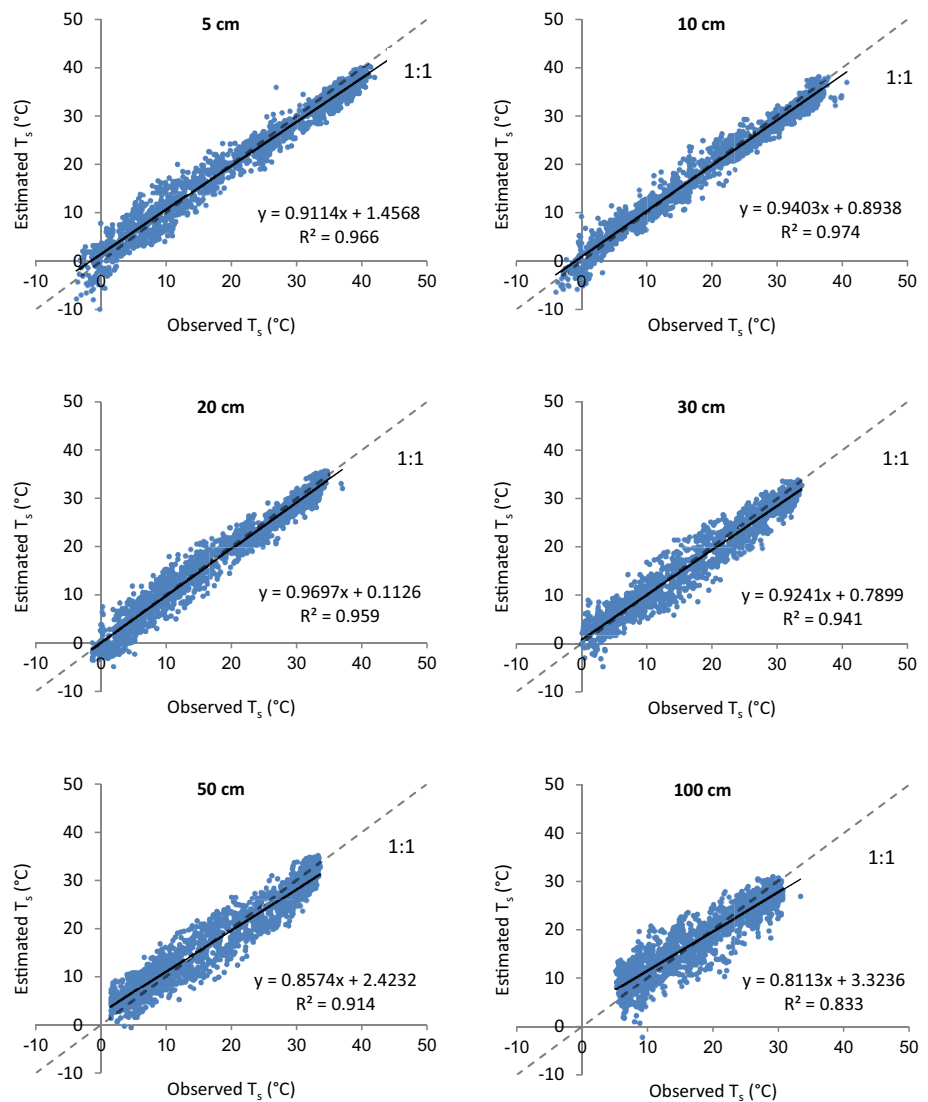
such as warmer days, the results of the model are lower than observed data.

As mentioned before, the GEP returns algebraic equations between input and output variables. The equations of the best model at each depth generated by GEP are presented in Table 6. It is clear from the table that in some cases, all used predictors in different scenarios (Table 2) are not seen. The reason of this fact can be explained from little effect of the ignored predictor, i.e., the coefficient of some variables is negligible. Moreover, the constant coefficients in algebraic equations (Table 6) are randomly values which are produced by GEP. As an example, the expression tree returned by GEP5 for 10 cm depth is shown in Fig. 7. This figure can be transformed into an equation which is seen in Table 6. The mentioned equation has been constructed from three components including Sub-ET 1, Sub-ET 2 and Sub-ET 3 (in Fig. 7), and each Sub-ET was seen as an algebraic sentence. For example, Sub-ET 1 can be written as $\text{Ln} \left[\left(\sqrt{e^{(T_{\min} + T_{\max})}} \right) \times R_a \right]$.

Results of the ANN models

The ANN in this study consists of one hidden layer and one output layer. The optimum number of nodes in the hidden layer was obtained by trial and error and was found to vary between 3 and 10 for the different depths. As for GEP, twelve different combinations of input variables were considered (Table 2). $RMSE$ and R^2 for the different scenarios and soil depths are presented in Table 7. The $RMSE$ ranges from 1.902 °C at the depth of 10 cm in ANN7 to 3.816 °C at 100 cm depth (in

Fig. 5 Scatter plots between observed and estimated soil temperature T_s for the best GEP scenario at 5, 10, 20, 30, 50 and 100 cm depths



ANN2). In general, ANN1 which uses only minimum and maximum air temperatures does not give accurate estimation of soil temperature. Out of all three inputs scenarios (ANN2 to ANN5), the ANN5 (with inputs of T_{min} , T_{max} and R_a) estimated T_s with the smallest error at all depths, except 20 cm depth. At 20 cm depth, ANN2 showed better results than the other three inputs scenarios. The ANN7 was the best scenario at all soil depths out of all four inputs scenarios (ANN6 to ANN8). Similar results were obtained by GEP model (except 30 cm depth). By adding U_2 to T_{min} , T_{max} and R_a , estimation accuracy of T_s improved in comparison with adding RH and n. Similar to the GEP models, in scenarios with five inputs (ANN9 to ANN11), different results were obtained. ANN10 at 5 and 20 cm depths, the ANN9 at 10 and 50 cm depths, the ANN11 at 30 and 100 cm depths gave the best results of all five input scenarios. In many cases, by considering the six inputs in ANN12, the ANN models' accuracy was increased.

With regard to the all statistical indices in Table 7, the ANN10 at 5 cm depth ($R^2 = 0.980$ and $RMSE = 2.191$ °C), the ANN7 at the depths of 10 cm ($R^2 = 0.980$ and $RMSE = 1.902$ °C), 20 cm ($R^2 = 0.970$ and $RMSE = 2.093$ °C) and 50 cm ($R^2 = 0.943$ and $RMSE = 2.663$ °C), the ANN5 at 30 cm depth ($R^2 = 0.957$ and $RMSE = 2.405$ °C) and the ANN12 at 100 cm depth ($R^2 = 0.909$ and $RMSE = 2.570$ °C) were the best scenarios.

Figures 8 and 9 show scatter and time series plots comparing observed and estimated soil temperature for the best scenarios at various depths. It is clear from the given R^2 of the fitted lines in the scatter plots (Fig. 8) that the performance of ANN7 is quite good with high correlation. A visual inspection of the estimated and observed soil temperature clearly demonstrates the potential of ANN modeling (Fig. 9).

Fig. 6 Time series of observed and estimated soil temperature T_s for the best GEP scenario at 5, 10, 20, 30, 50 and 100 cm depths

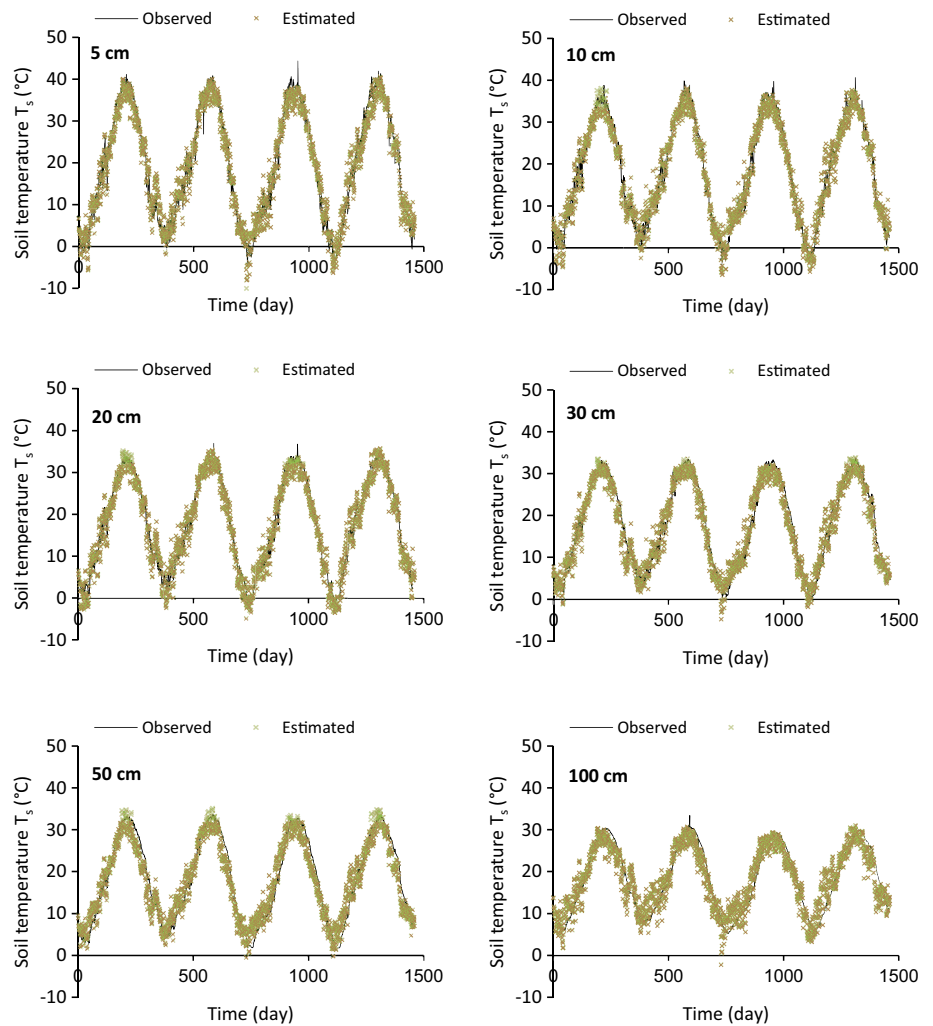


Table 6 Algebraic equations returned by the best GEP scenario for each depth

Soil depth (cm)	GEP model	Algebraic equations
5	GEP10	$T_s = A \tan(6.42 - T_{\max}) + (T_{\max} - 1.75) + A \tan(T_{\min} - [(0.08RH) + (T_{\max} - n) - (T_{\min} + 0.08)])$
10	GEP5	$T_s = \text{Ln}[\sqrt{e^{(T_{\min} + T_{\max})}} \times R_a] + A \tan[(-8.1) \times (T_{\max} - [A \tan(T_{\max}) \times (T_{\max} - 8.5)])] + 0.18$
20	GEP7	$T_s = \sqrt[3]{\frac{R_a - (T_{\max}) + 2.88}{-1.42}} + 0.79T_{\max} + \left[\frac{A \tan(T_{\max}) \times T_{\min}}{(R_a + 7.03)} \right] \times \sqrt{(R_a + 7.03)}$
30	GEP1	$T_s = -0.86 + 0.77T_{\max} + A \tan\left[A \tan\left(\frac{2T_{\min}}{T_{\max} - 2.82}\right)\right] + A \tan\left[\left(\sin(T_{\max}) - (8.68 - T_{\max})\right) \times (T_{\min} - 8.68) + 8.68\right]$
50	GEP7	$T_s = -0.43 \times (T_{\max} + u_2) + T_{\max} + \sqrt{9.42 + 0.22T_{\min}^2 + T_{\min}}$
100	GEP7	$T_s = 0.79 - \sqrt[3]{(T_{\max} + \sqrt{R_a})^2} + \left[3.1 - \left(\text{Ln} \frac{R_a}{9.59}\right)^2\right]^2 + T_{\max}$

Results of the MLR models

Beside the GEP and ANN methods, the MLR approach was used to estimate T_s . Statistical indices including RMSE and R^2 for this method are given in Table 8. As seen, considerable

differences are not observed between the models' accuracy at a specific depth, except for 100 cm depth. At this depth, increasing the predictors leads to improvement in the performance of the scenarios than upper layers. Also for all depths, the MLR12 with full inputs is the best model.

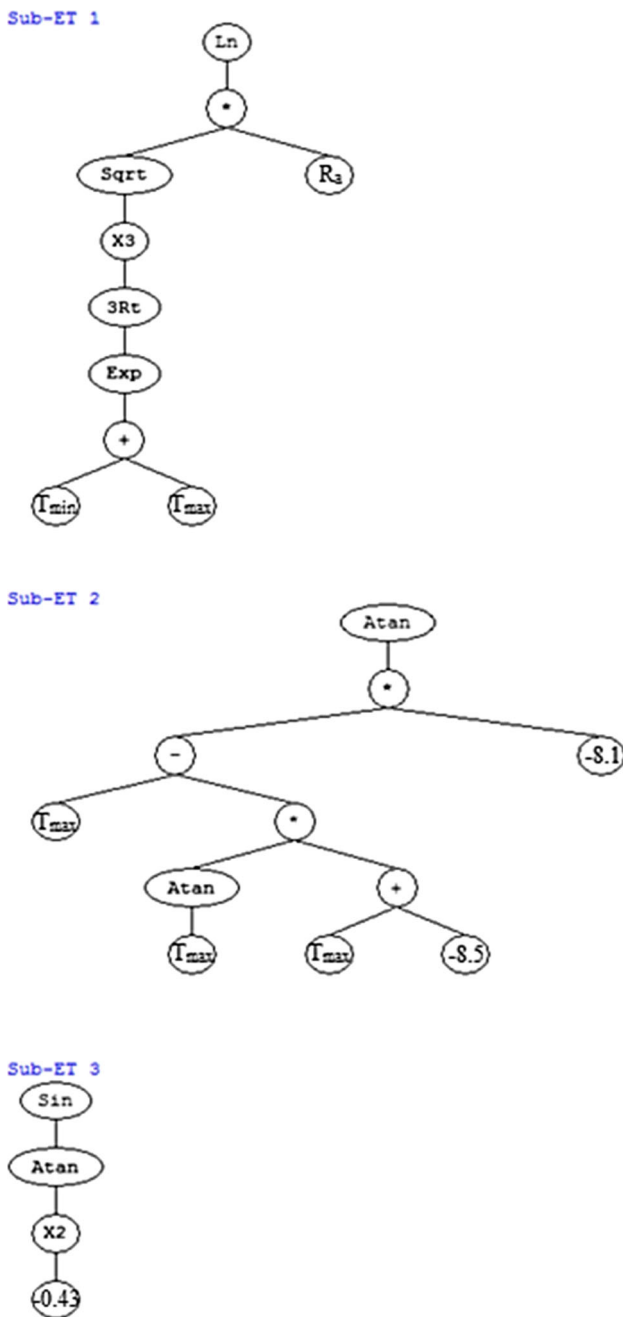


Fig. 7 Expression tree of the GEP5 model at the depth of 10 cm

Comparison of the GEP, ANN and MLR models

Comparing Tables 5 and 7 shows that in the scenarios with the same predictors (e.g., comparing GEP1 and ANN1 at a specific depth), the accuracy of ANN is better than for GEP. Similar to the results obtained from the present study, the superiority of the ANN to other models (e.g., multiple linear regression, multi-nonlinear regression, neuro-fuzzy and genetic programming) has been reported by Bilgili (2010), Tabari et al. (2011), Bilgili et al. (2013) and Kisi et al. (2016). Moreover, increasing depth and using the

scenarios with further predictors cause an improvement in the accuracy of the ANN models than GEP scenarios, especially at the depth of 100 cm (see Tables 5 and 7). The values of R^2 and RMSE (Table 5) indicate that the accuracy of the GEP models does not necessarily increase with increasing the number of input variables. For example, the GEP12 scenario considered the full predictor sets; however, this scenario performed the worst for 100 cm depth. Similar behavior was observed for GEP10 at 30 and 50 cm depths, i.e., increasing predictors in GEP does not cause the enhancement in the models' performance. For ANN, the situation is different. Here for scenarios considering larger numbers of input variables, the performance of the respective ANN tends to be better than for scenarios considering less input variables. At 100 cm depth for instance the ANN12 scenario worked the best. For all other depths, scenarios which use five input variables (i.e., ANN9-ANN11) show high performance. The ANN is an intelligent model. Therefore, the ANN itself considers the parameters which have high correlation with soil temperature in modeling process, whereas the GEP does not have this ability. However, in the case of the correct definition of predictors to the GEP, the solutions of the GEP can be satisfied.

The scenarios GEP2 to GEP5 and ANN2 to ANN5 might be of interest for regions where only the respective subsets of meteorological predictors are available. In such regions, the models generated by GEP2 using T_{min} , T_{max} and RH which yields reasonable results, especially between 20 and 100 cm depth, and ANN5 with input variables T_{min} , T_{max} and R_a are recommended.

The results of GEP, ANN and MLR methods (Tables 5, 7 and 8) reveal that the MLR estimations are generally better than GEP for investigated depths and scenarios. On the other hand, the ANN showed the best performance.

The tendency of used models in estimating T_s was investigated. In warmer days, underestimation was observed. For other days, constant trends were not seen. However, all methods generally showed underestimation.

From a depth of 5 to 10 cm, the accuracy for all scenarios increased for all used methods. 10 cm depth had the highest accuracy in estimating T_s in comparison with all other depths due to the vicinity of upper layers to the atmosphere. In general, from 20 to 100 cm depth, the error increased. Similar results were obtained by Tabari et al. (2011). This can be due to the meteorological conditions having a stronger effect on the soil temperature near the surface than deeper depths. The upper soil layers serve as an insulation between lower atmosphere and deeper soil layers. In other words, the deeper soil layers respond much slower to changes in the atmospheric conditions than the soil layers near the surface.

Table 7 Statistical indices for the ANN models at different depths for 12 different scenarios (see Table 2)

Depth (cm)	Indices	Scenario											
		ANN1	ANN2	ANN3	ANN4	ANN5	ANN6	ANN7	ANN8	ANN9	ANN10	ANN11	ANN12
5	R^2	0.977	0.978	0.981	0.977	0.979	0.982	0.981	0.981	0.982	0.980	0.982	0.984
	RMSE*	2.324	2.401	2.335	2.353	2.254	2.233	2.217	2.285	2.273	2.191	2.304	2.279
10	R^2	0.976	0.976	0.979	0.977	0.978	0.978	0.980	0.978	0.980	0.978	0.980	0.980
	RMSE	1.951	2.039	2.017	1.963	1.930	1.929	1.902	1.994	1.913	1.944	1.970	1.957
20	R^2	0.965	0.967	0.968	0.968	0.966	0.967	0.970	0.967	0.967	0.968	0.966	0.969
	RMSE	2.252	2.211	2.244	2.214	2.215	2.242	2.093	2.213	2.193	2.145	2.236	2.215
30	R^2	0.948	0.949	0.946	0.951	0.957	0.951	0.953	0.952	0.955	0.951	0.956	0.957
	RMSE	2.590	2.600	2.641	2.545	2.405	2.542	2.492	2.536	2.483	2.538	2.433	2.462
50	R^2	0.918	0.922	0.919	0.923	0.941	0.941	0.943	0.932	0.940	0.939	0.932	0.943
	RMSE	3.129	3.075	3.118	3.066	2.712	2.706	2.663	2.889	2.675	2.716	2.861	2.689
100	R^2	0.787	0.789	0.788	0.802	0.903	0.902	0.907	0.882	0.907	0.902	0.909	0.909
	RMSE	3.810	3.816	3.775	3.699	2.635	2.627	2.586	2.874	2.583	2.652	2.580	2.570

* RMSE is in °C

Bold values indicate statistical indices of the best scenario in each depth

Fig. 8 Scatter plots between observed and estimated soil temperature T_s for the best ANN scenario at 5, 10, 20, 30, 50 and 100 cm depths

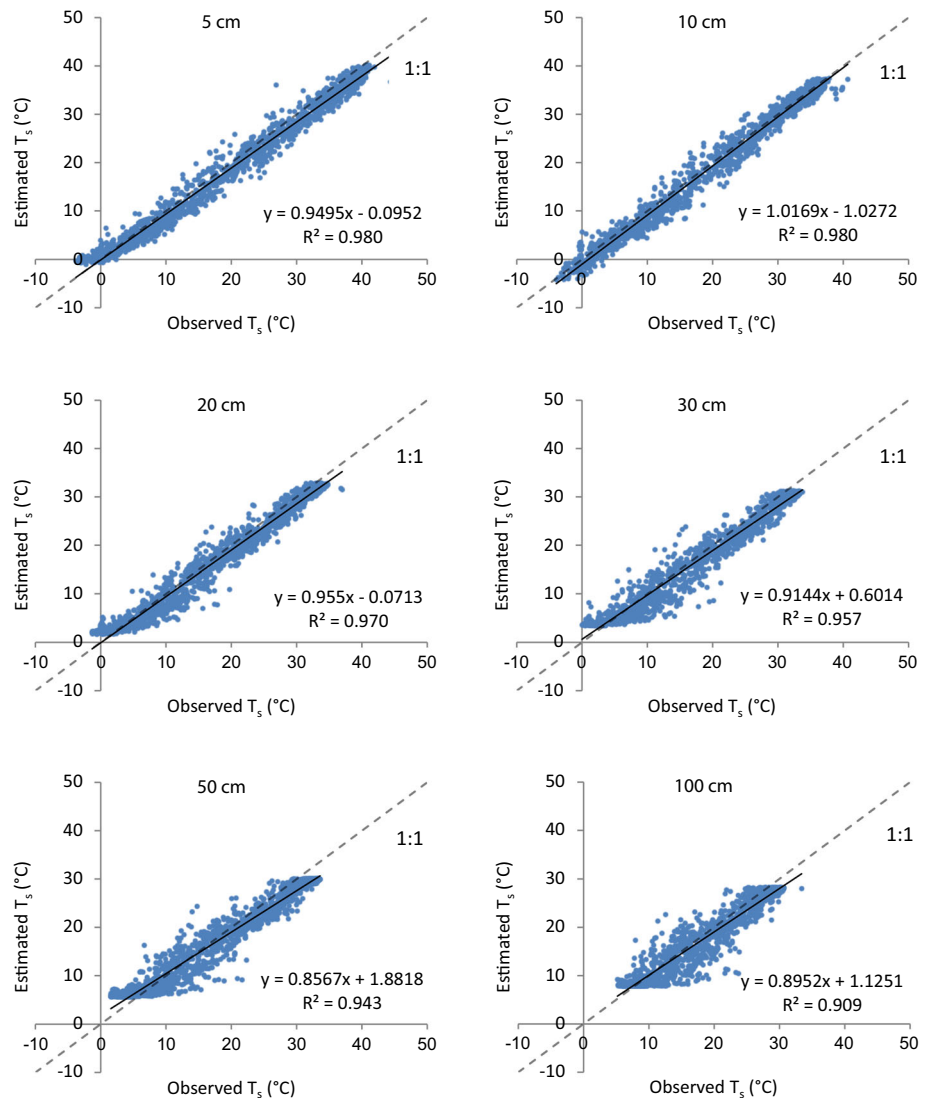


Fig. 9 Time series of observed and estimated soil temperature T_s for the best ANN scenario at 5, 10, 20, 30, 50 and 100 cm depths

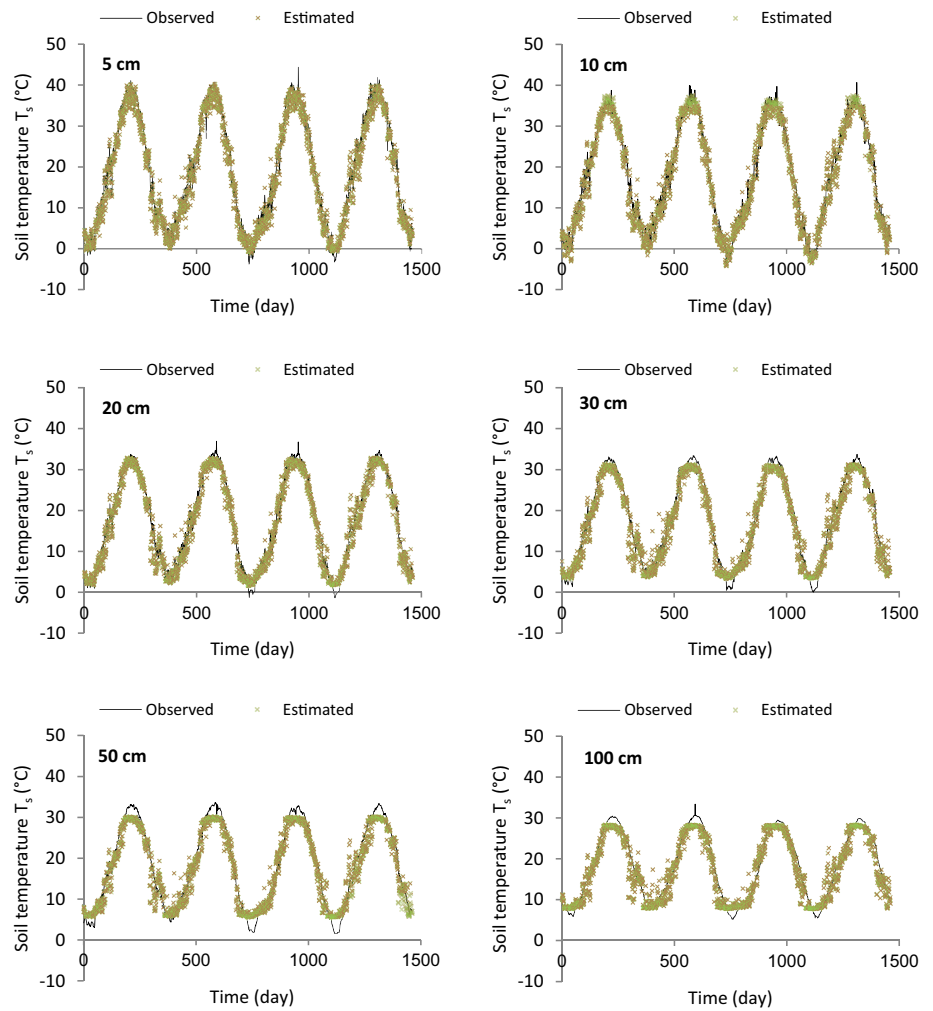


Table 8 Statistical indices for the MLR models at different depths for 12 different scenarios (see Table 2)

Depth (cm)	Indices	Scenario											
		MLR1	MLR2	MLR3	MLR4	MLR5	MLR6	MLR7	MLR8	MLR9	MLR10	MLR11	MLR12
5	R^2	0.962	0.963	0.967	0.962	0.967	0.968	0.968	0.970	0.970	0.970	0.971	0.972
	RMSE*	2.738	2.731	2.615	2.704	2.590	2.583	2.519	2.520	2.474	2.527	2.441	2.429
10	R^2	0.965	0.966	0.968	0.966	0.967	0.968	0.968	0.969	0.970	0.969	0.970	0.971
	RMSE	2.291	2.276	2.201	2.263	2.223	2.207	2.174	2.161	2.127	2.159	2.105	2.088
20	R^2	0.953	0.953	0.954	0.954	0.953	0.953	0.954	0.954	0.955	0.954	0.955	0.956
	RMSE	2.447	2.447	2.430	2.393	2.449	2.449	2.392	2.433	2.378	2.433	2.373	2.367
30	R^2	0.934	0.934	0.934	0.937	0.935	0.935	0.937	0.936	0.938	0.936	0.938	0.939
	RMSE	2.765	2.765	2.763	2.687	2.734	2.735	2.673	2.724	2.664	2.723	2.660	2.657
50	R^2	0.905	0.905	0.904	0.910	0.911	0.911	0.915	0.913	0.915	0.912	0.916	0.916
	RMSE	3.271	3.270	3.278	3.173	3.133	3.132	3.069	3.120	3.065	3.116	3.055	3.054
100	R^2	0.769	0.769	0.767	0.783	0.844	0.844	0.849	0.847	0.850	0.847	0.850	0.852
	RMSE	3.909	3.909	3.922	3.774	3.212	3.212	3.149	3.189	3.144	3.183	3.129	3.123

* RMSE is in °C

Bold values indicate statistical indices of the best scenario in each depth

Table 9 Pearson's correlation coefficients between meteorological parameters and soil temperature

Meteorological parameter	Soil depth (cm)					
	5	10	20	30	50	100
T_{\min}	0.917	0.916	0.912	0.897	0.874	0.809
T_{\max}	0.972	0.971	0.966	0.956	0.935	0.873
R_a	0.824	0.816	0.785	0.745	0.687	0.541
RH	-0.859	-0.857	-0.846	-0.835	-0.815	-0.760
U_2	0.085	0.085	0.072	0.053	0.027	-0.024
n	0.680	0.676	0.655	0.642	0.623	0.569

Effect of meteorological parameters on soil temperature

To investigate the effects of meteorological variables, such as minimum air temperature, maximum air temperature, extraterrestrial radiation, relative humidity, wind speed and sunshine hours on soil temperature, the correlation coefficients between mentioned parameters and soil temperature were obtained at all depths. Table 9 shows the mentioned correlation coefficients. As seen, the highest correlation is observed between T_{\max} and T_s . After T_{\max} , T_{\min} had the highest correlation with T_s . Air temperatures (T_{\max} and T_{\min}) appear to strongly affect soil temperature. This is in accordance with Mihalakakou (2002), Tabari et al. (2011) and Kisi et al. (2015) studies. They stated that soil temperature has the highest correlation with air temperature. Relative humidity has a strong negative correlation with T_s . This is rational because solar radiation causes the evaporation of soil moisture. Consequently, ambient relative humidity increases and radiation energy dose not increase T_s . After T_{\max} , T_{\min} and RH, the least informative parameters concerning T_s are R_a and n , respectively. The weakest correlation was observed between U_2 and T_s . This means that wind speed has only little effect on T_s in comparison with other meteorological parameters. The correlation between meteorological parameters and T_s decreases with depths. Also, a little difference is observed between the correlations of meteorological parameters and T_s at 5–50 cm depths. However, for 50 and 100 cm depths, this difference is slightly more considerable than upper depths. This is due to negligible influence of meteorological parameters on T_s at 100 cm depth in comparison with upper layers.

Conclusion

In the recent years, artificial intelligence models have been increasingly used to predict meteorological and environmental parameters such as soil temperature (T_s). In this study, gene expression programming (GEP), artificial neural networks (ANN) and multiple linear regression

(MLR) were used to estimate T_s at six different depths. The results indicated that used models are able to estimate soil temperature. However, the performance of ANN was better than for GEP and MLR. For used methods, the highest and lowest accuracy were observed at 10 and 100 cm depths, respectively. Between 20 and 100 cm depths, the accuracy decreased with depth. Between 5 and 10 cm depths, the accuracy increased. The results showed that the performance of GEP5 ($R^2 = 0.974$ and $RMSE = 2.027$ °C), ANN7 ($R^2 = 0.980$ and $RMSE = 1.902$ °C) and MLR12 ($R^2 = 0.971$ and $RMSE = 2.088$ °C) at 10 cm depth was superior to all other investigated scenarios and depths. Moreover, correlation coefficients between soil temperature and meteorological parameters revealed that the correlations coefficients decreased with increasing depth. As the surface layers of the soil act as insulation, the effect of the meteorological parameters on T_s is less apparent in the deeper layers. Of all atmospheric parameters considered, air temperatures (maximum and minimum air temperatures) showed the highest correlation to soil temperature. Furthermore, it was found that in the warmer days, the GEP and ANN models underestimated T_s . For other days, the special trends are not observed.

At the studied station (and other synoptic stations in Iran), night measurements of T_s were not carried out. This can cause the creation of errors in correct estimation of T_s . In the present study, T_{\min} , T_{\max} , R_a , RH, U_2 and n were used to estimate T_s . Future works may consider the effects of other variables such as precipitation, actual solar radiation and soil moisture on T_s .

References

- Aytek A, Asce M, Alp M (2008) An application of artificial intelligence for rainfall runoff modeling. *J Earth Syst Sci* 117:145–155
- Bilgili M (2010) Prediction of soil temperature using regression and artificial neural network models. *Meteorol Atmos Phys* 110:59–70
- Bilgili M, Sahin B, Sangun L (2013) Estimating soil temperature using neighboring station data via multi-nonlinear regression and

- artificial neural network models. *Environ Monit Assess* 185:347–358
- Chaturvedi DK (2008) *Soft computing: techniques and its applications in electrical engineering*. Springer, Heidelberg
- de Martonne E (1925) *Traité de Géographie Physique*, 3 tomes. Librairie Armand Colin, Paris
- Droulia F, Lykoudis S, Tsiros I, Alvertos N, Akylas E, Garofalakis I (2009) Ground temperature estimations using simplified analytical and semi-empirical approaches. *Sol Energy* 83:211–219
- Ferreira C (2001) Gene expression programming: a new adaptive algorithm for solving problems. *Complex Syst* 13:87–129
- Garcia-Suarez AM, Butler CJ (2006) Soil temperatures at Armagh observatory, Northern Ireland, from 1904 to 2002. *Int J Climatol* 26:1075–1089
- George RK (2001) Prediction of soil temperature by using artificial neural networks algorithms. *Nonlinear Anal* 47:1737–1748
- Ghuman BS, Lal R (1982) Temperature regime of a tropical soil in relation to surface condition and air temperature and its Fourier analysis. *Soil Sci* 134:133–140
- Guo Z, Wu Z, Dong X, Zhang K, Wang S, Li Y (2014) Component thermodynamical selection based gene expression programming for function finding. *Math Probl Eng* 16:6263–6285
- Güven A (2009) Linear genetic programming for time-series modeling of daily flow rate. *J Earth Syst Sci* 118:137–146
- Hanks RJ, Austin DD, Ondrechen WT (1971) Soil temperature estimation by a numerical method. *Soil Sci Soc Am J* 35:665–667
- Haykin S (1998) *Neural networks—a comprehensive foundation*, 2nd edn. Prentice-Hall, Upper Saddle River, pp 26–32
- Hosseinzadeh Talaei P (2014) Daily soil temperature modeling using neuro-fuzzy approach. *Theor Appl Climatol* 118:481–489
- Huang Y, Lan Y, Thomson SJ, Fang A, Hoffmann WC, Lacey RE (2010) Development of soft computing and applications in agricultural and biological engineering. *Comput Electron Agric* 71:107–127
- Jackson T, Mansfield K, Saafi M, Colman T, Romine P (2008) Measuring soil temperature and moisture using wireless MEMS sensors. *Meas* 41:381–390
- Jungqvist G, Oni SK, Teutschbein C, Futter MN (2014) Effect of climate change on soil temperature in Swedish boreal forests. *PLoS one* 9(4):e93957. doi:[10.1371/journal.pone.0093957](https://doi.org/10.1371/journal.pone.0093957)
- Kang S, Kim S, Oh S, Lee D (2000) Predicting spatial and temporal patterns of soil temperature based on topography, surface cover and air temperature. *For Ecol Manag* 136:173–184
- Kim S, Singh VP (2014) Modeling daily soil temperature using data-driven models and spatial distribution. *Theor Appl Climatol* 118:465–479
- Kisi O, Tombul M, Zounemat Kermani M (2015) Modeling soil temperatures at different depths by using three different neural computing techniques. *Theor Appl Climatol* 121:377–387
- Kisi O, Sanikhani H, Cobaner M (2016) Soil temperature modeling at different depths using neuro-fuzzy, neural network, and genetic programming techniques. *Theor Appl Climatol*. doi:[10.1007/s00704-016-1810-1](https://doi.org/10.1007/s00704-016-1810-1)
- Kocak K, Saylan L, Eitzinger J (2004) Nonlinear prediction of near-surface temperature via univariate and multivariate time series embedding. *Ecol Model* 173:1–7
- Kumar M, Raghuvanshi NS, Singh R (2011) Artificial neural networks approach in evapotranspiration modeling: a review. *Irrig Sci* 29:11–25
- Landeras G, Lopez JJ, Kisi O, Shiri J (2012) Comparison of Gene Expression Programming with neuro-fuzzy and neural network computing techniques in estimating daily incoming solar radiation in the Basque Country (Northern Spain). *Energy Convers Manag* 62:1–13
- Mehdizadeh S, Behmanesh J, Khalili K (2016) Comparison of artificial intelligence methods and empirical equations to estimate daily solar radiation. *J Atmos Sol-Terr Phys* 146:215–227
- Mihalakakou G (2002) On estimating soil surface temperature profiles. *Energy Build* 34:251–259
- Napagoda NADN, Tilakaratne CD (2012) Artificial neural network approach for modeling of soil temperature: a case study for Bathalagoda area. *Sri Lankan J Appl Stat* 13:39–59
- Paul KI, Polglase PJ, Smethurst PJ, O'Connell AM, Carlyle CJ, Khanna PK (2004) Soil temperature under forests: a simple model for predicting soil temperature under a range of forest types. *Agric For Meteorol* 121:167–182
- Prangnell J, McGowan G (2009) Soil temperature calculation for burial site analysis. *Forensic Sci Int* 191:104–109
- Rojas R (1996) *Neural networks*. Springer, Heidelberg
- Shaker F, Monadjemi AH, Yazdanpanah H (2014) Comparing artificial neural networks and linear regression model in predicting soil surface temperature. *Int J Sci Knowl* 5:12–19
- Shiri J, Kisi O, Yoon H, Lee KK, Nazemi AH (2013) Predicting groundwater level fluctuations with meteorological effect implications—a comparative study among soft computing techniques. *Comput Geosci* 56:32–44
- Shiri J, Sadraddini AA, Nazemi AH, Kisi O, Landeras G, Fakheri Fard A, Marti P (2014a) Generalizability of gene expression programming-based approaches for estimating daily reference evapotranspiration in coastal stations of Iran. *J Hydrol* 508:1–11
- Shiri J, Kim S, Kisi O (2014b) Estimation of daily dew point temperature using genetic programming and neural networks approaches. *Hydrol Res* 45:165–181
- Tabari H, Sabziparvar AA, Ahmadi M (2011) Comparison of artificial neural network and multivariate linear regression methods for estimation of daily soil temperature in an arid region. *Meteorol Atmos Phys* 110:135–142
- Tabari H, Hosseinzadeh Talaei P, Willems P (2014) Short-term forecasting of soil temperature using artificial neural network. *Meteorol Appl* 22:576–585
- Traore S, Güven A (2012) Regional-specific numerical models of evapotranspiration using gene-expression programming interface in Sahel. *Water Resour Manag* 26:4367–4380
- Usovicz B, Walczak R (1994) Soil temperature prediction by numerical model. *Pol J Soil Sci* 28:87–94
- Yilmaz T, Ozbek A, Yilmaz A, Buyukalaca O (2009) Influence of upper layer properties on the ground temperature distribution. *J Therm Sci Technol* 29:43–51
- Zadeh LA (1992) *Fuzzy logic, neural networks and soft computing*. One-page course announcement of CS 294–4. University of California, Berkeley
- Zorn CR, Shamseldin AY (2015) Peak flood estimation using gene expression programming. *J Hydrol* 531:1122–1128