The use of artificial neural networks for forecasting the monthly mean soil temperatures in Adana, Turkey

Mehmet BİLGİLİ*
Department of Electrical and Energy Divisions, Adana Vocational High School, Çukurova University, 01160 Adana - TURKEY

Received: 18.01.2010

Abstract: The objective of this paper was to develop an artificial neural network (ANN) model in order to predict monthly mean soil temperature for the present month by using various previous monthly mean meteorological variables. For this purpose, the measured soil temperature and other meteorological data between the years of 2000 and 2007 at Adana meteorological station were used. The soil temperatures were measured at depths of 5, 10, 20, 50, and 100 cm below the ground level by the Turkish State Meteorological Service (TSMS). A 3-layer feed-forward artificial neural network structure was constructed and a back-propagation algorithm was used for the training of ANNs. The models consisting of the combination of the input variables were constructed and the best fit input structure was investigated. The performances of ANN models in training and testing procedures were compared with the measured soil temperature values to identify the best fit forecasting model. The results show that the ANN approach is a reliable model for prediction of monthly mean soil temperature.

Key words: Artificial neural network, meteorological variables, prediction, soil temperature

Türkiye’nin Adana ilindeki aylık ortalama toprak sıcaklıklarının tahmini için yapay sinir ağılarının kullanımını


Anahtar sözcükler: Meteorolojik değişkenler, tahmin, toprak sıcaklığı, yapay sinir ağ

* E-mail: mbilgili@cu.edu.tr
Introduction

Soil temperature is an important meteorological parameter, especially for ground source heat pump applications, solar energy applications such as the passive heating and cooling of buildings, frost prediction, and other agricultural applications (Mihalakakou 2002; Koçak et al. 2004; Yılmaz et al. 2009). It determines the type and rate of different physical and chemical reactions in the soil. It also affects diffusion of nutrients in soil and their uptake by plants. It influences the rate of organic matter decomposition, which in turn affects soil structure and water movement in the soil. Seed germination, seedling emergence, and plant growth are more rapid as the soil temperature increases up to the optimum level. The functional activities of plant roots, such as absorption and translocation of water, are also related to the soil temperature. Crop species differ in their response to soil temperature, and each species has its optimum range of temperature for maximum growth (Tenge et al. 1998). Moreover, soil surface temperature is an important factor for calculating the thermal performance of buildings in direct contact with the soil as well as for predicting the efficiency of earth-to-air heat exchangers (Mihalakakou 2002).

It is clearly evident that soil temperature is an important parameter that directly affects the growth of plants and biological and physical processes occurring in the soil (García-Suárez and Butler 2006). Paul et al. (2004) stated that daily and annual fluctuations in soil temperature influence both biological and chemical processes in the soil, for example, rates of decomposition and mineralization of soil organic matter and release of CO₂. Soil temperature also affects plant growth directly and indirectly. Tenge et al. (1998) pointed out that extremely high soil temperatures, as observed in tropical climates, may result in seedling mortality, low plant stand, higher water demands, and high incidence and severity of plant diseases.

George (2001) stated that prediction of weather parameters such as soil temperature, air temperature, wind speed, relative humidity, and rainfall are useful for agricultural purposes, and all of these are highly correlated due to solar energy. Gao et al. (2008) pointed out that prediction of soil surface temperature plays an important role in numerical hydrological and atmospheric models. Yilmaz et al. (2009) stated that determination of ground surface temperature and ground temperature at different depths is very important for agricultural and ground source heat pump applications and for the calculation of heat losses from the parts of buildings that are buried in the ground. For these purposes, accurate soil temperature measurements or predictions are required. Soil temperature depends on a variety of environmental factors, including meteorological conditions such as surface global solar radiation and air temperature; soil physical parameters such as albedo of surface; 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; Paul et al. 2004; García-Suárez and Butler 2006). For this reason, prediction of soil temperature is rather difficult, especially near the ground surface where the soil temperature variations are the highest (Mihalakakou 2002).

In recent years, there have been several studies concerning the determination of soil temperatures using analytical models, numerical models, and experimental methods (Tenge et al. 1998; Enrique et al. 1999; Kang et al. 2000; George 2001; Mihalakakou 2002; Koçak et al. 2004; Paul et al. 2004; Gao et al. 2007; Gao et al. 2008; Droulia et al. 2009; Prangnell and McGowan 2009). In addition, models based on the Fourier technique and on artificial neural networks have been developed. The objective of this paper was to develop an ANN model that can be used to predict monthly soil temperature by using various meteorological variables of the previous month in the city of Adana, Turkey. The developed model provides a simple and accurate way to predict the soil temperature of the next month at any chosen depth.

Materials and methods

Location of the site

The monthly meteorological data used in this study were obtained from Adana meteorological station, located at 36°59’N, 35°18’E. It is located at an altitude of 28 m above sea level in the eastern Mediterranean region of Turkey. Adana is one of the first industrialized cities and currently one of the
more economically developed cities of Turkey. It is the fourth largest city of Turkey, and it is a major agricultural and commercial center. The Mediterranean climate is dominant in this region, usually hot and dry in the summer season and lukewarm and rainy in the winter season. Winters are about 13-15 °C and very humid, and summers are 34-39 °C. Climate properties vary depending on the level of the height above sea level. On the slope of a mountain looking at the sea, an increase of terrestrial effects on climate is observed. However, the weather in this region does not show an intense terrestrial climate, due to the effect of the Mediterranean Sea (Bilgili et al. 2007).

**Input and output data analysis**

Monthly meteorological variables were measured between the years of 2000 and 2007 by the Turkish State Meteorological Service (TSMS). These meteorological variables were soil temperature (S), atmospheric temperature (T), atmospheric pressure (P), wind speed (W), relative humidity (H), and rainfall (R).

One of the most important steps in developing a satisfactory forecasting model is the selection of the input variables, because these variables determine the structure of the forecasting model and affect the weighted coefficient and the results of the model. For this reason, cross-correlations between input and output variables were calculated to determine the best input structure. The obtained correlation coefficients are shown in Table 1. Here, input variables are the previous monthly mean soil temperature (S_{t-1}), the previous monthly mean atmospheric temperature (T_{t-1}), the previous monthly mean atmospheric pressure (P_{t-1}), the previous monthly mean wind speed (W_{t-1}), the previous monthly mean relative humidity (H_{t-1}), and the previous monthly mean rainfall (R_{t-1}), while the output variable is the monthly mean soil temperature of the present month (S_t). An adequate value of the cross-correlation function for an accurate simulation must be higher than 0.6 (Bechrakis and Sparis 2004). Therefore, as is shown in Table 1, significant correlation coefficients were indicated in bold. This means that these parameters had a strong correlation with each other. There was a high rate of correlation coefficient between the soil temperature and various meteorological variables, such as atmospheric temperature, atmospheric pressure, and soil temperature of the previous month. Soil temperature was positively correlated with atmospheric temperature and soil temperature of the previous month, while it was negatively correlated with atmospheric pressure. Because of that fact, in order to obtain a prediction model for the soil temperature of the present month (S_t), the previous monthly mean atmospheric temperature (T_{t-1}), the previous monthly mean atmospheric pressure (P_{t-1}), and previous monthly mean soil temperature (S_{t-1}) were selected as input (predictor) variables. In addition, there was not a high rate of correlation coefficient between the soil temperature and other meteorological variables such as rainfall, relative humidity, and wind speed. Therefore, they were not selected as input variables and could be eliminated.

Figure 1 shows the monthly mean soil temperature at a depth of 50 cm for the years 2000-2005. As seen from the Figure, during the winter and summer...
months, significant differences appeared. In addition, significant changes from year to year did not appear. The annual cycle of soil temperature had a peak in June and a minimum between December and January. For 2005, the monthly mean soil temperature varied drastically between 11 °C and 31.8 °C throughout the year. Figure 2 shows the monthly mean soil temperatures at the standard depths for the year 2000. During the summer season, soil temperature decreased with depth. Furthermore, the associated downward heat flux built up the soil’s heat store. On the other hand, during the winter season, the gradient reversed and the heat store was gradually depleted. The spring and autumn were transitional periods in which the soil temperature gradients reversed the sign. These reversals are important biological triggers to soil pathogens, soilborne insects, and many other chemical activities. This shows the importance of soil temperature and thus its estimation in agriculture.

### Artificial neural networks

A neural network consists of a large number of simple processing elements, called a neuron. Generally, an artificial neural network (ANN) can be defined as a system or mathematical model that consists of many nonlinear artificial neurons running in parallel, which may be generated as 1-layered or multiple-layered. Most ANNs have 3 layers: input, output, and hidden layers. In the literature, there are many types of ANNs, such as feed forward neural networks (FFNN), radial basis neural networks (RBNN), and generalized regression neural networks (GRNN) (Ustaoglu et al. 2008; Firat and Gungor 2009).

A 3-layer FFNN is shown in Figure 3. It has input, output, and hidden middle layers. Every neuron in each layer is connected to a neuron of an adjacent layer having a different weight. Each neuron receives signals from the neurons of the previous layer, weighted by the interconnect values between neurons, except the input layer. Neurons then produce an output signal by passing the summed signal through an activation function (Haykin 1994; Maqsood et al. 2005).

The process of determining ANN weights is called learning or training, and it is similar to the calibration of a mathematical model. The ANNs are trained with a training set of input and known output data. At the beginning of training, the weights are initialized, either with a set of random values or based on previous experience. Next, the weights are...
systematically changed by the learning algorithm such that, for a given input, the difference between the ANN output and the actual output is small. Many learning examples are repeatedly presented to the network, and the process is terminated when this difference is less than a specified value. At this stage, the ANN is considered trained (Kisi 2004).

The learning of ANNs is generally accomplished by a back-propagation algorithm. The back-propagation is the most commonly used supervised training algorithm in the multilayered feed-forward networks. In back-propagation networks, information is processed in the forward direction from the input layer to the hidden layer and then to the output layer. The objective of a back-propagation network is, by minimizing a predetermined error function, to find the optimal weights that would generate an output vector as close as possible to the target values of the output vector with a selected accuracy (Tayfur 2002).

ANNs can be trained to overcome the limitations of conventional approaches to solve complex problems that are difficult to model analytically (Sözen et al. 2005). Recently, there has been a substantial increase in the interest in artificial neural networks. Researchers have been applying the ANN method successfully in various fields of mathematics, engineering, medicine, economics, meteorology, psychology, neurology, the prediction of mineral exploration sites, electrical and thermal load predictions, adaptive and robotic control, and in many other subjects (Kalogirou 2001).

The fundamental processing element of a neural network is a neuron, which can process a local memory and carry out localized information (Elminir et al. 2007). Each neuron computes a weighted sum of the inputs it receives and adds it with a bias (b) to form the net input (x). The bias is included in the neurons to allow the activation function to be offset from 0 (Elminir et al. 2007):

\[ x = w_{11}P_{11} + w_{12}P_{2} + \ldots + w_{1j}P_{j} + b. \]  

The net input (x) is then passed to the subsequent layer through a nonlinear sigmoid function to form its own output, \( y_j \):

\[ y_j = 1 / (1 + e^{-x}). \]

Afterward, the output \( y_j \) is compared with the target output \( t_j \) using an error function of the following form:

\[ \delta_k = (t_j - y_j) y_j (1 - y_j). \]  

For the neuron in the hidden layer, the error term is given by the following equation (Elminir et al. 2007):

\[ \delta_j = y_j (1 - y_j) \sum_k \delta_i w_k \]

where \( \delta_i \) is the error term of the output layer and \( w_k \) is the weight between the hidden layer and the output layer. The error is then propagated backward from the output layer to the input layer to update the weight of each connection, as follows (Elminir et al. 2007):

\[ w_j(t+1) = w_j(t) + \eta \delta y_j + \alpha (w_j(t) - w_j(t-1)). \]  

Here, \( \eta \) is the learning rate, and the term \( \alpha \) is called the momentum factor, which determines the effect of past weight changes on the current direction of movement. Both of these constant terms are specified at the start of the training cycle and determine the speed and stability of the network.

**Results**

The monthly mean soil temperature of the present month (\( S_t \)) can be characterized as the function of the various previous monthly mean meteorological variables, such as soil temperature (\( S_{t-1} \)), atmospheric temperature (\( T_{t-1} \)), atmospheric pressure (\( P_{t-1} \)), depth (D), and month of the year (M). The relationship between soil temperature and input variables can be expressed as follows:

\[ S_t = f(S_{t-1}, T_{t-1}, P_{t-1}, D, M) \]

For the development of forecasting models, the total 480 data records (8 years \( \times \) 12 months \( \times \) 5 depths) for each variable were collected for the period 2000-2007 for the city of Adana, Turkey. The data set was divided into 2 subsets, a training and a testing data set. The training data set included a total of 360 data records from 2000-2005, which was 75% of the total data records. For more reliable evaluations and comparisons, the models were tested with the testing data set, which was not used during the training process. The testing data set consisted of a total of 120
data records, which was 25% of the total data, observed over the last 2 years. The values applied in the input and output layers were normalized by the following formula in the range of (0-1):

$$X_N = \frac{(X_R - X_{\min})}{(X_{\max} - X_{\min})}$$  \hspace{1cm} (7)

where $X_N$ is the normalized value, $X_R$ is the real value, $X_{\min}$ is the minimal value, and $X_{\max}$ is the maximal value. The minimum and maximum values of input and output variables are given in Table 2. The normalizing of the training inputs generally improves the quality of the training (Kraus et al. 1997). In order to determine the optimal network architecture, various structures of forecasting models were designed with MATLAB software. For this reason, different input structures were applied. The number of neurons in the input layer was changed. The predictions were performed by taking different numbers of hidden layer neurons, between 3 and 12. Different training algorithms were used. The different structures of forecasting models are given in Table 3. For each model, the mean absolute percentage error (MAPE) and the correlation coefficient (R) were used to see the convergence between the target values and the output values. Here, MAPE is defined as follows (Melesse and Hanley 2005):

$$\text{MAPE} = \frac{1}{n} \sum_{i=1}^{n} \text{abs} \left( \frac{o_i - t_i}{o_i} \right) \cdot 100$$  \hspace{1cm} (8)

where $n$ is the total number of months. In addition, the coefficient of correlation between the target value and output value is defined as follows (Bilgili and Sahin 2010):

$$R = \frac{\overline{t \cdot o} - \overline{t} \overline{o}}{\sqrt{\overline{t^2} - (\overline{t})^2} \sqrt{\overline{o^2} - (\overline{o})^2}}$$  \hspace{1cm} (9)

where $t$ is the target value and $o$ is the output value. The models given in Table 3 were trained and tested in order to compare and evaluate the performances of the ANN models. The training and testing results of the ANN models are presented in Figures 4-7. For the testing procedure, the MAPE values of the ANN models ranged from 1.62% to 21.95% different from the actual value of the monthly soil temperature. The maximum MAPE value appears to be 21.95% for the M5 model at a depth of 5 cm, while the M1 model provided the best result, 1.62%, for a depth of 100 cm. Moreover, the maximum correlation coefficient between the target value and output value was 0.9984 for the M1 model at a depth of 100 cm. As seen from the Figures, the results of the M1, M2, and M3 models were closer to each other. The performance values of these models were better than the other models, but the best fit result was obtained from the M1 model. In this model, the Levenberg-Marquardt (LM) learning algorithm was applied. Neurons in the input layer have no transfer function. The logistic sigmoid transfer function (logsig) and linear transfer function (purelin) were used in the hidden layers and output layer of the network as an activation function, respectively. The ANN architecture of the M1 model consists of an input layer, 1 hidden layer with 6 neurons, and an output layer.

According to the best fit result (M1 model), the new formulation, dependent on the previous monthly soil temperature ($S_{t-1}$), previous monthly atmospheric temperature ($T_{t-1}$), depth (D), and month of the year (M) for the outputs, is given with Eq. (10). The

<table>
<thead>
<tr>
<th>Input and output variables</th>
<th>$X_{\min}$</th>
<th>$X_{\max}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Monthly mean soil temperature ($S_t$) [°C]</td>
<td>7.7</td>
<td>37.6</td>
</tr>
<tr>
<td>Previous monthly mean soil temperature ($S_{t-1}$) [°C]</td>
<td>7.1</td>
<td>37.6</td>
</tr>
<tr>
<td>Monthly mean atmospheric temperature ($T_t$) [°C]</td>
<td>6.8</td>
<td>29.7</td>
</tr>
<tr>
<td>Monthly mean atmospheric pressure ($P_t$) [Bar]</td>
<td>1.0012</td>
<td>1.0210</td>
</tr>
<tr>
<td>Depth (D) [cm]</td>
<td>5</td>
<td>100</td>
</tr>
<tr>
<td>Month of the year (M)</td>
<td>1</td>
<td>12</td>
</tr>
</tbody>
</table>
The equation can be used for the prediction of the monthly mean soil temperature of the present month \( S_t \) in Adana, Turkey. 

\[
S_t = -0.067F_1 + 0.6727F_2 + 2.06832F_3 - 5.1115F_4 - 0.1396F_5 - 0.682F_6 + 0.72756 
\]  

Here, \( F_i (i = 1, 2, 3, 4, 5, 6) \) can be calculated by the sigmoid function according to Eq. (11). The formulation for the prediction of monthly mean soil temperature of the present month \( S_t \) in Adana is dependent on the previous monthly soil temperature \( (S_{t-1}) \), previous monthly atmospheric temperature \( (T_{t-1}) \), depth \( (D) \), and month of the year \( (M_t) \), as seen Eq. (12). The weights \( (W_{ij}) \) in Eq. (12) are given in Table 4.
The use of artificial neural networks for forecasting the monthly mean soil temperatures in Adana, Turkey

Figure 8. Comparison between prediction of ANN and actual results for training procedure; a) D = 5 cm, b) D = 10 cm, c) D = 20 cm, d) D = 50 cm, e) D = 100 cm.
Figure 9. Comparison between prediction of ANN and actual results for testing procedure; a) D = 5 cm, b) D = 10 cm, c) D = 20 cm, d) D = 50 cm, e) D = 100 cm.
\[ E_i = W_{1i}(S_{i,t-1}) + W_{2i}(T_{i,t-1}) + W_{3i}(D) + W_{4i}(M_t) + W_{5i} \]

The scatter diagrams of the network predictions against the actual values were drawn for different depths in order to indicate the performance of the M1 ANN model. As seen in Figures 8 and 9, the results of the prediction have fairly close agreement with the corresponding actual measurements.

Discussion

In this study, artificial neural network models were developed to predict the monthly mean soil temperature for the present month by using various previous monthly mean meteorological variables. The models, consisting of the combination of the input variables, were constructed in order to obtain the best fit input structure. From a series of ANN exercises, the M1 model, consisting of 4 input variables, the previous monthly soil temperature \( S_{i,t-1} \), previous monthly atmospheric temperature \( T_{i,t-1} \), depth \( D \), and month of the year \( M_t \), was found to be the best model for forecasting the monthly mean soil temperature of the city of Adana, Turkey. The results obtained with this model were compared with the measured data. Errors obtained were within the acceptable limits. The best result was found to be 1.62\% for the depth of 100 cm. The advantage of this model is that, having the required various previous monthly mean meteorological variables, the monthly mean soil temperature for the present month can be predicted quickly and satisfactorily without the use of any other parameters related to soil.

References


