EXCELLENT PUBLISHERS

Original Research Article   https://doi.org/10.20546/ijcmas.2019.807.014

Soil Temperature Prediction under Limited Data Condition

Debaditya Gupta*, Alivia Chowdhury and Md. Shamimur Rahaman

Soil and Water Engineering Department, Bidhan Chandra Krishi Viswavidyalaya, Mohanpur - 741252, Nadia, West Bengal, India

*Corresponding author

A B S T R A C T

Soil temperature plays a key role in crop water requirement and crop yield. The accurate field estimation of soil temperature is difficult and expensive. Therefore the present study focuses on the estimation of soil temperature in Mohanpur using Artificial Neural Network with input weather data such as maximum temperature, minimum temperature, wind speed, sunshine hours and the results shows that a good correlation exists between the maximum and minimum temperature with the soil temperature. The results statistics shows that with all the given input data condition model shows good results ($R^2 = 0.95$, RMSE = 1.54, MAE = 1.21) and also the model behaves well for sparse data condition i.e. only when maximum and minimum temperatures are available results ($R^2 = 0.91$, RMSE = 1.86, MAE = 1.46).

Introduction

Soil temperature is very important for various agricultural processes (Sattari *et al.,* 2017). The daily soil temperature in sub-tropical regions, particularly in India shows temporal fluctuations. Soil temperature fluctuations affect various processes within the soil such as microbial decomposition, p and k absorption, soil-moisture content (Elshorbagy and Parasuraman, 2008) and soil respiration (Gaumont-Guay *et al.,* 2006). Soil temperature measurements help agriculturists partially to decide and optimize crop water requirement; therefore, optimal design for the planning of irrigation and drainage systems. Plant root growth and biological soil activity directly influenced by soil temperature (Kang *et al.,* 2000). Soil temperature depends upon various climatic factors like ambient air temperature, relative humidity, solar radiation, rainfall, etc. Obtaining accurate measurements of soil temperature are very expensive. Therefore, the development of predictive models to predict soil temperature without using historical soil temperature data is essential (Napagoda and Tilakaratne). In recent years, there have been several studies concerning the determination of soil temperatures using traditional statistical models, other mathematical models and experimental methods (Tenge *et al.,* 1998;

Enrique *et al.,* 1999; Kang *et al.,* 2000; Shannon *et al.,* 2000; George, 2001; Mihalakakou, 2002; Timlin *et al.,* 2002; Koçak *et al.,* 2006; Gao *et al.,* 2007; Droulia *et al.,* 2009; Prangnell and McGowan, 2009). Artificial neural networks are well known effective tools for building mathematical models which deal with many issues with unknown variables (Mohammad *et al.,* 2014). Artificial neural network models have been widely used for solar radiation prediction (Benghanem *et al.,* 2009; Ozgoren *et al.,* 2012), Evapotranspiration estimation (Sudheer *et al.,* 2003; Khoob 2008; Chowdhury *et al.,* 2017, flood modeling (Mukerji *et al.,* 2009; Kia *et al.,* 2012;Rezaeianzadeh *et al.,* 2014) etc. As the soil temperature non-linearly related to the climatic variables, therefore, applications of artificial neural networks are carried out by various researchers for computing soil temperature using different climatic data conditions (Chun-Chieh *et al.,* 1997; Han and Felker, 1997; Tyronese *et al.,* 2008; Bilgili, 2011; Shiri and Kisi, 2011; Wu *et al.,* 2013; Tabari *et al.,* 2015). The direct measurements of soil temperatures are expensive. Therefore ANN models provide an alternative to statistical models and direct measurements for predicting soil temperature. Research works in Indian subtropical regions are limited regarding soil temperature prediction using the non-linear technique; therefore, the present study conducts the modeling of Soil Temperature in Mohanpur with limited weather data by using Artificial Neural Network approach.

Materials and Methods

Study area description and data collection

The experimental data collected from the Department of Agricultural Meteorology and physics, Bidhan Chandra Krishi Viswavidyalaya for the Mohanpur area from 2013 to 2018. Mohanpur is a village in Nadia district, West Bengal, India under Kalyani subdivision. The latitude and longitude of Mohanpur area are respectively 23.66^{o}N, 88.24^{o}E. It has an average elevation of 10 meters (33feet). It has a humid climate with an average rainfall of 1400 mm, approximately 75% of which occurs during June to September due to the onset of southwest monsoon. The mean temperature, relative humidity, and bright sunshine hours of the study area vary from 8°C to 40°C, 30 to 95% and 6.94 to 6.97, respectively. Daily Data of soil temperatures (measured at a depth of 5 cm), maximum air temperature (Tmax), minimum air temperature (Tmin), sunshine hours (n), relative humidity (RH) and wind speed (u_2) data were collected from Department of Agricultural Meteorology and Physics, Bidhan Chandra Krishi Viswavidyalaya. The ANN model was run with all available meteorological inputs to estimate soil temperature data. To develop the ANN models, MATLAB 2013a software was used.

ANN architecture

Artificial Neural Network (ANN) is an efficient computing system whose central theme borrowed from the analogy of biological neural networks. ANNs also named as "artificial neural systems," or "parallel distributed processing systems," or "connectionist systems." ANN execution inspired by the human brain (Haykin, 2001). ANN acquires a large collection of units that are interconnected in some pattern to allow communication between the units. These units also referred to as nodes or neurons, are simple processors which operate in parallel. Every neuron connected with another neuron through a connection link. Each connection link is associated with a weight that has information about the input signal. This is the most useful information for neurons to solve a particular problem because the weight usually

excites or inhibits the signal that is communicated. Each neuron has an internal state, which is called an activation signal. Output signals, which produced after combining the input signals and activation rule, may be sent to other units.

A Multi-Layer Perceptron neural network consists of mainly three main layers viz. Input layers, Hidden layers, and output layers. There may be single or many hidden layers that are why it is termed as the multi-layer perceptron. The data from the environment comes in the input layer as input.

Then this input data is processed in the hidden layer where the input layers are multiplied by some certain weights, and the sum of weights are processed in the hidden layer by some transfer function and then finally the hidden layers are also multiplied by some weights and finally moves to the output layers as a weighted sum. Transfer function is many types viz. logistic sigmoid, hyperbolic tangent sigmoid, linear, etc.

There are several ways to check the accuracy of our models, some are printed directly in R within the summary output, and others are just as easy to calculate with specific functions.

R-Squared (coefficient of determination) is probably the most commonly used statistics and allows us to understand the percentage of variance in the target variable explained by the model. It can be computed as a ratio of the regression sum of squares and the total sum of squares.

$$R^2 = \left[\frac{n \sum xy - \sum x \sum y}{\sqrt{[n \sum x^2 - (\sum x)^2][n \sum y^2 - (\sum y)^2]}} \right]^2$$

Where x = observed soil temperature, y = predicted soil temperature, n = number of data points

Root Mean Squared Deviation or Root

Mean Squared Error (RMSE)

The previous indexes measure the amount of variance in the target variable that can be explained by our model. This is a good indication, but in some cases, we are more interested in quantifying the error in the same measuring unit of the variable. In such cases, we need to compute indexes that average the residuals of the model. The problem is the residuals are both positive and negative, and their distribution should be fairly symmetrical. This means that their average will always be zero. So we need to find other indexes to quantify the average residuals, for example by averaging the squared residuals.

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

This is the square root of the mean the squared residuals, with being the estimated value at point i, being the observed value in i and being the sample size. The RMSE has the same measuring unit of the variable y. Mean Squared Deviation or Mean Squared Error (MAE) is simply the numerator of the previous equation, but it is not used often. The issue with both the RMSE and the MSE is that since they square the residuals, they tend to be more affected by large residuals. This means that even if our model explains the large majority of the variation in the data very well, with few exceptions; these exceptions will inflate the value of RMSE. To solve the problem with large residuals Mean Absolute Error (MAE) used, where the absolute value of the residuals is averaged

$$MAE = \frac{\sum_{i=1}^{n}|y_i - x_i|}{n}$$

This index is more robust against large residuals. Since RMSE is still widely used, even though its problems are well known, it is

always better to calculate and present both in a paper.

Results and Discussion

The soil temperature is a nonlinear function of maximum temperature, minimum temperature, relative humidity, wind speed, a sunshine hour. Since there is not any standard relationship between the soil temperature with the ambient temperature, wind speed, RH, sunshine hour, therefore in order to overcome the problem in the present study a MLP neural network as well as a multiple linear regression model is developed for our study are with different input data combination taken into account which are tabulated as below (Table 2) and the statistical results of the different model combinations are shown in the same table. A correlation matrix (Table 1) is created to show the dependency of the input variables with the soil temperature and based on the correlation, matrix the input variables arselected for the MLP and MLR model.

The transfer function used is logistic sigmoid (logsig) in the hidden layer and linear (purlin) in the output layer.

$$f(x) = \frac{1}{1+e^{-x}}$$

From the correlation matrix, it is found that good correlation exists between the maximum and minimum temperature with the soil temperature followed by wind speed and poor correlation exist between RH and sunshine hr. With the soil temperature. Therefore different input combinations are selected for the ANN model for estimate soil temperature.

M-I-H1-H2-O is the ANN model architecture used with a two layer feed forward neural network model.
Where,

M = Model

I = Number of input nodes (i.e. input variables)

H1 = Number of nodes in the first hidden layer

H2 = Number of nodes in the second hidden layer

O = Number of nodes in the output layer (i.e., one here (Ts))

Optimum number of hidden layers and hidden neurons selection is a very difficult job, and most commonly trial and error method is used to calculate hidden layer and hidden neuron optimization.

In general, most of the MLP-ANN models are constructed with one or two hidden layers.

Generally, the upper limit number of hidden nodes in the first hidden layers lies within (2I-1) and number of nodes in the second hidden layer lies within (2H1-1) (Rajeshkaran and Pi, 2004)

$$H1 \leq (2I - 1) \text{ and } H2 \leq (2H1 - 1)$$

Where $I, H1$ and $H2$ are the numbers of input

nodes, hidden nodes in the input layer, first and second hidden layer respectively.

The Matlab window for ANN is shown below (Fig. 3).

The scatter diagram for Actual vs. Predicted Soil temperature for different ANN architectures for different input combinations shown in the following figures (Fig. 1–4).

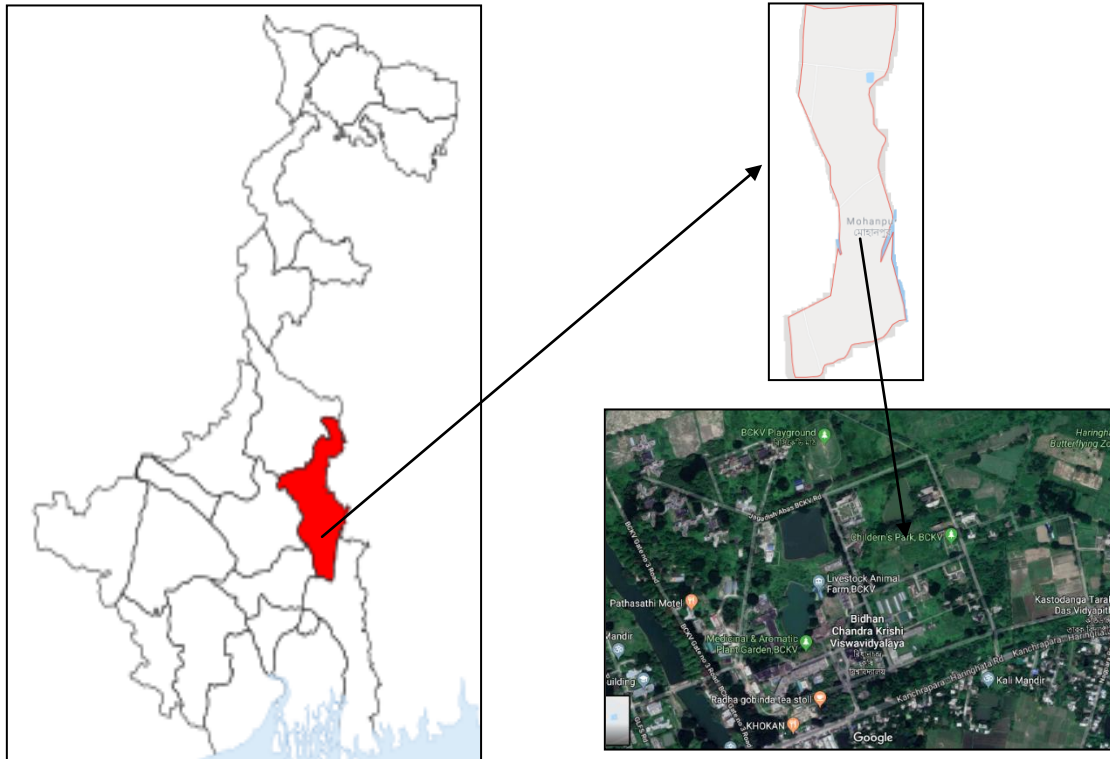**Fig.1** Areal view of Mohanpur Area (Source: Google Images)
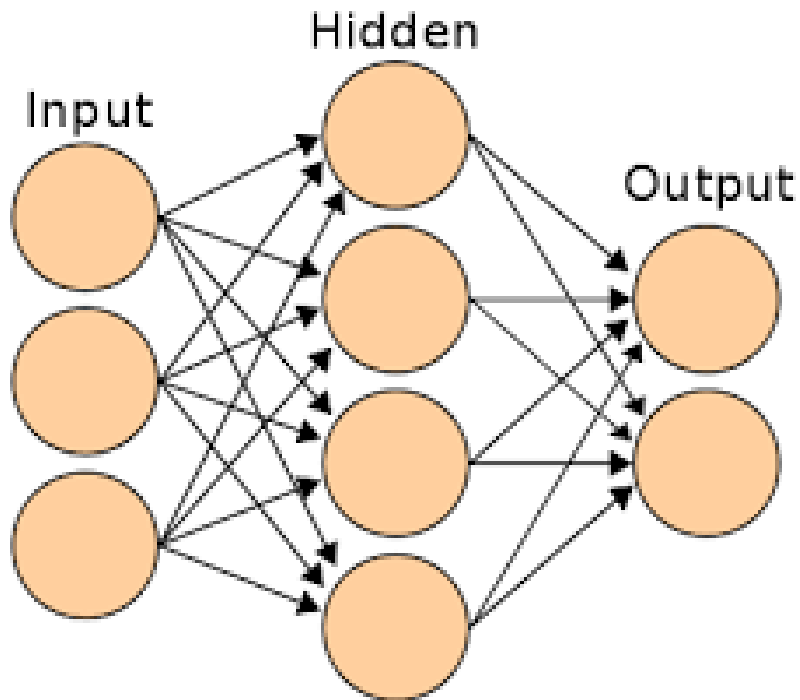

Fig.2 Diagram of Artificial Neural Network
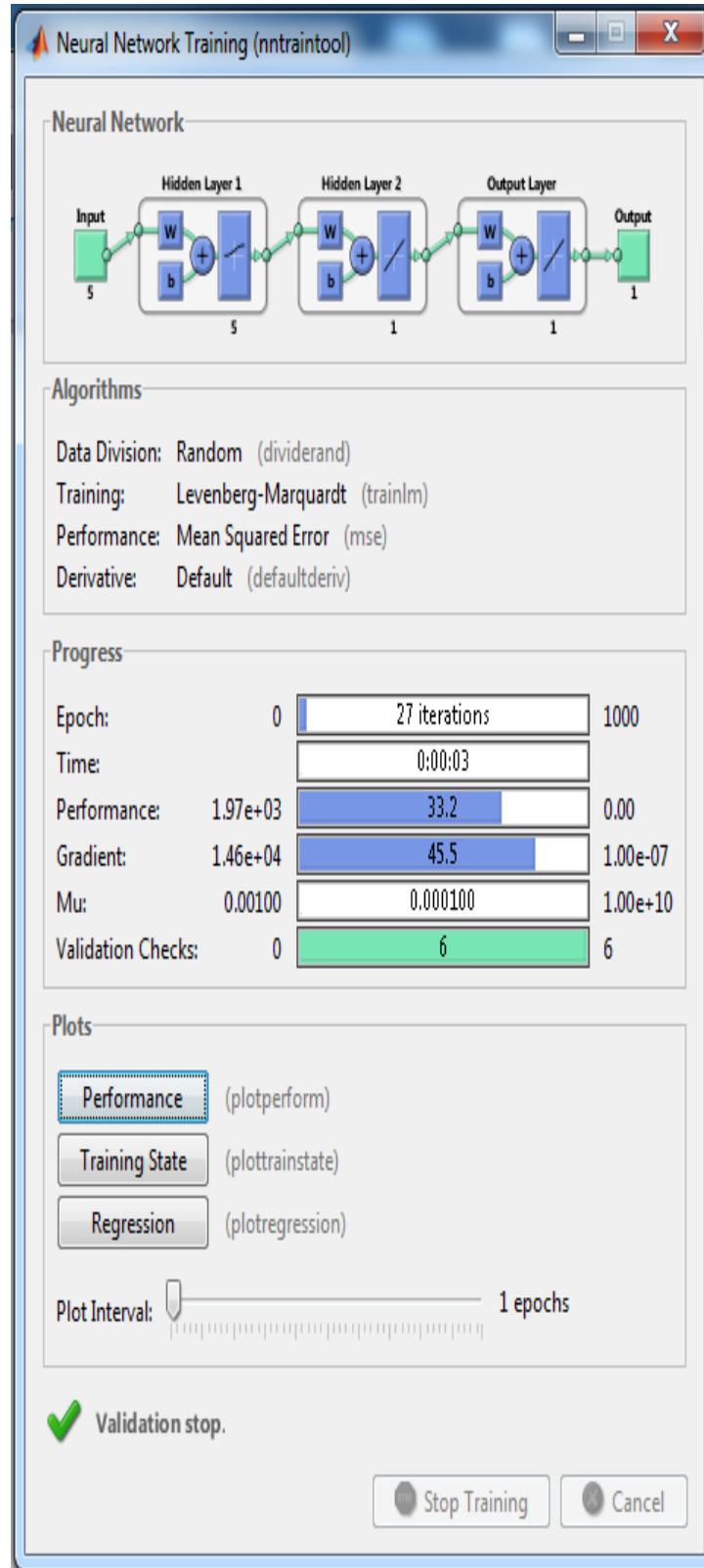
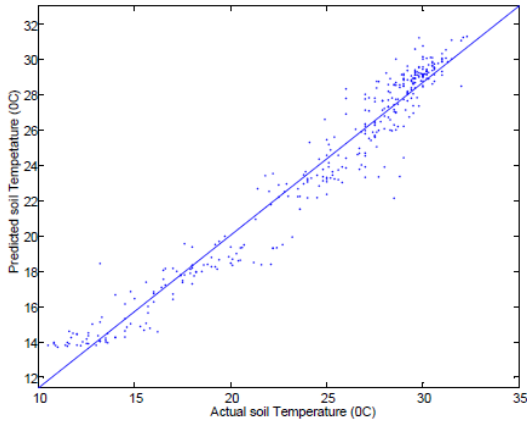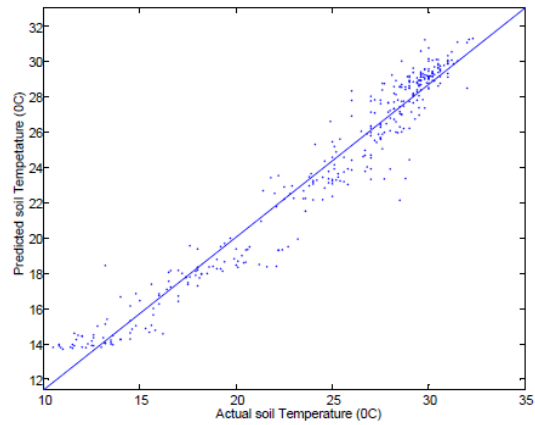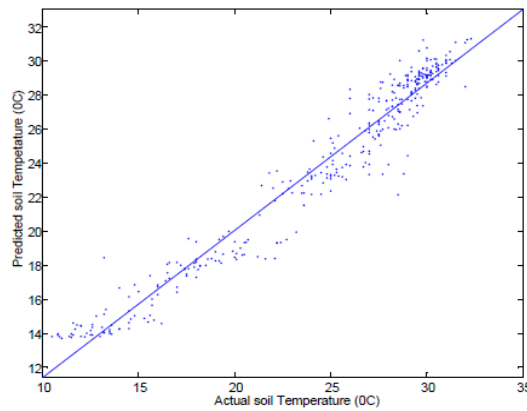**Fig.3** Matlab 2013 a window for training ANN

Fig.4 (a) – (h) Actual vs. predicted soil temperature for different input and neuron number combinations for ANN models (a) M-5-5-1-1 (b) M-5-5-1 (c) M-5-5-5-1 (d) M-2-5-5-1 (e) M-2-2-3-1 (f) M-3-4-5-1 (g) M-3-4-7-1
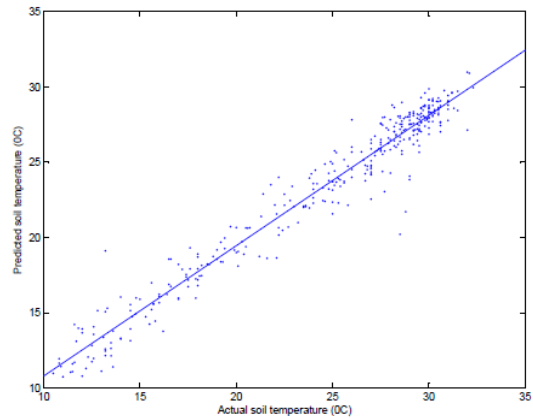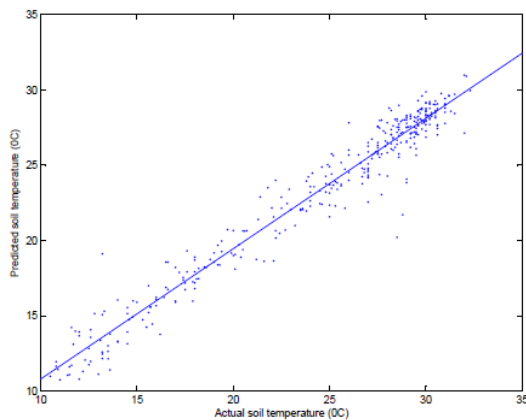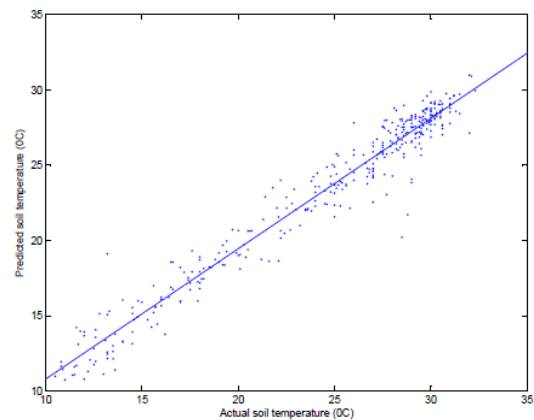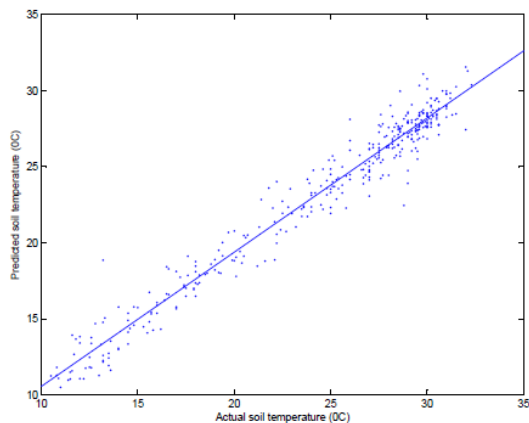

(a)

(b)

(c)

(d)

(e)

(f)

(g)
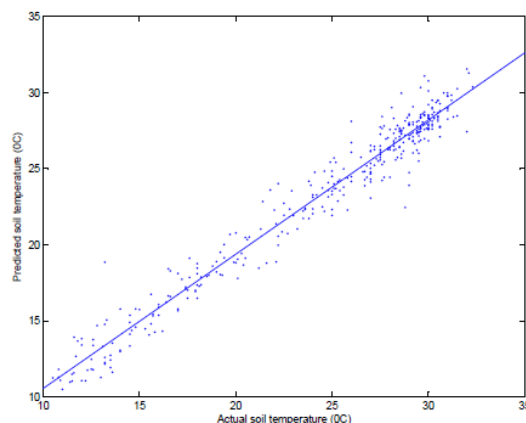
(h)

Table.1 Correlation coefficients between the input variables and soil temperature

	Tmax($^{\circ}$C)	Tmin($^{\circ}$C)	RH (%)	n	u_2(m/s)
Tsoil	0.64	0.77	-0.01	0.01	0.25

Table.2 The various input combinations for MLP models with their statistical results

Model	Inputs	RMSE	MAE	R^2
M-5-5-1-1	Tmax, Tmin, RH, u_2, n	1.54	1.21	0.95
M-5-5-1	Tmax, Tmin, RH, u_2, n	1.73	1.40	0.94
M-5-5-5-1	Tmax, Tmin, RH, u_2, n	1.86	1.54	0.95
M-2-5-5-1	Tmax, Tmin	1.82	1.50	0.90
M-2-2-3-1	Tmax, Tmin	1.81	1.46	0.91
M-3-4-5-1	Tmax, Tmin, u_2	1.73	1.45	0.96
M-3-4-7-1	Tmax, Tmin,u_2	1.70	1.35	0.94
M-3-4-6-1	Tmax, Tmin, n	2.5	2.15	0.82

From the table, it is clear that the MAE, MSE shows much less value as well as high R^2 value in case of MLP when all the input variables viz. Tmax, Tmin, RH, u2, n are provided.

Therefore MLP model can successfully be applied in case of estimation of the soil temperature at Mohanpur, BCKV. From the results, it is also shown that the model is also performing good (Low MAE, RMSE and high R^2) in the case when the model is built with only two input variable viz. Tmax and Tmin

and also when three input variables Tmax, Tmin, and u2 are taken as input for model construction. Therefore both the MLP model as well as the MLR model are suitably estimating the soil temperature under sparse data condition. There is a scope of further improvement of the ANN model by minimizing the model process time using optimizing some hyper model parameters as well as incorporating multiple nonlinear regression models. Therefore there is a huge scope for carrying further research work in this aspect.

In this study, ANN models were developed to predict monthly mean soil temperature at the 5cm depths by using daily maximum air temperature, minimum air temperature, relative humidity, wind speed and sunshine hours at Mohanpur, Nadia, West Bengal. The correlation coefficients between the ANN predictions and actual monthly mean soil temperatures were greater than 95%, indicating that the models can accurately estimate soil temperatures in 5 cm soil depth. All ANN models exhibited low RMSE values, implying that they have good long-term performance. The performance was obtained at the 5cm soil depth, as it provided the greatest determination coefficient (R^2>95%) and the lowest value for root mean square error. The results also suggest that ANNs can be used in place of mathematical simulation models for estimating soil temperature. Further study is needed to determine if the ANN is a promising approach for interpolating soil temperature in other soil depth also. Even with the just two meteorological inputs, maximum air temp and minimum air temperature, ANNs could simulate the seasonal fluctuation of soil temperatures with satisfactory accuracy. The input data for ANN models are easy to obtain and are measured routinely worldwide. Execution time can take as little as a few seconds for a 1-yr simulation. ANNs can produce accurate estimations while requiring only a few years of real data for training. In this study, ANNs were developed based on the short-term application of a 5-yr local-scale data set. The potential of ANNs for long-term application on larger scales is still unproven, but, because of all of their advantages, ANNs are easy and practical to apply from site to site. Their fast execution should also be helpful for simulation of spatial data on a large scale. Here the study was done for soil temperature for 5 cm depth. Hence further study can be carried out for prediction of soil temperature for other depths like 10 cm, 15 cm, 20 cm to justify the applicability of developed ANN model.

Acknowledgment

References

Benghanem, M., Mellit, A., and Alamri, S. N. (2009). ANN-based modeling and estimation of daily global solar radiation data: A case study. Energy conversion and management, 50(7), 1644-1655.

Bilgili, M., (2011). The use of artificial neural networks for forecasting the monthly mean soil temperatures in Adana, Turkey. Turkish Journal of Agriculture and Forestry, 35(1), 83-93.

Brown, S. E., Pregitzer, K. S., Reed, D. D., and Burton, A. J. (2000). Predicting daily mean soil temperature from daily mean air temperature in four northern hardwood forest stands. Forest Science, 46(2), 297-301.

Chowdhury, A., Gupta, D., Das, D. P., Bhowmick, A. (2017). Estimation of reference evapotranspiration using artificial neural network for Mohanpur, Nadia district, West Bengal: a case study. International Journal of Research in Engineering and Technology, 6, 125-130

Droulia, F., Lykoudis, S., Tsiros, I., Alvertos, N., Akylas, E., and Garofalakis, I. (2009). Ground temperature estimations using simplified analytical and semi-empirical approaches. Solar Energy, 83(2), 211-219.

Elshorbagy, A., and Parasuraman, K. (2008). On the relevance of using artificial neural networks for estimating soil moisture content. Journal of Hydrology, 362(1-2), 1-18.

Enrique, G. S., Braud, I., Jean-Louis, T., Michel, V., Pierre, B., and Jean-Christophe, C. (1999). Modeling heat and water exchanges of fallow land covered with plant-residue mulch. Agricultural and Forest Meteorology, 97(3), 151-169.

Gao, Z., Bian, L., Hu, Y., Wang, L., and Fan, J. (2007). Determination of soil temperature in an arid region. Journal of arid environments, 71(2), 157-168.

Gaumont-Guay, D., Black, T. A., Griffis, T. J., Barr, A. G., Jassal, R. S., and Nesic, Z. (2006). Interpreting the dependence of soil respiration on soil temperature and water content in a boreal aspen stand. Agricultural and Forest Meteorology, 140(1-4), 220-235.

George, R. K. (2001). Prediction of soil temperature by using artificial neural networks algorithms. Nonlinear Analysis: Theory, Methods and Applications, 47(3), 1737-1748.

Han, H., and Felker, P. (1997). Estimation of daily soil water evaporation using an artificial neural network. Journal of Arid Environments, 37(2), 251-260.

Haykin, S. S., and Haykin, S. S. (Eds.). (2001). Kalman filtering and neural networks (pp. 221-269). New York: Wiley.

Kang, S., Kim, S., Oh, S., and Lee, D. (2000). Predicting spatial and temporal patterns of soil temperature based on topography, surface cover, and air temperature. Forest Ecology and Management, 136(1-3), 173-184.

Khoob, A. R. (2008). Comparative study of Hargreaves's an artificial neural network's methodologies in estimating reference evapotranspiration in a semiarid environment. Irrigation Science, 26(3), 253-259.

Kia, M. B., Pirasteh, S., Pradhan, B., Mahmud, A. R., Sulaiman, W. N. A., and Moradi, A. (2012). An artificial neural network model for flood simulation using GIS: Johor River Basin, Malaysia. Environmental Earth Sciences, 67(1), 251-264.

Koçak, K., Şaylan, L., and Eitzinger, J. (2004). Nonlinear prediction of near-surface temperature via univariate and multivariate time series embedding. Ecological Modelling, 173(1), 1-7.

Mihalakakou, G. (2002). On estimating soil surface temperature profiles. Energy and Buildings, 34(3), 251-259.

Mohammad, E. M., Alwahab, S. A., Warttan, H. A. (2014). Modeling Soil Temperature Using Artificial Neural Network. 5th International Conference on Environmental Science and Technology, vol. 69.3.

Mukerji, A., Chatterjee, C., and Raghuwanshi, N. S. (2009). Flood forecasting using ANN, neuro-fuzzy, and neuro-GA models. Journal of Hydrologic Engineering, 14(6), 647-652.

Napagoda, N. A., and Tilakaratne, C. D. (2012). Artificial neural network approach for modeling of soil temperature: a case study for Bathalagoda area. Sri Lankan Journal of Applied Statistics, 13, 39-59.

Ozgoren, M., Bilgili, M., and Sahin, B. (2012). Estimation of global solar radiation using ANN over Turkey. Expert Systems with Applications, 39(5), 5043-5051.

Prangnell, J., and McGowan, G. (2009). Soil temperature calculation for burial site analysis. Forensic Science International, 191(1-3), 104-109.

Rajasekaran, S., and Pai, G. V. (2003). Neural

networks, fuzzy logic and genetic algorithm: synthesis and applications (with cd). PHI Learning Pvt. Ltd.

Rezaeianzadeh, M., Tabari, H., Yazdi, A. A., Isik, S., and Kalin, L. (2014). Flood flow forecasting using ANN, ANFIS and regression models. Neural Computing and Applications, 25(1), 25-37.

Sattari, M. T., Dodangeh, E., and Abraham, J. (2017). Estimation of daily soil temperature via data mining techniques in semi-arid climate conditions. Earth Sciences Research Journal, 21(2), 85-93.

Sudheer, K. P., Gosain, A. K., and Ramasastri, K. S. (2003). Estimating actual evapotranspiration from limited climatic data using neural computing technique. Journal of Irrigation and Drainage Engineering, 129(3), 214-218.

Tabari, H., Talaee, P. H., and Willems, P.

(2015). Short- term forecasting of soil temperature using artificial neural network. Meteorological Applications, 22(3), 576-585.

Tenge, A. J., Kaihura, F. B. S., Lal, R., and Singh, B. R. (1998). Diurnal soil temperature fluctuations for different erosion classes of an oxisol at Mlingano, Tanzania. Soil and Tillage Research, 49(3), 211-217.

Timlin, D. J., Pachepsky, Y., Acock, B. A., Šimunek, J., Flerchinger, G., and Whisler, F. (2002). Error analysis of soil temperature simulations using measured and estimated hourly weather data with 2DSOIL. Agricultural Systems, 72(3), 215-239.

Yang, C. C., Prasher, S. O., and Mehuys, G. R. (1997). An artificial neural network to estimate soil temperature. Canadian Journal of Soil Science, 77(3), 421-429.