MAUSAM, **53**, 4 (October 2002), 471-480

551.509.21 : 551.501.86

Short term prediction of atmospheric temperature using neural networks

S. PAL, J. DAS, P. SENGUPTA and S. K. BANERJEE*

Electronics and Communication Sciences Unit,

Indian Statistical Institute, Calcutta, India

**Cyclone Detection Radar,*

India Meteorological Department, Calcutta, India

(*Received 3 May 2001, Modified 19 March 2002*)

सार - इस शोध पत्र में भूतल के अधिकतम तथा न्यूनतम तापमान हेतु पूर्वानुमान निदर्श पर आधारित न्यूरल संजाल पर विचार किया गया है। इस शोध पत्र में न्यूरल संजाल के बहु-स्तरीय पर्सेप्ट्रान (एम.एल.पी) किस्म के प्रवणता शिष्ट जानकारी के पृष्ठ संचरण प्रक्रिया के केवल एक गुप्त परत पर विचार किया गया है। इस संजाल में 25 निवेश नोड तथा दो निर्गम नोड सम्मिलित हैं। इस संचाल में ट्रेनिंग निदर्श के सेट का उपयोग करते हुए गुप्त परत में विभिन्न प्रकार के बहुत से नोडों के साथ इसे ट्रेन्ड किया गया है तथा परीक्षण निदर्श के सेट के साथ प्रत्येक नोड का परीक्षण किया गया है। यह संजाल अधिकतम तापमान के आकलन के लिए निवेशों के रूप में तथा निर्गम के रूप में न्यूनतम तापमान की पिछले दो क्रमिक दिनों की सूचना (जैसे अवदाबों, तापमानों, सापेक्षिक आर्द्रताओं आदि) को ग्रहण करता है। गुप्त परत में 20 अथवा इससे कम न्यूरान्स के संचाल अनुकूलतम पाए गए हैं तथा परीक्षण मामलों के 80 प्रतितशत मामलों में अधिकतम तथा न्यूनतम 2° सै. की त्रुटि पाई गई है।

ABSTRACT. In this paper, a neural network based forecasting model for the maximum and the minimum temperature for the ground level is proposed. A backpropagation method of gradient-decent learning in multi-layer perceptron (MLP) type of neural network with only one hidden layer is considered. This network consists of 25 input nodes and two output nodes. The network is trained with a varying number of nodes in the hidden layer using a set of training sample and each of them is tested with a set of test sample. It accepts previous two consecutive days information (such as pressures, temperatures, relative humidities, etc.) as inputs for the estimation of the maximum and the minimum temperature as output. The network with 20 or less neurons in the hidden layer is found to be "optimum" and it produces an error within ±2° C for 80% of test cases.

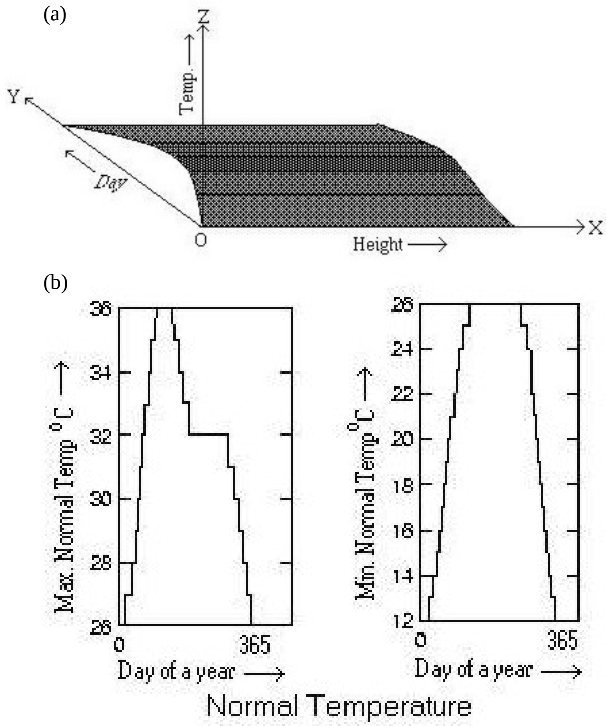
Key words – Neural networks, Backpropagation, Atmospheric science, Temperature forecasting.

1. Introduction

There are many real life problems in which future events must be predicted on the basis of past history. In such cases, prediction depends on (*i*) knowledge of underlying laws (a very powerful and accurate means of prediction) and (*ii*) the discovery of strong empirical regularizes in observations of a given system. However, the laws underlying the behaviour of a system are not easily discovered and the empirical regularities or periodicities are not always evident and can often be masked by noises.

Though perfect prediction is hardly even possible, neural networks can be used to obtain reasonably good prediction in a number of cases [Tsintikidis *et al*. 1997; Salehfare and Benson, 1998]. Weather forecasting is a technique for the estimation of temperature, rainfall, humidity, wind speed, wind direction and atmospheric pressure etc. well in advance. These parameters vary from place to place on the earth surface. So the latitude and longitude of a particular place on the earth surface will be another parameter in a weather forecasting model. Again this model may not provide accurate result because of many other factors like topography of a place, surrounding structures and environmental pollution etc. The dynamics of lower atmosphere are continuously changing. So the accuracy of the forecasting model can be improved if one considers all these factors.

The domain experts forecast maximum and minimum temperatures by his intuition based on previous



Figs. 1(a&b) . (a) Temperature profile of a particular place and time over the year upto certain height and (b) A cycle of normal temperature distribution over a year for Dum Dum station

weather conditions or by some statistical methods. This may not be always precise but it must help to design a model. If the model is a rule-based system then the formulation of an exhaustive rule-base is very difficult. Sometimes a purely statistical model works fine, but it does not take care of all the instantaneous fluctuations of different weather parameters. It is possible to put these interactions and expert rule-base in one model by using a neural net which is a concept of natural process of learning.

The domain expert uses previous days weather condition, that is, local parameters such as pressure, temperature, vapour pressure, relative humidity, wind speed, wind direction and radiation etc. The domain knowledge of an expert is the atmospheric normal conditions (*i.e.* normal temperature, pressure, humidity and temperature etc.).

In this paper, an attempt is made to describe a neural network based temperature forecasting model. This paper is organized in the following order. Section 2 describes the materials and method for the collection of data. Section 3 describes the proposed model for short-term temperature forecasting. The results and discussion is presented in Section 4. A conclusion is at the end in Section 5.

2. Method of data collection

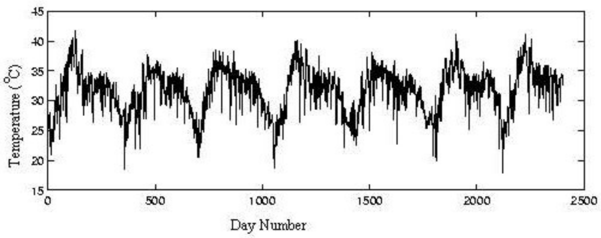
Normally, temperatures are measured twice in a day at different heights and places using radiosonde (*i.e*. balloon floating) technique. The parameters such as temperatures, wind direction and its velocity etc. are measured by the meteorologists of a country. Since our country is very large and we have a limited number of stations where radiosonde flights are made regularly. The measurements of lower atmospheric parameters are necessary for different applications such as avionics, pollution dispersal, communication and so on.

The accurate prediction of these meteorological parameters is therefore, very essential for agricultural and other social activities. Here the prediction or forecasting technique is an analysis of time series data with nearly regularized variations such as the sessional fluctuations etc. Also some other parameters like place or point on the earth surface or atmosphere, position of the Sun, clouds on the sky, pollution in the environment etc. are knowingly or unknowingly affect the accuracy of the forecasted value of the temperature. So it is extremely difficult to incorporate all these parameters (if known) for temperature forecasting in a purely mathematical model (it may be formal or empirical). The previous records of temperature shows that the maximum and minimum temperatures are measured at different heights at a particular place and also at particular time of a day. We observe that the profile of these temperature is a surface with respect to the month and height from the earth surface. In Fig. l(a), near the ground where the height is almost zero but pressure is largest, the temperature is increased from the month January to May, June etc. of a year then again decrease till December of this year and January of the next year, here this is following a cyclic curve at any height. As height increases, the atmospheric temperature and pressure also decreases. The trained of the profile becomes same in nature.

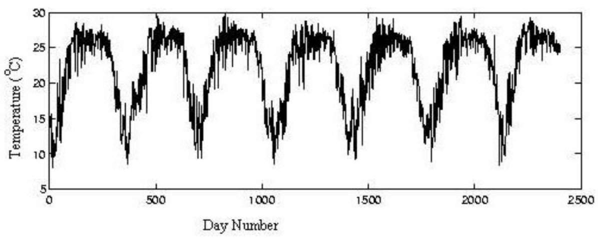
Further we observed that this temperature profile is not fixed, it changes slowly with time that means a significant changes may be observed after a period of certain years. The normal temperature for Dum Dum station both the maximum (*i.e*. dry bulb) and the minimum (*i.e*. wet bulb) are shown in Fig. l(b).

A sample for measured temperature (both maximum and minimum) for the period 1989-95 of Dum Dum station are shown in Fig. 2.

(a) Actual maximum temperature



(b) Actual minimum temperature



Figs. 2(a&b). Actual temperature distribution over the period 1989-95 for Dum Dum station (a) maximum temperature and (b) minimum temperature

Now our objective is to design a temperature forecasting model for the ground level of particular place. Assume that, the last two days information is able to forecast both the maximum and the minimum atmospheric temperature. But there is an effect on temperature variation by others such as pollution etc. Standard models are not able to forecast these situations. A neural network based model can automatically learn these kinds of effects. In multi-layer perceptron (MLP) type of neural network with only one hidden layer is suitable because it has an universal approximation capability. In neural network there are various learning schemes and here we are adopted the backpropagation method of gradient decent learning method for forecasting. We aim is to design an MLP so that the network forecast *n*th day using the information from (*n*-1)th and (*n*-2)th days.

The standard statistical models are best suited for time series analysis. Neural network, specially, MLP has a capability to learn the behavior of the given training sample as an universal approximation function. The training sample is used to train the network. The training sample may contain multiple inputs as well as multiple outputs. The empirical model try to set up an input-output relationship using some mathematical notations. A trained MLP acts as a function in the domain of training sample. Also a small fluctuation in the input side reflects to the output side. But in the standard statistical or empirical models, these small fluctuations may be smoothen, that is, these fluctuations are not reflected into the output side. So neural network based model is better than the conventional statistical techniques so far as the short term predictions are concerned.

In this problem we have collected the following information for a day : (1) pressure$_1$ (mean sea level pressure at 1730 hr), (2) pressure$_2$ (mean sea level pressure at 0830 hr), (3) vapour pressure at 1730 hr, (4) vapour pressure at 0830 hr, (5) relative humidity at 1730 hr, (6) relative humidity at 0830 hr, (7) maximum temperature at 1730 hr, (8) minimum temperature at 0830 hr, (9) rainfall, (10) direct radiation and (11) defuse radiation. These data are arranged so that they form a record. The record structure contains first 11 fields [field number 1 to 11] contains the day before yesterday information [such as (1) pressure$_1$, (2) pressure$_2$, (3) vapour pressure at 1730 hr, (4) vapour pressure at 0830 hr, (5) relative humidity at 1730 hr, (6) relative humidity at 0830 hr, (7) maximum temperature at 1730 hr, (8) minimum temperature at 0830 hr, (9) rainfall, (10) direct radiation and (11) defuse radiation] next 11 fields [field number 12 to 22] contains yesterdays information [such as (12) pressure$_1$, (13) pressure$_2$, (14) vapour pressure at 1730 hr, (15) vapour pressure at 0830 hr, (16) relative humidity at 1730 hr, (17) relative humidity at 0830 hr, (18) maximum temperature at 1730 hr, (19) minimum temperature at 0830 hr, (20) rainfall, (21) direct radiation and (22) defuse radiation] and last 5 fields are today's information [such as (23) day, (24) month, (25) year, (26) maximum temperature at 1730 hr, (27) minimum temperature at 0830 hr]. These records are divided into two parts; one part contains 2285 different records in random order for learning the neural network model known as training data and other 270 records is for testing the model known as test data. In case of training data first 25 fields of a record acts as an input data set for input nodes of the network and the remaining 2 fields of the record acts as an target data set for the output nodes in the output layer of the network.

3. Model

Several multilayered perceptrons architecture (Werbos 1974, Rumelhart *et al*. 1986, LeCum 1988) as well as other neural network structures with various numbers of hidden layers and hidden processing elements (PEs) were developed. In our proposed model, the number of input and output nodes (*i.e*. PE) was determined and fixed respectively by the number of components in input vector and output vector for each sample. The sample data are gathered by the actual measurement as described in Section 2. The number of components for an input is 25 (it includes two consecutive days information from field number 1 to 25 of the record such as atmospheric pressure, vapour pressure, relative humidity, temperature,

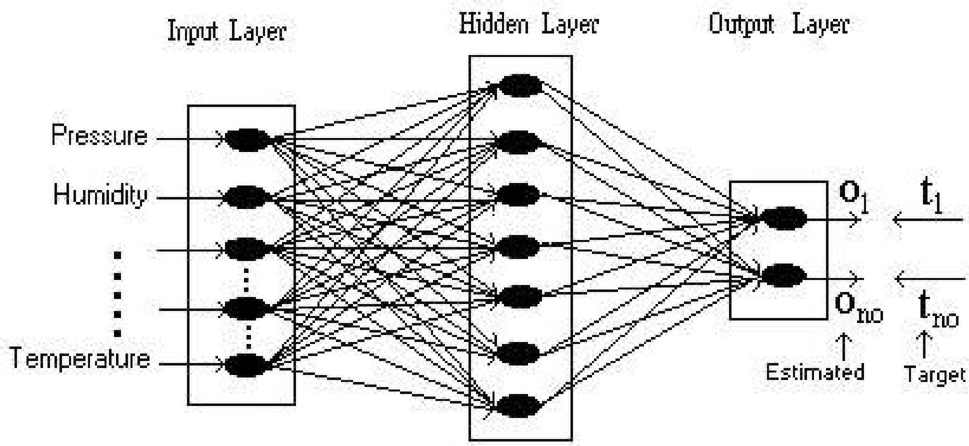


Fig. 3. A neural net for temperature forecasting

normal pressure and temperature, radiation etc.) and for an output (target) sample is 2 (field number 26 to 27 *i.e.* maximum temperature *i.e.*, dry bulb temperature (T_{dry}) and minimum temperature *i.e.*, wet bulb temperature (T_{wet}).

The network is trained with a varying number of nodes in the hidden layer using a set of training sample of size 2285 and each of them is tested with a set of test sample. It accepts previous two consecutive days information (such as pressures, temperatures relative humidities etc.) as inputs for the estimation of the current day's maximum and the minimum temperature as output.

The required neural network architecture was designed as shown in Fig. 3. Suppose a network consists of 25 nodes in the input layer, 15 nodes in the hidden layer and 2 nodes in the output layer. Then the neural net have total 42 nodes with $25 \times 15 + 15 \times 2 = 405$ synapse connections (weights). The best weights were determined by training the network several times using a multilayer perceptron algorithm.

3.1. *Multi-layer perceptron*

It is a layered neural network. This network consists of several layers of neurons of which the first one is the input layer and the last one is the output layer, remaining layers are called hidden layers. A typical MLP net is shown in Fig. 3. There are complete connections between the nodes in successive layers but there is no connection within a layer. Every node, except the input layer nodes, compute the weighted sum of its inputs and apply a sigmoidal function to compute its output, which is then transmitted to the nodes of the next layer. The objective of the MLP learning is to set a connection weights such that the error between the network output and the target output is the minimum. The network weights may be learnt by several methods of which the backpropagation technique is most popular one. In this work we use the backpropagation learning. It is known that a single hidden layer is sufficient for a multilayer perceptron to compute a uniform ε approximation to a given training set representation by the set of inputs and a desired (target) output [Haykin 1994]. Hence in this study we restrict ourselves to a three layers network.

3.1.1. *Algorithm for forecasting*

Let n_i = Number of neurons in input layer

n_h = Number of neurons in hidden layer

n_0 = Number of neurons in output layer

h = The output vector for n_h nodes in the hidden layer, $h = (h_1\ h_2\ h_3 \ldots h_{nh})$

i = The input vector for n_i nodes in the input layer, $i = (i_1\ i_2\ i_3 \ldots i_{ni})$

O = The output vector for n_o nodes in the output layer, $0 = (0_1\ 0_2\ 0_3 \ldots 0_{n0})$

TABLE 1

Cumulative percentage frequency (Training sample)

Range	% Frequency of temperature							
	h = 10		*h* = 15		*h* = 20		*h* = 25	
	Max	Min	Max	Min	Max	Min	Max	Min
±0.5	22.50	14.69	22.81	16.25	23.75	22.50	14.06	20.00
±1.0	41.56	36.25	44.06	37.19	43.12	48.12	29.38	37.19
±1.5	60.94	60.31	59.69	62.50	60.94	68.75	43.12	61.88
±2.0	70.94	77.19	70.94	78.75	71.56	85.00	57.81	80.94
±2.5	77.81	88.12	77.81	89.38	79.06	90.94	70.00	90.00
±3.0	85.94	93.44	85.94	94.06	84.69	95.00	79.38	94.38
±3.5	91.88	96.25	91.56	97.50	91.25	97.50	85.62	97.19
±4.0	95.00	98.12	94.69	98.75	94.06	99.06	89.69	98.44
±5.0	97.81	98.75	97.50	99.69	98.44	100.00	94.69	99.38

TABLE 2

Cumulative percentage frequency table (Test sample)

Range	% Frequency of temperature							
	h = 10		*h* = 15		*h* = 20		*h* = 25	
	Max	Min	Max	Min	Max	Min	Max	Min
±0.5	29.52	20.27	27.95	20.75	28.63	22.12	18.56	22.19
±1.0	52.40	40.75	50.48	40.41	51.37	45.21	34.38	40.68
±1.5	69.59	58.90	66.44	58.36	66.71	64.04	49.45	59.04
±2.0	80.68	76.16	79.04	74.59	80.34	78.84	62.74	76.85
±2.5	87.60	87.12	87.40	86.85	87.19	88.90	72.19	87.53
±3.0	91.99	93.84	90.75	93.56	91.78	94.52	80.41	93.49
±3.5	94.52	97.53	93.77	97.81	94.38	97.33	86.64	97.19
±4.0	96.51	98.97	96.03	98.97	96.30	98.84	91.16	98.56
±5.0	98.77	99.79	98.42	99.86	98.77	99.86	95.68	99.79

W^{ih} = A weight matrix $\left[W_{u,v}^{ih}\right]_{n_i \times n_h}$ for input to hidden layer

W^{ho} = A weight matrix $\left[W_{u,v}^{ho}\right]_{n_h \times n_o}$ for hidden to output layer

$F(x)$ = $\dfrac{1}{1+e^{-x}}$, an activation function.

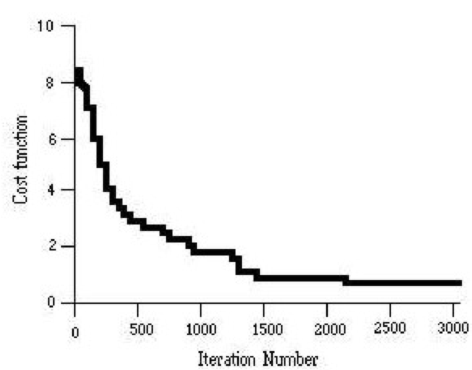
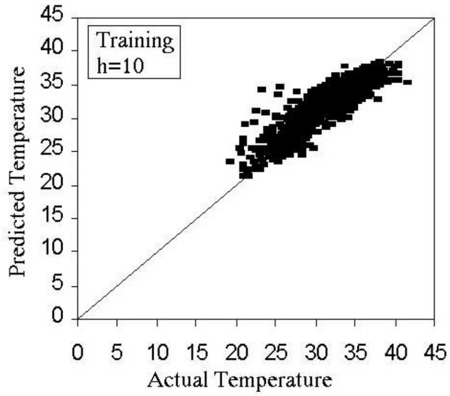
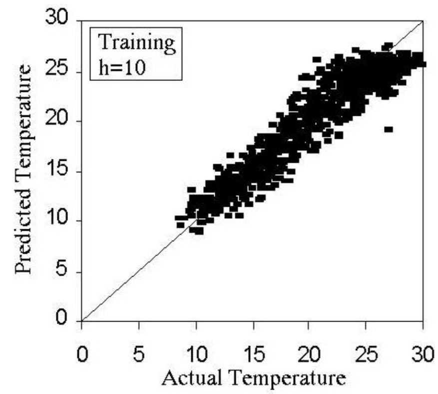


Fig. 4. Cost function [or sum of square errors] *versus* iteration number when the number of hidden layer nodes is 20

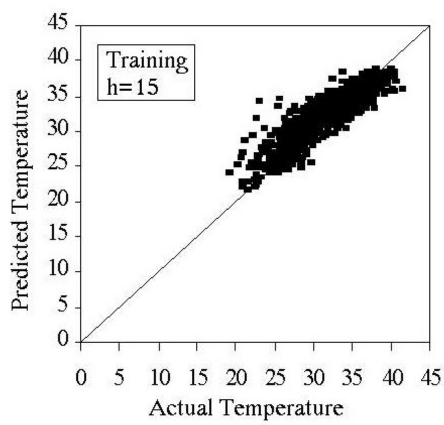


(a) Maximum

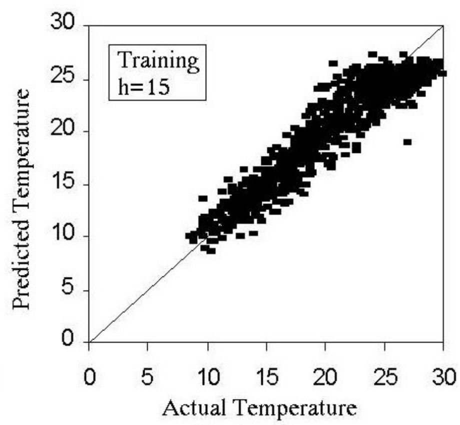


(b) Minimum

Figs. 5(a&b). Actual temperature *versus* its predicted temperature with training sample for h = 10 (a) Maximum temperature and (b) Minimum temperature

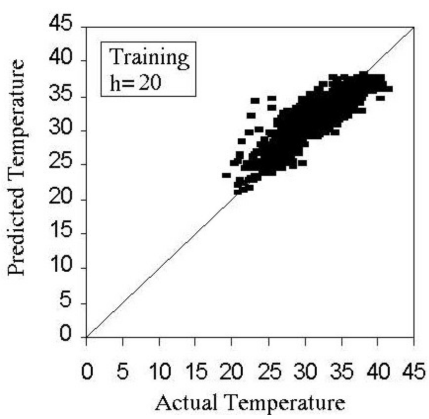


(a) Maximum

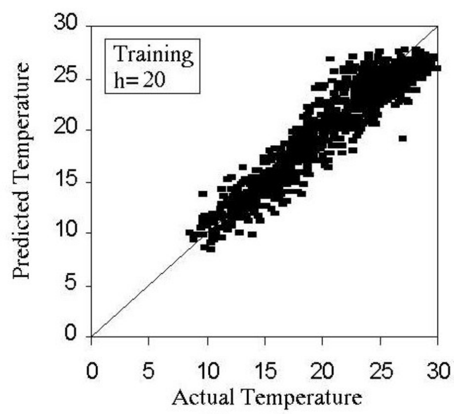


(b) Minimum

Figs. 6(a&b). Actual temperature *versus* its predicted temperature with training sample for h = 15 (a) Maximum temperature and (b) Minimum temperature

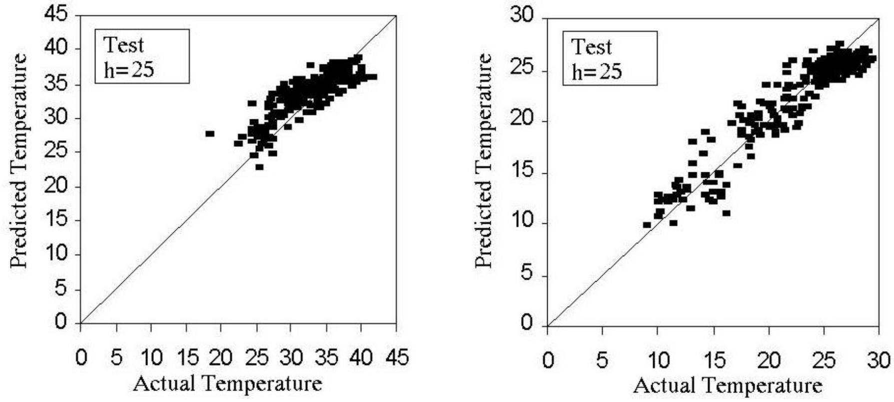


(a) Maximum



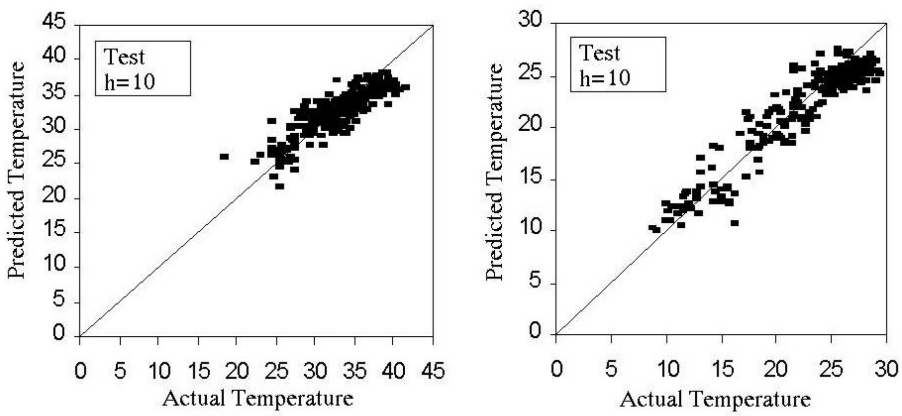
(b) Minimum

Figs. 7(a&b). Actual temperature *versus* its predicted temperature with training sample for h = 20 (a) Maximum temperature and (b) Minimum temperature



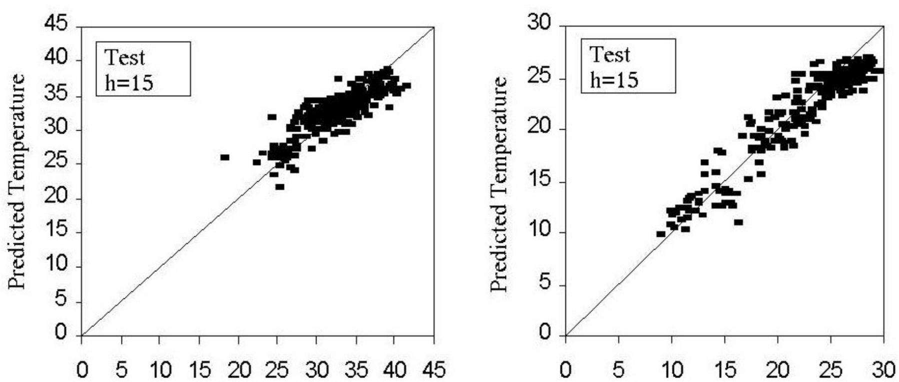
(a) Maximum (b) Minimum

Figs. 8(a&b). Actual temperature *versus* its predicted temperature with training sample for h = 25 (a) Maximum temperature and (b) Minimum temperature



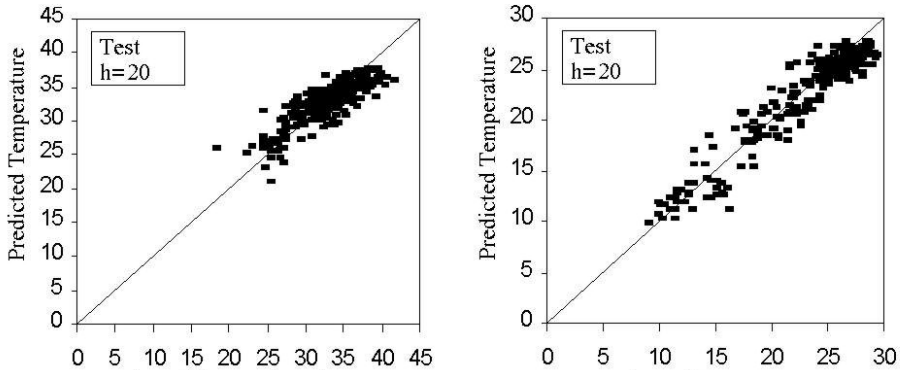
(a) Maximum (b) Minimum

Figs. 9(a&b). Actual temperature *versus* its predicted temperature with test sample for h = 10 (a) Maximum temperature and (b) Minimum temperature



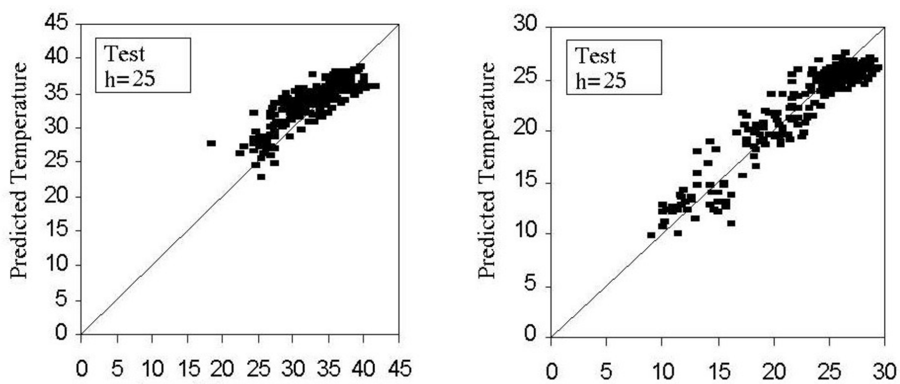
(a) Maximum (b) Minimum

Figs. 10(a&b). Actual temperature *versus* its predicted temperature with test sample for h = 15 (a) Maximum temperature and (b) Minimum temperature



(a) Maximum (b) Minimum

Figs. 11(a&b). Actual temperature *versus* its predicted temperature with test sample for h = 20 (a) Maximum temperature and (b) Minimum temperature



(a) Maximum (b) Minimum

Figs. 12(a&b). Actual temperature *versus* its predicted temperature with test sample for h = 25 (a) Maximum temperature and (b) Minimum temperature

After training the weight matrices W^{ih} and W^{ho} are used for prediction

Step 1 : Compute the hidden-layer output for forecasting as follows :

$h = F(iW^{ih})$.

Here, $h_k = F(a_k) = \dfrac{1}{1+e^{-ak}}$

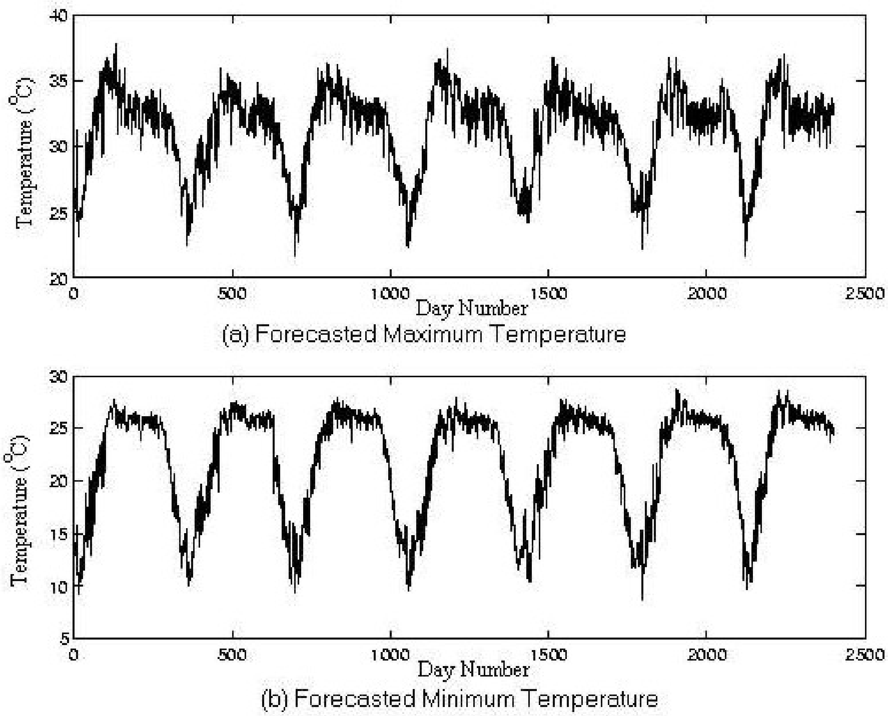
where $a_k = \sum_{l=1}^{n_i} W_{lk}^{ih} i_l$, for k = 1, 2, …., n_h

Step 2 : Compute normalized forecasted output as follows :

$O = F(hW^{ho})$

Here, $O_k = F(b_k) = \dfrac{1}{1+e^{-bk}}$, where,

$b_k = \sum_{l=1}^{nh} W_{lk}^{h_0} h_l$, for k = 1,2,..., n_o



Figs. 13(a&b). Forecasted temperature distribution over the period 1989-95 for Dum Dum station (a) Maximum temperature and (b) Minimum temperature

Step 3 : Compute forecasted output from the normalized forecasted output by multiplying its scale factor, where the scale factor is depends on maximum and minimum values of each output components.

4. Results and discussion

Initial weights between the layers of a network are random. A training sample is presented to the network. It computes the errors with respect to its target value and modifies the weights accordingly. This process continues for all training sample. Compute sum of square error for the set of training sample, if the sum of square error is in the limit then the process of training is over otherwise it continues with the modified weights.

The neural networks with number of neurons 10, 15, 20 and 25 respectively in the hidden layer are trained with the same set of training sample.

A set of test as well as training data are presented to each of the trained network to generate the prediction results. Cumulative percentage frequency tables using training and test data are presented in Tables 1 and 2 respectively. It is found that the neural net with 25 nodes in the hidden layer achieved a poor prediction results than a neural net with 10, 15 or 20 nodes in the hidden layer.

Also Tables 1 and 2 shows that the predicted minimum temperatures are more accurate than the maximum. The reason behind may be that, due to the maximum temperatures fluctuate more than the minimum ones. The results imply there may be a possibility of getting good results from neural networks used to learn the maximum and minimum temperature separately. Because in this case minimum temperature learning is affected by the maximum temperature learning to some extent.

Fig. 4 shows that the nature of changes of cost function (*i.e.* sum of square error) with the number of iterations.

Figs. 5-8 shows the scatter plot between the actual temperature those are used for training the neural net and the corresponding predicted temperature for h = 10 (Fig. 5), h = 15 (Fig. 6), h = 20 (Fig. 7) and h = 25 (Fig. 8). Here in each figure (Figs. 5-8) (a) is showing the case of maximum temperature and every figure (Figs. 5-8) (b) is showing the case of minimum temperature. Similarly the same kind of results are shown in Figs. 9-12 for test sample. The diagonal line in each Figs. 5-12 indicates the base line of the deviations between the actual and the prediction temperatures.

This model works well when number of nodes in the hidden layer is 20 and its average error is normally less than ± 2° for about 80% of test and training cases. Also we have implemented the traditional linear regression method and described that the neural network method is better because the sigmoid activation function in MLP has a higher capability of approximating a function, even if it is in non-linear nature.

Fig. 2 shows temporal variation of maximum and minimum temperatures in actual and Fig. 13 shows the predicted values for h = 15. Comparison of actual and predicted temperature distributions indicates that, predicted temperatures for drier months are much more accurate than that of the rainy or cloudy months. It is possible to improve the forecasting technique by a suitable training scheme and sufficient training data. Possible ways are (*i*) seasonal learning (*ii*) inclusion of upper air wind pattern and cloud coverage.

5. Conclusion

Meteorological prediction is highly complicated as it involves a large number of variables, some of which are not estimated directly. Moreover, the different variables are not always uniformly changed over the consecutive years but some are changed rapidly because of the different ecological and meteorological conditions. Better prediction accuracy may be achieved if one uses more and more of the above meteorological changes.

On the other side still a better prediction results may be achieved by changing the architecture of the neural network such as (*i*) Neural Network using radial basis function (*ii*) Partitioning the training sample (*iii*) significant feature selection by using a suitable feature analysis method and (*iv*) Network fusion [Ueda 2000].

Acknowledgements

The authors are grateful to the Director, ISI for providing infrastructure to carry out work and to the Director General, India Meteorological Department for providing data. Also the authors gratefully acknowledge the reviewers for their valuable comments.

References

Haykin S., 1994, "Neural networks a comprehensive foundation", Macmillan college publishing , New York.

LeGum, Y., 1988, "A theoretical framework for back-propagation", Proceedings of the 1988 Neural Network Model Summer School.

Rumelhart, David, E., Hinton, Geoffrey and Williams, Ronald, J., 1986, "Learning representations by backpropagating errors", *Nature,* 323.

Salehfar, H. and Benson, S. A., 1998, "Electric utility coal quality analysis using artificial neural network techniques", *Neurocomputing*, **23**, 195-206.

Tsintikidis, D., Haferman, J. L., Anagnostou, E. N., Krajewski, W. F. and Smith, T. F., 1997, "A neural network approach to estimating rainfall from spaceborne microwave data", *IEEE Geoscience and remote sensing,* **35**, 5, l079-1093.

Ueda N., 2000, "Optimal linear combination of neural networks for improving classification performance", IEEE Tr. on PAMI, **22**, 2, 207-215.

Werbos, Paul, 1974, "Beyond regression: New tools for prediction and analysis in the behavioral sciences", Ph.D. dis., Harvard University.