Monthly Mean Temperature Prediction Based on a Multi–level Mapping Model of Neural Network BP Type①

Yan Shaojin (严绍瑾), *Peng Yongqing* (彭永清) and *Guo Guang* (郭 光)
Nanjing Institute of Meteorology, Nanjing 210044
Received August 1, 1994; revised November 5, 1994

ABSTRACT

In terms of 34–year monthly mean temperature series in 1946–1979, the multi–level mapping model of neural network BP type was applied to calculate the system's fractual dimension $D_0 = 2.8$, leading to a three–level model of this type with $i \times j = 3 \times 2, k = 1$, and the 1980 monthly mean temperture prediction on a long–term basis were prepared by steadily modifying the weighting coefficient, making for the correlation coefficient of 97% with the measurements. Furthermore, the weighting parameter was modified for each month of 1980 by means of observations, therefore constrcuting monthly mean temperature forecasts from January to December of the year, reaching the correlation of 99.9% with the measurements. Likewise, the resulting 1981 monthly predictions on a long–range basis with 1946–1980 corresponding records yielded the correlation of 98% and the month–to month forecasts of 99.4%.

Key words: Neural network, BP–type multilevel mapping model, Monthly mean temperature prediction

I. INTRODUCTION

Meteorology is concerned with atmospheric state and its change and the ultimate goal is to offer weather information and prediction over a wide range of spatial / temporal scales through intensive and extensive research for socioeconomic prosperity. It has been a major concern of meteorologists to prepare element forecasts on a variety of scales, among which monthly mean temperature and monthly rainfall fall into the scope of long–term forecasting and bear intimate relation to possible drought and flood. The authors (1993) proposed a scheme of a dynamic system's reconstruction based on phase space continuation of one–dimensional time series for monthly mean temperature prediction. This scenario is no doubt applicable to the case of monthly precipitation. But in this article we investigate the problem along a new line. The description of an intricate system by nonlinear interaction between elements with the consequent extensions—results derived recently in neural network theory–undoubtedly provides a useful approach to nonlinear prediction research. For this reason, an attempt was made to explore monthly mean temperature forecasting by virtue of the neural network BP–type model.

II. OPERATION OF THE MODEL

Neural network is a large–scale dynamic system,characterized largely by the ability to show the overall effect of the network and to accomplish the parallel distribution of complicated

①This work was funded by the National Natural Science Foundation of China.

information on a large scale. There exists a multitude of such models and only the BP–type multi–level mapping version that is now in widespread use is considered in this work, consisting of a hidden–level element responsible for feeding back result onto the hidden level with the aid of least mean deviation to change the weighting coefficient matrix so as to realize the "information input to product output" operation as expected. The model is of usefulness to nonlinear forecasting. In Fig. 1, X stands for the signal input level with X_i being i–th component of the input; Y for the hidden level assuming j number of $(j < i)$; Z for the input level with k number of $(k < j)$, and k set to be unity for simplicity. We denote $W_{ij}(W_{jk})$ as the weighting coefficient between the input and hidden (the hidden and output) levels. It is seen from Fig. 1 that there exist nodes of the input, output and hidden levels. The input signal is first propagated forward to the hidden level node and, after nonlinear interaction (usually represented by squashing function), the signal runs from the hidden to output nodes which, subjected to nonlinear interaction again, comes out as the final result. The squashing function (SF) normally takes the form of S–type like $f(x) = \frac{1}{1+e^{-x}}$ which is differentiatible on a continuous basis, of monotonous increment and saturation.

The components of neural elements of the same level are not interrelated and the elements of adjacent levels are associated by error revision of the weighting coefficient, i.e., by

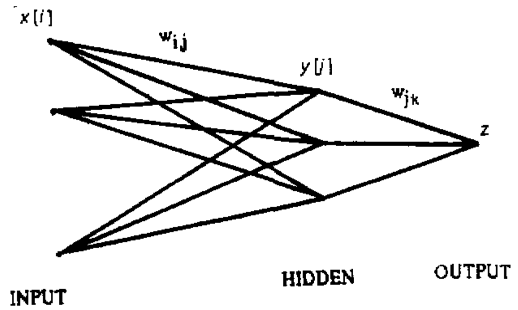

Fig. 1. Neural network 3–level BP–type model.

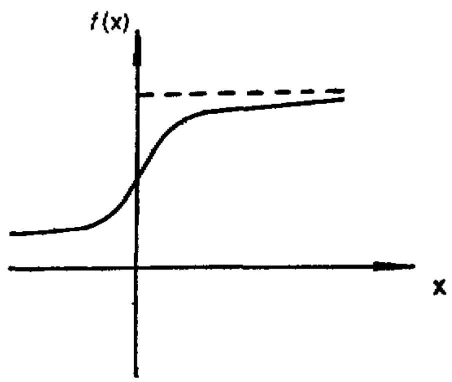

Fig. 2. Representation of S–type function.

using the difference between the expectation and actual output of the elements for modifying the parameter of the coefficient to reduce the difference.

Calculations are as shown in Fig. 1. The hidden level Y_j is found by the input level X_i multiplied by the weighting W_{ij} yielding

$$Y_j = \sum_i x_i W_{ij} \ , \tag{1}$$

and the SF of Y_j in the form $\tilde{Y}_j$

$$\tilde{Y}_j = \frac{1}{1 + e^{-y_j}} \ . \tag{2}$$

The information transmitted from the hidden to output level is obtained by the hidden–level $\tilde{Y}_j$ multiplied by the weighting W_{jk},

$$Z_k = \sum_j W_{jk} \tilde{y}_j \tag{3}$$

for which the SF is in the form

$$\tilde{Z}_k = \frac{1}{1 + e^{-z_k}} \tag{4}$$

with the expected output assumed to be d_k, we proceed as follows.

The weighting coefficient is revised using the back propagation of squared error. Set the squared error function to have the form

$$E_k = \frac{1}{2}(\tilde{Z}_k - d_k)^2 \ . \tag{5}$$

Now, let ΔW_1 be the revised weighting coefficient (WC) W_{ij} between the input and hideden layers

$$\triangle W_1 = -\frac{\partial E_k}{\partial W_1(i,j)} \tag{6}$$

and ΔW_2 be the coefficient W_{jk} between the hidden and output levels

$$\triangle W_2 = -\frac{\partial E_k}{\partial W_2(j,k)} \ . \tag{7}$$

For each calculation of the time series, the WC needs to be revised. Therefore, the WC of the n–th step has the revision as

$$\Delta W_1(i,j)_n = \sum_{n=1}^{n} \Delta W_1(i,j)_n \ , \tag{8}$$

$$\Delta W_2(j,k)_n = \sum_{n=1}^{n} \Delta W_2(j,k)_n \ , \tag{9}$$

where $n = 1, 2, \cdots N - d$, with N denoting the time series length and d the number of independent coordinates describing the study system, defined by its fractual dimensionality D_0. After $(n - d)$ calculations, N number of data in the series have passed through the operation, which is referred to as a cycle. Generally, hundreds, even thousands or more of such cycles are required to reach the needed value of the WC.

The WC has the form at the $(l + 1)$ cycle

$$W_1(i,j)_{l+1} = W_1(i,j)_l + \eta(\sum_{n=1}^{N-d} -\frac{\partial E_k}{\partial W_{ij}})_n \ , \tag{10}$$

$$W_2(j,k)_{l+1} = W_2(j,k)_l + \eta(\sum_{n=1}^{N-d} -\frac{\partial E_k}{\partial W_{kj}})_n \ , \tag{11}$$

where η represents the learning rate.

III. NEURAL NETWORK PREDICTION OF MONTHLY MEAN TEMPERATURE

1. Used in this study are the datasets of monthly mean temperature spanning 1946–1980 (Central Meteorological Bureau, 1974), with the 1946–1979 data (N = 408) as the basis for the model construction and WC revision, which are applied to forecast the monthly temperature of 1980, followed by comparison of the predictions with actual measurements to assess the forecasting.

The data $0 < x(t) < 1$ are used and part of it is illustrated in Fig. 3.

2. Following Yan and Peng (1993), the lagged continuation phase space $\tau = 2$ is used to compute the fractual dimensionality $D_0 = 2.8$ with the known basic time series (N = 408 for the first 34 years). Thus we can predict the number of degrees of freedom (i. e., independent coordinates) of the given dynamic system, with the bottom limit of $|D_0 + 1| = 3$ and the upper of $|2D_0 + 1| = 6$ so that the input layer of the model will assume $i = 3, 4, 5, 6$.

3. *Initial values of W_1 and W_2 ($n = 1$) are taken by means of a random generator*, with $(\Delta W_1)_{n=1} = 0$, $(\Delta W_2)_{n=1}$ and $\eta = 0.15$. After a large number of WC revision the relative error of 0.001 should be reached between the 408–th output and observation and then the resulting WC is utilized to predict the monthly mean temperature from January to December, 1980.

4. Since $k = 1$ and $i = 3, 4, 5, 6$ are used for the model for the study system, $i \times j$ can take 3×2, 4×2, 5×2, 5×3, 6×3, etc. From the calculations one can see that among these versions, the 3×2 model prediction is closest to the measurement. Table 1 presents the 1980 monthly predictions on a long–range basis in terms of the 3×2 model with different random initial values W_1 and W_2 applied.

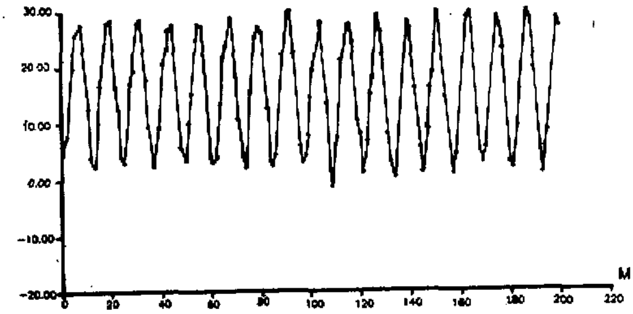

Fig. 3. Part of the 1946–80 monthly mean temperatures.

Table 1. 1980 January–December Monthly Mean Temperatures with Different Random Initial Values Used in the 3 × 2 Model

	J	F	M	A	M	J	J	A	S	O	N	D	number of cycles
obser.	2.1	3.2	7.0	13.9	20.2	24.4	26.3	24.7	21.3	16.7	11.9	2.1	
pred.1	2.0	2.9	6.6	14.0	21.2	26.5	26.8	23.7	18.1	14.6	10.4	4.6	402
pred.2	1.9	2.8	6.3	14.1	21.8	26.7	26.8	23.1	17.6	14.8	10.7	4.7	286
pred.3	2.0	2.9	6.5	13.9	21.3	26.7	26.8	23.9	17.7	14.2	10.0	4.7	367

(Cf. Fig. 4).

Note that for prediction 1, $W_1(i,j)=\begin{pmatrix} 3.28416, & -2.09867, & -5.37359 \\ -3.76703, & -0.01469, & 2.69419 \end{pmatrix}$

$$W_2(j,k)=(-5.24629,\ \ 1.58751)$$

for prediction 2, $W_1(i,j)=\begin{pmatrix} 3.16490, & -2.03741, & -5.29458 \\ -3.63203, & -0.01197, & 2.58302 \end{pmatrix}$

$$W_2(j,k)=(-5.20791,\ \ 1.48307)$$

and for prediction 3, $W_1(i,j)=\begin{pmatrix} 3.26419, & -2.08840, & -5.34298 \\ -3.74109, & -0.01371, & 2.69705 \end{pmatrix}$

$$W_2(j,k)=(-5.23790,\ \ 1.56939)$$

Note also that the solid line denotes a measurements–connecting curve and the three broken lines the results of predictions 1 through 3.

(Cf. Fig. 5)

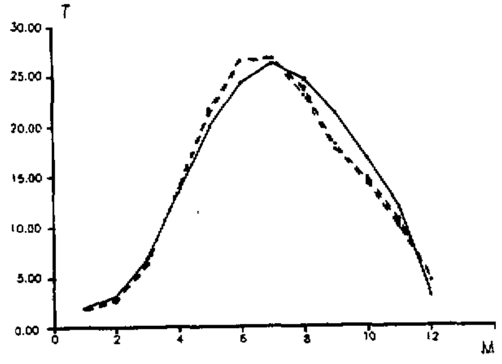

Fig. 4. Comparison of the 3 × 2 model predicted monthly temperatures to observations (solid line) with the results of predictions 1 to 3 shown by broken lines.

With the 35–years monthly temperatures in 1946–1980 as the basic dataset in virtue of the scheme just used, the monthly temperatures of 1981 are tabulated as follows.

Table 2. Long–term Prediction of Monthly Mean Temperatures of 1981

	J	F	M	A	M	J	J	A	S	O	N	D
obser.	0.0	4.0	10.3	15.0	20.7	25.0	28.4	27.4	22.0	14.8	8.8	3.3
pred.	0.3	3.5	9.6	14.6	20.2	24.4	27.3	25.8	19.7	13.5	9.9	4.8

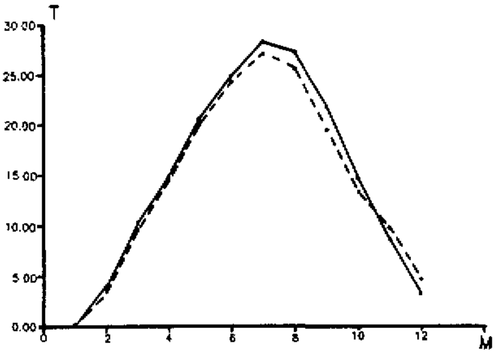

Fig. 5. Monthly temperatures of 1981 with the observed (predicted) magnitudes given by solid (broken) line.

5. To improve the accuracy of forecasting and serve as a correction, the 3 × 2 model is used with the WC modified on a monthly basis for monthly temperature prediction and the results are summarized in Table 3.

Table 3. 1980 Monthly Temperature Prediction with WC Modified on a Monthly Basis

	J	F	M	A	M	J	J	A	S	O	N	D
obser.	2.1	3.2	7.0	13.9	20.2	24.4	26.3	24.7	21.3	16.7	11.9	3.1
pred.4	2.0	3.2	7.0	13.9	19.9	24.3	26.5	24.6	21.5	16.9	12.3	3.5
pred.5	2.0	3.1	7.1	14.0	20.0	24.3	26.4	24.6	21.5	16.8	12.2	3.4
pred.6	2.0	3.2	7.0	13.9	20.2	24.4	26.3	24.7	21.3	16.7	11.9	3.1

These values of this table are utilized to prepare Fig. 6 where solid line denotes the observed temperatures and broken lines the results from predictions 4,5 and 6. Because of their close agreement these lines are hardly discernible.

Table 4. 1981 Monthly Mean Temperature Prediction

	J	F	M	A	M	J	J	A	S	O	N	D
obser.	0.0	4.0	10.3	15.0	20.7	25.0	28.4	27.4	22.0	14.8	8.8	3.3
pred.	0.3	3.8	10.2	14.7	20.6	24.8	28.1	27.2	21.7	15.0	9.1	3.6

(Cf. Fig. 7)

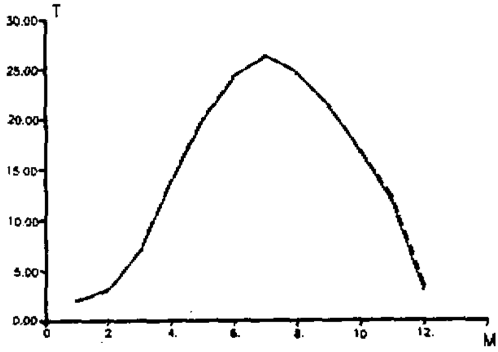

Fig. 6. Comparison of observations (solid line) to results from predictions 4,5 and 6 (broken lines).

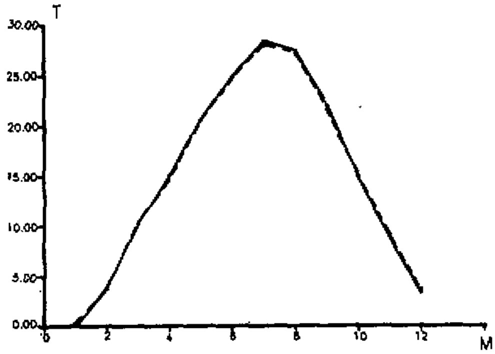

Fig. 7. Monthly temperature prediction (broken line) in comparison to the observed (solid line).

IV. ANALYSIS AND DISCUSSION

It is an attempt to predict monthly mean temperature with the neural network BP–type three layers mapping model. From the foregoing calculations we come to the following.

1. The fractual dimensionality $D_0 = 2.8$ is found from the monthly mean temperature series of Nanjing. With the determination of $k = 1$, the form of the three level models depends on $i \times j$, for which a range of versions (e. g., 3×2, 4×2, 5×2, 5×3, 6×3, etc.) is allowable. After a bulk of computations the 3×2 model is found optimal with its prediction closest to the observed, suggesting that the attractor of the system is set in a 3D model.

Table 5. Error Analysis of Long–range Prediction of 1980 Monthly Mean Temperatures

	MAE*	MRE	MSD	CC
pred.1	1.40000000	0.01861096	1.85472400	0.97933520
pred.2	1.74166700	0.03518327	2.24518000	0.97063240
pred.3	1.51666700	0.02230451	2.05588600	0.97490900

* MAE = mean absolute error; MRE = mean relative error;
MSD = mean squared deviation; CC = correlation coefficient.

2. For the BP three layer version, the number of components of the input layer can use its fractual dimension as the first guess. However, selection of the number of hidden nodes can be done from experience, not following any theory.

3. Error analyses are done of the results from predictions 1,2 and 3 (for long-term prediction of 1980 monthly mean temperatures) and 4,5 and 6 (for 1980 monthly mean temperature forecasting) and from both types of the 1981 predictions.

Table 6. As in Table 5 Except for 1981

MAE	MRE	MSD	CC
0.50833333	0.02472585	1.31750000	0.98161859

Table 7. Error Analysis of 1980 Monthly Mean Temperature Predictions

	MAE	MRE	MSD	CC
pred.4	0.16666690	0.0100832	0.21602500	0.99972590
pred.5	0.15000020	0.00547727	0.16832530	0.99983250
pred.6	0.15000010	0.00713229	0.16832520	0.99984070

Table 8. As in Table 7 but for 1981

MAE	MRE	MSD	CC
0.01500000	0.0269372	0.24949494	0.99447204

The above analyses indicate that the long-term predictions of 1980 and 1981 monthly mean temperatures have correlations of 97% and 98%, respectively, with the observations, and with the corresponding figures of 99.9% and 99.4% for the month-to month forecasts, which meet the needs of prediction.

4. For predictions 1, 2 and 3 the 3×2 version was based on the resulting sequences from W_1 and W_2 of different initial values ($n=1$) determined by a random generator. As shown in Table 1, initial values of the weighting coefficients were obtained through different number of cycles for modification to reach proper accuracy of forecast, but nevertheless their results were comparable. This suggests that initial weighting magnitudes are allowed to be taken stochastically if the model and forecasting accuracy are defined. And the same was true of predictions 4,5 and 6. Our study shows that the neural network BP-type multilevel model is a highly useful algorithm and a nonlinear optimization in which the usual gradient descent technique is used, to which the addition of hidden level nodes will give rise to the increase in the number of modifiable parameters, leading to still more accurate solutions. The neural network can be regarded as an input to output mapping that means a highly nonlinear process. If the number of input and output nodes is set to be i and k, respectively, then the network will be represented by the R^i to R^k mapping.

REFERENCES

Central Meteorological Bureau of China (1974), Compilation of Temperature and Rainfall over China 118-120 (in Chinese).

Jiao Licheng (1990), *Theory on Neural Network System*, Xi'an Electronics Sci / Tech. Univ. Press, Xi'an 34-36 (in Chinese).

Peng Yonqing, Zhu Yufeng and Yan Shaojin (1993), Preliminary study on dynamic systems reconstruction with 1D time series, *J. Appl. Meteor.*, 4: (suppl. iss.) 31-38 (in Chinese).

Yan Shaojin and Peng Yongqing (1993), *Chaos Theory and Atmospheric Sciences*, China Meteorological Press, Beijing, pp. 111-116 (in Chinese).