

Study on Ann-Based Multi-Step Prediction Model of Short-Term Climatic Variation

Jin Long (金龙), Ju Weimin (居为民)

Jiangsu Research Institute of Meteorological Sciences, Nanjing, 210008

and Miao Qilong (缪启龙)

Nanjing Institute of Meteorology, Nanjing, 210044

(Received December 21, 1998; revised August 10, 1999)

ABSTRACT

In the context of 1905–1995 series from Nanjing and Hangzhou, study is undertaken of establishing a predictive model of annual mean temperature in 1996–2005 to come over the Changjiang (Yangtze River) delta region through mean generating function and artificial neural network in combination. Results show that the established model yields mean error of 0.45°C for their absolute values of annual mean temperature from 10 yearly independent samples (1986–1995) and the difference between the mean predictions and related measurements is 0.156°C. The developed model is found superior to a mean generating function regression model both in historical data fitting and independent sample prediction.

Key words: Climate trend prediction, Mean generating function (MGF), Artificial neural network (ANN), Annual mean temperature (AMT)

1. Introduction

Owing to integrative effects of climate inherent variation and human activities, climate anomaly occurs frequently at a full range of spatial and temporal scales, exerting innegligible impacts on socioeconomic development on a sustainable basis of countries and regions all over the world. To alleviate the disastrous influence it is urgent to carry out the research of long-term climatic prediction schemes for operational purposes. In this respect study of numerical methods has been recently made enormous progresses and some predictive models have shown higher skills and great prospect. On the other hand, numerical prediction at greater than a monthly and a seasonal scale is under experiment and statistical techniques are predominant in long-range prediction on an operational basis in the meteorological community of the world (Zhou, 1993). The widely-accepted statistical techniques are confined to a monthly and a seasonal scale, beyond which (e.g., a yearly and a decade scale) climate trend prediction is just tentative in most studies. The present work is built on the artificial neural network (ANN) and mean generating function (MGF) schemes to make prediction of 1996–2005 annual mean temperature in the Yangtze River delta area to make progress in the research of predictions at still longer time scales.

2. Predictive scheme

It is generally difficult to find out meaningful predictors that aid in the prognosis

of annual mean temperature (AMT) on a long term basis such that the prediction has to depend essentially on the analysis of time series and periods in the main. At present, the MGF analysis of such series (Cao et al., 1993) is employed in multi-step prediction study for a decade ahead in addition to conventional autoregression models on temporal sequences which are, however, used dominantly in long-range prediction of monthly mean temperature due to the fact that they are normally based on a larger sample size for establishment but the length of AMT sequences is limited in general (Hu et al., 1994). Moreover, the establishment of multi-step prediction models on MGF on the series was largely by adopting linear regression of the MGFs selected (Wei and Cao, 1994). In many cases, however, the selection depends mainly on the calculation of the entropy of classified predictive information and linear correlation. In fact, whether there arises linear correlation alone between MGF and predictand may differ from one element to another. Instead, the ANN scheme has focus on nonlinear mapping relation between model input and output without the need to know the internal structure of a prediction system in advance. Therefore, an attempt is made here to predict AMT with the aid of ANN combined with MGF technique for model establishment.

2.1 MGF scheme

Given an observational sequence of n samples

$$x(t) = \{x(1), x(2), \dots, x(n)\}, \quad (1)$$

after (1) is normalized, we proceed to find its MGFs through

$$\bar{x}_l(i) = \frac{1}{n_l} \sum_{j=0}^{n_l-1} x(i+j), \quad (2)$$

where $i = 1, 2, \dots, l$ with $1 \leq l \leq M$ and $n_l = \text{INT}(\frac{n}{l})$ where generally $M = \text{INT}(\frac{n}{2})$ or $\text{INT}(\frac{n}{3})$ on the basis of a sample length in which INT stands for the integral to be taken. To get more MGFs for choice we deal with (1) by first- and second-order difference schemes for their corresponding sequences in terms of

$$\Delta x(t) = x(t+1) - x(t), \quad t = 1, 2, \dots, n-1 \quad (3)$$

$$\Delta \Delta x(t) = \Delta^2 x(t) = \Delta x(t+1) - \Delta x(t), \quad t = 1, 2, \dots, n-2 \quad (4)$$

followed by finding the MGFs of the difference series using (2) and the resulting MGF sequences (also the original series) are extended in a periodic manner by

$$f_l(t) = \bar{x}_l(i - \text{INT}(\frac{t-1}{l})l) \quad (5)$$

with $i = 1, 2, \dots, N$, and $l = 1, 2, \dots, M$. Such that we get $3 \times M$ extended series, consisting of the original, first- and second-order difference sequences. Then calculation is done of the entropy of classified information of these extension MGF sequences by

$$2I = 2[\sum_{i=1}^I \sum_{j=1}^I n_{ij} \ln n_{ij} + N \ln n - (\sum_{i=1}^I n_i \ln n_i + \sum_{j=1}^I n_j \ln n_j)], \quad (6)$$

where i denotes the number of categories of predicted trend, n_{ij} the number of possible events of i and j kind on a contingency table. Then MGFs with high entropy are sorted out to construct a learning matrix of ANN.

2.2 ANN Principle and learning algorithm

Today, the ANN scheme has been applied to more and more fields of prediction (Lee et al., 1992; McCann, 1992). It regards an ANN learning matrix from a prediction system as nonlinear mapping between input and output, and the mapping relation is realized without the need to know in advance the internal structure of the study system but by making its limited number of samples subjected to training in order to simulate the relationship between input and output related to the structure. The reader referred to Jin et al. (1997) for detailed derivation of forward feedback ANN learning algorithm used here, whose steps for computational purposes are introduced below.

Set the input and expected output to be A_k and C_k ($k=1,2,\dots,m$), respectively, of the learning sample and give stochastically a group of small values V_{hi} as initial connection weighing ones for the input to the hidden layer, a set of connection weighing functions W_{ij} for the hidden to the output layer, and the threshold values θ_i and γ_j of the hidden and output layer units, followed by finding A_k and C_k ($k=1,2,\dots,m$) as follows.

1) On the basis of a connection weighing matrix (initially a group of randomly small values) and the input of a learning sample we find new activation values of the hidden layer by

$$b_i = f\left(\sum_{h=1}^n a_h V_{hi} + \theta_i\right) \quad (7)$$

with $i=1,2,\dots,p$ and the function for activation is the sigmoid form

$$f(x) = \frac{1}{1 + e^{-x}}. \quad (8)$$

2) Activations of the output layer units are found by

$$C_j = f\left(\sum_{i=1}^p W_{ij} b_i + \gamma_j\right), \quad (9)$$

where $j=1,2,\dots,q$ (initial W_{ij} is given as a set of random small values).

3) General errors of output-layer units are found by

$$d_j = C_j(1 - C_j)(C_j^k - C_j), \quad (10)$$

in which $j=1,2,\dots,q$ and C_j^k = expected output of output layer unit j .

4) Error of a hidden-layer unit relative to d_j is found by

$$e_i = b_i(1 - b_i) \sum_{j=1}^q W_{ij} d_j \quad (11)$$

with $i=1,2,\dots,p$.

5) Connection weighing values for hidden to output layer units are adjusted by

$$\Delta W_{ij} = \alpha b_i d_j, \quad (12)$$

in which $i=1,2,\dots,p$, $j=1,2,\dots,q$ and α is the learning factor ($0 < \alpha < 1$).

6) Threshold values of output-layer units are regulated by

$$\Delta \gamma_j = \alpha d_j \quad (13)$$

with $j=1,2,\dots,q$.

7) Connection weighing values of input to hidden layer units are adjusted by

$$\Delta V_{hi} = \beta a_h e_i, \quad (14)$$

in which $h = 1, 2, \dots, n$, $i = 1, 2, \dots, p$ and β is the momentum factor ($0 < \beta < 1$).

8) Threshold values of hidden-layer units are regulated by

$$\Delta \theta_i = \beta e_i \quad (15)$$

with $i = 1, 2, \dots, p$.

9) Steps are repeated from (7) to (15) just mentioned until $j = 1, 2, \dots, q$ and $k = 1, 2, \dots, m$ to find the error between actual and expected output and training is ended when the output error of all samples is below a prescribed convergent level. In training, the ANN weighing is adjusted stepwise and the nonlinear relation is determined between model input and output in accordance with the trained sample. Now we can predict the future state of a predictand in terms of the connection weighing coefficients from the input to hidden layer and therefrom to the output layer, and corresponding threshold values.

3. Climatic prediction for a decade ahead (1996–2005) over the Changjiang (Yangtze River) delta plains

To uncover the variation of annual mean temperature in the delta region as close to reality as possible with the effect of urban thermal islands considered, investigation is performed of 1905–1995 AMT sequences, indicating that in 1905–1955 the AMT is 0.07°C higher (0.79°C lower) at Shanghai than at Nanjing (Hangzhou) while the corresponding figures in 1956–1995 are 0.53 and -0.41°C (see Table 1). Comparison shows that the urban thermal island effect of Shanghai is pronounced, a result that is similar to that of Zhou and Zhang (1985). It is known that AMT conditions in different parts of a climatic region are in higher correlation with each other. So we think that it is better to employ the Nanjing and Hangzhou AMT series in 1906–1995 as the sample sequences for model establishment. As stated earlier, our aim is to explore a model construction for predicting AMT in 1996–2005 in the delta area in terms of ANN-MGF techniques in combination. In general, climatic variation is highly uncertain over a period as short as a decade.

Table 1. 1906–1995 AMT($^\circ\text{C}$) at Shanghai versus those at Nanjing and Hangzhou

Year	Shanghai	Nanjing	ΔT	Shanghai	Hangzhou	ΔT
1906–1955	15.55	15.48	0.07	15.55	16.34	-0.79
1956–1995	15.84	15.31	0.53	15.84	16.25	-0.41

The first thing is to examine results from the ANN-MGF model. Therefore, the model establishment is just based on the Nanjing and Hangzhou AMT series in 1916–1985 ($n = 70$, with the exclusion of Shanghai because of its considerable thermal island effect) and their 1986–1995 series are utilized as independent samples, followed by verifying the predictive ability. The $n = 70$ series are dealt with for MGFs from themselves and the resulting first- and second-order difference sequences and for their periodic extension ($M = \frac{N}{2} \approx 34$) through (2) to (5), leading to 102 MGF extension series, each of which is made subjected to the calculation of entropy of classified information in terms of (6), from which MGFs with higher entropy are sorted out, covering all the three sequences used. Afterwards, 9 MGF sequences

selected are put to construct an ANN learning matrix, and to meet the needs of the conditions of ANN knot function the matrix is first normalized by

$$k_i = \frac{z_i - x}{y - x}, \quad (16)$$

where k_i represents normalized data for the matrix, z_i the measurement, x and y the coefficients to be determined. The forward-feedback ANN model used here for model establishment has its knot function of sigmoid form (8) with its values ranging from 0 to 1. But, considering the limited number of samples used for model establishment the domain is set between 0.1 and 0.9 for the normalized data. Thus, x and y are found by

$$\begin{aligned} a - x &= 0.1(y - x) \\ b - y &= 0.9(y - x), \end{aligned} \quad (17)$$

where a and b denote a maximum and a minimum in the sample series, respectively. As the ANN training is ended, the prediction value is given as

$$z_i = k_i(y - x) + x. \quad (18)$$

The learning matrix, after treated by (16), is loaded on a three-layer forward feedback ANN input end, and the input layer has 9 knots, with one output knot and 9 hidden knots. Then, we randomly give initial weighing values and corresponding threshold values to the output and hidden layers to make training of the normalized input matrix by virtue of (7) to (15), where the learning and momentum factors are evaluated as 0.9 and 0.7, respectively, with convergent error of the ANN model set to be 0.0001. As training operation numbers 14233, the set level of convergence error is reached and the operation is terminated. Historical data fittings and AMT predictions are obtained by the operation of addition and subtraction based on the ANN-defined connection weighing coefficients and related threshold values. The 1916–85 AMT measurements (solid) and its model-produced fittings (dotted line) for the research area is shown in Fig.1, wherefrom we see that the well-fitted predictions arrive at correlation of 0.9923 with measured AMT at $n = 70$, revealing identical high- and low-value

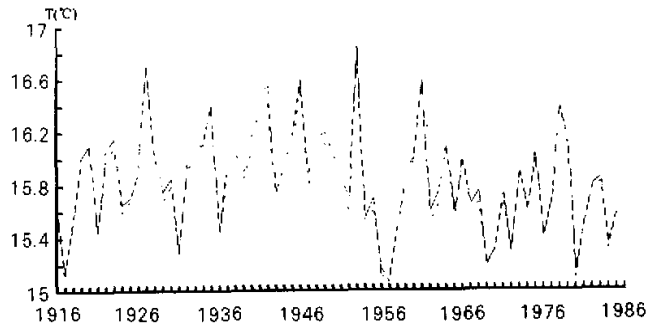


Fig. 1. Fittings (dotted) of the ANN-MGF predictions versus observed AMT (full line) for 1916–1986.

years. In the model context, 10 AMT independent samples (1986–1995) are employed to make predictions for testing the model output against the observed (see Table 2). It is evident therefrom that the averaged absolute error is 0.450°C for the period.

Table 2. ANN—MGF model AMT predictions based on 1986–1995 independent samples in comparison to the observed with their absolute values average given(AVA)

Year	1986	1987	1988	1989	1990	1991	1992	1993	1994	1995	AVA
Observed	15.60	15.85	15.90	15.80	16.65	15.95	15.95	15.60	17.15	16.15	
Predicted	15.92	16.52	15.54	15.78	15.73	16.35	15.78	15.72	16.04	15.72	
Difference	-0.32	-0.67	0.36	0.02	0.92	-0.40	0.17	-0.12	1.10	0.42	0.45

We also see that the averaged AMT differs by 0.156°C between the mean prediction and observation (15.904 compared to 16.06°C). It is an attempt to make AMT predictions for 1996–2005 using the developed model. Their mean is utilized to indicate the mean state in the period to follow, that is, they will differ by approximately 0.15°C, with the mean prediction lower, a result that may be satisfactory. Since the model AMT predictions are just tentative, it is necessary to examine the difference between the model output and results from the general MGF scheme (Cao et al., 1993). And to make objectively comparison, we utilized the same 9 MGF series and the usual regression establishing model scheme to formulate the prognostic equation of the form

$$\begin{aligned}\hat{Y}(t) = & 14.8137 - 1.4171f_{11}^{(0)} + 0.8591f_{22}^{(0)} + 0.6569f_{26}^{(0)} \\ & + 0.6822f_{29}^{(0)} + 1.7017f_{33}^{(0)} + 0.4327f_{13}^{(1)} \\ & - 0.5033f_{24}^{(1)} - 0.3691f_{19}^{(2)} + 0.2385f_{34}^{(2)},\end{aligned}\quad (19)$$

where $f_i^{(0)}$, $f_i^{(1)}$, and $f_i^{(2)}$ are the MGFs of the original sequence, first and second-order difference series for the ANN model establishment and which has a complex correlation coefficient of 0.8253. Fig. 2 presents (19)–given 1916–1985 AMT fittings (dotted) in comparison to the observed (full line).

Inspection of Fig. 2 shows better results from the ANN–MGF techniques than from the commonly-used regression in the context of the same MGFs. Then we drew on (19) to make prediction in terms of 1986–1995 independent samples for testing purpose (see Table 3). Comparison of the mean of difference in absolute values of Tables 2 and 3 (0.450 vs 0.481) shows that the ANN–MGF model is superior on the whole.

Table 3. AMT predictions with the MGF regression model based on 1986–1995 independent samples in comparison to the observations, with the absolute valued averaging(AVA) given

Year	1986	1987	1988	1989	1990	1991	1992	1993	1994	1995	AVA
Observation	15.60	15.85	15.90	15.80	16.65	15.95	15.95	15.60	17.15	16.15	
Prediction	16.32	15.59	15.76	15.56	15.68	15.79	15.84	16.34	16.04	15.70	
Difference	-0.63	-0.26	-0.14	-0.24	-0.97	-0.16	-0.11	0.74	-1.11	-0.45	0.481

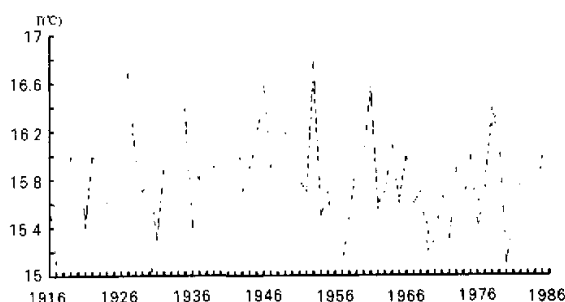


Fig. 2. Fittings (dotted) of the regression - MGF predictions versus observed AMT (full line) for 1916-1986.

From the above analysis we see that our ANN-MGF model allows to make prediction of the climate state for a decade to come, resulting in high closeness to the observed in the context of 1986-1995 independent samples. Hence, we have attempted to perform AMT prediction using the developed model for the study region, with the results summarized in Table 4. We see therefrom that the mean of 10 predictions is 16.078°C, a figure that is very close to the mean AMT in 1986-1995. On the other hand, both the mean of 1986-1995 AMTs and the mean of 1996-2005 predictions are higher than the AMT of 15.82°C averaged over 1916-1985. This indicates that our prediction for 1996-2005 is likely to exceed the 1916-1985 AMT on a mean basis.

Table 4. 1996-2005 ANN-MGF AMT predictions for the Yangtze River delta area

Year	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	mean
Prediction	15.87	16.06	16.08	15.79	16.03	16.35	16.07	16.13	16.02	16.38	16.078

4. Summary and discussion

In view of the fact that climatic variation is marked by pronounced uncertainty and stochastic character and the variation and mutual effects of factors, internal and external to the system are intricate, the mechanisms are almost unknown for long-term weather processes and, particularly, climatic variation at longer than annual scales, and predictive schemes are very few. For this reason, based on the AMT series, an attempt is made to perform multiple prediction of future ATMs in terms of an MGF method in combination with the ANN scheme, able to solve nonlinear problems. Our developed model gives quite small error between the mean prediction and observation for 1986-1995 from the independent samples, which suggests the utility of our model. But, when using the ANN for model establishment, no techniques nor principles are offered by its theoretical research for constructing a model learning matrix on which the model predictive ability depends strongly so that when MGFs are singled out for formulating a learning matrix, focus is on MGFs with higher entropy of classified information that are distributed in the three series (the original, first-

and second difference sequences). Moreover, we have calculated the complex correlation coefficient to investigate the MGF contribution to the model in order to provide an optimal learning matrix. The calculation of the complex correlation coefficient is to show linear correlativity of a predictand with multiple MGFs while the calculation of MGF information entropy aims at the conditional probability of events. As such, it is of importance to ANN model establishment to conduct further research of the problem. In addition, most of the excitation functions for the ANN scheme in wide use adopt the algorithm of sigmoid function and gradient descent, leading in some cases to the oscillation, no or slow convergence of error function, which would make model improvement difficult. Beyond that, there is no sophisticated theory on the optimization of ANN structure for model construction. For instance, it is a common practice to select an appropriate number of hidden knots through progressive increase or decrease, which will make it difficult to determine other parameters and to improve the model as well. This would limit its polarizing possibility. For this reason, the authors are making efforts to try a genetic algorithm of evolution type in study of such problems.

REFERENCES

- Cao Hongxing, Wei Fengying, and Liu Changsheng, 1993: Multiple prediction model with rainfall timeseries, *Quart. J. Applied Meteorology*, **4**(2), 198–204 (in Chinese).
- Hu Jifu, Jiang Hongchuan, Zhou Qingman et al., 1994: Optimal ordering of an autoregression model and its applications to long-range prediction, *Quart. J. Appl. Meteor.*, **5**(2), 196–201 (in Chinese).
- Jin Long, Luo Ying, Guo Guang, et al., 1997: Study on mixed model of neural network for farmland flood/drought prediction, *Acta Meteorologica Sinica*, **11**(3), 364–373.
- Lee, K. Y., Cha, Y. T., and Park, J. H., 1992: Short-term load forecasting using the artificial neural network, *IEEE transactions on Power System*, **7**, 124–132.
- McCann, D. W., 1992: A neural network short-term forecast of significant thunderstorms, *Weather Forecasting*, **7**, 525–534.
- Wei Fengying, and Cao Hongxing, 1990: *Mathematical Models with Application for Long-Term Prediction*, China Meteorological Press, Beijing (in Chinese) 49–90.
- Zhou Jiabin, 1993: Problems of statistical meteorology, *Scientia Atmospherica Sinica*, **17** (suppl. Issue), 1–6 (in Chinese).
- Zhou Shuzhen, and Zhang Chao, 1985: *Effects of 4 Thermal Islands on Urban Climate of Shanghai*, in *Urban Climate and Planning*, Science Press, Beijing (in Chinese) 128–137.