On Spatiotemporal Series Analysis and Its Application to Predict the Regional Short Term Climate Process

WANG Geli* (王革丽), YANG Peicai (杨培才), and LÜ Daren (吕达仁)

Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing 100029

(Received 4 March 2003; revised 12 November 2003)

ABSTRACT

Based on the theory of reconstructing state space, a technique for spatiotemporal series prediction is presented. By means of this technique and NCEP/NCAR data of the monthly mean geopotential height anomaly of the 500-hPa isobaric surface in the Northern Hemisphere, a regional prediction experiment is also carried out. If using the correlation coefficient R between the observed field and the prediction field to measure the prediction accuracy, the averaged R given by 48 prediction samples reaches 21%, which corresponds to the current prediction level for the short range climate process.

Key words: spatiotemporal series, regional climate prediction, nonlinear prediction

1. Introduction

In the last ten years, it has been a very active research field to use a chaotic time series for building a nonlinear prediction model. In numerous research works, two ways of building the model are usually used: the artificial neural network technique (Cybenko, 1989; Hsieh, 2001) and the state space reconstruction technique (Farmer and Sidorowich, 1987; Casdagli, 1989; Sugihara and May, 1990 ; and Sivakumar et al., 2001). Casdagli (1989) systematically reviewed the latter, and compared three different model building techniques: the global approximation, the local approximation, and the radiation function base approximation. Farmer and Sidorowich (1987) showed that, in the case of a not too high embedding dimension, the first-order local approximation can give better results for those classical chaotic systems.

However, most time series from the real world, especially those representing climate processes, are too short to satisfy the length requirement (Theiler et al., 1992). Too short a history cannot give a full description for the state distribution of the dynamics system. In other words, it cannot be expected to predict an event that occurs in one hundred years by using the data accumulated in a period of 50 years. This difficulty is usually referred to as the "data bottleneck" problem for short time series analysis.

In order to solve the above problem, some atmosphere scientists have suggested the reconstruction of the dynamic system with observation data from different spatial positions. They have once applied this idea to estimate the dimension of climate attractors and achieved some successes (Essex and Lookman, 1987; Keppenne and Nicolis, 1989; Yang et al., 1994). So, people may ask the question: could this idea be used to predict the spatiotemporal series? The answer is positive. In fact, Yang et al. (2000) have used the spatiotemporal neural network technique to predict the distribution of the monthly mean value of the column ozone over China, and obtained a prediction accuracy of over 0.43, which represents the correlation coefficient between the observed field and the prediction field. Their results show that the spatiotemporal series can effectively improve the ergodicity of the single-variable time series.

Based on the above idea, this paper carries out an experiment to predict the monthly mean geopotential height anomaly field of the 500-hPa isobaric surface (MMGHAF500) by using the state space reconstruction technique and NCEP/NCAR data. The organization of the paper is as follows. In the next section, the basic idea behind the technique used and the procedure for building the prediction model are described. Then, in section 3, to test the model ability, a prediction experiment for the geopotential height field of

*E-mail: wgl@mail.iap.ac.cn

500 hPa is given. Finally, in the last section, a brief summary is presented.

2. Basic idea

In this section, we will take the local approximation technique (Casdigli, 1989) as the pattern to introduce the basic idea and the key procedure to build the spatiotemporal series prediction model. In general, as an extension of the prediction theory of the single-variable time series, which has had more than 10 years of history, the technique for building the spatiotemporal series prediction model is still based on the reconstruction of state space, except the space distribution of the variable in the predicted region should be taken into account. Thus, in the description of the procedure for building the model, we will focus on the treatment of the space construction of the spatiotemporal series.

For simplicity, we suppose that the analyzed spatiotemporal series has a one-dimensional space distribution, i.e., it consists of K subsequences from K different observation stations. If all these subsequence sizes are equal to N, this dataset can be indicated as

$$x(i,j) = \{x(s_i, t_j)\} \qquad i = 1, 2, \cdots, K; \quad j = 1, 2, \cdots, N \tag{1}$$

where s_i and t_j represent the i-th observation station and the j-th observation time, respectively.

We also suppose that the spatiotemporal series is controlled by the same physical law, or that the subsequences observed at the different positions describe an identical dynamical system. Such an assumption means that all reconstructed sub-trajectories given by the subsequences should twine on an identical attractor. So, if we apply the same reconstruction parameters to all subsequences, then we can obtain a reconstructed sub-trajectory family:

$$\begin{aligned} y_m(i,j) = &\{x(s_i,t_j),\ \ x(s_i,t_j+\tau), \cdots,\\ &x[s_i, t_j + (m-1)\tau]\}|_{j=1,2,\cdots,N} \\ i =&1, 2, \cdots, K \end{aligned} \tag{2}$$

where m and τ stand for the embedding dimensionality and the delay time, respectively, and $t_j = t_N - (m-1)\tau$. Obviously, this sub-trajectory family describes the dynamics of the spatiotemporal series on the reconstructed attractor, and gives a much larger phase point set than that reconstructed only by one subsequence.

The prediction of the spatiotemporal series is carried out station by station in the given region (i=1, 2, $\ldots$, K). For the i-th observation station, if the current time and the corresponding current state are denoted as N and $y\ (i,N)$, respectively, then the nearest neighbor of $y\ (i,N)$, $y\ (q^*, j^*)$, should satisfy:

$$\|y\ (i,N) - y\ (q^*, j^*)\| = \min_{\substack{q=1,2,\cdots,K \\ j=1,2,\cdots,N-1}} \{\ \|y\ (i,N) - y\ (q,j)\ \|\ \} \tag{3}$$

where $||\bullet||$ stands for the Euclidean distance between two phase points on the embedded attractor. Expression (3) means that the nearest neighbor of $y\ (i,N)$ is not only with respect to those phase points lying in the i-th sub-trajectory, but also to the whole attractor. Obviously, doing so is helpful to improve the limitation due to the short subsequence.

According to Expression (3), we can find p nearest neighbors one by one, and denote them as

$$\Omega_1 = \{y\ (i, N_L)\}, \qquad L = 1, \cdots, p\,. \tag{4}$$

At the same time, we can also find out their mappings on the reconstructed attractor and denoted them as

$$\Omega_2 = \{y\ (i, N_L + 1)\}, \qquad L = 1, \cdots, p\,. \tag{5}$$

Then, according to the local approximation technique (Casdigli, 1989), we can build a prediction model by means of Ω_1 and Ω_2.

3. Prediction experiments

In this section, we will focus our attention on regional short-term climate prediction, which is one of the hot spots in the study of climate prediction theory. We believe that it should be such a field in which the spatiotemporal series prediction theory could play an important role.

The data used in our prediction experiments are the MMGHAF500 in the Northern Hemisphere, which are provided by NCEP/NCAR. Using such data is convenient for comparing with prediction results given by other techniques. The data used cover 480 months from January 1958 to December 1997; those of the first 432 months (from January 1958 to December 1993) are used to build up the prediction model, and those of the last 48 months (from January 1994 to December 1997) are used to test the prediction skill.

In terms of the Auto-Correlation Function Model (Abarbanel et al., 1993) and the False Neighbor Method (Fraser, 1986), we obtain the reconstruction parameters, the lagged time $\tau$and the embedded dimensionality m, which are 3 and 5, respectively. The capacity of the nearest neighboring set is taken as 200. The order of the prediction model is assumed to be dependent on the season, which are given as 0, 1, and 2 for the winter period (from December to February), for the spring period and autumn period (from March to May; and from September to November) and for the summer period (from June to August), respectively.

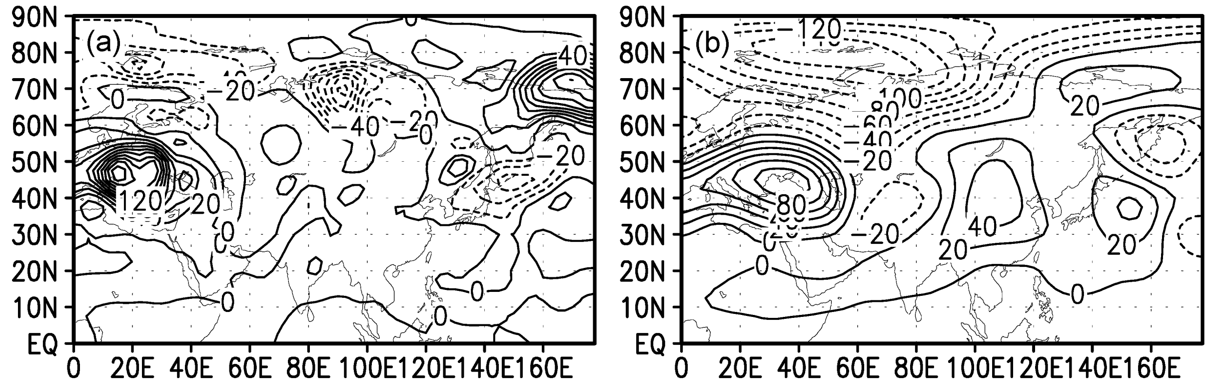

Fig. 1. Situations of the prediction field (a) and the observed field (b) for the 500 hPa monthly mean geopotential height anomaly in the Northern Hemisphere for April 1994.

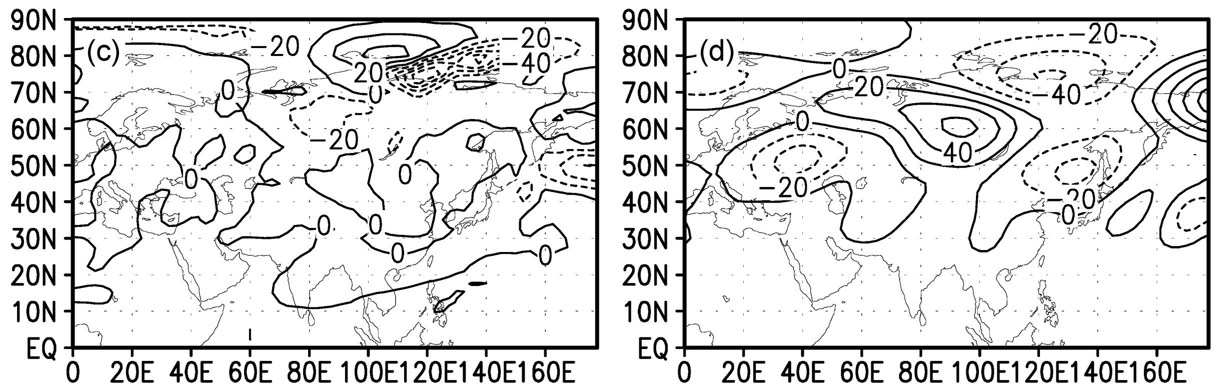

Fig. 2. The same as Fig. 1 except in July 1997.

Figures 1 and 2 give two prediction cases, the better one and the worse one. If using the correlation coefficient between the observed field and prediction field to measure the prediction accuracy, the mean value given by the 48 predicting cases (from January 1994 to December 1997) reaches 0.21, which corresponds to the prediction level of GCM models (Wang, 1996). This shows that, in regional short-term climate prediction, the spatiotemporal series prediction technique based on the theory of state space reconstruction has a credible prediction skill.

4. Summary

The theory and method for spatiotemporal series prediction are discussed in this paper by means of state space reconstruction theory. Essentially, the idea is a natural promotion of the prediction theory for a single-variable time series (Packard et al., 1980; Takens, 1981, Farmer and Sidorowich, 1987; and Casdigli, 1989). The basic assumption of this technique is that all the observed subsequences in the predicted region are under the control of the same dynamical law. This assumption ensures that those reconstructed sub-trajectories should twine on an identical attractor. Through studying this sub-trajectory family, we can get the statistical behaviors of the attractor, and predict the dynamics of the spatiotemporal series.

When using the local approximation technique (Casdigli, 1989) to build the spatiotemporal series prediction model, an important improvement consists in that, compared with the single-variable time series, the selection range of the nearest neighbor set is not only limited in the sub-trajectory in which the current point lies, but is also extended on the whole attractor. Obviously, this technique may be used to improve the ergodicity of the single-variable time series, and this then raises the prediction skill.

The prediction experiment for MMGHAF500 given in this paper presents a prediction accuracy corresponding to GCM (about 20%–30%). This result shows a valuable prospect in applying this technique to regional climate prediction, even though the experiment is preliminary.

The spatiotemporal series prediction technique seems to be more appropriate for predicting a field change process without a clear physical mechanism. As it appears, compared with the numerical methods such as GCM, it does not require building up the complete mathematical-physical model. This is because, in its view, all information describing the physical background has been included in the observed history data.

Acknowledgments. This work was supported by the National Key Program for Developing Basic Sciences in China (Grant No. G1999043405) and the National Natural Science Foundation of China (Grant No. 40035010).

REFERENCES

Abarbarnal, H. D. I., R. Brown, J. L. Sidorowich, and L. S. Tsimring, 1993:The analysis of observed chaotic in physical systems. *Reviews of Modern Physics*, **65**(4), 1331–1392.

Casdagli, M., 1989: Nonlinear Prediction of Chaotic time Series. *Physica D*, **35**, 335–356.

Cybenko, G., 1989: Continuous value neural networks with two layers are sufficient. *Mathematics of Control, Signals, and Systems*, **2**, 303–413.

Essex, C., and T. Lookman, 1987: The climate attractor over short timescales. *Nature*, **326**, 64–66.

Farmer, J. D., and J. Sidorowich, 1987: Predicting Chaotic Time Series. *Physical Review Letters*, **59**, 845–848.

Fraser, A. M., and H. L. Swinney, 1986: Independent coordinator for strange attractors from mutual information. *Phys. Rev. A*, **33**, 1134–1240.

Keppenne, C. L., and C. Nicolis, 1989: Global properties and local structure of the weather attractor over Western Europe. *J. Atmos. Sci.*, **46**, 2356–2370.

Packard, N. H., J. P. Crutchfield, J. D. Farmer, and R. S. Shaw, 1980: Geometry from a time series. *Phys. Rev. Lett.*, **45**, 712–715.

Sivakumar, B., R. Berndtsson, and R. Persson, 2001: Monthly runoff prediction using phase-space reconstruction. *Hydrological Sciences Journal*, **46**(2), 131–146.

Sugihara, G., and R. M. May, 1990: Nonlinear forecasting as a way of distinguishing chaos from measurement error in time series. *Nature*, **344**, 734–741.

Takens, F., 1981: Detecting strange attractors in turbulence. *Dynamical Systems and Turbulence*, Springer-Verlag, 366–381.

Theiler, J., S. Eubank, A. Longtin, and B. Galdrikian, 1992: Testing for nonlinearity in time series: The method of surrogate data. *Physica D*, **58**, 77–94.

Wang Shaowu, Wu Xiaohong, Ye Jinlin, and Chen Zhenhua, 1996: Annual prediction for 500hPa monthly mean height in Northern Hemisphere. *Research on Climate Prediction*, China Meteorological Press, 50–61. (in Chinese)

Wang Shaowu, Zhu Jinhong, 2001: A review on seasonal climate prediction. *Advances in Atmospheric Sciences*, **18** (2), 197–208.

Hsieh, W. W., 2001: Nonlinear principal component analysis by neural networks. *Tellus*, **53**A, 599–615.

Yang P. C., G. P. Brasseur, J. C. Gille, and S. Madronich, 1994: Dimensionalities of ozone attractors and their global distribution. *Physica D*, **76**, 331–343.

Yang P. C., X. J. Zhou, and J. C. Bian, 2000: A nonlinear regional prediction experiment on a short-range climatic process of the atmospheric ozone. *J. Geophys. Res.*, **105**(D10), 12253–12258.