ADVANCES IN ATMOSPHERIC SCIENCES, VOL. 30, NO. 6, 2013, 1601–1607

Nonstationary Time Series Prediction by Incorporating External Forces

WANG Geli*[1] (王革丽), YANG Peicai[1] (杨培才), and ZHOU Xiuji[2] (周秀骥)

[1]*Key Laboratory of Middle Atmosphere and Global Environment Observations, Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing* 100029

[2]*Chinese Academy of Meteorological Sciences, Beijing* 100081

(Received 21 July 2012; revised 24 October 2012; accepted 21 January 2013)

ABSTRACT

Almost all climate time series have some degree of nonstationarity due to external forces of the observed system. Therefore, these external forces should be taken into account when reconstructing the climate dynamics. This paper presents a novel technique in predicting nonstationary time series. The main difference of this new technique from some previous methods is that it incorporates the driving forces in the prediction model. To appraise its effectiveness, three prediction experiments were carried out using the data generated from some known classical dynamical models and a climate model with multiple external forces. Experimental results indicate that this technique is able to improve the prediction skill effectively.

Key words: external force, nonstationary system, climate prediction

Citation: Wang, G. L., P. C. Yang, and X. J. Zhou, 2013: Nonstationary time series prediction by incorporating external forces. *Adv. Atmos. Sci.*, **30**(6), 1601–1607, doi: 10.1007/s00376-013-2134-z.

1. Introduction

Recent studies have pointed out the nonstationarity character of the climate system. For example, Tsonis (1996) studied the decadal variation of global precipitation and noticed that its variance has changed over the last century, which implies that the precipitation process over this period has not been stationary. Also, Trenberth (1990) noted significant changes in the mean sea level pressure by the end of the 1970s, which indicated the nonstationary behavior of this system. In addition, studies have shown that although ENSO was depicted as a nonstationary phenomenon, it could be rectified by incorporating changes in the mean state (An et al., 2005; Boucharel et al., 2009). In addition, global warming is the topic of climate change-related research in the 21st century and several studies have focused on this topic (for instance, Lean and Rind, 2001; Moberg et al., 2005; Scafetta and West, 2007). Several possible causes of global warming have been suggested and investigated, including natural variation (such as volcanic emission and solar activity) and influences of human activities (such as greenhouse gases, aerosol radiative effects); however, irrespective of the cause, the findings related to global warming have depicted climate process as a nonstationary phenomenon.

However, according to almost all current theories of time series prediction, application of the ergodic theory requires the predicted process to be assumed as stationary, which has become one of the main barriers in climate prediction theories (Yang et al., 2003). To ravel out this problem, some studies on predictions of nonstationary time series have been presented recently (Hegger et al., 2000; Wang and Yang, 2005; Wan et al., 2005; Gong et al., 2006; Yang et al., 2010; Song and Li, 2012; He et al., 2012). The basic idea used in all these studies was to remove or reduce the nonstationarity of the predicted system using some mathematical techniques, thereby improving the prediction. In fact, the nonstationarity is generated because of the fact that the driving forces acting on the observed system change with time (Manuca and Savit, 1996). Thus, the most effective way to remove the nonstationarity may be to incorporate all the driving forces in the reconstructed dynamical system considering them as the

*Corresponding author: WANG Geli, wgl@mail.iap.ac.cn

state variables of that system. Based on this principle, we present an algorithm to incorporate driving forces to predict the nonstationary climate time series.

Following is a brief introduction of the algorithm for establishing the prediction model. To test its effectiveness, we carried out several prediction experiments on the given time series generated from some known classical dynamical systems and a climate model, which are discussed next. Finally, a brief discussion is provided.

2. Methods

Some of the basic principles to predict a nonstationary time series have been presented in a recent paper (Wang et al., 2011), which are presented here as well for convenience and completeness. Let us assume that we have a single variable time series $x(t), t = 1, \ldots, n$ from a dynamical system. It is common practice by now to use this time series and time delays to reconstruct its phase trajectory:

$$\begin{aligned} \boldsymbol{X}(t) &= \{x(t), x(t-\tau), \ldots, x(t-(m_1-1)\tau)\}t \\ &= 1, \ldots, N \,, \end{aligned} \tag{1}$$

where m_1 is the dimensionality of the embedding space, τ is the time lag, and $N = n - (m_1 - 1)\tau$ is the number of phase points of the reconstructed trajectory, which are required to reconstruct the underlying phase trajectory of this dynamical system using the embedding theorem of Takens (1981). Once this is done, we can then build a nonlinear prediction model of the following form:

$$\boldsymbol{X}(t+P) = f_P(\boldsymbol{X}(t)) \,, \tag{2}$$

where f is some appropriate mapping and P the prediction time step. The Takens embedding theorem is appropriate only for an autonomous dynamical system; therefore, we followed the method of Stark (1999) to embed the driving forces in the same state space for a nonstationary system.

Now let us consider the case of nonstationary time series. If the above system is driven by an external force $\alpha(t), t = 1, \ldots, n$, which is generated by a driving system and observed at the same time, then, with the same time lag τ and another embedding dimension m_2 for $\alpha(t)$, one can reconstruct the dynamics of the system into a $m_1 + m_2$ dimensional phase space and show it as

$$\begin{aligned} \boldsymbol{Y}(t) = \big\{ & x(t), x(t-\tau), \ldots, x[t-(m_1-1)\tau]; \alpha(t), \\ & \alpha(t-\tau), \ldots, \alpha[t-(m_2-1)\tau] \big\}, t = 1, K \,, \end{aligned} \tag{3}$$

where $K = n - [\max(m_1, m_2) - 1]\tau$ stands for the number of phase points on this trajectory. Based on Eq.(3), we may build a model similar to Eq.(2) to predict the above process, i.e.

$$\boldsymbol{X}(t+P) = f_P(\boldsymbol{X}(t); \alpha(t)) \,, \tag{4}$$

where the prediction step P, which was considered as 1 in this study, and f_P is a desired mapping assumed to be a quadratic polynomial; now, the task is to find the cost function $\eta = \sum_{k=1}^{N} \left[f(\boldsymbol{x}_k, \boldsymbol{\alpha}_k) - x_{k+1} \right]^2$, which is reached its minimum value. For more details, one can refer to the studies of Farmer and Sidorowich (1987) and Casdagli (1989).

3. Experiments

We applied the approach referred above to perform some prediction experiments using several given nonstationary time series. The first experiment was performed with data from Lorenz system:

$$\begin{cases} \dfrac{dx}{dt} = -\sigma x + \sigma y \\ \dfrac{dy}{dt} = r(t)x - y - xz \\ \dfrac{dz}{dt} = xy - bz \end{cases} , \tag{5}$$

σ was taken as 10 and b as 8/3 in this model, while the Rayleigh number $Ra(t)$ was regarded as a time-varying driving force factor. The Rayleigh number followed three different functions, increased linearly, decreased linearly, and U-shape functions, as shown in Fig. 1. Under the present case, the Rayleigh number $r(t)$ took values in the range 20–48, which indicated that the Lorenz system lied in a chaotic regime. Integrating this modified Eq.(5), a phase trajectory with

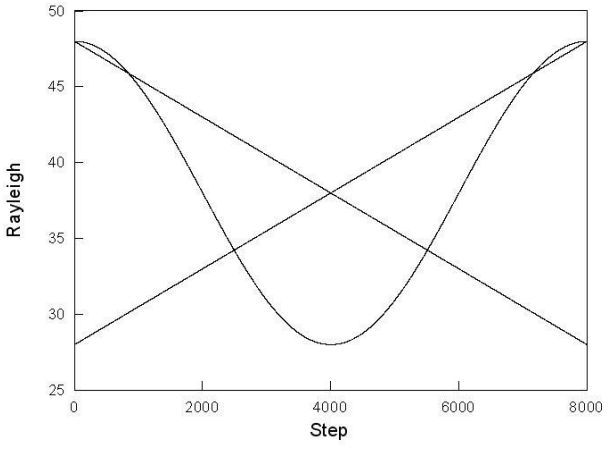

Fig. 1. Rayleigh values generated three different functions.

8000 state points could be obtained. Therefore, three nonstationary time series of the variable x were obtained. In all experiments, the time lag τ was always taken to be 1, and the embedding dimensions of the observations $\{x_t\}$ and m_1, were taken as 1 and 2, while those of the driving force $\{Ra_t\}$ and m_2 as 0 and 1, respectively. The case of $m_2 = 0$ implies that the driving forcing was not regarded in the prediction model, or, in other words, the predictions were based on the stationarity (hereafter called as "stationary model"); in contrast, the case of $m_2 = 1$ implies means that the driving force was taken into account (called as "forcing model"). In the experiments, the first 7200 data were applied to establish the prediction model, while the last 800 data were used for prediction and testing.

Figure 2 illustrates the error between prediction and observation of the variable x for r being the linearly increased function; the solid line is for the stationary model, while the dashed line is for the forcing model. Figure 3 shows the prediction accuracies provided by these experiments, which were expressed in terms of RMSE. In this figure, all the solid lines present the data for the stationary model, while the dashed lines for the forcing model. It was noted that all the prediction errors for the forcing model were lower than those for the stationary model, indicating that introducing the driving force into the prediction model could improve the predictive skill effectively.

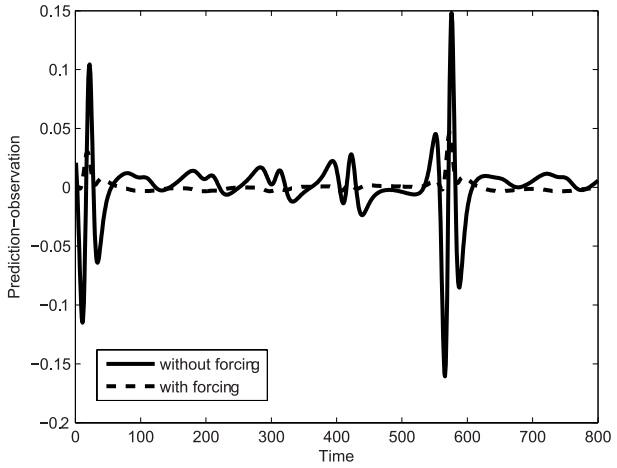

Fig. 2. Error between prediction and observation.

The second experiment also used data from the Lorenz system, but the Rayleigh number $r(t)$ was given by the logistic map $r(t+1) = \mu r(t)[1 - r(t)]$, where the value of μ was taken as 3.9, which implied chaotic behavior. We multiplied $r(t)$ by 32 to get a time series whose values ranged from 3.2 to 29.3, and assumed this time series to be the time-varying Rayleigh number to force the Lorenz system. Under the present case, the modified Lorenz system could obey the sates varying from state points to chaotic regimes (see Fig. 4); therefore, one nonstationary time series was obtained. The data set still consisted of 8000 values of the variable x.

Following the first experiment, out of a total 8000 data, the preceding 7200 data were applied to establish the predictive model, while the subsequent 800 points were used to test the prediction. We assumed that m_1 took values from 3 to 5 and m_2 either 0 or 1 (which corresponded to the stationary or forcing model). The experimental results for this case are listed in Table 1 and Fig. 5. From Table 1, it can be seen that: (1) all RMSE values given by the forcing model were much lower than those by the stationary one, and (2) the growth in error rate with prediction steps for the forcing model was lower than that the stationary one. From these findings it can be concluded that, in comparison with the stationary model, the forcing model had not only a higher prediction accuracy, but also a better predictability. Figure 5 presents the correlation

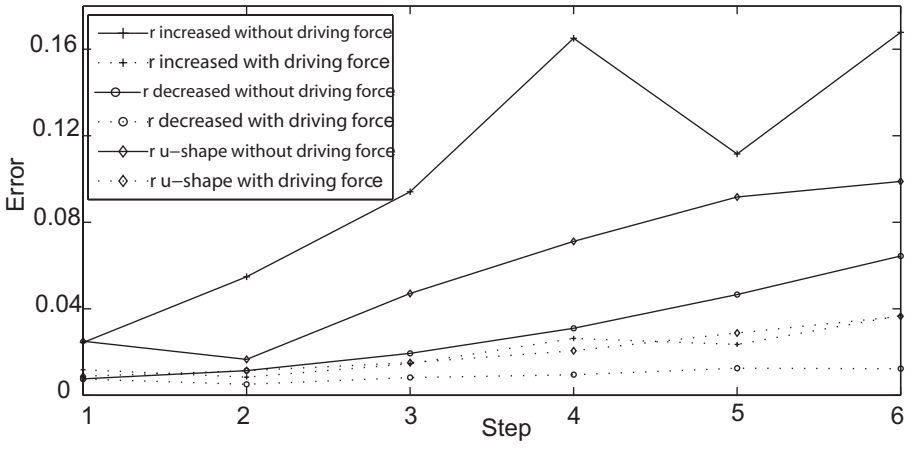

Fig. 3. Dependency of the prediction errors on the leading time step for experiment 1.

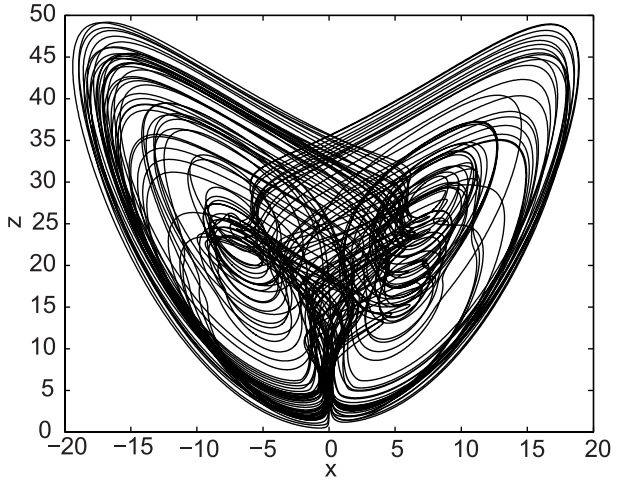

Fig. 4. Projection of the trajectory of Lorenz system in state plane (x, z) for the given time-varying Rayleigh number

Table 1. RMSE comparisons of the prediction experiments

m_2	$\varepsilon_{T=1}$	$\varepsilon_{T=2}$	$\varepsilon_{T=3}$	$\varepsilon_{T=4}$	$\varepsilon_{T=5}$	$\varepsilon_{T=6}$
0	1.21	1.66	5.52	2.47	5.64	12.22
1	0.58	0.88	0.99	2.29	2.44	2.79

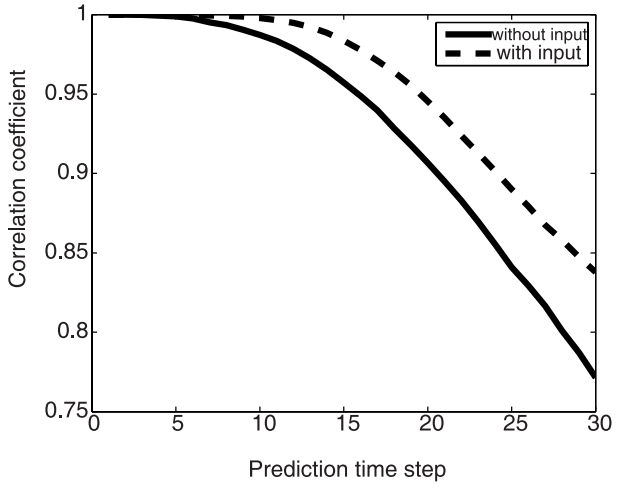

Fig. 5. Predictions of the stationary model and forcing model in experiment 2; all the results are averaged over the three appointed values of m_1.

coefficients between the actual and prediction values. It gives values similar to those in Table 1, indicating that the forcing model excelled the stationary model.

The third experiment was for predicting a nonstationary climate time series, which involved surrogate data of the Northern Hemisphere mean surface air temperature T_i generated by the Flexible Global Ocean-Atmosphere-Land-Sea ice model [FGOALS, refer to Zhou et al. (2008) for details]. FGOALS was driven by reconstructed natural forces (volcanic activity, denoted by α_i, and solar radiation, denoted as γ_i) and anthropogenic forces (greenhouse gas, denoted as β_i, and tropospheric aerosols, denoted as λ_i), data on which were taken from Crowley (2000) and Ammann et al. (2007), respectively. All the data were the annual anomaly values and covered the last 1000 years. Figure 6 shows their fluctuations in the last millennium; the red line indicates their trend over the period 1951–2000 in the time series of $\{\gamma_i\}$.

It should be noted that the referred four driving forces were considered synchronously in this experiment. A simple extension of Eqs. (3) and (4) should be as follows:

$$\begin{aligned}\boldsymbol{y}(i) = \big\{&T_i, T_{i-\tau}, \ldots, T_{i-(m_1-1)\tau};\\ &\alpha_i, \alpha_{i-\tau}, \ldots, \alpha_{i-(m_2-1)\tau};\\ &\beta_i, \beta_{i-\tau}, \ldots, \beta_{i-(m_2-1)\tau};\\ &\gamma_i, \gamma_{i-\tau}, \ldots, \gamma_{i-(m_2-1)\tau};\\ &\lambda_i, \lambda_{i-\tau}, \ldots, \lambda_{i-(m_2-1)\tau}\big\}_{i=1,2,\ldots,N}, \end{aligned} \tag{6}$$

and

$$T_{i+P} = f_P(\boldsymbol{T}_i;\ \boldsymbol{\alpha}_i;\ \boldsymbol{\beta}_i;\ \boldsymbol{\gamma}_i;\ \boldsymbol{\lambda}_i)\,. \tag{7}$$

Equations. (6) and (7) allowed us to establish a prediction model involving four driving forces. Following the above examples, data were divided into two parts: the preceding 950 points were applied to construct the predictive model and the following 50 points for testing the prediction accuracy. The Northern Hemisphere mean surface air temperature was predicted for the period 1951–2000. The parameters used for building

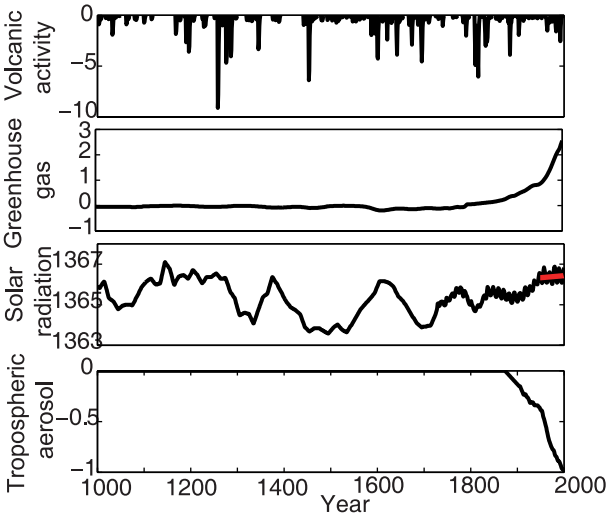

Fig. 6. Time series of $\{\alpha_i\}$, $\{\beta_i\}$, $\{\gamma_i\}$, and $\{\lambda_i\}$.

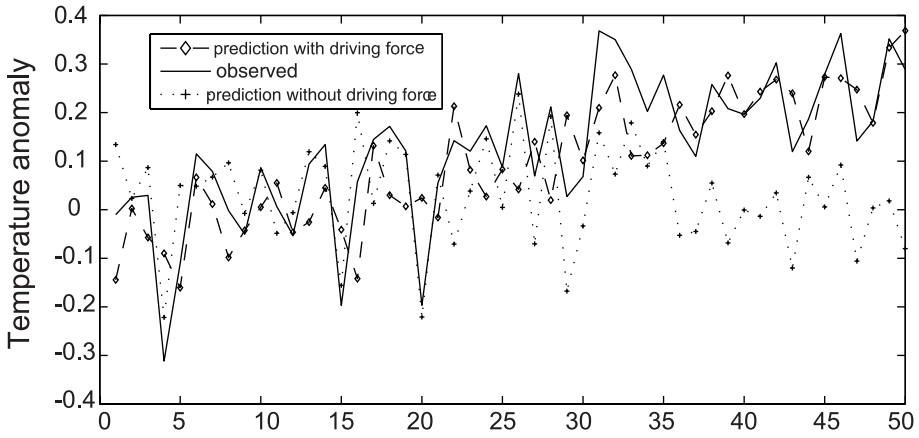

Fig. 7. Comparisons between the prediction abilities for the Northern Hemisphere mean surface air temperature anomaly.

the model were assigned the following values: the time lag τ was 1, the embedding dimensions of T_i of m_1 were taken from 3 to 7, the embedding dimensions of all the four driving forces of m_2 were set to be either 0 (for the stationary model) or in the range from 3 to 5 (for the forcing model).

All results were averaged over the embedding dimensions; the prediction results of the Northern Hemisphere mean surface air temperature anomaly with or without driving forces and observed values are shown in Fig. 7. We used the correlation coefficients between observation and prediction and the RMSE to denote the prediction accuracy, which were found to be 0.57 and 0.19 for the stationary model, and 0.67 and 0.12 for the force model, respectively. This illustrates again that the introduction of driving forces to prediction models can yield an obvious improvement in their accuracy.

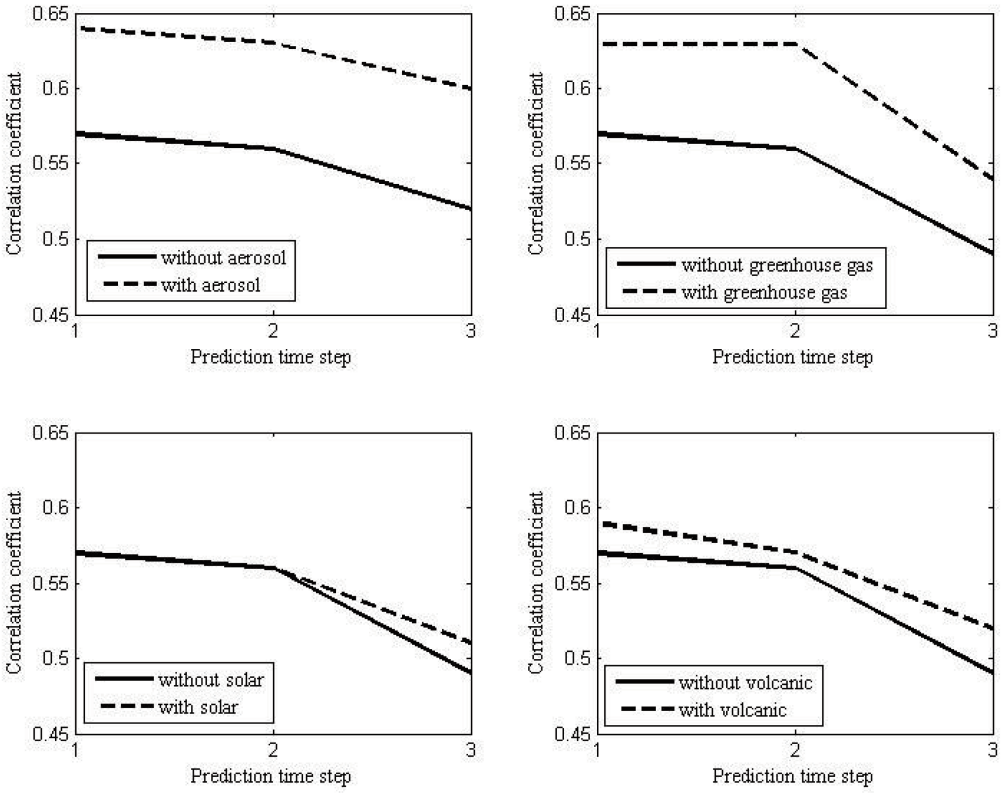

Fig. 8. Dependencies of prediction errors on the prediction step for using individual forcing.

In order to investigate the individual influences of each driving force factor on the prediction, we repeated the above experiment in four different cases, considering the solar radiation, volcanic aerosols, greenhouse gases, and sulfate aerosol forces, corresponding to the solar radiation model, volcanic aerosol model, greenhouse gas model, and sulfate aerosol model, respectively. The results are shown in Fig. 8. From this figure, it can be noted that in comparison with the stationary model, all models with any individual forcing incorporated can make improvement in predicting global temperature, especially the greenhouse gas and sulfate aerosol models. Note that solar activity was not particularly active for the forecast time period (1951–2000) here, as it experienced little variation during that period, as evident from its trend (red line) in the time series of solar activity shown in Fig. 6.

Again, we considered the overall influence, as shown in Fig. 9; as can be seen in the figure, the average correlation coefficient between observation and prediction depended on prediction time steps over the embedding. The solid line represent the result when no forcing was considered, the broken line is the average over the embedding and over the four forces acting individually, and the dash–dot line is the average for all four forces. Clearly, the position of dash–dot line above all other lines indicates that the improvement in predicting global temperature was the result of the collective behavior of the referred forces and might not be a result of an individual dominant forcing.

4. Discussion

Because time dependency of driving forces is the essential cause of nonstationarity, it is necessary to consider its influence on the prediction of time series. However, due to the lack of a complete theory to predict nonstationary process, no general and effective method has yet been developed. One can only choose from some available techniques to remove its nonstationarity for resetting it under the framework of stationary theory. As an attempt to improve the situation, we proposed a new technique and applied it to predict several artificial nonstationary time series with known external forces. The prediction results given by these experiments showed its effectiveness. In essence, the main idea of this technique was to consider all the driving forces as state variables and incorporate them into the prediction model. Therefore, the reconstructed system was changed to be an autonomous system, and thereby, the prediction returned to within the framework of the stationary theory. For the prediction of nonstationary time series with known driving forces, this technique can be used.

It should be noted that an assumption was implied in this technique: each driving force was assumed to be stationary. Otherwise, it would have been necessary to find the driving forces corresponding to this technique and incorporate these new increased driving forces into the prediction model.

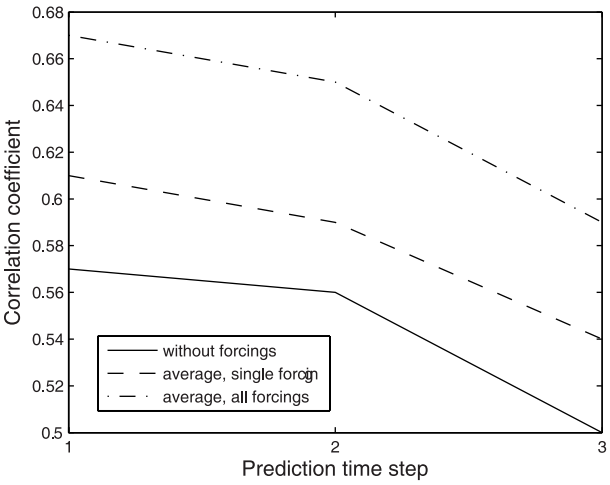

Fig. 9. Comparison of predictions for situations when no forcing was considered, individual forcing was considered, and all forces were considered.

Acknowledgements. This research was supported by the National Natural Science Foundation of China under Grant Nos. 40890052, 41075061, and 41275087. The authors are grateful to the anonymous referees and editors for their valuable suggestions.

REFERENCES

Ammann, C. M., F. Joos, D. S. Schimel, B. Otto-Bliesner, and R. A. Tomas, 2007: Solar influence on climate during the past millennium: Results from transient simulations with the NCAR climate system model. *Proc. of the National Academy of Sciences of the United States of America*, **104**, 3713–3718.

An, S. I., W. W. Haish, and F. F. Jin, 2005: A nonlinear analysis of ENSO cycle and its interdecadal changes. *J. Climate*, **18**, 3229–3239.

Boucharel, J., B. Dewitte, B. Garel, and Y. Penhoat, 2009: ENSO's non-stationary and non-Gaussian character: The role of climate shifts. *Nonlinear Processes in Geophysics*, **16**, 453–473.

Casdagli, M., 1989: Nonlinear prediction of chaotic time series. *Phys. D*, **35**, 335–356.

Crowley, T. J., 2000: Causes of climate change over the past 1000 years. *Science*, **289**, 270–277.

Farmer, J. D., and J. Sidorowich, 1987: Predicting chaotic time series. *Physical Review Letters*, **59**, 845–848.

Gong, Z. Q., G. L. Feng, W. J. Dong, and J. P. Li, 2006: The research of dynamic structure abrupt change of nonlinear time series. *Acta Physica Sinica*, **55**(6),

3181–3187. (in Chinese)

He, W. P., L. Wang, S. Q. Wan, L. J. Liao, and T. He, 2012: Evolutionary modeling for dryness and wetness prediction. *Acta Phys. Sin.*, **61**(11), 119201, 1–8. (in Chinese)

Hegger, R., H. Kantz, L. Matassini, and T. Schreiber, 2000: Coping with non-stationarity by over-embedding. *Physical Review Letters*, **84**, 4092–4101.

Lean, J., and D. Rind, 2001: Earth's response to a variable sun. *Science*, **292**, 234–236.

Manuca, R., and R. Savit, 1996: Stationarity and nonstationarity in time series analysis. *Phys. D*, **99**, 134–161.

Moberg, A., D. M. Sonechkin, K. Holmgren, N. M. Datsenko, and W. Karlen, 2005: Highly variable Northern Hemisphere temperatures reconstructed from low- and high-resolution proxy data. *Nature*, **443**, 613–617.

Scafetta, N., and B. J. West, 2007: Phenomenological reconstructions of the solar signature in the NH surface temperature records since 1600. *J. Geophys. Res.*, **112**, D24S03, doi: 10.1029/2007JD008437.

Song, T., and H. Li, 2012: Chaotic time series prediction based on wavelet echo state network. *Acta Phys. Sin.*, **61**(8), 080506, 1–7. (in Chinese)

Stark, J., 1999: Delay embeddings for forced systems: Deterministic forcing. *Journal of Nonlinear Science*, **9**, 255–332.

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

Trenberth, K. E., 1990: Recent observed inter-decadal climate changes in the northern hemisphere. *Bull. Amer. Meteor. Soc.*, **7**, 988–993.

Tsonis, A. A., 1996: Widespread increases in low-frequency variability of precipitation over the past century. *Nature*, **382**, 700–702.

Wan, S., G. Feng, G. Zhou, B. Wan, M. Qin, and X. Xu, 2005: Extracting useful information from the observations for the prediction based on EMD method. *Acta Meteorologica Sinica*, **63**, 516–525.

Wang, G., and P. Yang, 2005: A compound reconstructed prediction model for nonstationary climate process. *Inter. J. Climatol.*, **25**, 1265–1277.

Wang, G. L., P. C. Yang, J. C. Bian, and X. J. Zhou, 2011: A novel approach in predicting non-stationary time series by combining external forces. *Chinese Science Bulletin*, **56**, 3053–3056, doi: 10.1007/s11434-011-4638-1.

Yang, P., J. Bian, G. Wang, and X. Zhou, 2003: Hierarchies and nonstationarity in climate systems. *Chinese Science Bulletin*, **48**, 2148–2154, doi: 10.1360/03wd0175

Yang, P. C., G. L. Wang, J. C. Bian, and X. J. Zhou, 2010: The prediction of non-stationary climate series based on empirical mode decomposition. *Adv. Atmos. Sci.*, **27**(4), 845–854, doi: 10.1007/s00376-009-9128-x.

Zhou, T. J., B. Wu, and X. Y. Wen, 2008: A fast version of LASG/IAP climate system model and its 1000-year control integration. *Adv. Atmos. Sci.*, **25**, 655–672, doi:10.1007/s00376-008-0655-7.