

A novel approach in predicting non-stationary time series by combining external forces

WANG GeLi^{1*}, YANG PeiCai¹, BIAN JianChun¹ & ZHOU XiuJi^{2,3}

¹ *Laboratory for Middle Atmosphere and Global Environment Observation, Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing 100029, China;*

² *Chinese Academy of Meteorological Sciences, Beijing 100081, China;*

³ *State Key Laboratory of Severe Weather, Beijing 100081, China*

Received March 2, 2011; accepted June 15, 2011

In this paper, we investigate a novel technique that reconstructs the observed time series and incorporates driving forces. Furthermore, to illustrate and test the technique, we consider a couple of predictive experiments using ideal time series provided by the logistic and Lorenz systems with specific driving forces. The preliminary results show this approach can improve prediction proficiency to some extent, and the external forces play a similar role to that of state variables.

driving force, non-stationary system, climate prediction, time series prediction

Citation: Wang G L, Yang P C, Bian J C, et al. A novel approach in predicting non-stationary time series by combining external forces. *Chinese Sci Bull*, 2011, 56: 3053–3056, doi: 10.1007/s11434-011-4638-1

Short-term prediction is one of the difficult issues in climate change studies. Because of the complexity of short-term climate processes and lack of general knowledge underpinning its mechanisms, proficiency in prediction remains at a stand-still [1,2]. As for climate change in the 21st century, global warming is the most important factor, with many studies devoted to the topic. The possible causes range from change in natural forces (such as solar activity and volcanic emission) to human activities (such as green-house gases, sulfate emission in aerosols, and land use). In addition, the internal variability of climate systems should not be ignored either. However, no matter the cause of climate change, global warming in fact gives notice that climate processes are non-stationary.

So far many climate predictive theories formulated in statistics and in nonlinear science are based on the hypothesis that the process is stationary, thus ergodicity theory can be used under such conditions to reconstruct the dynamics from an observed time series and establish predictive equations constructed from it.

This reconstruction is at variance with the basic behavior of climate process. As is well-known for any weather or climate system, it is impossible to state the initial conditions incontrovertible. In fact, except for global warming, recent work has addressed the proper characterization of non-stationarity using weather and climate data. For instance, Tsonis [3] analyzed low-frequency (decadal to multi-decadal) variation in global precipitation over the past century and found that fluctuations about the global mean have increased significantly, implying that the second-order moment of the precipitation has changed on those scales, or that global precipitation process is non-stationary in the past century. In another case, Trenberth [4] found that the observed winter Pacific mean sea-level pressure underwent an abrupt change towards the end of the 70s; such a change in regime highlights the decline in the stationary behavior of the system. They noted that the behavior of these quantities is non-stationary. Most real world time series have some degree of non-stationarity due to external perturbations of the observed system [5].

However, there is as yet no general theory about non-stationary processes. What scientists can do for the moment

*Corresponding author (email: wgl@mail.iap.ac.cn)

is to remove the non-stationarity or reduce its degree using certain special processes via adding assumptions, and to re-express it within a stationary theory framework.

Recently, based on some climate observations, the concept of a “climate hierarchy” that implied climate is a cascade phenomenon originating from different hierarchies was proposed [6–9]. Techniques termed “compound reconstruction” and “segregated prediction” were applied to predict non-stationary time series; technically, though, these approximated the non-stationary contributions of the signals into stationary components.

Hegger et al. [10] presented an over-embedding technique, using increased embedding dimensions to treat time-varying parameters as state parameters which removed the non-stationarity on its physical cause. Wang et al. [11] used other higher-dimensional embedding such as ‘the support vector machine’ to predict some non-stationary processes and obtained better prediction proficiency.

In fact, the essential cause of non-stationarity is the time-dependent changes in the external forces, and the above embedding techniques are a patch up in the reconstructed dynamical system arising from lack of information from external forcing. Thus, from this point of view, a direct incorporation of external forces into a reconstructed model should fare better at predicting non-stationary time series. However, an important issue is obtaining the forcing data from time series. Certainly, we can find relationships among the various climatic factors, but these must be established through causal effects [12]; an incomprehensive correlation analysis can result in mistakes [13]. Based on this concern, this work presents an algorithm to predict these non-stationary processes by using state space reconstruction theory combined with time-dependent driving forces. Considering that natural climate dynamics are complex enough and many other factors such as noise or data length will influence prediction proficiency. Thus we begin with time series from ideal non-stationary systems, since the data length and precision can be controlled and guaranteed.

1 Methodology on establishing predictive model

In the field of nonlinear time series analysis, the most important aspects are the state space reconstruction theory [14] and the embedding theorem [15]. According to the latter, developed by Takens [15], for a given single variable time series, one can use a couple of appropriate values of the embedding dimension and the delay time to convert the series into a phase trajectory in state space. The dynamics on the reconstructed trajectory is equivalent to that of the original system that generated the time series. Based on this trajectory, one can establish a model to predict the future state of the system; for details see [16–19]. It should be noted that the Takens embedding theorem is only appropriate for an autonomous dynamical system. For a non-sta-

tionary system, however, we can still embed the external force components in the same state space [20].

For convenience, assuming a non-stationary process composed of two series $\{x_i\}_{i=1,2,\dots,n}$ and $\{\alpha_i\}_{i=1,2,\dots,n}$, the former being the state variable and the latter for an external forcing. With a selected time lag τ , we embed the time series in an $m_1 + m_2$ dimensional phase space and express the reconstructed state trajectory as

$$\bar{y}(i) = \left\{ x_i, x_{i-\tau}, \dots, x_{i-(m_1-1)\tau}; \alpha_i, \alpha_{i-\tau}, \dots, \alpha_{i-(m_2-1)\tau} \right\}_{i=1,2,\dots,N}, \quad (1)$$

or simply as

$$\bar{y}(i) = \{ \bar{x}_i, \bar{\alpha}_i \}_{i=1,2,\dots,N}. \quad (2)$$

Here m_1 and m_2 are the given embedding dimensions for $\{x_i\}_{i=1,2,\dots,n}$ and $\{\alpha_i\}_{i=1,2,\dots,n}$, respectively, $N = n - (\max(m_1, m_2) - 1)\tau$ is the number of phase points on the trajectory. We use the global approximation method [16] to establish a predictive model as:

$$x_{i+1} = \hat{f}(\bar{x}_i; \bar{\alpha}_i) + \varepsilon_i, \quad (3)$$

where \hat{f} is a desired function and ε_i are fitting errors. Our task is then to find such an \hat{f} that the cost function reaches its minimum. Here, \hat{f} is assumed to be a second-order polynomial,

$$\eta = \sum_{k=1}^N \left[\hat{f}(\bar{x}_k, \bar{\alpha}_k) - x_{k+1} \right]^2. \quad (4)$$

2 Results

For completeness, we describe the following prediction tests in the presence or not of external forcing which we call the “forcing model” and “stationary model”, respectively.

The above predictive approach is applied to two distinct groups of experiments. The first group is from the logistics map:

$$x_{t+1} = \mu_t x_t (1 - x_t). \quad (5)$$

Here μ_t represents a time-dependent external forcing. For the logistics model, if the value of μ_t varies between 3.57 and 4.0, the system should exhibit chaotic behavior. If we let μ_t change according to the following two cases:

$$\mu_t^{(1)} = 3.95 - 0.4e^{-2.5t}, \quad (5.1)$$

$$\mu_t^{(2)} = 3.45 + 0.5e^{-1.5t}, \quad (5.2)$$

we obtain two different non-stationary time series, denoted by $S^{(1)}$ and $S^{(2)}$ in Figure 1, that exhibit chaotic behavior; the length of both series is 2000.

In the following tests, the initial 1900 data points are

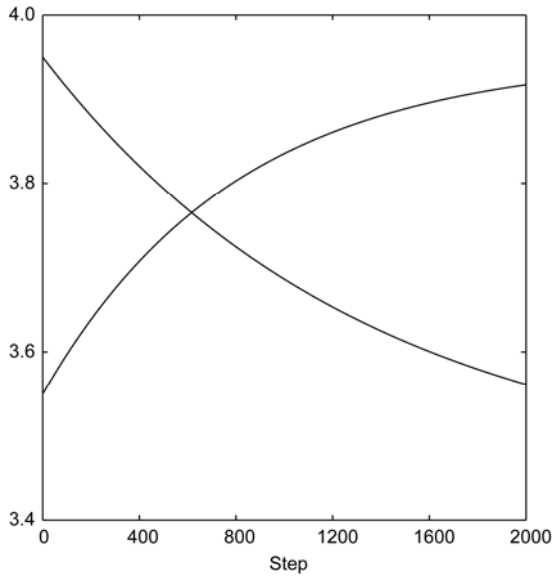


Figure 1 Two different driving forces.

used to build the predictive model, whereas the last 100 data points are used to test the prediction proficiency by using the root mean square error (RMSE). Here the lag τ is equal to one, while the embedding dimensions of the observations $\{x_t\}$ m_1 are set to 1 and 2, and of the external force $\{\mu_t\}$ m_2 are set to 0 and 1, respectively. For m_2 zero, the external force was not taken into account in the predictive model; i.e. the predictions were based on stationarity. Table 1 shows the various RMSEs resulting from experiments using different m_1 and m_2 values. It is noted that the RMSE is improved if an external force is considered. This indicates that introducing external forces into the predictive model can provide an effective means to predict non-stationary processes. For more details and cases see Wang and Yang [21].

The second experiment required predicting a non-stationary climate time series from the Lorenz system:

$$\begin{aligned} \dot{x} &= -\sigma x + \sigma y, \\ \dot{y} &= r(t)x - y - xz, \\ \dot{z} &= xy - bz, \end{aligned} \tag{6}$$

Table 1 RMSE comparison of the prediction experiments

m_2	m_1	$E_{S(1)} (\%)$	$E_{S(2)} (\%)$
0	1	2.12	2.89
0	2	1.39	1.69
0	mean*	1.76	2.29
1	1	0.61	0.63
1	2	0.63	0.62
1	mean	0.62	0.63

* indicates the averaged values for $m_1=1$ and $m_1=2$.

where $r(t)$ is the time-varying Rayleigh external forcing that is assumed to take values in the range (28, 48) during a cycle, and obeys one of the three different functions, see Figure 2. Here also σ and b take fixed values 10 and 8/3 respectively. Applying the fourth Runge-Kutta method to integrate eq. (6), we can get one trajectory with 8000 state numbers on its attractor, three different non-stationary time series denoted $T^{(1)}$, $T^{(2)}$, and $T^{(3)}$. In these tests, the first 7200 data points are used to build the predictive model, while the last 800 data points are used to test the predicted RMSE, except for the embedding dimension m_1 which ranges from 3 to 5, all other parameters are all set to the same values as the above experiment. Table 2 shows cases of RMSE resulting from these experiments. It is noted that (1) with the exception of special cases, the RMSE for the “forcing model” is less than that for the “stationary model”, indicating that introducing external forces into the predictive model can effectively improve prediction proficiency; and (2) increase in the speed of the RMSE for the “forcing model” caused by increasing the iterations is lower than that for the “stationary model”, again showing that introducing external forces can improve predictability of the system.

3 Discussion

We have considered the notion that non-stationarity for climate systems can have an important effect on investigating the underlying mechanisms of climate change and developed new climate prediction theories and means. In this paper, we presented a novel approach incorporating driving forces; several experiments from ideal time series show that this approach can improve prediction proficiency. If external forcing of a system is stationary, then the reconstructed system also has stationary characteristics; in other words,

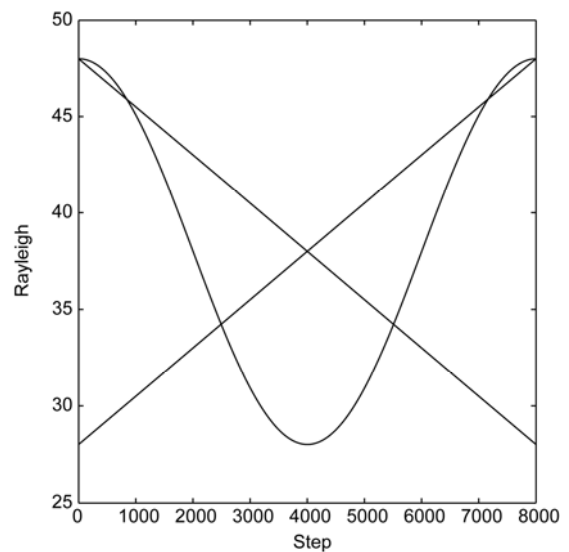


Figure 2 Rayleigh values with three different changes.

Table 2 RMSE comparison of the second prediction experiment

Prediction steps	m_2	m_1 (mean)	$E_{T^{(1)}}(\%)$	$E_{T^{(2)}}(\%)$	$E_{T^{(3)}}(\%)$
1	0	3–5	0.76	2.44	2.50
	1	3–5	0.75	1.18	0.88
2	0	3–5	1.14	5.48	1.66
	1	3–5	0.51	0.84	1.14
3	0	3–5	1.94	9.41	4.71
	1	3–5	0.82	1.46	1.51
4	0	3–5	3.10	16.51	7.12
	1	3–5	0.95	2.63	2.06
5	0	3–5	4.66	11.16	9.17
	1	3–5	1.25	2.35	2.88
6	0	3–5	6.44	16.78	9.89
	1	3–5	1.23	3.69	3.65

eq. (3) merely describes a stationary process composed of two hierarchal structures. By contrast, if external forcing is non-stationary, then other correlated external forces should be considered, adding consequentially a new hierarchy onto (3).

To perfect this technique, three additional studies should be considered. The first is the development of techniques to find the external forces based on causality. Obtaining the causal relationships among variables is one of the most important issues in time series prediction. However, to our knowledge, no effective approaches have currently been provided. A recommended method, provided by Verdes et al. [22], suggests using a measure called ‘transfer entropy’ to analyze causality. As an application to climate science, this analysis pointed out that causal relationship exists between both solar radiation and carbon dioxide and the global averaged surface air temperature. The second is an extension of the Takens embedding theorem to non-stationary time series. As stated before, this theorem holds only for autonomous system although many real systems are not autonomous. Thus, it is necessary to modify this theorem so that it can be applied to non-stationary time series. The third includes the need to treat an increased embedding dimensionality caused by external force factors and to apply and develop a proper model to deal with this kind of increased complexity.

Open Access This article is distributed under the terms of the Creative Commons Attribution License which permits any use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

This work was supported by the National Natural Science Foundation of China (40890052, 41075061 and 40940023).

- 1 Wang S W. Predictability and uncertainty in short-term climate prediction (in Chinese). *Adv Earth Sci*, 1998, 13: 8–14
- 2 Li W J, Zhang P Q, Li Q Q, et al. Research and operational application of dynamical climate model prediction system (in Chinese). *J Appl Meteorol Sci*, 2005, 16: z1
- 3 Tsonis A A. Widespread increases in low-frequency variability of precipitation over the past century. *Nature*, 1996, 382: 700–702
- 4 Trenberth K E. Recent observed inter-decadal climate changes in the northern hemisphere. *Bull Amer Meteor Soc*, 1990, 7: 988–993
- 5 Manuca R, Savit R. Stationarity and nonstationarity in time series analysis. *Physica D*, 1996, 99: 134–161
- 6 Chen B M, Ji L R, Yang P C, et al. An approach to improving the dynamical extended-range (monthly) prediction. *Chinese Sci Bull*, 2003, 48: 696–703
- 7 Yang P, Bian J, Wang G, et al. Hierarchy and non-stationarity in climate systems: Exploring the prediction of complex systems. *Chinese Sci Bull*, 2003, 48: 2148–2154
- 8 Wang G L, Yang P C. A compound reconstructed prediction model for non-stationary climate process. *Int J Climatol*, 2005, 25: 1265–1277
- 9 Yang P C, Wang G L, Bian J C, et al. The prediction of non-stationary climate series based on EMD. *Adv Atmos Sci*, 2010, 27: 845–854
- 10 Hegger R, Kantz H, Matassini L, et al. Coping with non-stationarity by over-embedding. *Phys Rev L*, 2000, 84: 4092–4101
- 11 Wang G L, Yang P C, Mao Y. On the application of non-stationary time series prediction based on the SVM method (in Chinese). *Acta Phys Sin*, 2008, 57: 714–719
- 12 Granger C W. Testing for causality: A personal viewpoint. *J Econ Dyn Control*, 1980, 2: 329–352
- 13 Geng Z. Causality and Simpson Paradox. <http://TJLT.chinajournal.net.cn>, 2000
- 14 Packard N H, Crutchfield J P, Farmer J D, et al. Geometry from a time series. *Phys Rev Lett*, 1980, 45: 712–715
- 15 Takens F. Detecting Strange Attractors in Turbulence. *Dynamical Systems and Turbulence*. Heidelberg: Springer-Verlag, 1981. 366–381
- 16 Farmer J D, Sidorowich J. Predicting chaotic time series. *Phys Rev Lett*, 1987, 59: 845–848
- 17 Casdagli M. Nonlinear prediction of chaotic time series. *Physica D*, 1989, 35: 335–356
- 18 Yang P C, Zhou X J, Bian J C. A nonlinear regional prediction experiment on a short-range climatic process of the atmospheric ozone. *J Geophys Res*, 2000, 105: 12253–12258
- 19 Elsner J B, Tsonis A A. Nonlinear prediction, chaos and noise. *Bull Amer Meteorol Soc*, 1992, 73: 49–60
- 20 Stark J. Delay embeddings for forced systems: Deterministic forcing. *J Nonlinear Sci*, 1999, 9: 255–332
- 21 Wang G L, Yang P C. A recent approach incorporating external forces to predict non-stationary processes. *Atmos Oceanic Sci Lett*, 2010, 3: 151–154
- 22 Verdes P F, Granitto P M, Ceccatto H A. Over-embedding method for modeling non-stationary systems. *Phys Rev Lett*, 2006, 96: 118701