Vol 14 No 3, March 2005
1009-1963/2005/14(03)/0628-06

Chinese Physics

On the climate prediction of nonlinear and non-stationary time series with the EMD method*

Wan Shi-Quan(万仕全)[a)], Feng Guo-Lin(封国林)[b)d)†], Dong Wen-Jie(董文杰)[c)],
Li Jian-Ping(李建平)[d)], Gao Xin-Quan(高新全)[e)], and He Wen-Ping(何文平)[b)]

[a)] *Department of Atmospheric Sciences, Nanjing University of Information Science & Technology, Nanjing* 210044, *China*
[b)] *Department of Physics, Yangzhou University, Yangzhou* 225009, *China*
[c)] *National Climate Centre of China, Beijing* 100081, *China*
[d)] *Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing* 100029, *China*
[e)] *College of Atmospheric Science, Lanzhou University, Lanzhou* 730000, *China*

(Received 13 September 2004; revised manuscript received 26 November 2004)

At present, most of the statistical prediction models are built on the basis of the hypothesis that the time series or the observation data are linear and stationary. However, the observations are ordinarily nonlinear and non-stationary in nature, which are very difficult to be predicted by those models. Aiming at the nonlinearity/non-stationarity of the observation data, we introduce a new prediction scheme in this paper, in which firstly using the empirical mode decomposition the observations are stationarized and a variety of intrinsic mode functions (IMF) are obtained; secondly the IMFs are predicted by the mean generating function model separately; finally the predictions are used as new samples to fit and predict the original series. Research results show that the individual IMF, especially the eigen-IMF (namely eigen-hierarchy), has more stable predictability than the traditional methods. The scheme may effectively provide a new approach for the climate prediction.

Keywords: empirical mode decomposition, nonlinear/non-stationary time series, hierarchy theory, climate prediction

PACC: 9260X

1. Introduction

Paying attention to the Hilbert spectrum analysis of nonlinear/non-stationary time series, Huang *et al*[1,2] presented a new method–empirical mode decomposition (EMD) in 1998 and improved it later, which could make the signals stationary and result in a variety of fluctuations with different timescales, in which each fluctuation is designated to one intrinsic mode function (IMF). That is to say, the original data will be decomposed by the EMD method and the various timescale stationary signals can be obtained; this process is called the shifting process of data. The first IMF represents high-frequency (small timescale or low hierarchy) component of the original series. As decomposition goes on, the IMF's frequency will be reduced gradually and becomes the lowest for the last one, which usually acts as the indication of the trend of the original series. Then, these IMFs, derived from the original series, can serve as the basis of that expansion which can be linear or nonlinear as dictated by the original series, and it is complete and almost orthogonal. A sum of these IMFs and a mean trend can be reconstructed into the original series. As mentioned above, the main purpose of the EMD method is to carry out the Hilbert transform and obtain the Hilbert spectrum, which may reflect a certain physical background of the original system. In recent years, the combination of the EMD method and the Hilbert spectrum analysis has offered a powerful tool for nonlinear/non-stationary time series, which has been gradually applied to many fields, such as biology, oceanography, meteorology, medicine, engineering and astronomy, etc.[3−10]

*Project supported by the National Natural Science Foundation of China (Grant Nos 40275031 and 40325015), the National Key Programme Development for Basic Research (Grant No 2004CB418300) and Jiangsu Province Key Laboratory of Meteorological Disaster and Environmental Variation (KJS0302).

†Corresponding author. E-mail: feng_gl@sina.com

As we know, the climate system is a typical nonlinear and non-stationary one; and short-term climate prediction has been a frontier in atmospheric science and also a difficult subject in multidisciplinary science fields, because of the interaction of subsystems—the atmosphere, the hydrosphere, the lithosphere and the biosphere, etc, making the climate system nonlinear and non-stationary. While most of all the present major prediction methods with the observation data for the short-term climate are statistical and built on the basis of the theory of correlation/analogue, such as the mean generating function (MGF) model and optimal climate normal (OCN) model, etc, designed mainly for the linear/stationary systems, these models have weak abilities to predict a nonlinear system. Due to the multi-spatial/temporal scale, the evolution of the climate system is so complicated that it is difficult to build an accurate and effective regression equation with the observation data.[11−17] In view of these facts, it is necessary that the observation data be stationarized and hierarchized. A different hierarchy has different predictability and stability for the climate system.[18,19] Therefore, it is significant to extract the most stable and the highest predictable hierarchy from the observation data for the climate prediction, which may improve the prediction skill. In this paper, a new scheme including the EMD method, the MGF model and the optimal subset regression (OSR) model is used to improve the climate prediction skill for the Yangzhou 530a (1470-1999) dryness/wetness indices (DWI).[20]

2. IMFs of the DWI

The process of the EMD for the time series is described as follows (for detailed principles about EMD refer to Ref.[1]).

If the number of maxima/minima of a series $X(t)$ is larger than the number of up-zero /down-zero crossing points by two, then the series is not stationary, and needs to be stationarized. The main procedures include: picking out all of the maxima of the series $X(t)$ and fitting them into the upper envelope of the original series with the cubic spline function; similarly, picking out all of the minima and fitting them into the lower envelope with the cubic spline function. The mean envelope $m_1(t)$ of the series $X(t)$ is the mean value of the upper and lower envelopes; thus a new series $h_1(t)$ with the low frequency removed can be calculated by subtracting the mean envelope from the series $X(t)$:

$$h_1(t) = X(t) - m_1(t). \tag{1}$$

Usually, $h_1(t)$ is still a non-stationary series, and the above procedures must be repeated k times until the mean envelope approaches zero, then the first IMF ($C_1(t)$) is obtained:

$$C_1(t) = h_{1k}(t) - m_{1k}(t). \tag{2}$$

The first IMF represents the highest frequency component of the original series. The second IMF ($C_2(t)$) can be obtained from the margin series $r_1(t)$, which is calculated by subtracting the first IMF from the series $X(t)$. Such a procedure needs to be repeated until the last margin series $r_n(t)$ cannot be decomposed further; here the series $r_n(t)$ represents the mean trend of the original series:

$$\begin{aligned} r_2(t) &= r_1(t) - C_2(t), \cdots, \\ r_n(t) &= r_{n-1}(t) - C_n(t). \end{aligned} \tag{3}$$

The original series can be obtained by a sum of the IMFs and the mean trend:

$$X(t) = \sum_{j=1}^{n-1} C_j(t) + r_n(t). \tag{4}$$

The main purpose of the EMD method is to carry out the Hilbert transform and obtain the Hilbert spectrum, or to make the wavelet analysis and obtain the wavelet coefficients, so that both can reflect the actual physics background, i.e. each IMF may have a corresponding physical process. Therefore, it is very significant to pick out the major IMF, defined as the eigen-IMF or eigen-hierarchy in this paper, with the main law of the climate change from the original series in order to make the long-term prediction.

Although there is a little distortion at the end of the original series, the EMD method can deal with the short data. The DWI series with 530 samples, containing various time-scales such as annual, interdecadal and centennial scale, etc information, is long enough to guarantee the reliability of the voluminous decomposition.

According to the previous climate diagnoses,[21,22] the Yangzhou 1470–1999 DWI (Fig.1(a)) is decomposed into six hierarchies (Fig.1(b)–(g), i.e. IMFs.1–6) by the EMD, and the wavelet analysis of the IMFs shows that this DWI contains many timescales such as less than 10a period (IMF1), 10a period (IMF2), 20–40a period (IMF3), 70a period (IMF4), 100a period

(IMF5) and more than 100a period (IMF6) etc. For example, the wavelet coefficient of the IMF3 reveals a clear and 30a quasi-period in Fig.2; as will be shown later, this hierarchy is a key prediction factor.

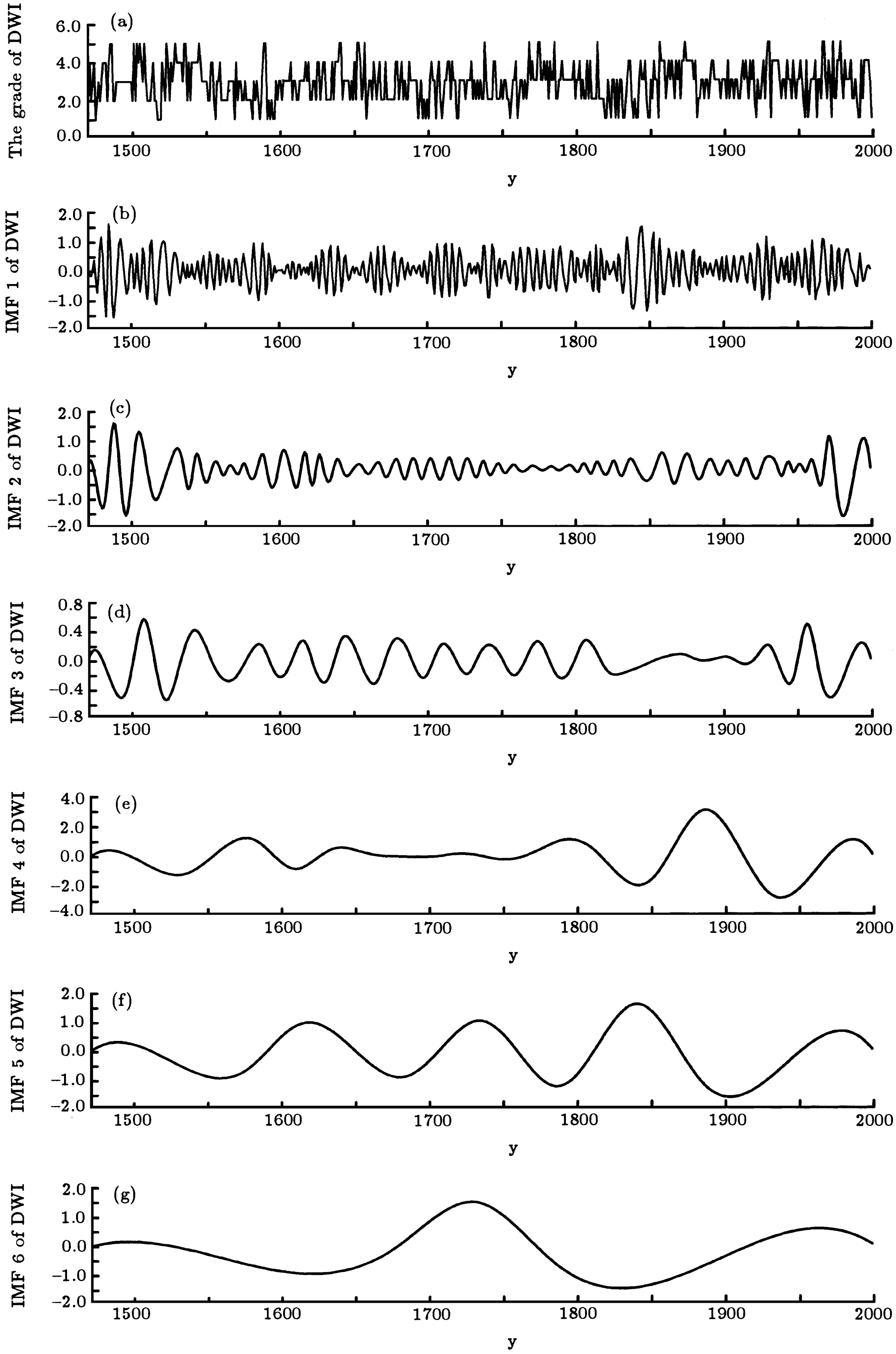

Fig.1. The DWI in Yangzhou (a) and its IMFs.1-6 (b–g).

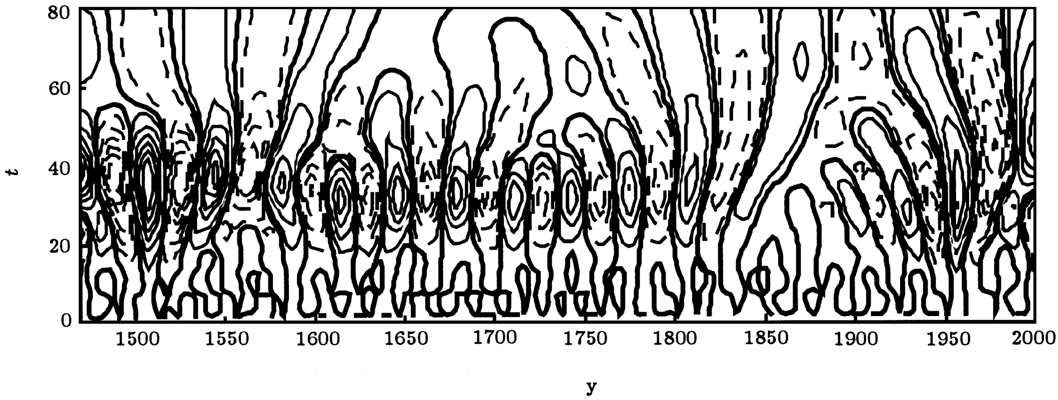

Fig.2. Wavelet coefficient of the IMF3.

3. Modelling of the climate prediction with IMFs

In order to make the climate prediction by using the EMD method, two groups of samples including 8 components in each group are selected (shown in Tables 1 and 2): (1) the length of 100a (1770–1869); (2) the length of 90a (1870–1959); and then allowing 30a (1870–1899) and 40a (1960–1999) to attain new samples separately. Then by using the MGF model, the first predictions with the two groups are performed; and the first 25 predicted values in group 1 and the 35 ones in group 2 are separately taken as the new samples to build the fitting equations. Allowing the last 5 values of the both for checking the second real predictions made by using the OSR model, the fitting and predicting are preformed in three kinds of cases: (a) an IMF, (b) the single DWI series (No.7) and (c) the association of IMFs (No.8). Finally, an eigen-IMF, having the largest correlation coefficient and the least root mean square error (RMSE), will be picked out to represent the original DWI series.

The results for group 1 are listed in Table 1. According to the second OSR fitting of the new samples predicted by the MGF model, here No.3 (IMF3) is just an eigen-hierarchy, namely a stationary component with a quasi-period of 20–40a, whose RMSE is the least (RMSE=0.81). And the largest (R=0.51) corresponding correlation coefficient of it, reaches a degree of confidence of 0.01 in t-test. Therefore, the IMF3 would be the most stable prediction factor and the best substitution of the trend of the DWI during 1770–1899. Meanwhile, it can be seen from Table 1 that the RMSE of the DWI series is 0.94 (No.7), and only 0.05 for the correlation coefficient, which does not pass the t-test. When the number of the IMFs added to the OSR fitting increases from 1 to 6, the corresponding RMSE and correlation coefficient are both improved obviously (RMSE=0.66 and R=0.71), the latter passes the t-test at 99.9%. Generally speaking, the fitting degree with one IMF is higher in the high-frequency range (low hierarchy, i.e. No.1), lower in the low-frequency range (high-hierarchy, i.e. IMFs.5–6) and the best is in the intermediate-frequency (i.e. IMFs.2–4). With so different an oscillation or a physics background as the decadal-scale oscillation for IMF2, the 20–40a oscillation for IMF3 and the 70a oscillation for IMF4, etc, the contribution of each IMF to the DWI prediction may be different in different periods. From Table 1 and Fig.3, we can see that the eigen-hierarchy (IMF3, i.e. No.3 in Fig.3) and the association of the IMFs (No.8 in Fig.3) have much bigger ability of prediction than that of the single DWI series (No.7 in Fig.3). The real predictions with the OSR model by means of both the eigen-IMF and the association are in good agreement with the general trend of the actual DWI, i.e. the trend in the period from 1895 to 1900. However, the real prediction for the single original DWI series only denotes a trend of average, and does not reflect the

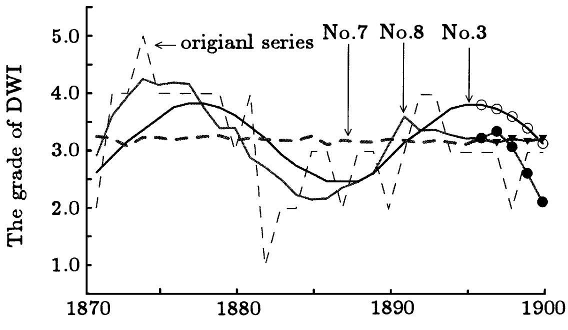

Fig.3. The curves of the fitting (1870–1894) and the prediction (1895–1899) in three kinds of cases (namely Nos 3, 8 and 7).

actual evolution of the observation data, such as the jump points. Based on the above analysis, there is an obvious gap between the prediction skills of the eigen-IMF (or of the association of the IMFs) and that of the single original DWI series.

Table 1. The results of the OSR fitting with the new samples predicted by MGF model (Sample size: 25a; i.e. 1870–1894).

Serial Number	Variable	IMF cycle/a	RMSE	Correlation Coefficient	*t*-test**
No.1	IMF1	< 10	0.94	0.04	
No.2	IMF2	10	0.89	0.33	
No.3	IMF3	30–40	0.81	0.51	0.01
No.4	IMF4	70	0.85	0.43	0.05
No.5	IMF5	100	0.92	0.17	
No.6	IMF6	> 100	0.88	0.33	
No.7	Original series		0.94	0.05	
No.8	IMFs.1–6*		0.66	0.71	0.001

*Denotes the association of IMFs.

** The blank denotes that the correlation coefficient does not pass the *t*-test.

In group 2, the IMF samples derived from the end of the data are distorted somewhat in the EMD process, but this test is of more practical significance for making the prediction. In order to reduce the distortion effects, the sample size of this group is decreased by 10, while the prediction step of the MGF model is increased by 10, and all the results are listed in Table 2. Although the abilities of the samples to fit the original DWI series are declined, the differences among their predictions are similar to those in group 1. For example, the RMSE for No.8 is 1.04, while the correlation coefficient can still reach 0.49, which passes the *t*-test at 99%. The association of IMFs predicts the trend of the actual DWI during 1995–1999 (No.8 in Fig.4) correctly. The IMF2 (No.2 in Fig.4) with a 10a oscillation during 1960–1999, whose correlation coefficient passes the *t*-test at 95%, can serve as the eigen-IMF of the original DWI series. To the secondary role, here is the IMF3 (No.3 in Fig.4) with a 30–40a oscillation, which also can basically predict the trend of the actual DWI in the same period (the figure is omitted); similar to that in group 1, No.7 can hardly be used to predict the future DWI.

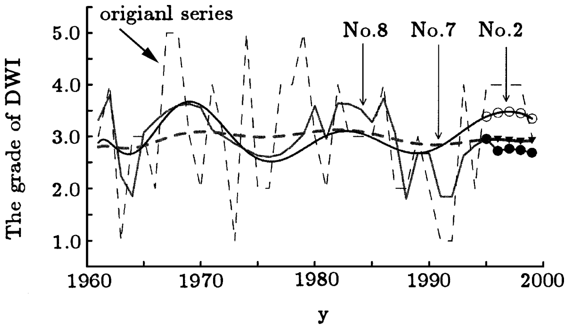

Fig.4. The curves of the fitting (1960–1994) and the prediction (1995–1999) in three kinds of cases (namely Nos. 2, 8, 7).

Table 2. The results of the OSR fitting with the new samples predicted by MGF model (Sample size: 35a; i.e. 1960–1994).

Serial Number	Variable	IMF cycle/a	RMSE	Correlation Coefficient	*t*-test**
No.1	IMF1	< 10	1.18	0.05	
No.2	IMF2	10	1.10	0.35	0.05
No.3	IMF3	30–40	1.17	0.16	
No.4	IMF4	70	1.18	0.05	
No.5	IMF5	100	1.18	0.11	
No.6	IMF6	> 100	1.17	0.14	
No.7	Original series		1.17	0.16	
No.8	IMFs.1–6*		1.04	0.49	0.01

*Denotes the association of IMFs.

** The blank denotes that the correlation coefficient does not pass the *t*-test.

The prediction of the both groups show that, for the nonlinear/non-stationary time series containing various hierarchies, making prediction with its eigen-IMF instead of itself will be more effective. As far as the Yangzhou 530a DWI, a nonlinear/non-stationary series, is concerned, its eigen-IMF is the hierarchy with an oscillation of 10–40a, which is the most stable and predictable, and can be used to represent the general evolution of the climate system.

4. Summary

Aiming at illuminating the nonlinearity/non-stationarity of the climate system and the deficiencies of the existing statistical prediction models, a new scheme based on the EMD method is proposed in this paper to improve the predictability of the climate. The scheme is designed such that by the EMD process the observation data are firstly stationarized and transformed into IMFs, i.e. multifarious stationary series, then with the IMFs, a prediction is done by the MGF model and new values are obtained. Finally, the values are used as new samples to fit and predict the original series (DWI in this paper). The test results also show that an individual IMF, especially for the eigen-IMF, has more stable predictability than that of its source, so the more the IMFs as fitting factors in the OSR model, the stronger the prediction ability of the model. To some extent, the eigen-IMF has more prediction skill and may replace the observation data for the climate prediction. This novel scheme can effectively improve the prediction ability of the present mathematical models, and may found wide applications in other relevant studies.

References

[1] Huang N E and Zheng S 1998 *Proc. R. Soc. Lond.* A **454** 903

[2] Huang N E, Zheng S and Long S R 1999 *Ann. Rev. Fluid Mech.* **31** 417

[3] Wei H, Zheng S, Norden E Huang *et al* 1998 *Proc. Natl. Acad. Sci.* **95** 4816

[4] Coughlin K T and Tung K K 2004 *Adv. Space Res.* **34** 323

[5] Li H S, Chen W W and Zhao Z. 2003 *Comput. Eng.* **29** 195 (in Chinese)

[6] Lin Z S and Wang S G 2004 *J. Trop. Meteorol.* **20** 90 (in Chinese)

[7] Yang Y, Yu D J, Chen J S *et al* 2003 *J. Hunan Univ. (Natural Sciences)* **30** 25 (in Chinese)

[8] Zhao L F, Zhang X L and Song J 2003 *J. Ocean Univ.* Qingdao **33** 759 (in Chinese)

[9] Hu J S, Yang S X, Zhou F J *et al* 2003 *Turbine Technology* **45** 308 (in Chinese)

[10] Zhao H W and Norden E Huang 2004 *Proc. Math. Phys. Eng. Sci.* **460** 1597

[11] Liu S D, Fu Z T *et al* 2002 *Acta Phys. Sin.* **51** 718 (in Chinese)

[12] Liu S S, Fu Z T, Chen H *et al* 2003 *Acta Phys. Sin.* **52** 1842 (in Chinese)

[13] Mo J Q and Lin W T 2004 *Acta Phys. Sin.* **53** 996 (in Chinese)

[14] Feng S D, Ren R C, Cui X P *et al* 2001*Chin. Phys.* **10** 1106

[15] Feng G L, Dai X G and Wang A H *2001 Acta Phys. Sin.* **50** 606 (in Chinese)

[16] Feng G L and Dong W J 2003 *Acta Phys. Sin.* **52** 2347 (in Chinese)

[17] Feng G L, Dong W J and Li J P 2004 *Chin. Phys.* **13** 1582

[18] Mu M, Li J P, Chou J F *et al* 2002 *Clim. Environ. Res.* **7** 227 (in Chinese)

[19] Li J P and Chou J F 2003 *Chin. Sci. Bull.* **48** 703

[20] Zhang D E, Li X Q and Liang Y Y 2003 *J. Appl. Meteorol. Sci.* **14** 379 (in Chinese)

[21] Zhu J H and Wang S W 2001 *Adv. Atmos. Sci.* **18** 1043

[22] Zhu Y M and Yang X Q 2003 *Adv. Atmos. Sci.* **20** 694