

# Application of EOF and Cao Method to the Research on Decadal Variability of ENSO Predictability

PENG Yue-hua<sup>1, 2+</sup>, DUAN Zheng-yu<sup>3</sup>, ZHOU Bo<sup>3</sup>, ZHU Zuo-xin<sup>3</sup>

<sup>1</sup> Dalian Naval Academy, Dalian, China

<sup>2</sup> PLA University of Science and Technology, Nanjing, China

<sup>3</sup> Dalian University of Technology, Dalian, China

**Abstract.** In order to research the decadal variability of El Niño-Southern Oscillation (ENSO) predictability with observational data, nonlinear signal analysis method is used to investigate the extent of certainty in different decades. Due to the conundrum of measuring the extent of certainty, Empirical Orthogonal Function (EOF) and Cao method are combined to research the data with spatio-temporal structure. The results show that: (1) Cao method can be used to roughly estimate the extent of certainty and randomness but it is not precise enough; (2) for data with spatio-temporal structure, Cao method can be applied to the primary component of EOF decomposition and the effect is significantly improved; (3) the decadal variability of ENSO predictability is not simply caused by observational data and further investigation is necessary.

**Keywords:** nonlinear signal; Cao method; EOF; predictability

## 1. Introduction

At present, the discussion that whether the signals give priority to determinability or randomness and what are their proportions is still a large problem internationally. Essentially, that is a modeling problem: is it better to use a deterministic or stochastic model, and in what proportion? One way to address this is with deterministic modeling procedures that are robust against overfitting. The opposite approach would be to apply surrogates to determine how much of the system variability can be explained as a stochastic source [1]. These schemes are used in theory research right now and have limitations separately in disposal of actual data. Cao method which is used to determine the minimum embedding dimension can also distinguish deterministic signals from stochastic signals and it can be used to qualitatively estimate the noise level but the result is not precise enough [2].

In meteorology, results of predicting ENSO have been showing considerable uncertainties [3]. The measurement of ENSO usually uses the Sea Surface Temperature Anomaly (SSTA) and the series of SSTA evolving with time on certain grid forms a signal. The observation series usually contain determinate component and random component. Generally, under the same model condition, more random component means more predictability. The results of many models show that the predictability of ENSO does not increase with the enhancement of model performance but presents definite decadal variability [4]. Whether the analysis of observation data supports this conclusion? For the data had temporal and spatial structure like SSTA, what is the result of judgment with Cao method? Is there any improved method? A method called Empirical Orthogonal Function (EOF) can dispose data with temporal and spatial structure and suitably analyze the principal component [5]. Then what about compounding of Cao method and EOF method? What is the outcome of this combining method? This paper tries to probe into these problems.

## 2. Model Results Of Decadal Enso Predictability And Data

### 2.1. Modeling results of decadal ENSO predictability

---

<sup>+</sup> Corresponding author. *E-mail address:* pengyuehua@hotmail.com.

Climatologists found out the predictability of ENSO had a decadal variability when using numerical models to predict. Youming Tang, Ziwang Deng et al experimented with three coupled models, including Zebiak-Cane model (ZC-model) and two hybrid coupled models — HCM1 and HCM2 [4]. We are concerned about the results in the second half of 20th century so those from 1950 to 2000 are illuminated briefly. The results of three models are almost the same, that is: the correlation coefficients are small and the mean square errors are large between 1950 and 1960 (i.e. 5th decade) so the predictability in 5th decade is small, from the same reason, the predictability is being larger in 6th decade while being smaller in 7th decade and being larger again in 8th and 9th decade.

## **2.2. Data and the source**

The data in this paper are the observational time series of sea surface temperature anomaly (SSTA) which are from the Extended Reconstructed Sea Surface Temperature Dataset (ERSST) of National Oceanic and Atmospheric Administration (NOAA). The data in spatial grid are ordered globally by 2\*2 and their temporal periods are arranged from January 1854 to December 2008 totaling 1860 months. Due to too many grids globally and the SSTA in the equator Pacific having significant physical meanings, we only selected the data in the equator Pacific.

In order to contrast with the decadal variability of model results, the dataset can be segmented. The data from January 1854 to December 1949 totaling 1152 months were referred to as the base data, hereafter data of 120 months (i.e. 10 years) were added to make the time series longer to reflect more characteristics of the whole. In this way, the symbol “1152” signified the 5th decade of 20th century ago, and “1272” signified 5th decade, and “1392” signified 6th decade, and “1512” signified 7th decade, and “1632” signified 8th decade, and “1752” signified 9th decade, and “1860” signified the period from January to the end.

## **3. Results Using Cao Method Alone**

### **3.1. The delay time calculated by mutual information method**

The Cao method requires knowing the delay time of time series beforehand [6]. There are many approaches to get the delay time. In this paper mutual information method was chosen. As the judgments of ENSO were usually the variability of SSTA in eastern Pacific of the equator, we chose 8 points ranged from 104oW~90oW to calculate their delay time and then drew figures with Cao method.

### **3.2. The results of Cao methods**

There were two important criterion values in Cao method---E1 and E2, and the evolvement of E1 and E2 with embedding dimensions could represent the results [6]. Since signals of different length were pretty much the same thing, we just showed two signals in 8 points whose time period were the 5th decade (signed “t1152”) and the 9th decade (signed “t1752”) in Fig. 1. The figure showed that values of E1 were tend to be stable around abscissa “dimension=10”, that is to say, the minimum embedding dimension could be 10; moreover, the values of E2 were basically close to 1 and fluctuated in a small range so the randomness is comparatively strong. However, we could not see the extent of certainty and randomness in different decades so we used simple statistics to analyze the values of E2.

### **3.3. Statistical analysis of E2 (m)**

The measurement of certainty was decided by the distance between E2 and 1. Closer distance meant smaller fluctuation and larger randomness. Therefore we could subtract 1 from E2 and then calculate the maximum, the minimum, the mean, the standard deviation of 8 points and draw figures to contrast. For convenience, the data from “t1152” to “t1860” could be marked with sequence number from “1” to “7”. For the sake of visualization, these statistical values could be plotted. The standard deviation could be done individually because its difference is very small and it may measure the fluctuating extent. Fig. 2 shows the results. We can see that: (1) with the signal longer and longer, the maximums, minimums and means are neither nearer and nearer nor farther and farther, so do the standard deviations. The deviations are mostly increasing except the one from data marked “1” to data marked “2” and that from data marked “4” to data marked “5”. It may be the reasons that on the one hand there are noises in the signal but longer signal doesn't mean stronger noises, on the other hand the observation precision is enhancing with the time forward but it

could not certainly weaken the noises. (2) The minimum and mean of signal marked “4” are smaller than those of adjacent signal marked “3” and “5” and the standard deviation is larger so signal “4” could be deemed to have larger certainty, that is, the data in the 7th decade contain less noises which means larger predictability under the same modeling condition. (3) The minimum of signal marked “5” is distinctly larger than those of adjacent signal marked “4” and “6” and the standard deviation is smaller so signal “5” could be deemed to have larger randomness, that is, the predictability of the data in the 8th decade is smaller. For the same reason, the predictability of the data in the 5th decade is also smaller.

There was much discrepancy between this result and the model result. The cause was either unsuited method or iffy data. We improved the method in the first instance.

## **4. Results Of Compounding E of And Cao Method**

### **4.1. Improvement of the method**

Owing to the poor results of using Cao method individually, other method was required. In view of the temporal-spatial structure in the SSTA data, the EOF method could be used to combine with Cao method. The way was to compartmentalize the data according to different decades to apply EOF method and then take the primary mode using Cao method to analyze. The following section shows the results of EOF method firstly and then the results of applying Cao method to time coefficients series afterward.

### **4.2. Results of EOF method**

Applying EOF decomposition to the data segments, 8 time coefficients (i.e. expansion coefficients of 8 points) series and corresponding variance contributions could be gained. From the variance contributions we can see that those of the first mode occupied the most proportion in all the modes.

### **4.3. Statistical analysis of Cao method towards the primary mode**

Similar to section 3, Cao method was used to analyze the primary time coefficient series because of their largest variance contributions. In this section the results of delay time were omitted because they are meaningless. The evolvments of E1 and E2 with embedding dimensions were plotted in Fig. 3 like Fig. 1 but there were a remarkable distinction that different line represented signal of different length in Fig. 3 but series of different spatial position in Fig. 1. We could see that, values of E2 were not far away from 1 so the randomness was still strong. This figure was not enough to arrive conclusion therefore the statistical results of E2 was required.

Since the line figure was visual, we just showed the figure of statistical values in this section in Fig. 4. The figure indicated that: (1) with the increasing data, the evolutive tendency of the maximums, minimums and means were consilient which was noticeable distinct with section 3. In view of that, the judgments of predictability in this section could mainly see the evolvment of the standard deviation. (2) The standard deviation of signal marked “4” were smaller than those of adjacent signal marked “3” and “5” so signal “4” could be deemed to have smaller certainty, that is, the data in the 7th decade contain more noises which means smaller predictability. (3) The standard deviation of signal marked “5” were larger than those of adjacent signal marked “4” and “6” so signal “5” could be deemed to have smaller randomness, that is, the data in the 8th decade contain less noises which means larger predictability. These conclusions are contrary with those in section 3 but coincided with the model results in section 2 which means our method is remarkably improved.

Although this result is close to the model result but not the same (i.e. not perfect). There are probably two reasons. On the one hand, the base data may interfere with the judgments of whole signal. To prevent this interfering, comparison among data of different decades purely may be better. The daily data could be used to ensure enough length of the signal. Though more intensive time usually meant stronger randomness, the random extent of individual series would not affect the comparison among different decades. On the other hand, the decadal variability of ENSO predictability is not simply caused by observation. Further investigation is necessary.

## **5. Summary**

A nonlinear signal analysis approach, the Cao method, and its compounding with EOF method were applied to the SSTA series in the late half of 20 century in order to research the relation between decadal variability of ENSO predictability and observation data, main conclusions could be as follows:

Cao method could be used to judge the extent of certainty and randomness roughly but was not precise;

For data with temporal-spatial structure, the results of using EOF method to decompose and then applying Cao method to the primary mode afterwards were remarkably better, but not perfect;

The decadal variability of ENSO predictability was not simply caused by observation, so further investigation was necessary.

In this paper the values of E2 gained by Cao method were statistically analyzed. Whether this approach is appropriate and universal needs further discussion. However, under the circumstance of difficult dilemma in judgment to the extent of certainty and randomness right now, the approach in this paper has some sense. Especially for the data with temporal-spatial structure, it is exemplary to use EOF method to decompose and then apply Cao method to the primary mode afterwards.

## 6. References

- [1] M. Small, "Applied Nonlinear Time Series Analysis: Applications in Physics, Physiology and Finance," Singapore: World Scientific, 2005.
- [2] X. K. Xu, "Sea Clutter Processing and Target Detecting Based on Nonlinear Analysis," Dissertation for the degree of Doctor in Dalian Maritime University, 2008:41-60
- [3] C. Y. Li, M. Mu, G. Q. Zhou, H. Yang, "Mechanism and Prediction Studies of the ENSO," Chinese Journal of Atmospheric Sciences, 2008, 32(4): 761-781
- [4] Y. M. Tang, Z. W. Deng, X. B. Zhou, Y. J. Yan, "Interdecadal variation of ENSO predictability in multiple models," Journal of climate. 2008, 4811-4832
- [5] M. Zhang, J. An and M. Zhu, "EOF Expansion in One Rainstorm," Chinese Journal of Atmospheric Sciences, 2007, 31(2): 321-328
- [6] L. Y. Cao, "Practical method for determining the minimum embedding dimension of a scalar time series," Physica D: Nonlinear Phenomena, 1997, 110(1-2): 43-50.

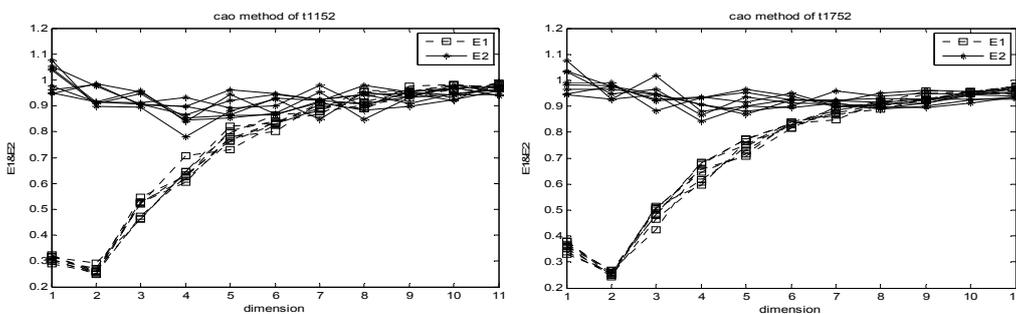


Fig.1. The evolutions of E1 and E2 with embedding dimensions in 8 points: (a) for signal signed "t1152"; (b) for signal signed "t1752"

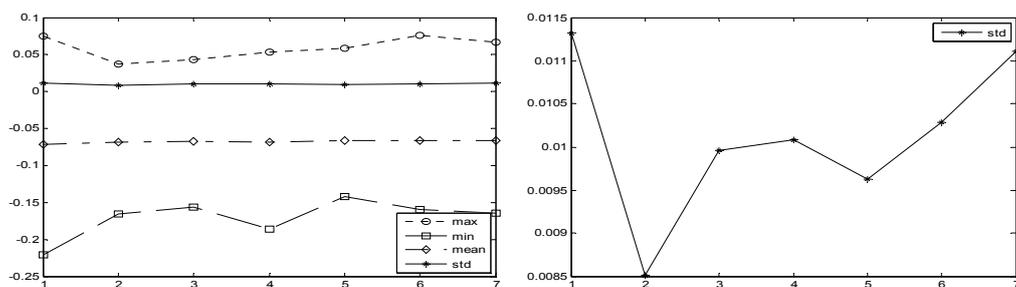


Fig.2. The statistical values of E2 with signal sequence number of different length by using Cao method alone: (a) maximums, minimums, means, standard deviations; (b) magnifying standard deviations.

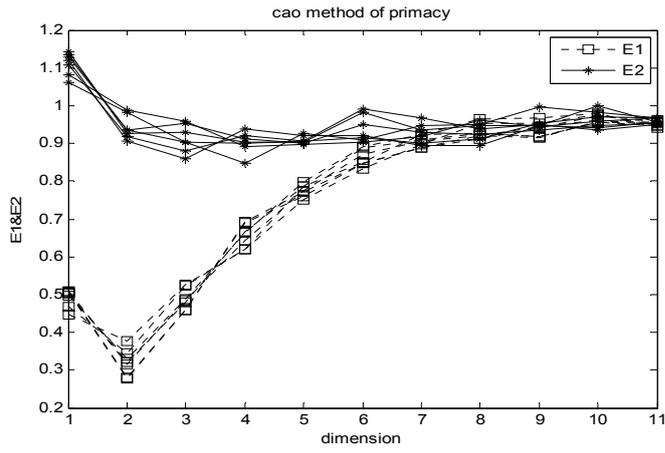


Fig.3. The evolutions of E1 and E2 with embedding dimensions in the primary mode for data of different length

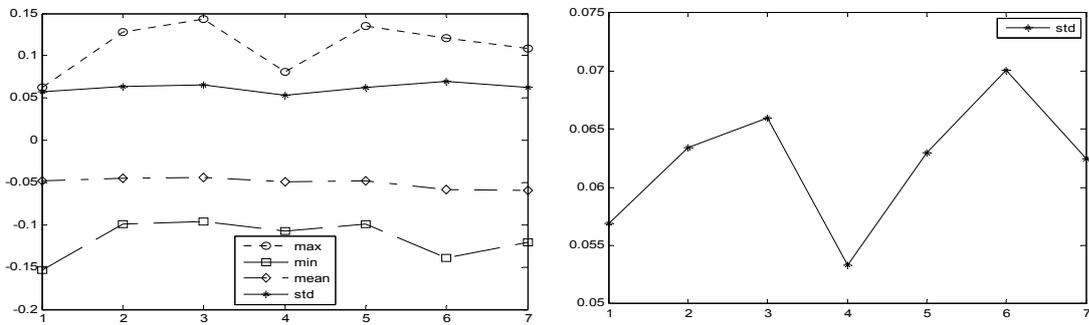


Fig.4. The statistical values of E2 with signal sequence number of different length by combining EOF method Cao method: (a) maximums, minimums, means, standard deviations; (b) magnifying standard deviations.