Short-range Climate Prediction Experiment of the Southern Oscillation Index Based on the Singular Spectrum Analysis①

Liu Jianwen (刘健文) and Dong Peiming (董佩明)
Beijing Aviation Institute of Meteorology, Beijing 100085

(Received September 1, 2000)

ABSTRACT

The Southern Oscillation Index (SOI) time series is analyzed by means of the singular spectrum analysis (SSA) method with 60-month window length. Two major oscillatory pairs are found in the series whose periods are quasi-four and quasi-two years respectively.

The auto-regressive model, which is developed on the basis of the Maximum Entropy Spectrum Analysis, is fitted to each of the 9 leading components including the oscillatory pairs. The prediction of SOI with the 36-month lead is obtained from the reconstruction of these extrapolated series. Correlation coefficient between predicted series and 5 months running mean of observed series is up to 0.8. The model can successfully predict the peak and duration of the strong ENSO event from 1997 to 1998.

It's also shown that the proper choice of reconstructed components is the key to improve the model prediction.

Key word: Southern Oscillation Index, Singular Spectrum Analysis, Principal component, Reconstruction

1. Introduction and background

As a complicated dynamic system, climate change possesses a wide range of spatial and temporal scales. Due to its direct social and economic impacts, climate variability has become one of the most interesting subjects and has drawn more and more attentions from many governments. The ENSO cycle is one of the most dominant global climate fluctuations with time scales of a few months to several years. It is remarkably associated with a coherent pattern of ocean and atmospheric anomalies.

Although it is well known that most El Niño events share similar characteristics, each one has special traits of its own. How to predict ENSO event is a very important item to climate study.

Dynamical and statistical models are now the two main methods on ENSO cycle prediction. Although the former has gained great progress due to deepened understanding of the coupled tropical ocean-atmosphere interaction mechanism, some unsolved problems, such as " climate drift", still remain. In many cases the latter works better. And, it has been shown that statistical forecast models can also provide valuable skill on interannual scales.

The singular spectrum analysis (SSA) (Vautard et al., 1992) is a linear analysis and prediction technique. It has been applied for years, mainly in digital signal processing, non-linear dynamics, and climatic diagnosis. Recently it has also been used in climatic diagnosis and prediction in China and a series of results have been gotten (Dong et al., 1998;

①This work was supported by the " National Key Project Studies on Short-Range Climate Prediction System in China" (96-908-04-02).

Ding et al., 1998; Jiang and Ding, 1998).

It was shown that SSA was particularly successful in forecasting the Southern Oscillation Index (SOI) with considerable skill before the 1990s (Keppencne and Ghil, 1992). In this paper, SSA is also applied to the time series of SOI with a dataset up to 1997. On the basis of analysis, a short–range climatic prediction model of SOI was developed and some experiments on it were carried out.

2. Data and methods

The basic dataset used in this study is the Southern Oscillation Index time series computed from monthly mean sea level pressure of Darwin and Tahiti provided by Chinese National Climate Center (CNCC). The index series ranges from 1951 to 1997. The series and its 5 months running mean are shown in Fig. 1. The correlation coefficient of two time series is 0.85.

SSA can be as an analytic tool. Vautard and Ghil (1989), and Plaut and Vautard (1994) pointed out that SSA was particularly effective in isolating harmonic oscillations with fluctuating amplitudes from noisy data. The fundamental property of SSA lies in the fact that

(1) two consecutive eigenvalues are nearly equal;

(2) the two corresponding time sequences described by the time empirical orthogonal functions (T–EOFs) are nearly periodic, with the same period, and in quadrature;

(3) the associated time principal components (T–PCs) are in quadrature.

There is then an oscillation in the series whose period is the same as that of T–EOFs themselves.

SSA is used in the study to isolate the T–PCs corresponding to ENSO activity from the remaining variability and noise. Since the T–PCs are the filtered versions of the raw SOI, their behavior is more regular than that of the original signal and more predictable accordingly. In practice we can take advantage of each significant component and reconstruct them with reconstruction algorithm. Obviously, the original signal is exactly the summation of all the reconstruction components.

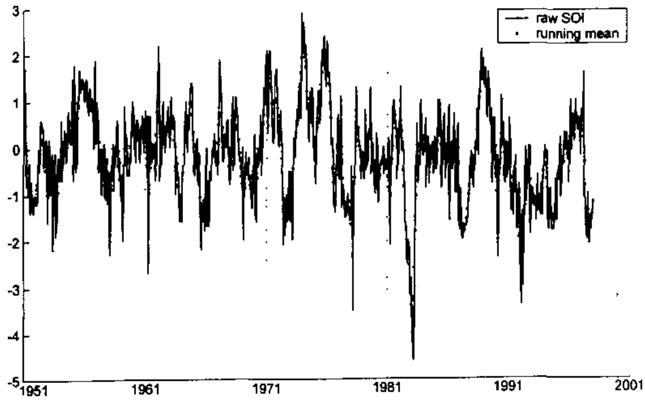

Fig. 1. Time series of SOI index (solid line) and its 5 months running mean (dotted line) from January 1951 to December 1997.

Furthermore, based on the above–prefiltered SOI dataset, forecasts can be made. The prediction procedures are as follows: For a given final year, SSA yields a set of T–EOFs. On the basis of these T–EOFs, a set of T–PCs are built. An Auto–Regression(AR) model is fitted to each leading T–PC so that these T–PCs are extrapolated. The reconstruction of those extrapolated T–PCs is regarded as a forecast result.

3. SSA results

SSA is applied to the SOI dataset with a window width of M = 60 months. Eigenvalues 2–3,4–5 form two pairs (see Fig. 2). Their associated T–EOFs are nearly periodic and in quadrature (see Fig. 3). The T–EOFs 2 and 3 have a period of 4–5 years, the next pair (T–EOFs 4 and 5) corresponds to a period of 2–3 years. Their associated T–PCs are in quadrature, too (see Fig. 4). This result is similar to that of Keppenne and Ghil (1992).

Spectral estimation of T–PCs with the maximum entropy method shows that the periods

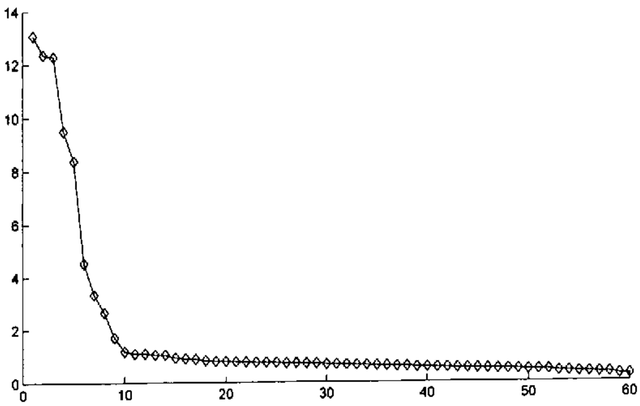

Fig. 2. Spectrum of the lag–covariance matrix. Eigenvalues are shown as percentages of the total variance.

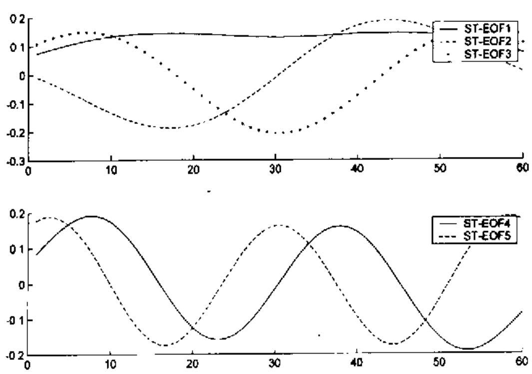

Fig. 3. The T–EOFs from 1 to 5.

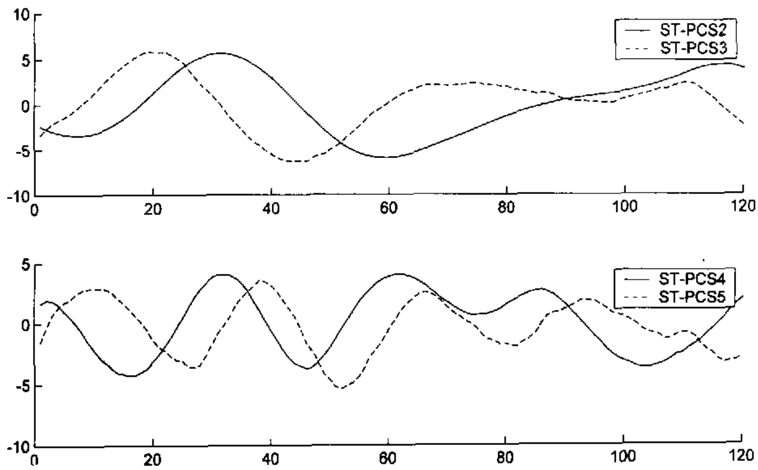

Fig. 4. The T–PCs of the last ten years (for period 1988–1997).

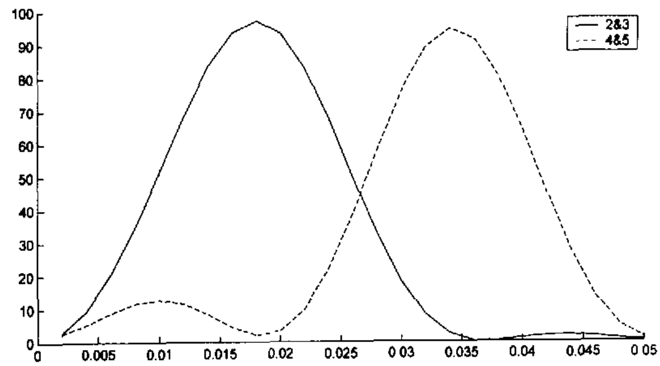

Fig. 5. Spectral estimation of the T–PCs for pairs 2 and 3, 4 and 5.

are more obvious. The maximal spectral intensities of the two T–EOFs pairs have peaks near 4–5 and 2–3 years respectively, with significance over 90% (Fig. 5).

It can be inferred that the two oscillatory modes are in the series, whose periods are equal to those of T–EOFs pairs respectively. They correspond to the low frequency (LF) component of ENSO and its high frequency (HF) variability, including the quasi–biennial oscillation.

It is noticed that there exist some differences when different time length data are used. Data set before the 1990s yields T–EOFs pairs at 1—2 and 3—4, which also stand for the LF and HF components of ENSO, but they are more obvious than those obtained from data till the 1990s.

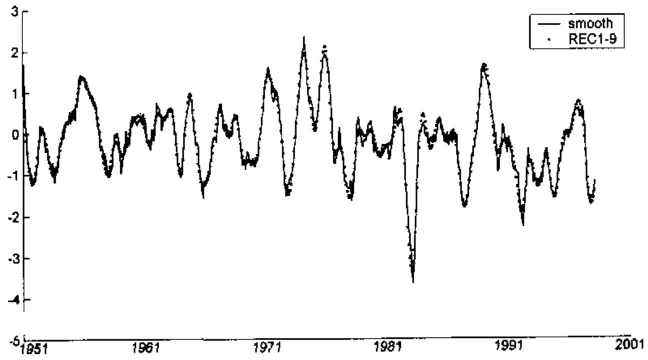

Fig. 6. Time series of reconstruction of T–PCs 1–9 (dotted line) and 5 months running mean (solid line).

A break appears in the spectrum of the lag–covariance matrix at order 9, which indicates that the remaining components after order 9 are close to or within the noise floor. The reconstruction of T–PCs 1–9 is compared with the 5 months running mean of the original SOI time series (see Fig. 6). Their correlation is up to 0.98. So the reconstruction, as well as the running mean can be both regarded as noise reduction and enhancement of the ENSO cycle.

Reconstructed time series obtained by combining the variance associated with two pairs of temporal principal components and 5 months running mean of original SOI are shown in Fig. 7. The correlation coefficient between these two time series is 0.82. The eigenvectors associated with oscillations represent 42.5% of the total variance. The LF component associated with pair 2—3 carries 24.6% of the variance and the other 17.9% is carried by the HF component associated with pair 4—5. The HF and LF components of ENSO are of comparable magnitude. They take decisive positions in the ENSO event.

The reconstruction of T–PCs 1–5 has a correlation coefficient of 0.92 with the running mean series. It reflects the main characteristic of ENSO as the reconstruction of T–PCs 2–5 does (Fig. 8). But it is also noticed that the T–PCs 1 has much effect on the reconstructed time series, especially those in the 1990s.

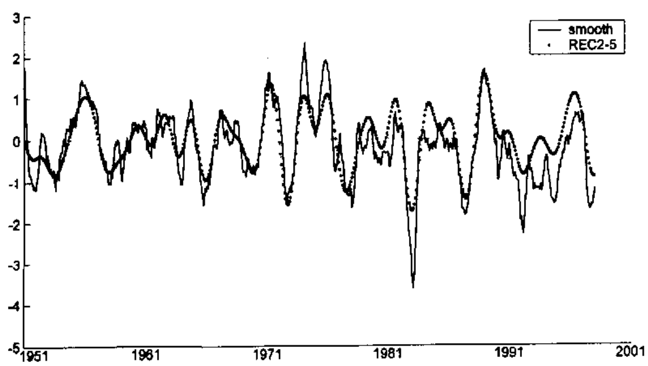

Fig. 7. Reconstructed time series obtained by combining the variance associated with two pairs of T–PCs (dotted line)and 5 months running mean (solid line).

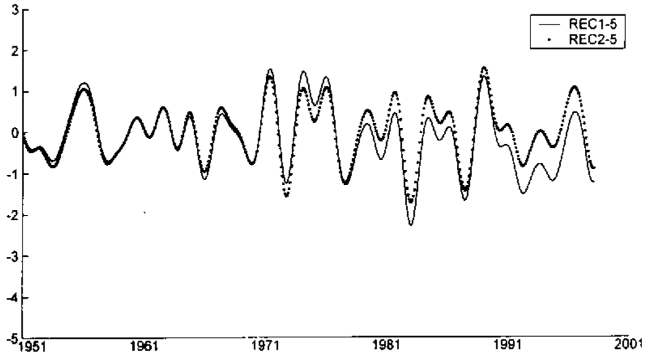

Fig. 8. Time series of reconstruction of T–PCs and T–EOFs 1–5 (solid line) and time series of reconstruction of T–PCs and T–EOFs 2–5(dashed line).

4. Prediction experiments

An AR model is fitted to each individual T–PC. The main idea lies in that the harmonic content of these time series makes them more predictable than the original one. It should be pointed out that the forecast is based on a set of pre–filtered time series, rather than on the raw SOI itself. Individual T–PC is not pure sine wave, but it has a very limited harmonic content. So AR model performs better in predicting the individual T–PC than the time series itself.

The reconstruction algorithm is used after the forecasts of individual T–PC. The SOI forecasts are obtained by summing up the forecasts corresponding to the leading T–PCs.

Firstly, the AR model is applied separately to the leading 1–5 T–PCs to make a 36–month forecast for each of them. Once a new SOI value is added to the unfiltered data, the T–EOFs and T–PCs are recomputed. AR model is established again and another 36–month forecast is made.

To estimate the prediction skill, we have performed two groups of prediction experiment with starting point from January 1986 to December 1988, and from January 1991 to December 1993, respectively. In these two groups of experiment, the entire procedure is repeated 36 times to make the 36 36–month SOI forecasts individually.

Figure 9 illustrates the forecast skill of the first group experiment. Figure 9a gives the correlation coefficients between SSA–AR,persistence forecast and the 5–month running mean time series. We find out that the correlation coefficient of persistence forecast is higher than the forecast in the first 3–month. However, after that, the forecast of SSA method is better than that of the persistence forecast. The correlation coefficient between SSA–AR forecasts and 5–month running mean series excesses 0.6 till 16–month of lead forecast.

Figure 9b shows the mean square errors of SSA–AR, climatology and persistence forecast. Among these three forecasts, the error of climatology is the largest even though it has the smallest value in the first 3–month of lead forecast. SSA–AR model is better than climatology up to 20–month of lead forecast ,and has nearly same error after that time.

The second group experiment, however, gives a different result (see Figs. 10a and 10b).

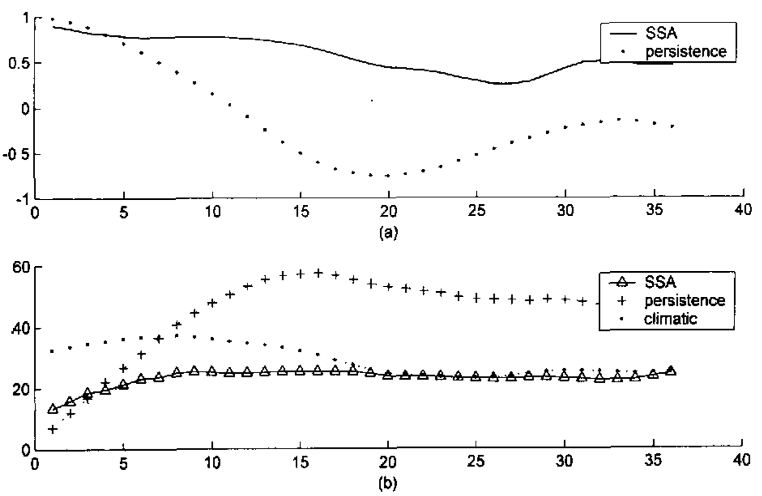

Fig. 9. The forecast skill of the first group experiment. (a) The correlation coefficients between the SSA-AR, persistence forecast and the 5-month running mean data, (b) the mean square errors of the SSA-AR, climatology and persistence forecast.

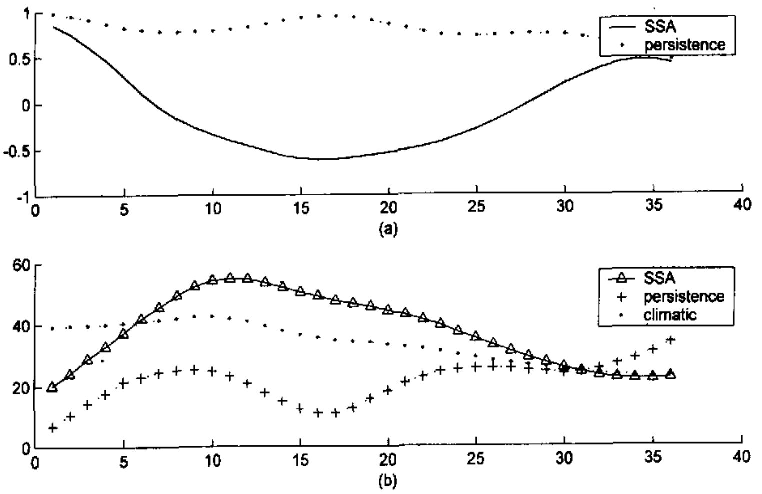

Fig. 10. The forecast skill of the second group experiment. The content is the same as Fig. 9.

The persistence forecast performs better than SSA-AR and climatology method for the period of January 1991 to December 1993. It is suggested that the ENSO cycle might have gotten some changes in the 1990s. It cannot be made certain yet because of the limited time span of data. Indeed, a warm phase of the ENSO cycle ususlly occurs at intervals of two to six years, and typically evolves over a period of roughly one year.

SSA–AR model is re–formed using forecasts of 1–9 leading components,which explain over 70% of the total variance. Inclusion of 6 to 9 components is to take account of the 1–2 years frequency variability that they carry. It performs well for the prediction of the 1997–1998 ENSO event. By use of data up to March 1997, low SOI values corresponding to an ENSO cycle warm episode from June 1997 to July 1998 are predicted, the lowest is at the end of 1997. Using data up to February, it is predicted that there is a weak cold episode of the ENSO cycle after the warm episode. It turns out, in this case, to be more skillful than the two other techniques used in CNCC, called analog (ANA) model and optimum filtering assembly (OFA) model respectively (Guo, 2000).

5. Summary

The singular spectrum analysis method is applied to the time series of the Southern Oscillation index (SOI). The time series is decomposed into terms of temporal principal component and empirical orthogonal functions (T–PCs and T–EOFs) so as to separate the deterministic oscillations from noise. Two major oscillatory pairs are found in the series whose periods are 4–5 and 2–3 years respectively.

The AR model is established for each of the leading T–PCs and each T–PC is extrapolated successively. With these extrapolated PCs reconstructed, a SOI forecast is obtained. Prediction experiments show that this method has certain predictive skill in the forecast of ENSO cycle up to 20 months. Experiments also show that the proper choice of reconstructed components is one of the keys to improve the model's forecast ability.

We are grateful to Professor Robert Vautard of Laboratoire de Meteorologie Dynamique(LMD),France,for his kind advisement and support during the visit at LMD of Mr. Liu Jianwen.

REFERENCES

Ding Yuguo, Jiang Zhihong, and Zhu Yanfeng, 1998: Experiment on short–term climatic prediction to SSTA over the Nino oceanic region. *Journal of Tropical Meteorology*, **14**, 289–296 (in Chinese).

Dong Peiming, Liu Jianwen, and Dong Jie, 1998: Diagnosis of low–frequency oscillation periods of the surface temperature field in Southeast China. *Meteorological Monthly*, **23(11)**, 7–10 (in Chinese).

Guo Yanyun, 2000: The ensemble technique study of ENSO forecast, in *The Study of ENSO Monitoring and Prediction*, edited by Zhai Panmao, China Meteor. Press, 115–124 (in Chinese).

Jiang Zhihong, and Ding Yuguo, 1998: Generality and applied features for singular spectrum analysis. *ACTA Meteorologica Sinica*, **56**, 736–745 (in Chinese).

Keppenne, C. L, and M. Ghil, 1992: Adaptive filtering and prediction of the Southern Oscillation Index. *J. Geophys. Res.*, **97**, 20449–20454.

Plaut, G., and R. Vautard, 1994: Spells of low–frequency oscillations and weather regimes in the Northern Hemisphere, *J. Atmos. Sci.*, **51**, 210–236.

Vautard, R., and M. Ghil, 1989: Singular spectrum analysis in nonlinear dynamics with applications to paleoclimatic time series. *Physica D*, **35**, 395–424.

Vautard, R., P. Yiou, and M. Ghil, 1992: Singular spectrum analysis:A toolkit for short,noise chaotic signals. *Physica D*, **58**, 95–126.

基于奇异谱分析的南方涛动指数短期气候预测试验

刘健文 董佩明

摘 要

采用奇异谱分析(SSA)技术,取60个月的窗口长度分析了1951−1997年的南方涛动指数(SOI)序列. 结果表明,序列中存在二个主要的振荡,其周期分别为准四年和准二年. 以最大熵谱分析为基础,对包括振荡分量在内的前9个分量建立自回归模型. 对自回归外推预报序列进行重构,获得超前36个月的SOI预测值. 预测序列和观测序列5个月滑动平均之间的相关系数达到了0.8. 这一奇异谱分析和自回归相结合的预报模型对1997—1998年强ENSO事件的持续时间和峰值作出了很好地预测. 试验还表明,合适地选择重构分量是提高模型预测能力的关键点之一.

关键词: 南方涛动指数,奇异谱分析,振荡分量,重构