ADVANCES IN ATMOSPHERIC SCIENCES, VOL. 24, NO. 2, 2007, 199–212

An Intercomparison of Rules for Testing the Significance of Coupled Modes of Singular Value Decomposition Analysis

LI Fang* [1,2] (李 芳) and ZENG Qingcun[1] (曾庆存)

[1] *Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing* 100080

[2] *Graduate University of the Chinese Academy of Sciences, Beijing* 100039

(Received 17 January 2006; revised 20 July 2006)

ABSTRACT

This paper clarifies the essence of the significance test of singular value decomposition analysis (SVD), and investigates four rules for testing the significance of coupled modes of SVD, including parallel analysis, nonparametric bootstrap, random-phase test, and a new rule named modified parallel analysis. A numerical experiment is conducted to quantitatively compare the performance of the four rules in judging whether a coupled mode of SVD is significant as parameters such as the sample size, the number of grid points, and the signal-to-noise ratio vary.

The results show that the four rules perform better with lower ratio of the number of grid points to sample size. Modified parallel analysis and nonparametric bootstrap perform best to abandon the spurious coupled modes, but the latter is better than the former to retain the significant coupled modes when the sample size is not much larger than the number of grid points. Parallel analysis and random-phase test are robust to abandon the spurious coupled modes only when either (1) the observations at the grid points are spatially uncorrelated, or (2) the coupled signal is very strong for parallel analysis and is not weak for random-phase test. The reasons affecting the accuracy of the test rules are discussed.

Key words: singular value decomposition analysis, significance test

DOI: 10.1007/s00376-007-0199-2

1. Introduction

Singular value decomposition analysis (SVD) is one of several multivariate methods for isolating linear combinations of variables within two geophysical fields $\boldsymbol{S}$ and $\boldsymbol{Z}$ that vary in both space and time. It involves performing the singular value decomposition of a temporal cross-covariance matrix of $\boldsymbol{S}$ and $\boldsymbol{Z}$ as $\boldsymbol{C_{sz}} = \boldsymbol{P\delta Q'}$, hence its name. In this paper, the notation ($'$) means transpose of a vector or matrix, N_s and N_z denote the number of grid points for fields $\boldsymbol{S}$ and $\boldsymbol{Z}$ respectively, and let rank($\boldsymbol{A}$) and cov($\boldsymbol{a}$, $\boldsymbol{b}$), respectively denote the rank of the matrix $\boldsymbol{A}$ and covariance of two time series $\boldsymbol{a}$, $\boldsymbol{b}$. The matrices $\boldsymbol{P}$ and $\boldsymbol{Q}$, whose columns $\boldsymbol{p}_i$ and $\boldsymbol{q}_i$ ($i = 1, 2, \cdots, r$) are r =rank($\boldsymbol{C_{sz}}$) pairs of coupled patterns or singular vectors, are orthogonal matrices. The matrix $\boldsymbol{\delta}$ is a diagonal matrix comprising positive singular values δ_i ($i = 1, 2, \cdots, r$) as its diagonal elements. It is easily shown that the left and right singular vectors are eigenvectors of the symmetrical matrices $\boldsymbol{C_{sz}C'_{sz}} = \boldsymbol{P\delta^2P'}$ and $\boldsymbol{C'_{sz}C_{sz}} = \boldsymbol{Q\delta^2Q'}$ respectively. When $\boldsymbol{S}$ and $\boldsymbol{Z}$ are projected onto the ith pair of coupled patterns, one obtains singular variables or weight vectors $\boldsymbol{u}_i$ and $\boldsymbol{v}_i$ whose covariance is C_i=cov($\boldsymbol{u}_i, \boldsymbol{v}_i$)=cov($\boldsymbol{p'}_i\boldsymbol{S}, \boldsymbol{q'}_i\boldsymbol{Z}$) = δ_i. Note that, for $i \neq j$, $\boldsymbol{u}_i(\boldsymbol{v}_i)$ may be correlated with $\boldsymbol{u}_j(\boldsymbol{v}_j)$, but $\boldsymbol{u}_i$ is uncorrelated with $\boldsymbol{v}_j$. Correlation coefficients between singular variables are positive but not necessarily ordered. That is to say, although $C_1 \geqslant C_2 \geqslant \ldots \geqslant C_r > 0$, it need not be true, and in general will not be true, that $R_1 \geqslant R_2 \geqslant \ldots \geqslant R_r$. Here R_i is the correlation coefficient between $\boldsymbol{v}_i$ and $\boldsymbol{u}_i$.

SVD is widely used in meteorology for two aims: one is to describe the links between two geophysical fields in a symmetrical manner through pairs of patterns or the correlation between twin singular variables (Prohaska, 1976; Lanzante, 1984; Dymnikov and Filin, 1985; Wallace et al., 1992; Hsu, 1994; Shen and Lau, 1995; Iwasaka and Wallace, 1995; Peng and Fyfe,

*E-mail: lifang@mail.iap.ac.cn

1996; Guo et al., 2000; Lau and Weng, 2001; Lau and Weng, 2001; Rodríguez-Fonseca and Serrano, 2002; Wang and Fu, 2002; Liu, 2003; Shabbar and Skinner, 2004; Chang, C. P.; Terray and Dominiak, 2005, to name only some of the representative studies); and the other, which is used frequently for statistical downscaling or climate reconstructing, is to estimate one field from the other (Feddersen et al., 1999; Feddersen, 2003; Feddersen and Ersen, 2005; Zhang and Zebiak, 2004; Widmann et al., 2003; Widmann et al., 2005; Li et al., 2005, to name only a few recent studies).

Although SVD is powerful as a method to extract important coupled modes of variability between time series of two geophysical fields, it has high potential for spurious coupled patterns and correlation due to sampling error, especially for small sample size. This can be easily understood from the following example. For an infinite time record, if any two grid points each from geophysical fields $\boldsymbol{S}$ and $\boldsymbol{Z}$ are uncorrelated, the matrix $\boldsymbol{C_{sz}}$ is zero matrix. Therefore, there are actually no coupled modes and correlation between the two fields, and it is pointless to conduct SVD. In practice, however, one generally has a finite time record of the two geophysical fields, and the matrices $\hat{\boldsymbol{C}}_{\boldsymbol{sz}}$ and $\hat{\boldsymbol{\delta}}$ as the estimated matrices of $\boldsymbol{C_{sz}}$ and $\boldsymbol{\delta}$ based on a sample of finite size are usually nonzero matrices, especially for small sample size. Hence, one will obtain the spurious coupled modes and wrongly consider the two fields to be correlated through conducting SVD for $\hat{\boldsymbol{C}}_{\boldsymbol{sz}}$. In this paper, matrices, vectors and variables with finite temporal dimensions, or estimated from a sample of finite size are marked with the notation ($\hat{}$). The phrase "spurious coupled modes" means the coupled modes that are not significant and should have been abandoned. There have been some earlier studies concerning how well SVD can extract coupled modes for climate data. By analyzing and comparing the errors (including sampling error and systematic error) in the first pair of coupled patterns of various methods based on an ideal experiment, Bretherton et al. (1992) pointed out: the sampling error in the first pair of coupled patterns of SVD was considerable when the sample size is not much larger than the number of grid points, though the simple method SVD could yield results similar to the more elaborate method BPCCA (Barnett and Preisendorfer, 1987). BPCCA is the method which applies canonical correlation analysis (CCA) to the principal components of the two analyzed fields. Cherry (1996) derived an experiment in which there were no spatial correlations and coupled modes between two analyzed fields. However, high correlation coefficients between singular variables from the leading five coupled modes led Cherry (1996) to question the utility of the SVD procedure. In section 4 of this paper, based on an ideal experiment similar to that of Bretherton et al. (1992) and Cherry (1996), we will quantitatively show the high sampling error of both patterns and correlation between singular variables of SVD when sample size is not too large.

Because SVD is widely used in meteorology and has high potential for spurious patterns and correlation due to sampling error, it is important and necessary to test the significance of coupled modes of SVD. Several rules have been applied to meteorological data; we will concentrate on three of them. Lanzante (1984) applied parallel analysis to test the significance of coupled modes between the 700 hPa heights and sea surface temperature in the Pacific and Atlantic; Wallace et al. (1992) applied nonparametric bootstrap to test the significance of coupled modes between winter sea surface temperature and 500 hPa height anomalies; and Terray and Dominiak (2005) applied random-phase test to test the significance of coupled modes between sea surface temperature over the Indian Ocean and ENSO. Although these rules have been used, there have been no studies to investigate and compare them.

The outline of this paper is as follows. In section 2, we clarify the essence of the significance test of SVD. In section 3, the essence, process and properties of the four rules, including parallel analysis, nonparametric bootstrap, random-phase test, and a new rule named modified parallel analysis, proposed in this paper, are investigated. In section 4, we quantitatively compare the performance of the four rules in the significance test of SVD in a numerical experiment with various parameters, such as sample size, the number of grid points, and the ratio of the coupled signal amplitude to uncoupled variability. Finally, section 5 presents our conclusions.

2. Essence of the significance test of SVD

In statistics theory, the approach to avoid incorrect conclusions regarding a characteristic of a population due to sampling error is to build a null hypothesis about the characteristic, and then decide whether the null hypothesis is accepted through analyzing a sample of finite size from the population, i.e., statistical hypothesis testing. There are two types of errors in hypothesis testing: (1) reject the null hypothesis H_0 when H_0 is true (i.e., type I error); and (2) accept H_0 when H_0 is false (i.e., type II error). Ideally, one would like both types of errors to have probability 0. Unfortunately, however, both these probabilities cannot be simultaneously less than any positive value for a finite sample size. One therefore has to content oneself by trying to control the probability of a type I error no more than a significance level α ($0 < \alpha < 1$)

without considering type II error, and this procedure is defined as a significance test (Rohatgi, 1984; Sheng et al., 1989).

The statistical hypotheses built in the significance test of SVD are based on the singular values δ_j ($j = 1, 2, \cdots, r$) of the cross-covariance matrix between the two analyzed fields. The null hypothesis for the first singular value is H_0: $\boldsymbol{C_{sz}}$ is zero matrix. When H_0 is accepted, there is no correlation between any pair of points from the two fields respectively. Consequently, it is pointless to interpret the coupled modes of SVD. If the null hypothesis H_0 is rejected, it is natural to examine the significance of higher order coupled modes. Since the singular values are positive and ordered from the largest to the smallest, one can begin by assuming that there is only one singular value for $\boldsymbol{C_{sz}}$. If this null hypothesis is rejected, one can assume that $\boldsymbol{C_{sz}}$ just has two singular values, and so forth. That is, the implied sequence of hypotheses is as follows:

H_0^j: $\boldsymbol{C_{sz}}$ just has $j-1$ singular values, $j = 1, 2, \cdots, r$

H_1^j: $\boldsymbol{C_{sz}}$ has at least j singular values, where H_1^j is the alternative hypothesis of H_0^j. According to the sequence of hypotheses above, the significance test of SVD can control the probability of retaining the spurious coupled modes no more than α.

However, the singular values δ_j ($j = 1, 2, \cdots, r$) of a population are generally unknown for actual meteorological fields. Therefore $\hat{\delta}_j$, the estimator of δ_j based on a sample of finite size from this population, is used to decide whether to reject the null hypothesis H_0^j. When $\boldsymbol{C_{sz}}$ has at least j singular value, the significance statistic

$$\widehat{\mathrm{SCF}}_j = \frac{\hat{\delta}_j^2}{\sum_{i=1}^{\hat{r}} \hat{\delta}_i^2} \times 100\% \,,$$

which is defined as the squared covariance fraction explained by the jth coupled mode of the two fields $\hat{\boldsymbol{S}}$ and $\hat{\boldsymbol{Z}}$, should be larger than the corresponding value when $\boldsymbol{C_{sz}}$ just has $j-1$ singular values. Hence, one can select a critical squared covariance fraction $\widehat{\mathrm{SCF}}_j^{\alpha}$ and a significance level α to let $P\{\widehat{\mathrm{SCF}}_j \geqslant \widehat{\mathrm{SCF}}_j^{\alpha}\} \leqslant \alpha$ as H_0^j is true. In practice, the critical fraction $\widehat{\mathrm{SCF}}_j^{\alpha}$ is only required to satisfy that $P\{\widehat{\mathrm{SCF}}_j \geqslant \widehat{\mathrm{SCF}}_j^{\alpha}\} = \alpha$ as H_0^j is true for the uniqueness of $\widehat{\mathrm{SCF}}_j^{\alpha}$, and $\{\widehat{\mathrm{SCF}}_j \geqslant \widehat{\mathrm{SCF}}_j^{\alpha}\}$ is defined as the critical region or rejection region for H_0^j. In other words, we just keep the probability of retaining the spurious coupled modes at significance level α. When $\widehat{\mathrm{SCF}}_j \geqslant \widehat{\mathrm{SCF}}_j^{\alpha}$ we reject the null hypothesis to judge the jth mode of SVD to be significant and keep on testing the significance of the $(j+1)$th mode, otherwise we accept the null hypothesis to judge only the first $j-1$ coupled modes to be significant and abandon the remaining $\hat{r} - j + 1$ modes. It is the purpose of the rules for testing the significance of coupled modes of SVD to estimate the critical fractions $\widehat{\mathrm{SCF}}_j^{\alpha}$ ($j = 1, 2, \cdots, \hat{r}$).

3. Rules for testing the significance of coupled modes of SVD

Principal component analysis (PCA), SVD, and CCA are matrix methods for analysis of structure in data sets. Their common ground may suggest that some rules used in the significance tests of PCA and CCA, such as parallel analysis and nonparametric bootstrap, can be modified to be used in the significance test of SVD.

However, the rules of maximum likelihood hypothesis testing, which have been developed for PCA [e.g., Sphericity test (Bartlett, 1950); Bartlett's test for the first principal component (Bartlett, 1954); Lawley's test for the second principal component (Lawley, 1956)] and CCA [e.g. Bartlett's test (Bartlett, 1939)], have not been modified to test the significance of coupled modes of SVD for two main reasons. The first reason is that they are classical parametric testing rules. The reference distributions (i.e., the approximate distribution of the test statistic when the null hypothesis is true) of these rules are derived analytically from the normal distribution function and based on sampling assumptions (i.e., observations are taken independently and identically from the same distribution, and the sample size is large enough). Earlier studies have pointed out that the performance of these rules is very sensitive to departure from the assumption of a normal distribution (Seber, 1984) and to a large sample size (Crawford, 1975). These rules do not work well when the observations are not independent and cannot be approximately normally distributed, or the sample size is small. The second reason is that the derivations of these methods are very complicated. In general, the observations of the analyzed fields are not independent or the sample size is small, and sometimes the observations are not approximately normally distributed. Furthermore, in these rules, only the Sphericity test (Bartlett, 1950) has been developed for a covariance matrix; the others are for correlation matrices. Consequently, how to modify the original test statistics in these rules to approximately fit the chi-square distribution or other classical distributions is an intractable work and sometimes impossible.

Moreover, based on matrix theory, the three propositions, which are (1) the data matrix of the field $\boldsymbol{S}$ is $\boldsymbol{0}$, (2) the covariance matrix $\boldsymbol{C_{ss}}$ of the data field $\boldsymbol{S}$

is $\mathbf{0}$, and (3) all eigenvalues of $\boldsymbol{C_{ss}}$ are equal to zero, are equivalent. Therefore, the rule in the significance test of PCA with the assumption that at least one eigenvalue of $\boldsymbol{C_{ss}}$ is nonzero (North et al., 1982), or at least two eigenvalues of $\boldsymbol{C_{ss}}$ are nonzero (Quadrelli et al., 2005), is meaningful, and its assumption is easily satisfied by most meteorological fields. However, any nonzero data matrices $\boldsymbol{S}$ and $\boldsymbol{Z}$ are in general uncorrelated, i.e., $\boldsymbol{C_{sz}} = \mathbf{0}$, though $\hat{\boldsymbol{C}}_{\boldsymbol{sz}}$ and $\hat{\boldsymbol{\delta}}$ are usually nonzero matrices due to sampling error. Therefore, the two parametric testing rules are unsuitable to be modified as the rules of significance test of SVD.

Here we consider four Monte Carlo methods, namely parallel analysis, nonparametric bootstrap, random-phase test and modified parallel analysis, and investigate their essences, procedures, and properties. Hereafter, the superscript "k" indicates specifically the kth pair of simulated fields built in these rules.

3.1 *Parallel analysis*

The rule is actually a Monte Carlo method. It generates a large number of squared covariance fractions explained by each coupled mode of the simulated fields and then uses these fractions to build confidence intervals for corresponding coupled modes. The simulated fields built in the rule are equivalent in size to the analyzed fields, but comprise the independent elements identically taken from a standard normal distribution population. The procedure of parallel analysis is: (1) Generate two simulated fields $\hat{\boldsymbol{S}}^k$ and $\hat{\boldsymbol{Z}}^k$ that are equivalent in size to the two analyzed fields but comprise the independent elements taken from $N(0,1)$ (i.e. Standard Normal distribution). (2) Perform SVD on the cross-covariance matrix of the two simulated fields and retain the singular values. Denote $\hat{d}_j^k$ as the jth ($j = 1, 2, \cdots, \hat{r}^k$) singular value in the kth simulation. (3) Repeat steps 1 and 2 a total of m times and define

$$\hat{b}_j^k = \frac{(\hat{d}_j^k)^2}{\sum_{i=1}^{\hat{r}^k} (\hat{d}_i^k)^2} \times 100\%$$

as the squared covariance fraction explained by the jth mode in the kth simulation. When $\hat{r} - \hat{r}^k = \Delta^k > 0$ ($k = 1, 2, \cdots, m$), let $\hat{b}^k_{(\hat{r}^k+1)} = \cdots = \hat{b}^k_{(\hat{r}^k+\Delta^k)} = 0$. Subsequently, order the squared covariance fractions explained by the jth ($j = 1, 2, \cdots, \hat{r}$) mode in the m simulations to satisfy $\hat{b}_j^1 \geqslant \hat{b}_j^2 \geqslant \cdots \geqslant \hat{b}_j^m$. Select a significance level α, then $\hat{b}_j^{\alpha\times m}$ is the critical value for the jth mode ($j = 1, 2, \cdots, \hat{r}$). (4) If the squared covariance fraction $\widehat{\mathrm{SCF}}_1$ explained by the first coupled mode of the two analyzed fields $\widehat{\boldsymbol{S}}$ and $\widehat{\boldsymbol{Z}}$ is less than $\hat{b}_1^{\alpha\times m}$, one can accept H_0^1 to conclude that the first coupled mode is spurious and the two fields $\boldsymbol{S}$ and $\boldsymbol{Z}$ are uncorrelated. Otherwise, one needs to consider whether $\widehat{\mathrm{SCF}}_2$ is less than $\hat{b}_2^{\alpha\times m}$. If the answer is "yes", then the two analyzed fields have only one significant coupled mode; or else they have at least two significant coupled modes, and the significance of remaining coupled modes needs to be further tested in the same way.

The procedure of parallel analysis is essentially parametric and assumes that the observations of analyzed fields are spatially and temporally independent, and identically taken from a normally distributed population. Therefore, it is robust to test the significance of coupled modes between two analyzed fields that satisfy the above assumptions. When the assumptions are unfit for the two analyzed fields, such as the observations at the grid points are dependent or not normally distributed, its ability to test the null hypotheses may be low (shown in subsections 3.4 and section 4).

3.2 *Nonparametric bootstrap*

Nonparametric bootstrap, one of the resampling methods first proposed by Efron (1979, 1987) to estimate the bias, standard deviation, prediction error, and confidence interval, was used to test the statistical hypotheses of SVD by Wallace et al. (1992). The essence of the rule is to randomly order one field in the time domain so that most of them were wrongly paired with the other field. The procedure to order one field randomly can be stated as follows: let the T vectors $\boldsymbol{X}_1, \boldsymbol{X}_2, \cdots, \boldsymbol{X}_T$ with size $N_s \times 1$ be the T independent observations of field $\boldsymbol{S}$; putting probability mass $\frac{1}{T}$ on each $\boldsymbol{X}_i$, if $\boldsymbol{X}_i^*$ ($i = 1, 2, \cdots, T$) is a random sample drawn with replacement from $\boldsymbol{X}_1, \boldsymbol{X}_2, \cdots, \boldsymbol{X}_T$, then $\boldsymbol{X}_1^*, \boldsymbol{X}_2^*, \cdots, \boldsymbol{X}_T^*$ construct the kth simulated field $\hat{\boldsymbol{S}}^k$. The Monte Carlo protocol used here is: (1) generate the simulated field $\hat{\boldsymbol{S}}^k$ by randomly ordering one field in the time domain; (2) perform SVD on the cross-covariance matrix of the simulated field $\hat{\boldsymbol{S}}^k$ and analyzed field $\hat{\boldsymbol{Z}}$, retaining the singular values $d_j^k (j = 1, 2, \cdots, \hat{r}^k)$; the following steps are similar to steps 3 and 4 of parallel analysis.

Nonparametric bootstrap need not assume the distribution of the observations in the analyzed fields. Thus, it can be conducted in situations where it is impossible or very intractable to derive reference distributions analytically as classical parametric methods. However, the rule assumes that observations of $\hat{\boldsymbol{S}}$ are temporally independent. Whether the temporal dependency of observations in the analyzed field $\hat{\boldsymbol{S}}$ affects the accuracy of the rule will be discussed in subsection 3.4.

3.3 *Random-phase test*

This rule was proposed firstly by Ebisuzaki (1997)

to test the significance of correlation between two time series with temporally correlation and could be considered as a rule of resampling time series in the frequency domain. Ebisuzaki (1997) put forward the rule to replace nonparametric bootstrap because the formal needn't assume that the analyzed time series is temporal independent. Recently, Terray and Dominiak (2005) introduced the method into the significance test of SVD based on the implied assumption that the temporal correlation of analyzed fields would markedly affect the critical fraction of the analyzed cross-covariance explained by each coupled mode of SVD.

Step 1 of random-phase test is to build the kth simulated fields $\hat{\boldsymbol{S}}^k$ by repeating the following process for the time series at each grid point of $\hat{\boldsymbol{S}}$: (a) compute the discrete Fourier transform of $\hat{\boldsymbol{s}}_h (h = 1, 2, \cdots, N_s)$ by

$$\hat{a}_{h(l-1)} = \frac{2 - \hat{\sigma}_{l-1}}{T} \sum_{t=1}^{T} \hat{s}_{ht} e^{2\pi i (t-1)(l-1)/T} \tag{1}$$

where $l = 1, 2, \cdots, T/2 + 1$ for even $T, l = 1, 2, \cdots, (T+1)/2$ for odd T, and $\hat{\sigma}_l = 0$ except for $l = 1$ and $l = T/2 + 1$ (for even T), in which case $\hat{\sigma}_l = 1$; (b) generate a Fourier series $\hat{\boldsymbol{y}}_h$ with random phase and the same power spectrum as the original series $\hat{\boldsymbol{s}}_h$ by

$$\begin{cases} \hat{y}_{h1} = 0\,, \\ \hat{y}_{hl} = |\hat{a}_{hl}| e^{i\hat{\theta}_l} \ \text{ for } \ 1 < l < \dfrac{T}{2} + 1 \ \text{ (for even } T)\,, \\ \qquad\qquad \text{and } \ 1 < l < \dfrac{T+1}{2} \ \text{ (for odd } T)\,, \\ \hat{y}_{h(\frac{T}{2}+1)} = 2^{\frac{1}{2}} \left| \hat{a}_{h(\frac{T}{2}+1)} \right| \cos(\hat{\theta}_{(\frac{T}{2}+1)}) \ \text{ for even } T \end{cases} \tag{2}$$

where $\hat{\theta}_l$ is a uniform random variable from $(0, 2\pi)$; (c) gain the simulated series $\hat{\boldsymbol{s}}_h^k$ through transformation of the series by

$$\hat{\boldsymbol{s}}_{ht}^k = \mathrm{Re} \sum_{l=1}^{n} \hat{y}_{hl} e^{-2\pi i t (l-1)/T} \tag{3}$$

where $n = T/2 + 1$ and $(T+1)/2$ for even and odd T respectively. The following steps are similar to steps 2, 3, and 4 of nonparametric bootstrap.

Even though the random-phase test is better than nonparametric bootstrap to test the significance of correlation between two time series with serial correlation, it has three weak points in the significance test of SVD. First, the random-phase test does not do well at resolving the low frequencies (periods greater than or on the order of the length of the time series) because the new series $\hat{\boldsymbol{s}}_h^k (k = 1, 2, \cdots, m; h = 1, 2, \cdots, N_s)$ built by the rule are periodically constructed and usually have no trends. The weak point was first pointed out and discussed by Ebisuzaki (1997) who used the rule to estimate the critical correlation coefficients of two pairs of time series generated by the models AR(1) and AR(2) respectively (first- and second-order autoregressive models). Second, Ebisuzaki (1997) recommended the random-phase test because it could make the simulated time series preserve the power spectrum of the original series. However, we have noted that each new series $\hat{\boldsymbol{s}}_h^k (k = 1, 2, \cdots, m; h = 1, 2, \cdots, N_s)$ built by the rule retains only the real part of series generated by the inverse Fourier transformation of $\hat{\boldsymbol{y}}_h$ in formula (3) above. In other words, it is not $\hat{\boldsymbol{s}}_h^k$ but the complex series $\{\sum_{l=1}^{n} \hat{y}_{ht} e^{-2\pi i t (l-1)/T}, t = 1, 2, \cdots, T\}$ that has the same power spectrum as the original series $\hat{\boldsymbol{s}}_h$. Third, the critical fractions $\widehat{\mathrm{SCF}}_j^{\alpha}$ $(j = 1, 2, \cdots, \hat{r})$ are insensitive to temporal correlation of analyzed fields $\hat{\boldsymbol{S}}$ (shown in subsection 3.4), and hence the precondition that random-phase test is better than nonparametric bootstrap cannot be satisfied in the significance test of SVD.

3.4 *Modified parallel analysis*

The optimal Monte Carlo methods require the assumptions, which are based to build the simulated fields, to be fitted exactly by the analyzed fields, or affect the critical values $\widehat{\mathrm{SCF}}_j^{\alpha} (j = 1, 2, \cdots, \hat{r})$ as little as possible. That is, the difference between $\widehat{\mathrm{SCF}}_j^{\alpha}$ and $\hat{b}_j^{\alpha \times m}$ due to unapt assumptions should be negligible for the optimal test rules. The procedure of parallel analysis assumes that the analyzed fields comprise elements that are independent and normally distributed. The assumptions are usually unfit for geophysical fields since the observations at the grid points are usually spatially and temporally correlated or sometimes not approximately normally distributed. The spatial correlation, serial correlation and distribution form affect the accuracy of the test rule in different degrees.

The spatial correlation and serial correlation decrease the effective spatial dimensions and effective sample size respectively. There are various distribution forms, such as normal distribution with kurtosis (known as the fourth moment of normalized principal component) equal to 3 (i.e. k=3), uniform distributions with k=1.8, exponential distribution with k=9, and so on. The influence of spatial correlation, temporal correlation, and distribution form on critical values $\widehat{\mathrm{SCF}}_j^{\alpha} (j = 1, 2, \cdots, \hat{r})$ can be reflected by the influence of the number of grid points, sample size, and distribution form on the critical values in the model with independent observations at grid points. In the follow-

Table 1. The means and standard deviations (std) of the critical squared covariance fractions explained by the first three modes with different spatial dimension N of the two analyzed fields and sample size T. The means and standard deviations are based on 100 independent realizations of the model with independent observations taken from a standard normal distribution population.

	N=(11,11)		N=(11,81)		N=(81,81)	
mode	mean	std	mean	std	mean	std
$T = 20$						
1	0.4528	0.0150	0.3002	0.0084	0.1542	0.0032
2	0.2849	0.0070	0.2177	0.0046	0.1259	0.0016
3	0.1963	0.0051	0.1699	0.0034	0.1086	0.0013
$T = 50$						
1	0.4027	0.0123	0.2353	0.0058	0.1005	0.0019
2	0.2649	0.0054	0.1810	0.0031	0.0848	0.0010
3	0.1918	0.0045	0.1495	0.0023	0.0752	0.0008
$T = 100$						
1	0.3842	0.0112	0.2085	0.0048	0.0789	0.0012
2	0.2597	0.0053	0.1658	0.0023	0.0684	0.0007
3	0.1913	0.0043	0.1408	0.0020	0.0616	0.0006

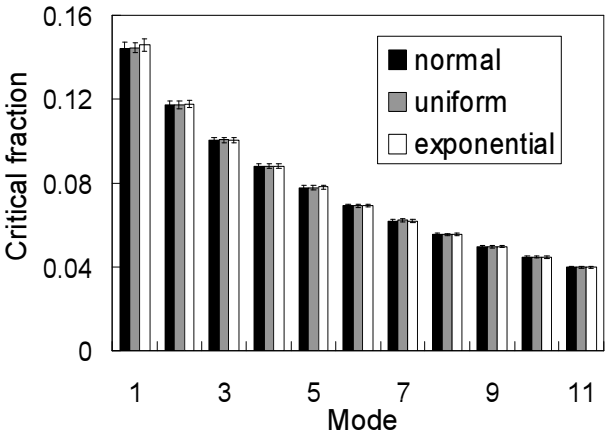

Fig. 1. The means (histograms) and standard deviations (error bars) of the critical squared covariance fractions explained by the first eleven modes. The means and standard deviations are based on 100 realizations of the model comprising independent observations with sample size T=100, spatial dimension N=36 and various distribution forms.

ing text, we concentrate on the first three coupled modes, which are usually used.

For 100 independent realizations of the model comprising independent observations taken from a standard normal distribution population, Table 1 shows that the influence of sample size on the means of critical fractions for the first three coupled modes is much smaller than that of the number of grid points. The standard deviations are much smaller than the corresponding means. Thus we can believe that the number of grid points affects the critical fractions more remarkably than sample size for the important first three coupled modes. Here, $N = (N_s, N_z)$. Note that, the critical fractions for the first three modes will increase steeply as long as N_s or N_z decreases.

For 100 independent realizations of the model comprising independent observations with $N = (36, 36)$ and T=100, we compare the influence of various distribution forms (normal distribution, uniform distribution and exponential distribution) on the critical fractions. The results (Fig. 1) show that the means of critical values for various distribution forms are almost equivalent and the standard deviations for the three distribution forms are much smaller than their corresponding means in the first three modes. Thus the critical fractions for the first three modes are quite insensitive to distribution form.

In conclusion, the assumption that the observations in the analyzed fields are spatially independent affects the critical value $\widehat{\mathrm{SCF}}_j^{\alpha}$ far more significantly than that of temporal independence and normal distribution. In fact, as long as the spatial correlation of either analyzed field increases, the critical values $\widehat{\mathrm{SCF}}_j^{\alpha}$ $(j = 1, 2, \cdots, \hat{r})$ will increase abruptly, and the differences between $\widehat{\mathrm{SCF}}_j^{\alpha}$ and $\hat{b}_j^{\alpha\times m}$ will be more obvious if the rules assume that the observations in the analyzed fields are spatially independent.

For the reason that the influence of spatial correlation is remarkable on the critical values $\widehat{\mathrm{SCF}}_j^{\alpha}$ $(j = 1, 2, \cdots, \hat{r})$, we modify the parallel analysis and require the number of grid points of simulated fields to be equal to the estimated effective number of spatial degrees of freedom (ESDOF) of analyzed fields so as to take out the assumption of spatial independence for analyzed fields in parallel analysis. The new rule is named as modified parallel analysis. ESDOF is used to estimate the number of grid points required to rep-

resent the field and can be easily calculated according to the formula proposed by Bagrov (1969) and Ter-Megreditchian (1969) as follows:

$$\hat{N}^*_{\rm ef} = \frac{\left(\sum\limits_{i=1}^{N}\hat{C}_{ii}\right)^2}{\sum\limits_{i,j=1}^{N}\hat{C}^2_{ij}} = \frac{\left(\sum\limits_{k=1}^{N}\hat{\lambda}_k\right)^2}{\sum\limits_{k=1}^{N}\hat{\lambda}^2_k}\ . \tag{4}$$

Here, $\hat{\boldsymbol{C}}$ is the estimated covariance matrix of the analyzed field, and $\hat{\lambda}_k$ is the kth eigenvalue of $\hat{\boldsymbol{C}}$. Formula (4) is based on two assumptions: (1) the observations in analyzed fields are normally distributed and (2) $\hat{\boldsymbol{C}}$ is as the estimated matrix of true $\hat{\boldsymbol{C}}$ with sufficient accuracy (i.e., the effective sample size is relatively large). However, Bretherton (1999) pointed out that $\hat{N}^*_{\rm ef}$ is insensitive to assumption (1) since it depends only on the partition of the variance between the EOFs, and the estimate error of $\hat{N}^*_{\rm ef}$ is small when the effective sample size is many times of $N^*_{\rm ef}$. The procedures of modified parallel analysis and parallel analysis are similar except for step 1, which requires spatial dimension of the simulated fields built in the two rules to be equal to $\hat{N}^*_{\rm ef}$ and spatial dimension of the analysis fields respectively.

One may ask why not modify the parallel analysis further by replacing sample size with effective sample size (ESS) to abolish the assumption that the observations are temporally independent for analyzed fields in parallel analysis? In our opinion, there are two reasons. First, the influence of temporal correlation on critical fractions is small and negligible relative to that of the spatial correlation for the first three modes. Moreover, modified parallel analysis, which requires only the number of grid points of the simulated fields to be equal to $\widehat{N^*_{\rm ef}}$ of the analyzed fields, can perform satisfactorily (shown in section 4). Second, Thiébaux and Zwiers (1984) pointed out the formula of ESS proposed by Davis (1976) was severely biased and should be considered only as a diagnostic quantity.

4. An experiment to compare the four rules

4.1 *Description of the three models*

The models in this paper are similar to those conducted by Bretherton et al. (1992) and Cherry (1996) to compare how well various methods can isolate coupled modes in climate data. The two fields $\boldsymbol{S}$ and $\boldsymbol{Z}$ of interest are also called left and right fields that have N_s and N_z grid points respectively. Let $\boldsymbol{S} = (\boldsymbol{s}_1, \boldsymbol{s}_2, \cdots, \boldsymbol{s}_{N_s})'$ and $\boldsymbol{Z} = (\boldsymbol{z}_1, \boldsymbol{z}_2, \cdots, \boldsymbol{z}_{N_z})'$ be $N_s \times 1$ and $N_z \times 1$ random vectors. For convenience, we assume that $\boldsymbol{S}$ and $\boldsymbol{Z}$ are anomaly fields, and $N_s = N_z = N$.

Model 1: The observations at the grid points are spatially and temporally uncorrelated and identically distributed. There is no coupling between the two analyzed fields. Since the critical fractions for the first three modes are insensitive to the distribution form according to the analysis in subsection 3.4, we assume that the observations are identically taken from N(0,1). Therefore the time series have the form

$$\boldsymbol{S}(t) = \boldsymbol{W}_s(t)\ ,\ \ \boldsymbol{Z}(t) = \boldsymbol{W}_z(t)\ ,\quad t = 1, 2, \cdots, T \tag{5}$$

Here, $w_{ij} \sim N(0,1)$. When the number T of observations of the two analyzed fields is finite, we denote the two fields as $\hat{\boldsymbol{S}}$ and $\hat{\boldsymbol{Z}}$.

Model 2: The observations at the grid points are spatially correlated but temporally uncorrelated. There is no coupling between the two fields. The spatial correlation structure is described by what geostatisticians refer to as the exponential covariance function $C(h) = v\mathrm{e}^{-5h/(N-1)L}$ (Isaaks and Srivastava, 1989). Here, h is the spatial lag, v is the variance of red noise generated, and L is the noise redness length (the separation at which the red noise correlation drops to e^{-1}). Let v=1 and L=1. Following the same method as Cherry (1996), we can generate the red noise $\boldsymbol{R}_s(t)$ and $\boldsymbol{R}_z(t)$ with mean 0 and variance 1, and $\boldsymbol{C}_{\boldsymbol{RR}} = \boldsymbol{V}$, where $V_{ij} = e^{-5|i-j|/(N-1)}$. Therefore, the time series have the form

$$\boldsymbol{S}(t) = \boldsymbol{R}_s(t)\ ,\ \ \boldsymbol{Z}(t) = \boldsymbol{R}_z(t)\ ,\ \ t = 1, 2, \cdots, T\ . \tag{6}$$

When the number T of observations of the two analyzed fields is finite, we denote the two fields as $\hat{\boldsymbol{S}}$ and $\hat{\boldsymbol{Z}}$.

Model 3: The observations at the grid points are spatially correlated and there is also a deterministic signal shared by the two analyzed fields. The signal is incorporated in each field by adding a signal matrix to the data matrix of Model 2. Hence, the time series have the form

$$\begin{cases} \boldsymbol{S}(t) = \boldsymbol{R}_s(t) + \eta\boldsymbol{\Phi}f(t)\ , \\ \boldsymbol{Z}(t) = \boldsymbol{R}_z(t) - \eta\boldsymbol{\Phi}f(t)\ ,\ \ t = 1, 2, \cdots, T \end{cases} \tag{7}$$

where η is a positive scalar, and

$$f(t) = 2^{\frac{1}{2}}\sin\left(\frac{4\pi}{T}t\right)\ . \tag{8}$$

While $\boldsymbol{\Phi}$ is a $N \times 1$ vector with elements given by

$$\Phi_i = c_b e^{\frac{-y_i^2}{2}}\ , \tag{9}$$

where

$$y_i = \frac{5(i-1)}{N-1} - 1,\quad i = 1, 2, \cdots, N \tag{10}$$

Here, c_b is a constant for normalizing $\boldsymbol{\Phi}$. This coupled signal is the bell signal proposed by Bretherton et al. (1992). When the number T of observations of two analyzed fields are finite, we denote the two fields as $\hat{\boldsymbol{S}}$ and $\hat{\boldsymbol{Z}}$.

4.2 *Analysis of the models*

For the models described above, we can calculate the exact covariance matrices for an infinite time record ($T \to \infty$) as follows:
for model 1, 2,

$$\boldsymbol{C_{sz}} = \mathrm{cov}(\boldsymbol{S}(t), \boldsymbol{Z}(t)) = 0 \,, \tag{11}$$

for model 3,

$$\begin{aligned}
\boldsymbol{C_{ss}} = \mathrm{cov}(\boldsymbol{S}(t), \boldsymbol{S}(t)) &= \mathrm{cov}(\boldsymbol{R}_s(t), \boldsymbol{R}_s(t)) + \\
&\quad 2\mathrm{cov}(\boldsymbol{R}_s(t), \eta\boldsymbol{\Phi} f(t)) + \\
&\quad \mathrm{cov}(\eta\boldsymbol{\Phi} f(t), \eta\boldsymbol{\Phi} f(t)) \\
&= \boldsymbol{V} + \eta^2 \boldsymbol{\Phi}\boldsymbol{\Phi}' \,, \\
\boldsymbol{C_{zz}} = \mathrm{cov}(\boldsymbol{Z}(t), \boldsymbol{Z}(t)) &= \mathrm{cov}(\boldsymbol{R}_z(t), \boldsymbol{R}_z(t)) - \\
&\quad 2\mathrm{cov}(\boldsymbol{R}_z(t), \eta\boldsymbol{\Phi} f(t)) + \\
&\quad \mathrm{cov}(\eta\boldsymbol{\Phi} f(t), \eta\boldsymbol{\Phi} f(t)) \\
&= \boldsymbol{V} + \eta^2 \boldsymbol{\Phi}\boldsymbol{\Phi}' \,, \\
\boldsymbol{C_{sz}} = \mathrm{cov}(\boldsymbol{S}(t), \boldsymbol{Z}(t)) &= \mathrm{cov}(\boldsymbol{R}_s(t), \boldsymbol{R}_z(t)) - \\
&\quad \mathrm{cov}(\boldsymbol{R}_s(t), \eta\boldsymbol{\Phi} f(t)) + \\
&\quad \mathrm{cov}(\eta\boldsymbol{\Phi} f(t), \boldsymbol{R}_z(t)) - \\
&\quad \mathrm{cov}(\eta\boldsymbol{\Phi} f(t), \eta\boldsymbol{\Phi} f(t)) \\
&= -\eta^2 \boldsymbol{\Phi}\boldsymbol{\Phi}' \,.
\end{aligned} \tag{12}$$

According to matrix theory, the rank of a zero matrix is zero, and the rank of a nonzero matrix is a positive integer. Thus, $\mathrm{rank}(\boldsymbol{C_{sz}}) = 0$ for models 1 and 2. For model 3, because $\mathrm{rank}(\boldsymbol{C_{sz}}) \leqslant \min(\mathrm{rank}(\boldsymbol{\Phi}), \mathrm{rank}(\boldsymbol{\Phi}'))$ and $\mathrm{rank}(\boldsymbol{\Phi}) = \mathrm{rank}(\boldsymbol{\Phi}') = 1$, $\mathrm{rank}(\boldsymbol{C_{sz}}) = 1$. That is, for models 1 and 2, there are actually no pairs of coupled patterns and no singular values of $\boldsymbol{C_{sz}}$; for model 3, there is only one pair of patterns and one singular value of $\boldsymbol{C_{sz}}$.

For climate data fields, the sample size T is finite and may not greatly exceed the number N of grid points. The sampling fluctuation in the cross-covariance matrix can greatly reduce the skill of SVD for isolating the coupled signal. To show the high potential of the spurious patterns and correlation in SVD due to sampling fluctuation more clearly than Bretherton et al. (1992) and Cherry (1996), we apply an ensemble of d (=100) independent realizations of each model to determine the sampling errors $(\varepsilon_{R_1}, \varepsilon_{R_2}, \varepsilon_{R_3})$ and (ε_p) for correlation coefficients between the singular variables from the first three modes and the normalized first pair of coupled patterns respectively. The sampling errors $\varepsilon_{R_1}, \varepsilon_{R_2}, \varepsilon_{R_3}$ and ε_p are defined as:

$$\varepsilon_{R_j} = \sqrt{\frac{1}{d}\sum_{k=1}^{d}(\hat{R}_{jk} - R_j)^2}, \quad j = 1, 2, 3 \tag{13}$$

$$\varepsilon_p = \sqrt{\frac{1}{2}\cdot\frac{1}{d}\cdot\frac{1}{N}\sum_{k=1}^{d}\sum_{i=1}^{N}\left[(\hat{p}_{ik} - p_i)^2 + (\hat{q}_{ik} - q_i)^2\right]} \,. \tag{14}$$

Here, $\hat{R}_{jk}$ is the correlation coefficient between the singular variables from the jth mode in the kth realization for a finite time record, while R_j is that from the jth mode for an infinite time record in models 1, 2, and 3; $\hat{p}_{ik}$ and $\hat{q}_{ik}$ are the first left and right coupled patterns respectively at the ith grid point in the kth realization for a finite time record, while p_i and q_i are that at the ith grid point for an infinite time record in model 3. Because $0 \leqslant R \leqslant 1$ and the patterns are normalized, the sampling errors are non-negative and no more than 1, i.e., $0 \leqslant \varepsilon_{R_j} \leqslant 1$ ($j = 1, 2, 3$), and $0 \leqslant \varepsilon_p \leqslant 1$.

According to the change of the sampling errors for 100 realizations with $\eta = 0.4$ and $N = 36$ as the sample size T varies (Fig. 2), ε_{R_2} and ε_{R_3} are almost the same and much larger than ε_{R_1}. For small $T, \varepsilon_{R_1}, \varepsilon_{R_2}, \varepsilon_{R_3}$ and ε_p are large. With $T = 50, \varepsilon_{R_2}, \varepsilon_{R_3}$ and ε_p are about 0.4. As T increases, all of the sampling errors decrease as expected. If instead the number N of grid points in each field varies with fixed T=100 and $\eta = 0.4$, the results (Fig. 3) indicate that the sampling errors is insensitive to N. With fixed N=36 and T=100, we compare the sampling errors over 100 realizations for the models 1 and 2, and for model 3 with various η. The results (Fig. 4) indicate that the sampling errors of correlation between the singular variables from the first three modes in model 1 are a lot larger than those in model 2. ε_{R_1} and ε_p decrease abruptly as η increases, but ε_{R_2} and ε_{R_3} are large and insensitive to η. Note that model 1 has the sampling errors $\varepsilon_{R_j} \approx 0.8$ ($j = 1, 2, 3$) and is the model with no spatial and temporal correlation, and no signal. There is, however, a great deal of spurious linear structure for SVD to exploit. On account of the high potential of the spurious coupled mode and correlation in SVD due to sampling error, it is essential to conduct the significance test of SVD.

4.3 *Experiments results*

According to the analysis in the former subsection, there is actually no coupled mode for models 1 and 2 and only one for model 3. If we only test the first three modes for d (=100) independent realizations of each model with various N, T (and η in model 3), the

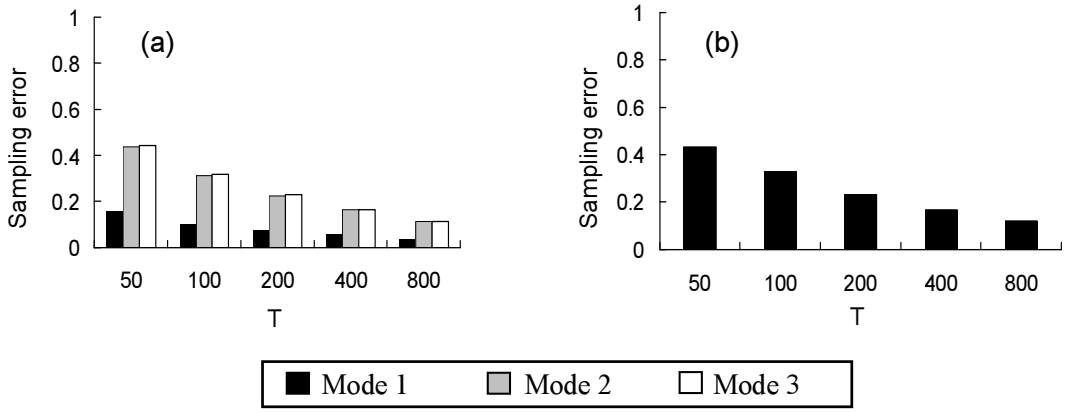

Fig. 2. Sampling errors for (a) the correlation coefficients between the singular variables from the first three modes and (b) the normalized first pair of coupled patterns respectively based on 100 independent realizations of model 3 with spatial dimension N=36, the signal-to-noise ratio η=0.4 and various sample size T.

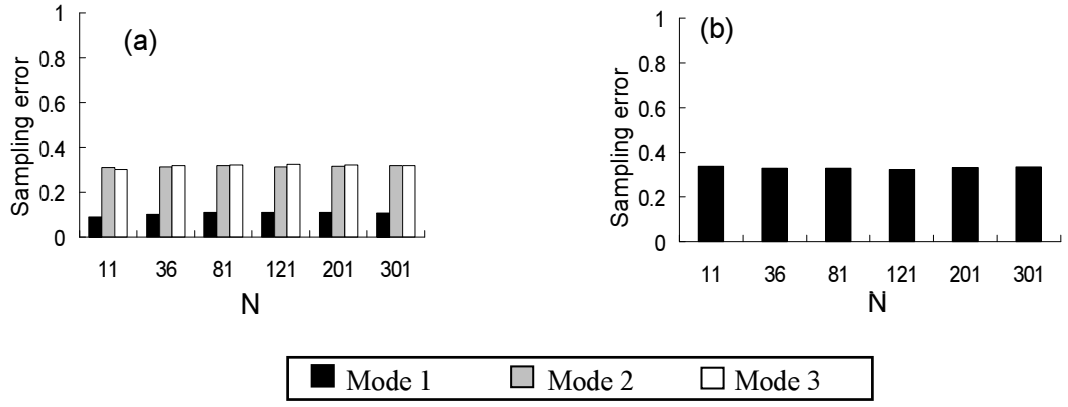

Fig. 3. As in Fig. 2 except with T=100, η=0.4 and various N.

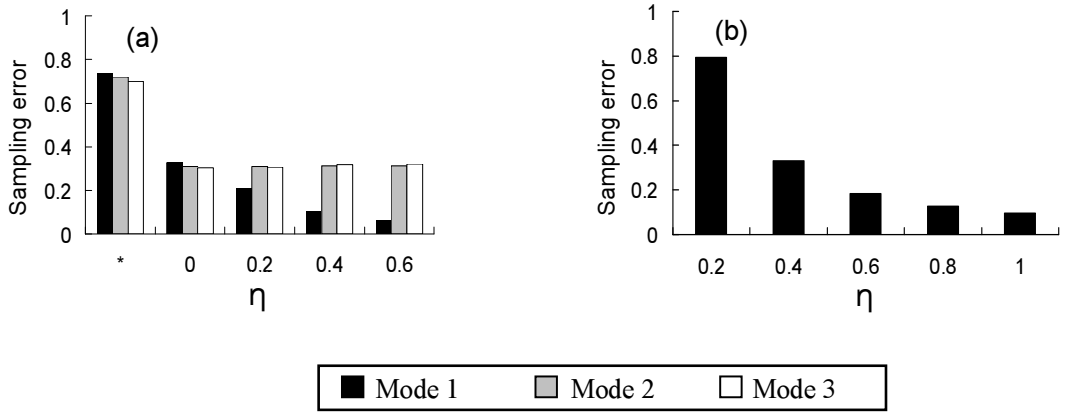

Fig. 4. As in Fig. 2 except with T=100, N=36 and various η. η="$*$" represents the situation in model 1, and η=0 represents that in model 2.

Table 2. The number of realizations rejecting the null hypothesis at each of the first three coupled modes for 100 independent realizations of the three models. The expected values for numbers in bold are 0, while 100 for those in normal font. The signal-to-noise ratio η="$*$" represents the situation in model 1, and $\eta = 0$ represents that in model 2.

		Parallel analysis			Nonparametric bootstrap			Random-phase test			Modified parallel analysis		
Mode		1	2	3	1	2	3	1	2	3	1	2	3
$\eta = 0.4, N = 36$													
T	50	100	**100**	**94**	60	**0**	**0**	100	**34**	**2**	40	**0**	**0**
	100	100	**100**	**94**	98	**0**	**0**	100	**16**	**2**	69	**0**	**0**
	200	100	**99**	**76**	100	**0**	**0**	100	**7**	**0**	100	**0**	**0**
	400	100	**98**	**41**	100	**0**	**0**	100	**6**	**0**	100	**0**	**0**
	800	100	**86**	**10**	100	**0**	**0**	100	**2**	**0**	100	**0**	**0**
$T = 100, N = 36$													
η	$*$	**9**	**2**	**1**	**0**	**0**	**0**	**5**	**1**	**0**	**0**	**0**	**0**
	0	**100**	**100**	**100**	**4**	**0**	**0**	**98**	**69**	**45**	**2**	**0**	**0**
	0.2	100	**100**	**100**	12	**0**	**0**	96	**53**	**26**	5	**0**	**0**
	0.4	100	**100**	**94**	98	**0**	**0**	100	**16**	**2**	69	**0**	**0**
	0.6	100	**79**	**15**	100	**0**	**0**	100	**1**	**0**	100	**0**	**0**
	0.8	100	**18**	**0**	100	**0**	**0**	100	**0**	**0**	100	**0**	**0**
	1	100	**0**	**0**	100	**0**	**0**	100	**0**	**0**	100	**0**	**0**
$T = 100, \eta = 0.4$													
N	11	100	**16**	**0**	91	**1**	**0**	95	**3**	**0**	91	**0**	**0**
	36	100	**100**	**94**	98	**0**	**0**	100	**16**	**2**	69	**0**	**0**
	81	100	**100**	**100**	97	**2**	**0**	100	**48**	**12**	53	**0**	**0**
	121	100	**100**	**100**	96	**2**	**0**	100	**64**	**27**	48	**0**	**0**
	201	100	**100**	**100**	94	**1**	**0**	100	**75**	**39**	57	**0**	**0**
	301	100	**100**	**100**	92	**1**	**0**	100	**72**	**46**	52	**0**	**0**

expected values for numbers in bold are 0, while 100 for those in normal font.

With fixed N=36 and η=0.4, all of these rules perform better with a larger sample size. Nonparametric bootstrap and modified parallel analysis are the best rules of the four to abandon the actual spurious coupled modes, and perform better to retain the significant coupled mode with a larger sample size. Parallel analysis works worst in abandoning the spurious coupled modes. It rejects the null hypothesis to consider the second mode, which is not significant and should have been abandoned, to be true for 86% realizations, even when T=800. Random-phase test performs worse than nonparametric bootstrap and modified parallel analysis to abandon the spurious modes, but much better than parallel analysis.

As the signal-to-noise ratio η varies with N=36 and T=100, all of the rules perform better with a stronger coupled signal. Note that the parallel analysis is robust only in the special cases when either (1) the observations at the grid points are spatially uncorrelated or (2) the coupled signal shared by the two analyzed fields is very strong. In other situations, the rule is apt to retain the spurious coupled modes. Random-phase test works well in model 1, and in model 3 with $\eta \geqslant 0.4$. When there is no coupled signal or weak signal ($\eta < 0.4$) shared by the two analyzed fields in which the observations are spatially correlated, the rule is unable to perform satisfyingly, especially in the situation of model 2.

Unlike the sampling errors that are insensitive to the number N of grid points with fixed T and η [Bretherton et al. (1992) considered the sampling errors to be independent of N based on experiment results], some rules are sensitive to N. The modified parallel analysis performs best to abandon the spurious modes, but its capability of retaining the significant mode is low when $N/T > 0.2$. As the number of grid points increases, parallel analysis and random-phase test perform worse to abandon the spurious coupled modes, especially the former, which does well only when N is much smaller than T. Nonparametric bootstrap performs satisfactorily with various N to control not only the probability of a type I error, but also the probability of a type II error. In other words, it is the most powerful one of the four rules to judge how many modes should be retained.

4.4 *Analysis of the experiment results*

In the procedures of the four test rules, every step is similar except for step 1 that illustrates how to build the simulated fields. There are three main factors that impact upon the accuracy of the four test rules: (1) the simulated fields are built inappropriately according to

step 1; (2) in step 3, the frequency (=α) of a type I error based on m independent pairs of simulated fields is used to replace the probability (=μ) of a type I error for any pair of simulated fields; and (3) the significance level α is unequal to 0. Let $\hat{b}_j^{\alpha}$ ($j = 1, 2, 3$) denote the critical squared covariance fractions explained by the first three modes for any simulated fields with significance level α. Then $\widehat{\mathrm{SCF}}_j^{\alpha} \neq \widehat{b}_j^{\alpha}$ and $\hat{b}_j^{\alpha} \neq \hat{b}_j^{\alpha\times m}$ are due to the first two reasons respectively. For d (=100) independent realizations of a particular model, even though $\widehat{\mathrm{SCF}}_j^{\alpha} = \hat{b}_j^{\alpha\times m}$, the number of realizations reserving spurious coupled modes may be unequal to 0 on account of the third reason. Here, we will discuss the first two reasons in detail.

The unapt simulated fields built to find out $\widehat{\mathrm{SCF}}_j^{\alpha}$ is the main reason why parallel analysis and random-phase test perform worse than nonparametric bootstrap and modified parallel analysis to abandon the spurious coupled modes, when the observations in the analysis fields are spatially correlated. The $\hat{N}_{\mathrm{ef}}^*$ of the unapt simulated fields built in the two test rules is larger than that of the analyzed fields, and hence the critical fractions $\hat{b}_j^{\alpha}$ ($j = 1, 2, 3$) of the simulated fields are smaller than the critical fractions $\widehat{\mathrm{SCF}}_j^{\alpha}$ of the analyzed fields based on the discussion in subsection 3.4. If we replace $\widehat{\mathrm{SCF}}_j^{\alpha}$ with $\hat{b}_j^{\alpha}$ ($j = 1, 2, 3$), there are many realizations whose fractions $\widehat{\mathrm{SCF}}_j^{\alpha}$ ($j = 1, 2, 3$), which should have been out of the rejection regions of null hypotheses, are in the rejection regions of null hypotheses now. The underlying reason, why $\hat{N}_{\mathrm{ef}}^*$ is different between simulated fields built according to parallel analysis and corresponding analyzed fields, is the improper assumption that the observations of analyzed fields are spatially independent. While the underlying reason for random-phase test is that the course of performing step (b) to build simulated fields makes every grid point have random phase and consequently induces simulated fields to have large $\hat{N}_{\mathrm{ef}}^*$. For example, for one realization of model 2 with T=100, N=36, the $\hat{N}_{\mathrm{ef}}^*$ is about 25 for 100 pairs of independent simulated fields $\hat{S}^k$ ($k = 1, 2, \cdots, 100$) built according to random-phase test but about 6 for the corresponding analyzed field (Fig. 5).

Step 3 in the procedure of the four test rules does not keep the probability, but only the frequency of reserving the actual spurious modes at the small valueα for finite m pairs of independent simulated fields, so the significance level of the four test rules is unequal toα but μ. In terms of probability theory, the sample space of a significance test with significance level μ contains only two sample points: 0 (accepting the null hypothesis) with probability equal to $1 - \mu$, and 1(rejecting the null hypothesis) with probability equal to μ. That is, a significance test is a Bernoulli trial. Let ε denote the error of estimating the probability μ with frequency (=α). According to the Weak Law of Large Numbers, also known as Bernoulli Theorem, ε converges to 0 as $m \to \infty$. The bound ε_b of ε for m pairs of simulated fields with fixed N/T can be derived from Chebyshev Inequality:

$$P(|\mu - \alpha| \geqslant \varepsilon_b) \leqslant \frac{\sigma^2}{m\varepsilon_b^2} . \tag{15}$$

If we let $\sigma^2/(m\varepsilon_b^2)$=0.05, then $P(|\mu - \alpha| \geqslant \varepsilon_b) \leqslant 0.05$. For Bernoulli distribution, the variance $\sigma^2 = \mu(1 - \mu)$, so

$$\varepsilon_b^2 = \frac{\mu(1-\mu)}{0.05m} \leqslant \frac{0.5(1-0.5)}{0.05m} .$$

Hence, for m (=100) independent pairs of simulation fields with fixed N/T, the possible value of μ ranges from 0 to 0.27. For fixed $N/T, \varepsilon$ is smaller and hence the gap between $\hat{b}_j^{\alpha}$ and $\hat{b}_j^{\alpha\times m}$ is narrower as m increases. For fixed m (=100) independent pairs of simulation fields, as N/T is lower, the fractions $\hat{b}_j^{\alpha}$ of the simulated fields are estimated more accurately, and hence $\hat{b}_j^{\alpha\times m}$ is closer to $\hat{b}_j^{\alpha}$.

For d (=100) independent realizations of one of the three models introduced in subsection 4.1, the number g of realizations rejecting the null hypothesis with probability μ at the spurious modes fits the binomial distribution, that is, $P(g) = C_d^g \mu^g (1 - \mu)^{d-g}$ ($g = 0, 1, 2, \cdots, d$). When μ=0, g is determinately equal to 0. While the possible value of g almost ranges from 0 to 15 with μ=0.05 (Fig. 6) and almost ranges from 15 to 40 with μ=0.27 (Fig. 7). Therefore, the possible number of realizations rejecting the null hypothesis at the spurious modes almost ranges from 0 to 40. In other words, for all modes of models 1 and 2 and remaining $\hat{r} - 1$ modes of model 3, when the small value

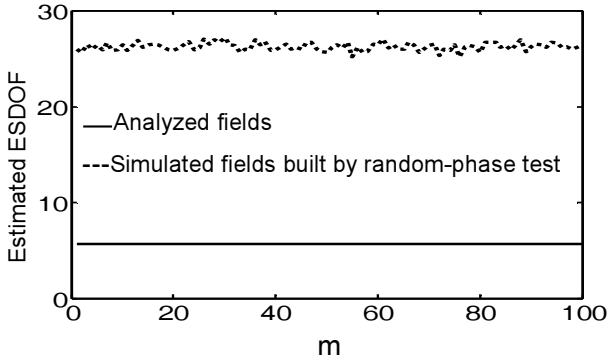

Fig. 5. The estimated effective number of spatial degrees of freedom (ESDOF) for m pairs of independent simulated fields built following random-phase test and for the corresponding analyzed field in one realization of model 2.

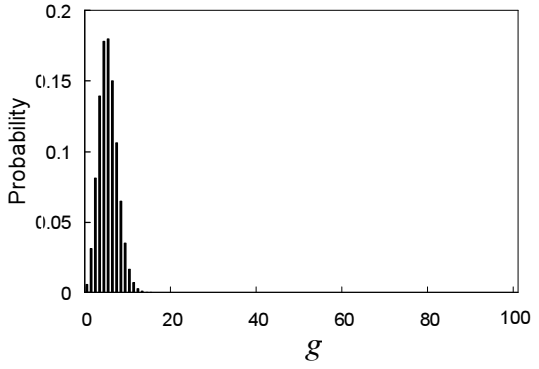

Fig. 6. The probability $P(g)$ of the number g of realizations rejecting the null hypothesis with significant level μ=0.05 at the spurious mode where the null hypothesis is true.

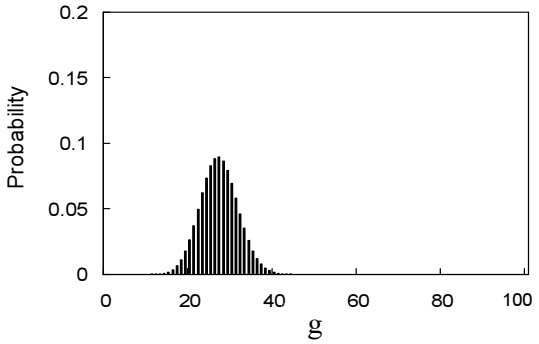

Fig. 7. As Fig. 6 except with significant level μ=0.27.

α is equal to 0.05 and $\hat{b}_j^{\alpha} = \widehat{\mathrm{SCF}}_j^{\alpha}$, the range of the possible value of g is from 0 to 40 and narrower as N/T is lower or m increases. When α =0 and $\hat{b}_j^{\alpha} = \widehat{\mathrm{SCF}}_j^{\alpha}, g$ converges to the expected value 0 as $N/T \to 0$ or $m \to \infty$.

5. Conclusions

Due to SVD having a high potential for spurious coupled patterns and correlation due to sampling error, it is essential to test the significance of coupled modes of SVD, especially for geophysical fields, in which the sample size is generally not large enough. In this paper, we have clarified the essence of the significance test of SVD, and investigated the essence, process, and properties of four significance test rules: parallel analysis, nonparametric bootstrap, random-phase test, and modified parallel analysis. The new rule (modified parallel analysis) has been proposed because the critical fractions for the first three coupled modes of the two analyzed fields are quite sensitive to spatial correlation.

Each of the rules is applied to a numerical experiment including three models: model 1 is that the observations at the grid points are spatially and temporally uncorrelated, and there is no coupled signal between the two fields; model 2 is that the observations at the grid points are spatially correlated but temporally uncorrelated, and there is also no coupling between the two fields; and model 3 is that there is a perfect anti-correlated coupled signal adding to the two data fields in model 2.

Results from the experiment show that the four rules are more accurate for testing the significance of coupled modes of SVD with a lower ratio of the number of grid points N to sample size T. Modified parallel analysis is the best rule to abandon the spurious coupled modes, and performs better to retain the significant modes with lower N/T. Nonparametric bootstrap performs quite satisfactorily to abandon the spurious modes and retain the significant modes. In particular, when $N/T > 0.2$, it is the best rule to control two types of errors. Parallel analysis is robust only in special cases when the observations at the grid points are spatially uncorrelated or there are very strong coupled signals shared by the two analyzed fields. If there are no coupled signals or a small signal ($\eta < 0.4$) between the two analyzed fields, random-phase test is unable to perform satisfactorily, especially in model 2. As the number of grid points increases, parallel analysis and random-phase test perform worse to abandon the spurious modes, especially the former which only does well when N is much smaller than T.

Three reasons that may affect the accuracy of the test rules have been discussed in this paper: (1) the simulated fields are built inappropriately; (2) in the step which selects the critical fractions, the frequency (=α) of a type I error based on m independent pairs of simulated fields is used to replace the probability μ of a type I error for any pair of simulated fields; and (3) the significance level α is unequal to 0.

Acknowledgements. The authors are grateful to the two anonymous reviewers for their helpful comments and suggestions. This work was supported by the National Natural Science Foundation of China under Grant No. 40233027.

REFERENCES

Bagrov, N. A., 1969: On the equivalent number of independent data. *Tr. Gidrometeor. Cent.*, **44**, 3–11. (in Russian)

Barnett, T. P., and R. W. Preisendorfer, 1987: Origins and levels of monthly and seasonal forecast skill for United States surface air temperatures determined by canonical correlation analysis. *Mon. Wea. Rev.*,

115, 1825–1850.

Bartlett, M. S., 1939: A note on tests of significance in multivariate analysis. *Mathematical Proceedings of the Cambridge Philosophical Society*, **35**, 180–185.

Bartlett, M. S., 1950: Tests of significance in factor analysis. *The British Journal of Psychiatry* (*Statistical Section*), **3**, 77–85.

Bartlett, M. S., 1954: A note on the multiplying factors for various X^2 approximations. *Journal of the Royal Statistical Society Series B—Statistical Methodology*, **16**, 296–298.

Bretherton, C. S., C. Smith, and J. M. Wallace, 1992: An intercomparison of methods for finding coupled patterns in climate data. *J. Climate*, **5**, 541–560.

Bretherton, C. S., C. Smith, and J. M. Wallace, M. Widmann, V. P. Dymnikov, J. M. Wallace, and I. Bladé, 1999: The effective number of spatial degrees of freedom of a time-varying field. *J. Climate*, **12**, 1990–2009.

Chang, C. P., P. A. Harr, and H. J. Chen, 2005: Synoptic disturbances over the equatorial South China Sea and western Maritime Continent during boreal winter. *Mon. Wea. Rev.*, **133,** 489–503.

Cherry, S., 1996: Singular value decomposition analysis and canonical correlation analysis. *J. Climate*, **9**, 2003–2009.

Crawford, C. B., 1975: Determining the number of interpretable factors. *Psychological Bulletin*, **82**, 226–237.

Davis, R. E., 1976: Predictability of sea surface temperature and sea level pressure anomalies over the North Pacific Ocean. *J. Phys. Oceanogr.*, **6**, 249–466.

Dymnikov, V. P., and S. K. Filin, 1985: A study of correlations between sea-surface temperature anomalies in mid-latitudes and anomalies in heating, based on data from the First GARP Global Experiment. Reprint of the Department of Numerical Mathematics of the U. S. S. R. Academy of Sciences. [Available from Leninskjj Prosp., 14; 117901 Moscow B-71].

Ebisuzaki, W., 1997: A method to estimate the statistical significance of a correlation when the data are serially correlated. *J. Climate*, **10**, 2147–2153.

Efron, B., 1979: Bootstrap methods: another look at the jackknife. *Annals of Statistics*, **7**, 1–26.

Efron, B., 1987: Better bootstrap confidence intervals. *Journal of the American Statistical Association*, **82**, 171–185.

Feddersen, H., 2003: Predictability of seasonal precipitation in the Nordic region. *Tellus*, **55A**, 385–400.

Feddersen, H., and U. Ersen, 2005: A method for statistical downscaling of seasonal ensemble predictions. *Tellus*, **57A**, 398–408.

Feddersen, H., A. Navarra, and M. N. Ward, 1999: Reduction of model systematic error by statistical correction for dynamical seasonal predictions. *J. Climate*, **12**, 1974–1989.

Guo Pinwen, Tian Hong, and Liu Xuanfei, 2000: Tropical convective activities related to summer rainfall anomaly in China. *Adv. Atmos. Sci.*, **17**, 121–128.

Hsu, H., 1994: Relationship between tropical heating and global circulation: Interannual variability. *J. Geophys. Res.*, **99**, 10473–10489.

Isaaks, E. H., and R. M. Srivastava, 1989: *An Introduction to Applied Geostatistics*. Oxford University Press, 561pp.

Iwasaka, N., and J. M. Wallace, 1995: Large scale air sea interaction in the northern hemisphere from a view point of variations of surface heat flux by SVD analysis. *J. Meteor. Soc. Japan*, **73**, 780–793.

Lanzante, J. R., 1984: A rotated eigenanalysis of the correlation between 700 mb heights and sea surface temperatures in the Pacific and Atlantic. *Mon. Wea. Rev.*, **112**, 2270–2280.

Lau, K. M., and H. Y. Weng, 2001: Coherence modes of global SST nd summer rainfall over China: An assessment of the regional impacts of the 1997–98 El Niño. *J. Climate*, **14**, 1294–1308.

Lau, K. M., H. Y. Weng, and H. T. Wu, 2001: Principal modes of rainfall-SST variability of the Asian summer monsoon: a reassessment of the monsoon-ENSO relationship. *J. Climate*, **14**, 2880–2895.

Lawley, D. N., 1956: Tests of the significance for the latent roots of covariance and correlation matrices. *Biometrika*, **43**, 128–136.

Li Fang, Lin Zhongda, Zuo Ruiting, and Zeng Qingcun, 2005: The method for correcting the summer precipitation anomaly predicted extraseasonally over East Asian monsoon region based on EOF and SVD. *Climatic and Environment Research*, **10**(3), 658–668. (in Chinese)

Liu, Y., 2003: Spatial patterns of soil moisture connected to monthly-seasonal precipitation variability in a monsoon region. *J. Geophys. Res.*, **108**, 8856–8869.

North, G. R., T. L. Bell, R. F. Cahalan, and F. J. Moeng, 1982: Sampling errors in the estimation of empirical orthogonal functions. *Mon. Wea. Rev.*, **110**, 699–706.

Peng, M., and J. Fyfe, 1996: The coupled patterns between sea level pressure and sea surface temperature in the midlatitude north atlantic. *J. Climate*, **9**, 1824–1839.

Prohaska, J., 1976: A technique for analyzing the linear relationships between two meteorological fields. *Mon. Wea. Rev.*, **104**, 1345–1353.

Quadrelli, R., C. S. Bretherton, and J. M. Wallace, 2005: On sampling error in empirical orthogonal functions. *J. Climate*, **18**, 3704–3711.

Rodríguez-Fonseca, B., and E. Serrano, 2002: Winter 10-day coupled patterns between geopotential height and Iberian Peninsula rainfall using the ECMWF precipitation reanalysis. *J. Climate*, **15**, 1309–1321.

Rohatgi, V. K., 1984: *Statistical Inference*. John Wiley & Son, Inc., 214–215.

Shabbar, A., and W. Skinner, 2004: Summer drought patterns in Canada and the relationship to global sea surface temperatures. *J. Climate*, **17**, 2866–2880.

Seber, G. A. F., 1984: *Multivariate Observations.* Wiley, New York, 686pp.

Shen, S. H., and K. M. Lau, 1995: Biennial oscillation

associated with the East Asian summer monsoon and tropical sea surface temperature. *J. Meteor. Soc. Japan*, **73**, 105–124.

Sheng Ju, Xie Shiqian, and Pan Chengyi, 1989: *Probability Theory and Mathematical Statistics*. Heigher Education Press, Beijing, 189–195.

TerMegreditchian, M. G., 1969: On the determination of the number of independent stations which are equivalent to prescribed systems of correlated stations. *Meteor. Hydrol.*, **2**, 24–36.

Terray, P., and S. Dominiak, 2005: Indian Ocean sea surface temperature and El Niño–Southern Oscillation: a new perspective. *J. Climate*, **18**, 1351–1368.

Thiébaux, H. J., and F. W. Zwiers, 1984: The interpretation and estimation of effective sample size. *J. Climate Appl. Meteor.*, **23**, 800–811.

Wallace, J. M., C. Smith, and C. S. Bretherton, 1992: Singular value decomposition of wintertime sea surface temperature and 500-mb height anomalies. *J. Climate*, **5**, 561–576.

Wang Hui, and Fu Rong, 2002: Cross-equatorial flow and seasonal cycle of precipitation over South America. *J. Climate*, **15**, 1591–1608.

Widmann, M., C. S. Bretherton, and E. P. Salathé Jr., 2003: Statistical precipitation downscaling over the northwestern United States using numerically simulated precipitation as a predictor. *J. Climate*, **16**, 799–816.

Widmann, M., 2005: One-dimensional CCA and SVD, and their relationship to regression maps. *J. Climate*, **18**, 2785–2792.

Zhang, R. H., and S. E. Zebiak, 2004: An embedding method for improving interannual variability simulations in a hybrid coupled model of the tropical Pacific ocean–atmosphere system. *J. Climate*, **17**, 2794–2812.