Theor. Appl. Climatol. (2008)
DOI 10.1007/s00704-007-0351-z
Printed in The Netherlands

Theoretical and Applied Climatology

[1] Department of Statistics, Columbia University, New York, USA
[2] Lamont-Doherty Earth Observatory of Columbia University, Palisades, New York, USA

Detecting shifts in correlation and variability with application to ENSO-monsoon rainfall relationships

L. F. Robinson[1], V. H. de la Peña[1], Y. Kushnir[2]

With 3 Figures

Received 29 January 2007; Accepted 23 July 2007; Published online 6 February 2008

Summary

This paper addresses the retrospective detection of step changes at unknown time points in the correlation structure of two or more climate times series. Both the variance of individual series and the covariance between series are addressed. For a sequence of vector-valued observations with an approximate multivariate normal distribution, the proposed method is a parametric likelihood ratio test of the hypothesis of constant covariance against the hypothesis of at least one shift in covariance. The formulation of the test statistic and its asymptotic distribution are taken from Chen and Gupta (2000). This test is applied to the series comprised of the mean summer NINO3 index and the Indian monsoon rainfall index for the years 1871–2003. The most likely change point year was found to be 1980, with a resulting p-value of 0.12. The same test was applied to the series of NINO3 and Northeast Brazil rainfall observations from the years 1856–2001. A shift was detected in 1982 which is significant at the 1% level. Some or all of this shift in the covariance matrix can be attributed to a change in the variance of the Northeast Brazil rainfall. A variation of this methodology designed to increase power under certain multiple change point alternatives, specificallly when a shift is followed by a reversal, is also presented. Simulations to assess the power of the test under various alternatives are also included, in addition to a review of the literature on alternative methods.

Correspondence: Lucy F. Robinson, Department of Statistics, Columbia University, 1255 Amsterdam Ave. 10th flr, MC 4409, New York, NY 10027, USA, e-mail: lfr24@columbia.edu

1. Introduction

Assessing the stability over time of climate processes and the connections between them is crucial to our understanding of a changing climate. Changes in variability or connections between processes, if robust, can profoundly change our assessment of climate impacts and affect climate predictability. An area of great recent concern is the relationship between the Indian monsoon rainfall (IMR) and the El Niño/Southern Oscillation (ENSO) phenomenon. The existence of a significant negative correlation between time series has been long been observed (Walker and Bliss 1937), but whether the strength of the relationship has decreased in recent decades is a subject of current debate.

Running correlation analysis, in which correlations are computed in overlapping moving windows, has frequently been used in an attempt to document and understand changes in the correlation between two climate indices. In particular, the existence of low-frequency modes of variability is of current interest in many areas of climate research, and running correlations have been used to represent the multi-decadal evolution of the relationship between two processes.

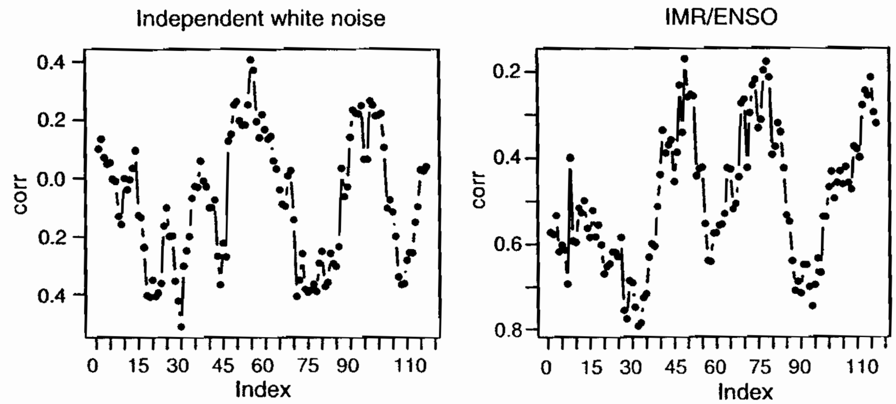

Fig. 1. Comparing the 21-year windowed running correlations of the IMR/ENSO time series with those of two uncorrelated simulated white noise processes illustrates Gershunov et al.'s (2001) observation that apparent periodic fluctuations in running correlations are not reliable indicators of a changing underlying correlation structure, as these fluctuations exist even in stable, uncorrelated processes

Among others, Krishnamurthy and Goswami (2000) have used running correlations to argue for the existence of low-frequency (15–25 year) oscillations in the relationship between the IMR and ENSO. Parthasarathy et al. (1991) used similar techniques to examine the relationships between monsoon rainfall and other climate variables.

However, Gershunov et al. (2001) have shown that there are serious problems in the physical interpretation of the results of a running-correlation analysis. These problems stem from the fact that a running correlation analysis applied to any two processes, even independent processes, produces what appears to be a low-frequency periodic evolution in the correlation. This however, is merely an artifact of the method itself and does not reflect any characteristic of the relationship between the processes. Sample correlations are inherently subject to random fluctuations, and the overlapping nature of the running correlations turns these fluctuations into smooth trends. Figure 1 (and similar figures in Gershunov et al. (2001)) compares the results of running correlation analysis of the ENSO/IMR relationship and of two uncorrelated white noise processes.

Gershunov et al. (2001) propose a method of determining whether observed fluctuations in running correlations are different from what would be expected by chance. They suggest comparing the standard deviation (SD) of an observed series of running correlations with upper and lower confidence bounds computed from the bootstrapped SDs of simulated processes with stationary correlations.

In their scheme, the SD of the running correlations of the ENSO/IMR series is compared to simulations of bivariate Gaussian observations with a correlation of 0.6 (the correlation of the entire ENSO/IMR series is about −0.6). They find that the ENSO/IMR series is actually significantly *less* variable than the simulations, with the observed SD below 5th percentile of the bootstrapped SDs of the simulations. They suggest that there is a physical process moderating the fluctuations of the sliding correlations.

While Gershunov et al.'s simulations help to illuminate the distribution of a running correlation series with constant correlation, the use of the SDs of the running correlation to characterize the evolution of the process is an indirect way to address the issue of a potentially changing relationship. The hypotheses being tested using their proposed method are not clearly related to the behavior of the processes themselves. Rather, they refer only to their running correlations, statistics whose variability does not give clear insight into the underlying correlation structure.

Kwon et al. (2005) use running correlation analysis and empirical orthogonal functions to examine the connection between ENSO and the Western North Pacific (WNP) summer monsoon. They apply the significance test suggested by Gershunov and find that the variation in the sliding correlations is significant at the 10% confidence level. Based on a comparison of the first two leading empirical orthogonal functions (EOF) of WNP summer-mean precipitation (based on station data), they conclude that the relationship in the period from 1994–2003 is weaker than in 1979–1993. In the first time period they find that the first mode of variation is one which is highly correlated with ENSO, and the second mode is highly correlated with another precipitation index, WNP Monsoon index (WNPMI). In the latter period, they find the same 2 dominant modes, but the order is reversed. In other words, the ENSO mode is the first dominant mode in the 1979–1993 period, and drops to the second dominant mode in the 1994–2004 period. The authors conclude from this that the relationship with ENSO has weakened. This is clearly an interesting observation, but it is difficult to firmly distinguish from chance variability without knowing the probability of such a reversal happening by chance.

Maraun and Kurths (2005) use nonlinear time-series methods to investigate the evolution of the phase coherence between ENSO and IMR series over the 1871–2004 time period. They decompose the interannual oscillation dynamics of the two series into amplitude and phase, assessing the relationship between them in terms of phase coherence irrespective of the amplitude. They find periods (1886–1908 and 1964–1980) in which the phases are strongly coupled in comparison to the rest of the time period. They also develop a simulation scheme by which to judge statistical significance. Empirical probabilities of typical lengths of interannual oscillations are computed from the ENSO and IMR series and used to create 1,000,000 pairs of annually resolved 150-year time series. Based on the simulations, the observed periods of phase coherence are found to be highly significant.

Kumar et al. (1999) use resampling methods to estimate the 95% upper confidence bound for 21-year sliding correlations and conclude that a change in the behavior of the ENSO/IMR correlations has occurred. The series is resampled 1000 times in random 21-year chunks, and 5th and 95th percentiles of the 1000 sample correlation coefficients are computed. When the series of observed running correlation is compared to the bootstrapped 90% confidence range they find that in recent decades the sliding correlations have exceeded the upper confidence bound (i.e. are closer to zero than would be expected under the hypothesis of constant correlation) and conclude that the ENSO/IMR relationship has become weaker. Implicitly, the authors have examined each of the 121 individual values of the running correlations. This creates multiple testing issues: even when all observations are drawn from the same distribution, we expect that 10% will fall outside of a 90% confidence range purely by chance. In light of these issues, the statistical significance of the exceedance of the 95% upper confidence bound in 1980 is unclear.

There appears to be no clear consensus on the best way to attach statistical significance to observed changes in correlation. A formal statistical test with clearly defined hypotheses could be useful. Parametric methods for detecting change points in a variety of contexts can be found in Chen and Gupta (2000). Their parametric likelihood ratio test for detecting change points will be presented with applications to the covariance relationship between IMR and ENSO, and for comparison, that between the Northeast Brazilian Rainfall and ENSO (see Chiang et al. 2000) for a discussion of this relationship.) In contrast to previous approaches, we will use the covariance matrix Σ rather than the correlation coefficient $\rho_{xy} = \sigma_{xy}/\sigma_x\sigma_y$ as the parameter of interest. A change in ρ can reflect changes in the covariance of the two processes, a change in the variance of one or both of the processes, or both. To detect a shift in variance rather than covariance, a univariate version of Chen and Gupta's test will be used.

In the applications presented the climate processes are slightly auto-correlated. However, the results of our analysis are virtually unchanged after removing the autoregressive components of the time series. The methods presented are intended for use on independent sequences of observations, but are also appropriate for the residuals of an ARIMA model. Local change point detection, a variation of the change point detection algorithm (Mercurio and Spokoiny 2004; Giacomini et al. 2006) is also presented, with the intent to increase power under multiple change point alternatives, for example in situations where a shift is followed by a reversal to the original state, a situation that is important in the long term study of ENSO and IMR.

2. Methodology

Likelihood ratio tests are a fundamental part of classical statistical hypothesis testing, and the literature on their general properties is extensive. Lehmann (1997) is a good resource for many aspects of hypothesis testing.

Given n independent observations $\mathbf{x}_1 \cdots \mathbf{x}_n$ observed in order, the general null hypothesis for a change point problem is that the probability distribution of the observations remains constant. If F_i is the distribution of $\mathbf{x}_i$, the null hypothesis is

$$H_0\colon F_1 = F_2 = \cdots F_{(n-1)} = F_n \tag{1}$$

and the alternative is

$$\begin{aligned} H_1\colon F_1 = \cdots = F_{k_1} \neq F_{(k_1+1)} = \cdots = F_{k_2} \neq F_{k_2+1} \\ = \cdots = F_{k_q} \neq F_{(k_q+1)} = \cdots = F_n, \end{aligned} \tag{2}$$

where q is the unknown number of change points and $1 < k_1 < \cdots k_p < n$ are the unknown positions of the change points. If $\mathbf{x}_1 \cdots \mathbf{x}_n$ come from a common parametric family of distributions, then the problem is one of detecting changes in the parameters of $F_1 \cdots F_n$, and the relevant hypotheses become $H_0\colon \theta_1 = \cdots = \theta_n$ and $H_1\colon \theta_1 = \cdots \theta_{k_1} \neq \theta_{(k_1+1)} = \cdots \theta_{k_2} \neq \theta_{k_2+1} = \cdots \theta_{k_q} \neq \theta_{(k_q+1)} = \cdots = \theta_n$ where θ_i is the vector of parameters for F_i.

The basic test procedure is to formulate the likelihood ratio (LR) based on maximum likelihood estimates of the parameters under the null and alternative hypotheses, as well as the m.l.e. of the change points,

$$\mathrm{LR} = \frac{\text{Likelihood of data under alternative}}{\text{Likelihood of data under null}} \tag{3}$$

and compute a p-value by comparing the observed LR to its distribution under the null hypothesis. In practice $\lambda = \log(\mathrm{LR})$ is used instead of LR. The global procedure outlined by Chen and Gupta (2000) for finding multiple change points is to look for the most significant change point k by testing $\mathbf{x}_1 \cdots \mathbf{x}_n$ using an alternative hypothesis of one change point, and then apply the same test on $\mathbf{x}_1 \cdots \mathbf{x}_k$ and $\mathbf{x}_{k+1} \cdots \mathbf{x}_n$ iteratively until the null hypothesis is no longer rejected. However, under some multiple change point alternatives the global procedure may lack power, and local change point detection maybe more appropriate. Chen and Gupta have derived the asymptotic distribution of λ for several distributions, including univariate and multivariate normal, gamma, exponential, Poisson and binomial, making the method widely applicable.

In the examples to be presented the data are yearly observations of vector-valued climate indices, and the parameter of interest is the covariance matrix. Specifically, we will test for significant changes in the covariance structure of the ENSO-precipitation relationship in India and Brazil in the last 130/150 years. The ENSO/IMR and ENSO/Brazilian rainfall series are modeled as multivariate normal. One can test for changes in the mean vector of their distributions, in the covariance matrix, or for a simultaneous change in both parameters. When the mean is known, it can be removed from the series which can then be modeled as mean zero. In this case, the null and alternative hypotheses are $H_0\colon \Sigma_1 = \cdots \Sigma_n$ and $H_1\colon \Sigma_1 = \cdots = \Sigma_k \neq \Sigma_{k+1} = \cdots = \Sigma_n$ where k is the position of the single change point at each iteration. The observations are $\mathbf{x}_1 \cdots \mathbf{x}_n$, each a vector of length m. In this case $m = 2$. Under H_0, the joint likelihood function of $\mathbf{x}_1 \cdots \mathbf{x}_n$ is

$$L_0(\Sigma) = \frac{1}{2\pi}^{mn/2} |\Sigma|^n \exp\left\{-\frac{1}{2}\sum_{i=1}^{n} \mathbf{x}_i{}'\Sigma^{-1}\mathbf{x}_i\right\}, \tag{4}$$

so the log-likelihood is

$$\begin{aligned} \log(L_0(\Sigma)) = -\frac{mn}{2}\log 2\pi - n\log|\Sigma| \\ -\frac{1}{2}\sum_{i=1}^{n} \mathbf{x}_i{}'\Sigma^{-1}\mathbf{x}_i. \end{aligned} \tag{5}$$

Σ is unknown so the maximum likelihood estimate $\widehat{\Sigma} = \frac{1}{n}\sum_{i=1}^{n} \mathbf{x}_i{}'\mathbf{x}_i$ is used, making the maximum log likelihood function

$$\begin{aligned} \log L_0(\widehat{\Sigma}) = -\frac{mn}{2}\log 2\pi \\ -\frac{n}{2}\log\left|\frac{1}{n}\sum_{i=1}^{n} \mathbf{x}_i{}'\mathbf{x}_i\right| - \frac{n}{2}. \end{aligned} \tag{6}$$

Under H_1, $\mathbf{x}_1 \cdots \mathbf{x}_k$ are iid $N_m(0, \Sigma_1)$ and $\mathbf{x}_{k+1} \cdots \mathbf{x}_n$ are iid $N_m(0, \Sigma_2)$. The MLE's for Σ_1 and Σ_2 are

$$\widehat{\Sigma}_1 = \frac{1}{k}\sum_{i=1}^{k} \mathbf{x}_i{}'\mathbf{x}_i \quad \text{and} \quad \widehat{\Sigma}_2 = \frac{1}{n-k}\sum_{i=k+1}^{n} \mathbf{x}_i{}'\mathbf{x}_i \tag{7}$$

respectively, making the maximum log likelihood function under H_1

$$\log L_1(\widehat{\Sigma}_1, \widehat{\Sigma}_2) = -\frac{mn}{2}\log 2\pi - \frac{k}{2}\log\left|\frac{1}{k}\sum_{i=1}^{k}\mathbf{x}_i{}'\mathbf{x}_i\right| - \frac{n-k}{2}\log\left|\frac{1}{n-k}\sum_{i=k+1}^{n}\mathbf{x}_i{}'\mathbf{x}_i\right| - \frac{n}{2}, \tag{8}$$

where $|\Sigma|$ is the determinant of Σ. The position of the change point k must also be estimated, and the MLE k is the value which maximizes $\log(L_1)$. The MLE's can only be obtained for $m \leq k \leq n-m$, so the maximum log likelihood ratio is

$$\lambda_n = \max_{m<k<n-m}\left(\log\left|\frac{1}{n}\sum_{i=1}^{n}\mathbf{x}_i{}'\mathbf{x}_i\right|^n - \log\left|\frac{1}{k}\sum_{i=1}^{k}\mathbf{x}_i{}'\mathbf{x}_i\right|^k - \log\left|\frac{1}{n-k}\sum_{i=k+1}^{n}\mathbf{x}_i{}'\mathbf{x}_i\right|^{n-k}\right)^{\frac{1}{2}}. \tag{9}$$

Chen and Gupta (2000) have calculated the limiting distribution of λ_n under H_0:

$$\lim_{n\to\infty} P(a_n\lambda_n - b_{mn} \leq x) = e^{-2e^{-x}} \quad \text{for all } x \in \mathbf{R},$$

with

$$a_n = (2\log\log n)^{\frac{1}{2}} \quad \text{and}$$

$$b_{mn} = 2\log\log n + \frac{m}{2}\log\log\log(n) - \log\left(\Gamma\left(\frac{m}{2}\right)\right), \tag{10}$$

where m is the dimension of the multivariate normal distribution.

This distribution is used to calculate the approximate p-value of an observed λ.

Perhaps a more common case is one in which the mean is unknown but the same under the null and alternative hypotheses. In this case, the maximum likelihood estimates for $\boldsymbol{\mu}, \Sigma_1$ and Σ_n can be found numerically by maximizing the log-likelihood function under the alternative,

$$-\frac{mn}{2}\log(2\pi) - \frac{k}{2}\log|\Sigma_1| - \frac{n-k}{2}\log|\Sigma_n| - \frac{1}{2}\left[\sum_{i=1}^{k}(\mathbf{x}_i - \boldsymbol{\mu})'\Sigma_1^{-1}(\mathbf{x}_i - \boldsymbol{\mu}) + \sum_{i=k+1}^{n}(\mathbf{x}_i - \boldsymbol{\mu})'\Sigma_n^{-1}(\mathbf{x}_i - \boldsymbol{\mu})\right], \tag{11}$$

for a specific k. Estimates for all possible change points must be computed to find the maximum of all likelihood ratios. In practice, this can be tedious and may present difficulty for large m. Simulation studies indicate that this numerical optimization may not be necessary, however. For n as small as 25 no substantive difference in the distribution of the test statistic was found between the case where a process was truly mean-zero and the one in the sample average was removed from a process with non-zero mean. Asymptotically, removing the sample mean is justified by the law of large numbers, which states that as the sample size increases $(\mathbf{x}_i - \bar{\mathbf{x}}) \to (\mathbf{x}_i - \boldsymbol{\mu})$ almost surely. This is the approach taken in the examples below. Independence between observations is preserved after removing the sample mean under the assumption of i.i.d. normality, thus it is important to confirm that this assumption is reasonable before proceeding.

The logic behind the univariate test for homogeneity of variance in the case of known mean is the same. The likelihood ratio test statistic is

$$\lambda_n = \max_{1<k<n-1}\left(n\log\hat{\sigma}_1^2 - k\log\hat{\sigma}_1^2 - (n-k)\log\hat{\sigma}_n^2 - \frac{n}{2}\right). \tag{12}$$

where

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^{n}(x_i - \mu)^2}{n}, \tag{13}$$

$$\hat{\sigma}_1^2 = \frac{\sum_{i=1}^{k}(x_i - \mu)^2}{k}, \quad \text{and}$$

$$\hat{\sigma}_n^2 = \frac{\sum_{i=k+1}^{n}(x_i - \mu)^2}{n-k}, \tag{14}$$

The asymptotic distribution under H_0 is

$$\lim_{n\to\infty} P(a_n\lambda_n - b_n \leq x) = e^{-2e^{-x}} \quad \text{for all } x \in \mathbf{R}, \tag{15}$$

with

$$a_n = (2\log\log(n))^{\frac{1}{2}} \quad \text{and}$$

$$b_n = \frac{1}{2}\log\log\log(n) + 2\log\log(n) - \log\left(\Gamma\left(\frac{1}{2}\right)\right). \tag{16}$$

In practice, if there is doubt as to whether the large sample distribution of the test statistic is appropriate, critical values can be computed via simulation.

The above test procedures all assume that the mean of the process does not change. If one wishes to test the hypothesis

$$H_0\colon \Sigma_1 = \cdots = \Sigma_n, \boldsymbol{\mu}_1 = \cdots = \boldsymbol{\mu}_n \tag{17}$$

against

$$H_1\colon \Sigma_1 = \cdots = \Sigma_k \neq \Sigma_{k+1} = \cdots \Sigma_n, \boldsymbol{\mu}_1 = \cdots = \boldsymbol{\mu}_k \neq \boldsymbol{\mu}_{k+1} = \cdots = \boldsymbol{\mu}_n, \tag{18}$$

the relevant test statistic as proposed by Chen and Gupta (2000) is

$$\max_{m<k<n-m} (n \log|\widehat{\Sigma}| - k \log|\widehat{\Sigma}_1| - (n-k)\log \widehat{\Sigma}_n)^{\frac{1}{2}}, \tag{19}$$

where

$$\widehat{\Sigma} = \frac{1}{n}\sum_{i=1}^{n}(\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})', \tag{20}$$

$$\widehat{\Sigma}_1 = \frac{1}{k}\sum_{i=1}^{k}(\mathbf{x}_i - \bar{\mathbf{x}}_k)(\mathbf{x}_i - \bar{x}_k)', \widehat{\Sigma}_n = \frac{1}{n-k}\sum_{i=k+1}^{n}(\mathbf{x}_i - \bar{\mathbf{x}}_{n-k})(\mathbf{x}_i - \bar{\mathbf{x}}_{n-k})', \tag{21}$$

$$\bar{\mathbf{x}}_k = \frac{1}{k}\sum_{i=1}^{k}\mathbf{x}_i, \bar{\mathbf{x}}_{n-k} = \frac{1}{n-k}\sum_{i=k+1}^{n}\mathbf{x}_i \tag{22}$$

It is important to note that the individual log-likelihood ratios are unreliable near the ends of the time series and typically produce very high values at near $k = m$ and $k = n - m$. We suggest including only the values roughly between $k = m + 3$ and $k = n - m - 3$ in the maximum above.

The limiting distribution is

$$\lim_{n\to\infty} P(a_n\lambda_n - b_{2m} \leq x) = e^{-2e^{-x}} \quad \text{for all } x \in \mathbf{R}, \tag{23}$$

with

$$a_n = (2\log\log n)^{\frac{1}{2}} \quad \text{and}$$
$$b_{2m} = 2\log\log n + m\log\log\log(n) - \log\Gamma(m). \tag{24}$$

The above procedures are valid in the case where observations are independent between time points. In the presence of autocorrelation, the same analysis can be applied to the process after the autoregressive components are removed (prewhitening). In practice, the components removed will be based on sample estimates of the autoregressive parameters, and the sensitivity of the test to this extra source of variability may need to be explored.

Local change point detection is a stepwise procedure which begins by testing an interval subset of the data for homogeneity and increases the size of the interval until a change point is detected or the interval being tested reaches the length of the entire series. At each stage of the testing procedure, the test statistic is the one outlined above. To begin, a family of intervals $I = \{I_j, j = 0, 1 \cdots\}$ is defined. Each interval is of the form $I_j = [n - m_j, n]$, with m: $m_0 < m_1 < \cdots n$ where n is the length of the series. Beginning with $I = I_0$, the procedure is to test I for homogeneity against the alternative of one change point as above. If the hypothesis of homogeneity is not rejected, the next larger interval is tested until a change point is detected or the largest possible interval is tested. If, for some interval, a change point is detected at some point k, the procedure begins again using intervals of the form $I_j = [k - m_j, k]$. Because multiple tests are being performed, the critical values at each stage are adjusted using the Bonferonni method, which is to replace the significcance level α with α/J where J is the number of tests being performed.

Following Giacomini et al. (2006), we set $m_j = m_0 c^j$, where $c = 1.5$ and $m_0 = 10$. For a time series of a given length n, this will yield J intervals contained in $[1, n]$, which lead to J different tests of homogeneity. To control the probability of H_0 being rejected falsely (type I error) for at least one interval at α, we set the rejection level for each interval at α/J.

The goal of this adjustment procedure is to increase power under some multiple change point alternatives. Imagine a 150-year time series in which there is a change in a parameter θ at years 50 and 100, and that θ has value θ_1 in the intervals [1, 50] and [101, 150] and θ_2 in the interval [51, 100] as shown in Fig. 2. The global approach is to begin by testing the entire series for homogeneity using the test statistic

$$\lambda_n = \max_{m<k<n-m} \lambda_k. \tag{25}$$

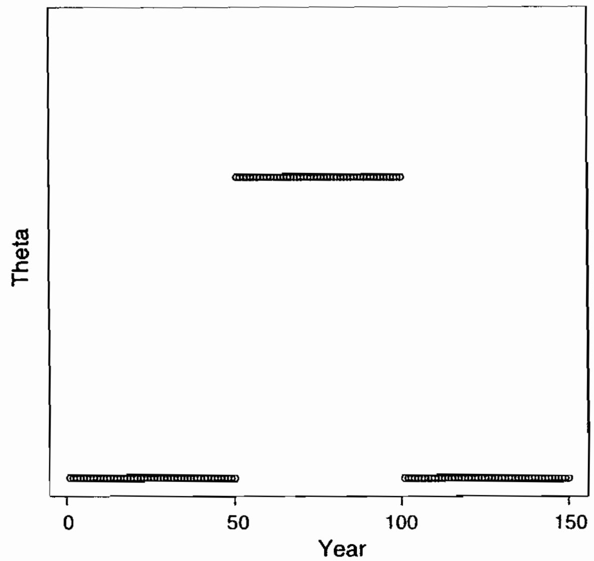

Fig. 2. Local change point detection maybe more powerful than a global test when a shift in any parameter, here designated as theta, is followed by a reversal

The maximum should occur at either year 50 or year 100. Supposing it is at year 50, the test statistic depends on two maximum likelihood estimates, $\hat{\theta}_1$ computed from the years [1, 50], and $\hat{\theta}_2$ based on the years [51, 150]. The size of the test statistic (and thus the probability of rejecting H_0) increases with the difference between $\hat{\theta}_1$ and $\hat{\theta}_2$. $\hat{\theta}_1$ should be close to θ_1, but $\hat{\theta}_2$ will be a compromise between θ_2 and θ_1. If the test for homogeneity were to be performed locally on the interval [1, 100] or [51, 150], the MLEs would not be distorted by the intervals [101, 150] or [1, 50], respectively. A greater difference between $\hat{\theta}_1$ and $\hat{\theta}_2$ should be expected, increasing the probability of rejection.

Disadvantages of the local procedure as compared to the global method include decreased power under single change point alternatives due to the adjusted significance levels, and the somewhat arbitrary nature of the interval selection process, which may influence results. This modified procedure is potentially important in long-term studies of climate variability, where several changes and reversals may be present.

3. Application

Two relationships were examined for a significant change in covariance structure, the ENSO/IMR series and an ENSO/Brazil rainfall series. The latter was studied by Chiang et al. (2000), who found that the generally weak negative correlation peaked in the mid 20th century and, more significantly, after 1980 or so. For the ENSO/IMR series, monthly rainfall totals and Pacific SST observations from 1871 to 2003 were both averaged over the months July to September. For the ENSO/Brazil series, monthly rainfall totals and SST were averaged over the months April to June, from 1856 to 2001. Each of the three individual series was tested for normality and homogeneous mean, and each assumption appears reasonable.

The ENSO series were slightly autocorrelated. The best fitting ARMA model, as chosen

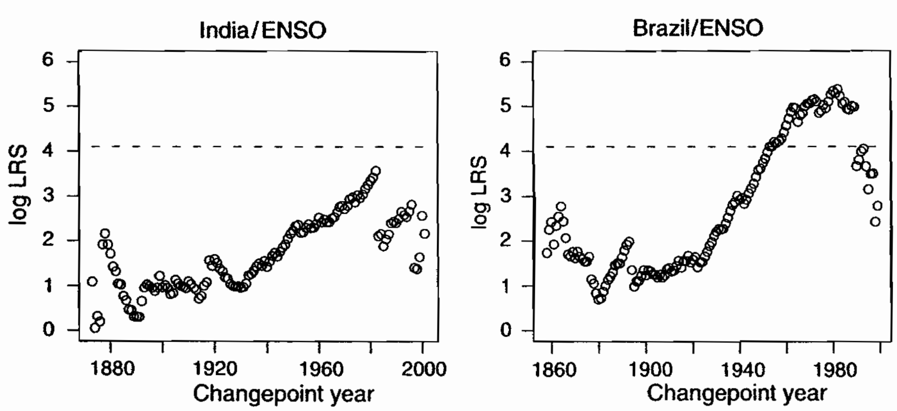

Fig. 3. The likelihood ratio test statistics at each possible change point year. 95% significance is indicated by the dotted line The test statistic for the Brazil series is at a maximum in 1982, with a p-value of less than 1%. The test statistic for the India sereis is maximized in 1980, with a p-value of 0.12. Although the significance of the change point for the India series is less clear than in the Brazilian series case, the similarity between the two series is suggestive

using the Akaike information criterion (AIC) was AR(2). The raw data were tested for change-points, as was a pre-whitened series from which the AR component had been removed. The results were virtually identical for both the raw and pre-whitened data.

The global analysis for the ENSO/IMR series may suggest an event in 1980 with a corresponding p-value of 0.12. Figure 3 shows the graph of the log-likelihood functions versus change point year k. Peaks indicate years where a change point is relatively likely (although not necessarily statistically significant). The dashed line is the critical value at the 5% level of significance. Approximate critical values obtained via simulation rather than the asymptotic distribution of the test statistic give a p-value of 0.14.

The sample covariance matrix in the time period from 1871 to 1980 was

$$\widehat{\Sigma}_1 = \begin{pmatrix} 4.4 & -0.659 \\ -0.659 & 0.27 \end{pmatrix}. \tag{26}$$

From 1980 to 2003 it was

$$\widehat{\Sigma}_2 = \begin{pmatrix} 3.76 & -0.207 \\ -0.207 & 0.404 \end{pmatrix}. \tag{27}$$

The local and global analysis yielded the same conclusions, although in the next section it will be shown that in some situations the results can differ.

The univariate version of the test designed to detect changes in variance was performed on the ENSO series, finding no significant changes in variance. For the ENSO/Brazil series, a significant change ($p = 0.005$) in the covariance matrix was detected in 1982. A test for equality of variance on the Northeast Brazil Rainfall series reveals that there is a shift in the variance of the univariate process which is significant at the 1% level. Thus, there is a significant change in the covariance structure in the ENSO/Brazil relationship, all or part of which can be explained by an increase in the variance of the Brazilian rainfall. The observed covariance matrices were

$$\widehat{\Sigma}_1 = \begin{pmatrix} 0.59 & -0.075 \\ -0.075 & 0.27 \end{pmatrix} \tag{28}$$

pre-1982, and

$$\widehat{\Sigma}_2 = \begin{pmatrix} 2.43 & -0.27 \\ -0.27 & 0.46 \end{pmatrix} \tag{29}$$

from 1982 to 2001. It should be noted that an increase in the variance of the Brazilian rainfall process results in decreased predictability using ENSO, since $\rho_{xy} = \sigma_{xy}/\sigma_x\sigma_y$. This is consistent with the findings of Chiang et al. (2000), although not with the reasons proposed in that paper.

As can be seen in Fig. 3, the first time the likelihood ratio crosses the 5% threshold is around 1960, and it continues to increase until the peak in 1982. Unlike under a sequential analysis framework, the estimated change point is not at the point of first crossing the significance threshold, but rather the point at which the test statistic is maximized, i.e. the MLE for the change point.

4. Power

Simulations were run to assess the power (the probability of rejection when the null hypotheses is false) of the global and local methods under specific alternatives. The power of the global test under one-change point alternatives is assessed using series of 150 simulated bivariate normal observations, the first 75 of which are generated using one covariance matrix, and the last 75 using a different covariance matrix. Thousand series of length 150 are generated and tested for homogeneity. The percentage of simulations in which the null hypothesis is rejected is an estimate of the power of the test. The results from these simulations are shown in Table 1 for $\alpha = 0.05$.

Some findings based on simulations can be stated in a general manner. In a situation of constant variance and changing covariance, the magnitude of the change in covariance must be rather large to achieve reasonable power. If both variance and covariance are changing, power increases with the magnitude of the absolute difference in the determinants of the covariance matrices. The power of the test decreases steadily as the change point approaches the beginning or end of the time series. The power of the global test appears to be greater for 1 change point than for 2 or more.

Because the local test comprises multiple individual hypothesis tests, the interpretation of the p-values is somewhat more difficult. To compare the power of the local test, using the intervals defined in Sect. 2, in comparison to the global

Table 1. The results of simulations to study the power of the change point detection method are above. For each combination of pre and post-change covariance matrices, 1000 simulations of length 150 were created with a change point after 75 observations. The percentage of the 1000 simulations in which the p-value fell below 5% is the observed power of the test

Σ before change	Σ after change	Power	Σ before	Σ after	power
$\begin{pmatrix}1 & 0\\0 & 1\end{pmatrix}$	$\begin{pmatrix}1 & 0.2\\0.2 & 1\end{pmatrix}$	0.08	$\begin{pmatrix}1 & 0.6\\0.6 & 1\end{pmatrix}$	$\begin{pmatrix}1 & 0.2\\0.2 & 1\end{pmatrix}$	0.45
$\begin{pmatrix}1 & 0\\0 & 1\end{pmatrix}$	$\begin{pmatrix}1 & 0.4\\0.4 & 1\end{pmatrix}$	0.27	$\begin{pmatrix}1 & 0.6\\0.6 & 1\end{pmatrix}$	$\begin{pmatrix}1 & 0.4\\0.4 & 1\end{pmatrix}$	0.16
$\begin{pmatrix}1 & 0\\0 & 1\end{pmatrix}$	$\begin{pmatrix}1 & 0.6\\0.6 & 1\end{pmatrix}$	0.81	$\begin{pmatrix}1 & 0.6\\0.6 & 1\end{pmatrix}$	$\begin{pmatrix}1 & 0.6\\0.6 & 1\end{pmatrix}$	0.07
$\begin{pmatrix}1 & 0\\0 & 1\end{pmatrix}$	$\begin{pmatrix}1 & 0.8\\0.8 & 1\end{pmatrix}$	1	$\begin{pmatrix}1 & 0.6\\0.6 & 1\end{pmatrix}$	$\begin{pmatrix}1 & 0.8\\0.8 & 1\end{pmatrix}$	0.35
$\begin{pmatrix}1 & 0\\0 & 1\end{pmatrix}$	$\begin{pmatrix}1 & 1\\1 & 1\end{pmatrix}$	1	$\begin{pmatrix}1 & 0.6\\0.6 & 1\end{pmatrix}$	$\begin{pmatrix}1 & 1\\1 & 1\end{pmatrix}$	1
$\begin{pmatrix}1 & 0\\0 & 1\end{pmatrix}$	$\begin{pmatrix}1.5 & 0\\0 & 1\end{pmatrix}$	0.16	$\begin{pmatrix}1 & 0.6\\0.6 & 1\end{pmatrix}$	$\begin{pmatrix}1.5 & 0\\0 & 1\end{pmatrix}$	0.94
$\begin{pmatrix}1 & 0\\0 & 1\end{pmatrix}$	$\begin{pmatrix}1.5 & 0.245\\0.245 & 1\end{pmatrix}$	0.38	$\begin{pmatrix}1 & 0.6\\0.6 & 1\end{pmatrix}$	$\begin{pmatrix}1.5 & 0.245\\0.245 & 1\end{pmatrix}$	0.72
$\begin{pmatrix}1 & 0\\0 & 1\end{pmatrix}$	$\begin{pmatrix}1.5 & 0.5\\0.5 & 1\end{pmatrix}$	0.83	$\begin{pmatrix}1 & 0.6\\0.6 & 1\end{pmatrix}$	$\begin{pmatrix}1.5 & 0.5\\0.5 & 1\end{pmatrix}$	0.41
$\begin{pmatrix}1 & 0\\0 & 1\end{pmatrix}$	$\begin{pmatrix}1.5 & 0.74\\0.74 & \end{pmatrix}1$	1	$\begin{pmatrix}1 & 0.6\\0.6 & 1\end{pmatrix}$	$\begin{pmatrix}1.5 & 0.74\\0.74 & 1\end{pmatrix}$	0.16
$\begin{pmatrix}1 & 0\\0 & 1\end{pmatrix}$	$\begin{pmatrix}1.5 & 1\\1 & 1\end{pmatrix}$	0.83	$\begin{pmatrix}1 & 0.6\\0.6 & 1\end{pmatrix}$	$\begin{pmatrix}1.5 & 1\\1 & 1\end{pmatrix}$	0.37
$\begin{pmatrix}1 & 0\\0 & 1\end{pmatrix}$	$\begin{pmatrix}1.5 & 1.5\\1.5 & 1\end{pmatrix}$	1	$\begin{pmatrix}1 & 0.6\\0.6 & 1\end{pmatrix}$	$\begin{pmatrix}1.5 & 1.5\\1.5 & 1\end{pmatrix}$	1

test in a multiple change point situation, 100 series were generated using

$$\begin{pmatrix}1 & 0\\0 & 1\end{pmatrix} \tag{30}$$

as the covariance matrix for observations $1 \cdots 50$, and $101 \cdots 150$, and

$$\begin{pmatrix}1 & 0.6\\0.6 & 1\end{pmatrix} \tag{31}$$

for observations $51 \cdots 100$. The observed power in detecting at least one change at a significance level of 5% for the local and global tests were 68% and 55%, respectively, suggesting that the local test is more powerful, 26% more powerful in this case, under some alternatives.

5. Summary and discussion

We have presented a parametric test for retrospective detection of change points in covariance matrices which we have not previously seen in analysis of climate data. The test assumes the observations are multivariate normal and independent in time. A hypothesis of homogeneous covariance is compared to one of at least one change point using a likelihood ratio. If a change point is detected, the data is split at the estimated change point and the two segments are tested for additional change points. The procedure is repeated until no more change points are found. In situations where a shift is followed by a reversal, a more powerful test maybe created by segmenting the data and testing segments of increasing size.

Two series were tested for changes in covariance: ENSO/Indian Monsoon Rainfall and ENSO/Northeast Brazil Rainfall. In the former, the resulting p-value was 0.12. This finding does not lend strong support to the claim that the ENSO/Monsoon relationship has recently changed. If one exists,the most likely year for a change point is 1980. For the ENSO/Northeast Brazil series, a significant change (p-value <0.01)

was detected in 1982. All or part of the latter change can be attributed to a change in the variance of the Brazil series. This finding differs from the conclusion reached by Chiang et al. (2000), who argued that a change in the frequency of strong El Niño is an explanation for a change in the correlation between the two processes. Additional research is necessary to sort out this inconsistency.

The proposed method is designed to detect abrupt shifts in the probability distributions of the observed processes, but obviously in some situations inhomogeneities would be better modeled by continuous trends. Sveinsson and Salas (2003) explore probability models for climate processes in the presence of shifts, trends and oscillatory behavior. Regression methods can be used to detect and model trends in the mean of a process, and the evolution of variance can be modeled using ARCH (autoregressive conditional heteroskedastic) or GARCH (generalized autoregressive conditional heteroskedasticity, see Bellerslev 1986) methodology. When trends are not constant over the entire observed record, a change point framework may still be needed to detect the beginnings, ends or reversals of trends. Likelihood ratios could be constructed in the above manner, with regression or ARCH parameters as the quantities of interest.

The interconnection between changes in the mean and variance of the distribution makes inference more difficult when both types of inhomogeneity exist. Changes in mean can disguise changes in variance and vice versa. The procedure outlined above is constrained to detect only simultaneous shifts in mean and variance, and may not perform well in other situations. A Bayesian approach in which uncertainties in both location and variance are addressed separately would be useful in creating a more flexible, realistic model.

The test used in this analysis is designed for a fixed sample size and does not assume *a priori* that any period in the observed record is without changes. Alternatively, when a stable reference period is available and the aim is to detect changes as new data is accumulated, methods from statistical quality control, such as sequential probability ratio test (SPRT) or cumulative sum (CUSUM), procedures, can be employed. A review of recent developments in using control charts for monitoring covariance matrices can be found in Yeh et al. (2006). If the assumption of known starting values for the parameters of interest is added to the analysis, a more powerful test maybe available.

Acknowledgments

Lucy Robinson and Victor de la Peña acknowledge support from the US National Science Foundation, through the IGERT Joint Program in Applied Mathematics and Earth and Environmental Science at Columbia University (DGE-02-21041), and through grant DMS-05-05949. Victor de la Peña and Yochanan Kushnir acknowledge support from Initiatives in Science and Engineering (ISE) through the Vice-Provost for Research of Columbia University. The authors thank Upmanu Lall for his helpful comments.

References

Bollerslev T (1986) Generalized autoregressive conditional heteroscedasticity. J Econometrics 31: 307–327

Chen J, Gupta AK (2000) Parametric statistical changepoint analysis. Birkhäuser Press, Boston

Chiang J, Kushnir Y, Zebiak S (2000) Interdecadal changes in eastern Pacfic ITCZ variability and its influence on the Atlantic ITCZ. Geophys Res Lett 27: 3687–3690

Gershunov A, Schneider N, Barnett T (2001) Low-frequency modulation of the ENSO-Indian monsoon rainfall relationship: signal or noise? J Climate 14: 2486–2492

Giacomini E, Hardle WK, Ignatieva E, Spokoiny V (2006) Inhomogeneous dependency modeling with time varying copulaeling with time varying copulae. SFB 649 Discussion Papers Humboldt University, Berlin, Germany

Krishnamurthy V, Boswani BN (2000) Indian Monsoon-ENSO relationship on interdecadal timescale. J Climate 13: 570–595

Kumar KK, Rajagopalan B, Cane M (1999) On the weakening relationship between the Indian monsoon and ENSO. Science 284: 2156–2159

Lehmann EL (2000) Testing statistical hypotheses. Springer, New York

Mercurio D, Spokoiny V (2004) Statistical inference for time-inhomogeneous volatility models. Ann Stat 32: 577–602

Parthasarathy B, Kumar KR, Munot AA (1991) Evidence of secular variations in Indian monsoon rainfall-circulation relationships. J Climate 4: 927–938

Sveinsson OGB, Salas JD, Boes DC, Pielke RA (2003) Modeling the dynamics of long-term variability of hydroclimatic processes. J Hydrometeorol 4: 489–506

Yeh A, Lin D, McGrath R (2006) Multivariate control charts for monitoring covariance matrix: a review. Quality Technology Quantitative Management 3: 415–436