Tellus (1997), 49A, 513–527
Printed in UK – all rights reserved

Copyright © Munksgaard, 1997
TELLUS
ISSN 0280–6495

Correlation of spatial climate/weather maps and the advantages of using the Mahalanobis metric in predictions

By D. B. STEPHENSON*, *Météo-France CNRM, 42 Avenue Gaspard Coriolis, 31057 Toulouse Cedex, France*

(Manuscript received 27 December 1996; in final form 16 May 1997)

ABSTRACT

The skill in predicting spatially varying weather/climate maps depends on the definition of the measure of similarity between the maps. Under the justifiable approximation that the anomaly maps are distributed multinormally, it is shown analytically that the choice of weighting metric, used in defining the anomaly correlation between spatial maps, can change the resulting probability distribution of the correlation coefficient. The estimate of the numbers of degrees of freedom based on the variance of the correlation distribution can vary from unity up to the number of grid points depending on the choice of weighting metric. The (pseudo-)inverse of the sample covariance matrix acts as a special choice for the metric in that it gives a correlation distribution which has minimal kurtosis and maximum dimension. Minimal kurtosis suggests that the average predictive skill might be improved due to the rarer occurrence of troublesome outlier patterns far from the mean state. Maximum dimension has a disadvantage for analogue prediction schemes in that it gives the minimum number of analogue states. This metric also has an advantage in that it allows one to powerfully test the null hypothesis of multinormality by examining the second and third moments of the correlation coefficient which were introduced by Mardia as invariant measures of multivariate kurtosis and skewness. For these reasons, it is suggested that this metric could be usefully employed in the prediction of weather/climate and in fingerprinting anthropogenic climate change. The ideas are illustrated using the bivariate example of the observed monthly mean sea-level pressures at Darwin and Tahitifrom 1866–1995.

1. Introduction

When predicting the weather and climate, it is often necessary to have an objective measure of similarity for comparing two spatial weather or climate maps. Numerous approaches have been used to define such similarity measures, and they can be broadly categorised into three classes: statistical, dynamical, and geographical. The statistical approach often uses correlations and root mean square measures, the dynamical approach uses physical quantities such as energy and potential vorticity believed to be quasi-conserved, and the geographical approach compares quantities deemed to be of human interest such as rainfall averaged over a certain geographical region. This article presents an analytical statistical approach to this problem and suggests a similarity measure which may be useful in future prediction studies. In weather and short-term climate prediction, skill is often judged by the spatial correlation (on the grid) between the predicted and the observed maps at various lead times (Miyakoda et al. 1972; Simmons et al. 1995; Palmer and Anderson 1994; and references therein). And in anthropogenic climate change studies, model predictions are also now being used as fingerprints which are then correlated with recent observed interdecadal climate changes (Barnett and Schlesinger 1987; Karoly et al. 1994; Santer et al. 1995; Tett et al.

e-mail: David.Stephenson@meteo.fr

1996; and references therein). In both these approaches, in order to evaluate the significance of the correlation, it is necessary to know how much correlation could occur solely by chance. In other words, in assessing predictive skill, it is necessary to have some idea about the probability distribution of the correlation coefficient. This is the focus of this study where it is shown that the distribution is dependent on the choice of metric used in defining the correlation coefficient and that this choice has some important implications for predictability.

There is an extensive literature on similarity measures especially in the context of the predictability of Northern Hemisphere (NH) extra-tropical atmospheric flow patterns. Lorenz (1969) applied a root mean square (RMS) measure to the geopotential height in order to search in past flow patterns for close states, known as analogues. Using such analogues, it was hoped that it would be possible to use their evolution to predict the future evolution of the current state. However, Lorenz (1969) found that the probability distribution of the RMS distances was best approximated by a χ^2 distribution with 44 degrees of freedom (d.o.f.), and hence there was only a very small chance of a state in such a high dimensional space being close to the current state, represented by the origin. Gutzler and Shukla (1984) managed to increase the number of analogues by spatial and temporal filtering of the data, yet such states gave poor predictions perhaps partly due to the loss of useful information caused by the filtering. They also tried two other similarity measures, COV and COR, defined as the anomaly pattern COVariance and CORrelation, respectively. Horel (1985) used a COR measure in his persistence study and from the variance of the COR he deduced the number of *statistically independent grid points* to be between 20–37 depending on the filtering. This method of calculating the *spatial degrees of freedom* was also applied to observed and model data by van den Dool and Chervin (1986). Toth (1991a) performed an objective comparison of 9 different similarity measures looking at how many analogues they gave and how good the predictions were using such analogues. He concluded that the RMS measure produced more analogues than the COR measure but that all the measures gave *practically no performance differences* when used to make analogue predictions (albeit slightly better results were obtained when using the difference in the gradient of height.) Wallace et al. (1991) found that with the COV, COR and RMS measures, the probability density of the measure was was well-fitted to distributions having about 20 degrees of freedom. Fraedrich et al. (1995) developed a general expression for estimating the spatial d.o.f. by relating the variance of the COR measure to the eigenvalues of the the correlation matrix. This was then used to investigate the seasonal variation of the spatial d.o.f.. A comprehensive review of estimating spatial d.o.f. by spatial and temporal embedding methods can be found in Toth (1995).

In this study, a generalised definition is made of the covariance measure by introducing an arbitrary Euclidean weighting metric in the definition. The probability distribution for the correlation coefficient is then derived analytically assuming a null hypothesis of multinormally distributed anomaly patterns. The distribution and its moments are shown to be metric dependent. Interesting extremal properties are shown to occur when the metric is the (pseudo-)inverse of the covariance matrix. Some generalised estimates of the number of spatial degrees of freedom are also proposed. The same analytical approach can also be applied to the generalised RMS similarity measure and gives similar conclusions at the expense of more complicated mathematical expressions. For this reason, we prefer to focus on the covariance measure in this study.

2. Basic definitions

Define the anomaly pattern/map at time t_k by the p-vector of variables:

$$\mathbf{z}_k = \begin{pmatrix} z_1(t_k) \\ z_2(t_k) \\ \cdots \\ z_p(t_k) \end{pmatrix}, \tag{1}$$

where the p variables are the anomalies at each of the p grid points. For example, $\mathbf{z}_k$ could represent a map of 500 hPa geopotential height anomalies at time t_k. Non-scalar vector and matrix symbols appear in boldface in this article. The sample mean over n sample times is denoted by

the overbar:

$$\bar{z} = \frac{1}{n} \sum_{k=1}^{n} z_k, \tag{2}$$

and is zero for the anomalies which are by definition centered. The $p \times p$ sample anomaly covariance matrix is defined as $\boldsymbol{S} = \overline{\boldsymbol{z}\boldsymbol{z}^{\mathrm{T}}}$, where T denotes the matrix transpose operation.The sample covariance matrix, $\boldsymbol{S}$, is an estimate of the population (true) covariance matrix, $\boldsymbol{\Sigma}$, and is frequently used in climate variability studies since it contains much information about the spatial structure of the major modes of variability (teleconnections). If the variables are first scaled (reduced) by dividing by their standard deviations, the sample covariance becomes the sample correlation matrix with unity along the diagonal. In what follows, we will refer to covariances with the understanding that we mean correlations if the data have been pre-scaled by their standard deviations. The generalised Anomaly pattern Correlation Coefficient (ACC), between anomaly patterns at times t_k and t_l, is defined as the weighted correlation

$$R_{kl} = \boldsymbol{z}_k^{\mathrm{T}} \boldsymbol{M} \boldsymbol{z}_l, \tag{3}$$

where $\boldsymbol{M}$ is a symmetric weighting matrix defining the metric. Commonly used choices for $\boldsymbol{M}$ have been the covariance metric $\boldsymbol{M} = \boldsymbol{I}$ (COV) and the correlation (scaled covariance) metric $M_{ij} = (S_i S_{jj})^{-1/2}$ (COR). The covariance metric has the property that the ACC is dominated by regions with large temporal variance. The correlation metric alleviates this problem but has the disadvantage of sometimes introducing more noise from regions with low temporal variance. A metric which has not been widely used in ACC studies but which is widely used in the study of multivariate statistics is the Mahalanobis metric $\boldsymbol{M} = \boldsymbol{S}^{-1}$ (MAH) (Mardia et al., 1979). This positive definite metric is equivalent to performing the Euclidean scalar product of the principal components (PCs), after all have been scaled to have unit variance. Under non-singular linear transformations of the variables, $\boldsymbol{z} \rightarrow \boldsymbol{z}' = \boldsymbol{A}\boldsymbol{z}$, the Mahalanobis ACC has the virtue of remaining invariant: $\boldsymbol{z}_k'^{\mathrm{T}} \boldsymbol{S}_{z'}^{-1} \boldsymbol{z}_l' = \boldsymbol{z}_k^{\mathrm{T}} \boldsymbol{S}_z^{-1} \boldsymbol{z}_l$ since $\mathbf{S}_z' = \boldsymbol{A}\boldsymbol{S}_z\boldsymbol{A}^{\mathrm{T}}$. For example, the numerical value of the Mahalanobis ACC remains constant even if the data is linearly interpolated onto another grid having the same number of grid points. This is not the case when using other metrics and led Zwiers (1987) to conclude that simple tests based on non-invariant Euclidean ACCs were capable of being manipulated. The inverse of the covariance matrix is used ubiquitously as a weighting factor (the Gauss-Markov weights) in ordinary least squares inverse problems such as in the assimilation of observed data into atmosphere and ocean models (Menke, 1989; Thacker and Lewandowicz, 1996).

3. Probability distribution of the anomalies

The probability distribution of the ACC depends on the p-dimensional probability density function (p.d.f.) of the anomalies, assuming of course that such a unique probability density function exists. This may not be the case due to intransitivity and non-stationarity in climate data but, for the sake of simplicity, as in many previous studies, a unique probability density function will be assumed to exist. For typical global climatic data on a 5° grid, the phase space dimension, p, would be of the order of 10^3 if all the grid points were independent. However, not all the grid points are independent in the sense that there exist non-zero spatial correlations between the variables at different grid points, and hence the total number of d.o.f. is much less. For example, NH wintertime daily 500 hPa geopotential height data has been estimated to have about 20–40 d.o.f. (independent dimensions). Because of the high dimensionality of the phase space, an impossibly long time series is necessary to visit all the space and thereby estimate the p.d.f.. For example, to estimate the p.d.f. in 10 bins for each of (say) 30 dimensions of the daily NH 500 hPa geopotential height data, requires at least 10^{30} sampling times, or 10^{17} times the age of planet Earth for daily data! Hence, it is necessary to make a null hypothesis about the distribution of the anomalies.

A reasonable and tractable null hypothesis, is that the anomalies are distributed multinormally with a p.d.f. given by:

$$\rho[z] = (\det 2\pi \boldsymbol{\Sigma})^{-1/2} \exp\left(-\frac{1}{2} z^{\mathrm{T}} \boldsymbol{\Sigma}^{-1} z \right) \tag{4}$$

where $\det 2\pi\boldsymbol{\Sigma}$ is the matrix determinant of $2\pi\boldsymbol{\Sigma}$. This distribution is widely used in making inferences and testing hypotheses in multivariate statistics (Mardia et al. 1979).The universality of the

multinormal distribution is partly explained by themultivariate generalisation of the central limit theorem which states that the limiting result of many independent multivariate processes yields a multinormal distribution. A good representation of this is afforded by the class of stochastic climate models proposed by Hasselmann (1976) in which a linear system is forced by Gaussian random shocks. Such systems generate Gaussian red noise whose probability distribution is multinormal. The key idea is that the system has many degrees of freedom which interact only weakly with one another, as is the case in many physical systems in statistical equilibrium. Although there are regions of the planet having strong correlations of climate/weather with elsewhere, many regions are only weakly correlated especially if one considers correlations only over short forecast times. Statistical mechanics could therefore be appropriate for explaining some aspects of climate/weather variability. The multinormal distribution is completely determined by the first and second moments, and if fit in this way, perfectly reproduces sampled second order moments such as the energy spectrum and the teleconnections.

Previous studies have assumed that the grid point anomalies were normally and independently distributed which is equivalent to making the assumption that the covariance matrix is diagonal (e.g., Wallace et al. 1991). For NH data, the existence of non-local correlations between the anomalies at differing grid points (Wallace and Gutzler, 1981) implies that the covariance matrix is not diagonal and invalidates the null hypothesis used in these previous studies. Due to the presence of teleconnections, the anomalies at different grid points are correlated and therefore it is not the grid point variables but the principal components which are uncorrelated and independently distributed. The teleconnection structure should be taken into account when discussing the correlations of weather and climate maps.

In pioneering studies, Toth (1991b) and Wallace et al. (1991) started to investigate whether the anomalies in NH daily 500 hPa geopotential height were Gaussian distributed (i.e. multinormal). By correlating the RMS distances between anomaly patterns with their absolute distances from the mean state of zero anomaly, Toth (1991b) concluded that the wintertime extra-tropical NH geopotential height data was *phase-averaged multinormal* and then used this to deduce certain fundamental consequences for prediction. The multinormal p.d.f. is a Gamma function of the squared radial distance from the origin in PC space and the radial symmetry can be clearly seen in Figures 13 and 14 of Marshall and Molteni (1993) which depict the p.d.f. of model generated 500 hPa geopotential height as a function of the two leading PCs. Wallace et al. (1991) were interested in investigating the existence of multi-modality in the probability distribution, which they did by examining the distributions of COV, COR and RMS measures of similarity. They concluded that they did not see evidence of *outright* bimodality but that they did see the existence of skewness in the distribution which was suggested to be associated with persistent blocking anomalies. The probability distribution of the NH extra-tropical wintertime height field is still very much a matter of open debate. It will be shown later how the moments of the ACC calculated using the Mahalanobis metric can be used a posteriori to test the multinormal null hypothesis.

4. Generalised inverse covariance: the pseudo-inverse

Because of the existence of spatial correlations having well-defined scales, such as the Rossby scale in mid-latitudes, there are typically less degrees of freedom than grid points. Hence, not all the rows in the $p \times p$ covariance matrix can be independent and the matrix is therefore rank deficient with a rank less than the dimension of the matrix ($r < p$). Such a matrix has a non-trivial null space and no unique inverse exists — the population covariance matrix is singular*. The multinormal null hypothesis requires the existence of an inverse covariance matrix despite the fact that in general a unique inverse does not exist. Fortunately, there is a simple solution to the problem by truncating the system to the complement of the null space. By the spectral decomposition theorem, the real symmetric covariance

* Furthermore, the rank of the sampled covariance matrix cannot exceed the number of independent samples used to construct the matrix ($r < n$), and in general $n < p$ and this is therefore an additional cause of rank-deficiency.

matrix can be written as $\mathbf{\Sigma}=\boldsymbol{Q}\mathbf{\Lambda}\boldsymbol{Q}^{\mathrm{T}}$ where $\boldsymbol{Q}$ is an orthogonal matrix and $\mathbf{\Lambda}$ is a diagonal matrix with real positive eigenvalues $\{\lambda_1, \lambda_2, \ldots, \lambda_p\}$ ranked in order of magnitude: $\lambda_1 \geqslant \lambda_2 \ldots \geqslant \lambda_p$. Because of the rank deficiency, only the leading r eigenvalues are non-zero, whereas the trailing $p-r$ are zero and give rise to the non-invertibility. The pseudo-inverse, $\mathbf{\Sigma}^-$ is obtained by truncating the space to the first r components, and is given by $\mathbf{\Sigma}^-=\boldsymbol{Q}\mathbf{\Lambda}^-\boldsymbol{Q}^{\mathrm{T}}$ where $\mathbf{\Lambda}^-=diag(\lambda_1^{-1}, \lambda_2^{-1}, \ldots, \lambda_r^{-1}, 0, 0, \ldots, 0)$. The pseudo-inverse satisfies the Moore-Penrose conditions $\mathbf{\Sigma}\mathbf{\Sigma}^-\mathbf{\Sigma}=\mathbf{\Sigma}$ and gives the minimum length solution in inverse problems (Menke 1989). The multinormal distribution using the pseudo-inverse is referred to as the *singular multinormal distribution* (Mardia et al., 1979).

This procedure corresponds to working in principal component space truncated to the first r leading modes and can be considered as a data-adaptive pre-filtering of the data (Barnett and Preisendorfer, 1987). The r leading PCs are associated with eigenvalues distinguishable from zero and represent the *signal*, whereas the trailing $(p-r)$ modes have eigenvalues statistically consistent with being zero and can be considered to represent *noise*. In general, the trailing eigenvalues are never numerically zero, and so it is necessary to have an algorithm for determining the rank r. A computational approach is to reject eigenvalues which fall below a tolerance determined by the machine accuracy. A review of more statistical approaches can be found in Jolliffe (1986). Perhaps the most reliable methods for efficiently separating the subspaces are the jackknife methods and their analytic extensions (Besse, 1992). In what follows, we assume that the p-dimensional space has been truncated in this manner so that the covariance matrix is non-singular and has $r \times r$ elements — in other words $p=r$ and the inverse of the covariance matrix has been replaced by its pseudo-inverse.

5. Probability distribution of the ACC

This section presents an analytic derivation of the probability distribution of the generalised anomaly correlation coefficient, defined using the arbitrary weighting metric $\boldsymbol{M}$ (not necessarily equal to the inverse covariance matrix, $\boldsymbol{S}^{-1}$). The low order moments of the distribution will be discussed and a method for testing the multinormal null hypothesis a posteriori will be presented.

5.1. *Analytical derivation*

To calculate analytically the probability distribution of the generalised anomaly correlation coefficients, it will be assumed that the anomaly patterns are multinormally distributed. In the following calculation, it is important to distinguish between the p.d.f. of correlations made simultaneously $\{R_{kk};\ k=1, 2, \ldots\}$ and the correlations made at different times $\{R_{kl};\ k, l=1, 2, \ldots;\ k \neq l\}$. As in Wallace et al. (1991), it will be assumed that there are an infinite number of serially uncorrelated samples (the asymptotic limit $n \to \infty$). Instead of considering the p.d.f, it is more convenient to work with its Fourier transform

$$\phi(t)=\int \mathrm{e}^{itR}\rho(R)\mathrm{d}R=E(\mathrm{e}^{itR}) \tag{5}$$

referred to as the characteristic function (c.f.). The expectation operator, denoted by $E()$, is given by the multiple integral over z space using the multinormal distribution as the measure:

$$E(\psi)=(\det 2\pi\,\mathbf{\Sigma})^{-1/2}\int \ldots \int \times\left(\prod_{r=1}^{p}\mathrm{d}z_r\right)\exp\left(-\frac{1}{2}\,z^{\mathrm{T}}\,\mathbf{\Sigma}^{-1}\,z\right)\psi. \tag{6}$$

For simultaneous correlations, R_{kk}, the p-dimensional integration can be performed by changing variables and then using the identity

$$\int \ldots \int\left(\prod_{r=1}^{p}\mathrm{d}z_r\right)\exp\left(-\frac{1}{2}\,z^{\mathrm{T}}\boldsymbol{A}^{-1}z\right)=(\det 2\pi\boldsymbol{A})^{1/2} \tag{7}$$

to give $\phi(t)=(\det(1-2it\mathbf{\Sigma}\boldsymbol{M}))^{-1/2}$. Using the relationship $\log\det \boldsymbol{A}=\mathrm{Tr}\log \boldsymbol{A}$, this can be rewritten as:

$$\phi(t)=\exp\left(-\frac{1}{2}\,\mathrm{Tr}\log(\boldsymbol{I}-2it\mathbf{\Sigma}\boldsymbol{M})\right) \tag{8}$$

and for small t, the logarithm can be expanded about unity to give

$$\phi(t)=\exp\left(+\frac{1}{2}\sum_{m=1}^{\infty}\frac{\beta_m(2it)^m}{m}\right), \tag{9}$$

where the coefficients are defined as $\beta_m=\mathrm{Tr}(\mathbf{\Sigma}\boldsymbol{M})^m$

and are all positive when the metric $\boldsymbol{M}$ is positive definite. The cumulants are given by $\kappa_m = 2^{m-1}(m-1)!\beta_m$.

For the case of maps at different times, R_{kl} is a bilinear combination of the anomalies z_k and z_l, and it is necessary to perform a double integration over the anomaly fields z. By completing the square in the exponent, changing variables and using the previous identities, the $2p$-dimensional integration can be performed to give

$$\phi(t) = \exp\left(-\frac{1}{2} \operatorname{Tr} \log(\boldsymbol{I} + t^2 \boldsymbol{\Sigma M \Sigma M}) \right), \tag{10}$$

which results in

$$\phi(t) = \exp\left(+\frac{1}{2} \sum_{m=1}^{\infty} \frac{\beta_{2m}(it)^{2m}}{m} \right), \tag{11}$$

when the logarithm is expanded about unity.

5.2. *A special case: the Mahalanobis metric*

For the Mahalanobis case, the matrix product $\boldsymbol{\Sigma M}$ is equal to $\boldsymbol{\Sigma S}^{-1}$ and tends to the identity matrix in the asymptotic limit, where the unbiased sample covariance becomes equal to the population covariance ($\boldsymbol{S} \to \boldsymbol{\Sigma}$). In the asymptotic limit, the MAH metric is unique in that all the coefficients have the same value $\beta_m = \operatorname{Tr}(\boldsymbol{I})^m = p$, and this allows the p.d.f. of the correlation coefficient, R_{kk}, to be written in the closed form

$$\rho(R_{kk}) = \frac{1}{2\pi} \int \frac{e^{-itR} dt}{(1-2it)^{p/2}}. \tag{12}$$

This is the p.d.f. of a χ^2 distribution having p degrees of freedom, as to be expected for the Mahalanobis case, where R_{kk} becomes the sum of the squares of p independent normally distributed variables (the PCs). Furthermore, this is the asymptotic limit of the Hotelling T^2 distribution when the number of samples tends to infinity (Kendall et al. 1983).

The distribution of the correlation coefficient, R_{kl} is less tractable but can be written in the closed form

$$\rho(R_{kl}) = \frac{1}{2\pi} \int \frac{e^{-itR} dt}{(1+t^2)^{p/2}}, \tag{13}$$

which can be evaluated analytically for every value of p. For $p=1$, the solution is $\rho(R) = K_0(R)/\pi$ where K_0 is the modified Bessel function of the second kind of zero order, also known as the Macdonald function (Bronshtein and Semendyayev, 1985; Page 584). This is singular as $R \to 0$ and asymptotes to $e^{-R}/\sqrt{R}$ for large R. For $p=2$, Cauchy integration yields the solution $\rho(R) = e^{-|R|}/2$. For larger values of p, the distribution becomes less exponential and more Gaussian and tends to $\rho(R) = (2\pi p)^{-1/2} \exp(-R^2/2p)$ — a normal distribution with zero mean and variance equal to the dimension p.

5.3. *Product moments of the correlation R_{kl}*

A probability density function can often be usefully characterised by its low order moments which are obtainable from the derivatives of the c.f. as follows

$$E(R^m) = (-i)^m \left[\frac{d^m \phi}{dt^m} \right]_{t=0}. \tag{14}$$

Under the null hypothesis of multinormality, an anomaly pattern z_k and its antilogue $-z_k$ are equally probable and hence $\rho(R_{kl})$ is perfectly symmetric about the origin and all its odd moments are by definition zero. This in turn implies that the Maclaurin expansion of the c.f. must contain only even powers of t as can be seen to be the case in eq. (11). In this section, we will discuss the second and fourth moments of R_{kl} which from eqs (11) and (14) are given by $E(R_{kl}^2) = \beta_2$ and $E(R_{kl}^4) = 3\beta_2^2 + 6\beta_4$. The moments depend on the constants $\{\beta_m\}$ and hence on the eigenvalue spectrum of the matrix product $\boldsymbol{\Sigma M}$.

The variance of the ACC distribution, $E(R_{kl}^2) = \beta_2$, has been previously used to estimate the number of degrees of freedom in NH geopotential height data. In Wallace et al. (1991), the number of d.o.f., N_D, was estimated with a COV metric by using $E(R^2) = \bar{s}^4 N_D^{-1}$ and was found to be 20. $\bar{s}^2$ is the square of the time mean of the spatial standard deviation and is used as a scale to make the covariances dimensionless. The null hypothesis in Wallace et al. (1991) is that the different grid points have anomalies which are independently normally distributed with variances all equal to $\bar{s}^2$. Under this hypothesis, one can write $\bar{s}^2 = E(\boldsymbol{z}_k^{\mathrm{T}} \boldsymbol{M} \boldsymbol{z}_k)$ which can be re-expressed as $\bar{s}^2 = \operatorname{Tr}(\boldsymbol{\Sigma M}) = \beta_1$. Substituting this into the expression for N_D then gives the more general expression

$$N_D = \beta_1^2 / \beta_2. \tag{15}$$

It can be seen that the dimension estimate depends on the eigenvalue spectrum of the ΣM matrix and hence depends on the choice of metric. The estimate can give values from 1 to p depending on the choice of metric. With the MAH metric in the asymptotic limit, the matrix $\Sigma M = I$ has a perfectly flat eigenvalue spectrum and hence $N_D = p$. For a singular metric with ΣM having only one non-zero eigenvalue, N_D becomes 1. For COV or COR metrics, N_D is a measure of the flatness of the eigenvalue spectrum of the covariance or correlation matrix respectively. For the COR metric, $\mathrm{Tr}(\Sigma M) = \mathrm{Tr}(C) = p$ and $\mathrm{Tr}(\Sigma M \Sigma M) = \mathrm{Tr}(C^2)$ where C is the correlation matrix and this yields $N_D = p^2/\mathrm{Tr}C^2$ which is the generalised eq. (2.5) in Fraedrich et al. (1995).

Kurtosis is a measure of the flatness of a distribution and its deviation from 3 is a measure of how many values lie outside the Gaussian distribution. The kurtosis is defined as $E(R^4)/(E(R^2))^2 = 3 + 6\beta_4\beta_2^{-2}$ and can be shown to be between $3 + 6p^{-1}$ and 9 for positive definite metrics. For all positive definite metrics, the kurtosis exceeds 3 and the distribution is more peaked than normal (keptokurtic) — minimal kurtosis occurs for the Mahalanobis metric and tends to 3 for large p. The deviation of the kurtosis from 3 can be used to define a kurtosis dimension estimate, $N_D^{(2)} = \beta_2^2/\beta_4$ where $N_D^{(2)}$ is in the range 1 to p. For example, the COV ACC frequency distribution in column 4 of Table 1 of Wallace et al. (1991), derived from 500 hPa wintertime geopotential height data on a grid with 445 grid points, has a kurtosis of 3.364 which exceeds the minimum bound of $3\frac{6}{445}$ obtained with the Mahalanobis metric assuming the upper bound of $p = 445$. The kurtosis dimension estimate then gives $N_D^{(2)} = 16.5$ which is close to the dimension estimate of $N_D = 20$ calculated using the variance of the distribution. A distribution with a large kurtosis is characterised by having non-negligible numbers of extreme events lying several standard deviations from the mean (outliers), and such events can be difficult to predict and can have a detrimental effect on the the average predictive skill (Simmons et al. 1995). The number of outliers depend inversely on the dimensionality of the phase space since, as has been shown, the deviation of the kurtosis from 3 is inversely related to the kurtosis dimension estimate. This could be one of the reasons why reduction of the dimension by filtering and limiting the spatial domain in Gutzler and Shukla (1984) led to poorer predictive skill.

5.4. *Generalised dimension estimates*

The estimate of the number of d.o.f. can be generalised to

$$N_D^{(m)} = \beta_m^2/\beta_{2m}, \tag{16}$$

where $m = 1, 2, 3, \ldots$ and $m = 1, 2$ give, respectively, the variance and kurtosis estimates previously discussed. All these estimates are in the range 1 to p and by the spectral decomposition of ΣM can be shown to satisfy $N_D^{(m+1)} \geqslant N_D^{(m)}$ for all m. Maximum dimensions are obtained for the Mahalanobis metric which has $N_D^{(m)} = p$ for all m. For the COR metric, one obtains $N_D^{(m)} = (TrC^m)^2/\mathrm{Tr}C^{2m}$. It may be of interest to use some of these estimates to corroborate the previous variance estimates of dimension. However, it should be remembered that the trace of any power of the correlation/covariance matrix is only invariant under orthogonal rotations of the variables in p-space, transformations where $AA^T = I$, and hence these estimates of dimension are not invariant under non-singular linear transformations of the data. For example, linear interpolations from one grid to another can cause these estimates to change. Only the Mahalanobis metric gives a dimension estimate, equal to the rank of the covariance matrix, which remains invariant under all invertible linear transformations. Without resorting to topological approaches, the most invariant measure of dimension is that obtained by estimating the rank of the covariance matrix using the procedures described in Jolliffe (1986). The rank of the covariance matrix is of use in deciding how many leading PCs are important and it therefore merits further attention in observational and modelling studies.

5.5. *Testing the multinormal null hypothesis*

The null hypothesis of multinormally distributed anomalies was made in deriving the distribution of the ACC. This hypothesis can be tested a posteriori by comparing the higher moments of the ACC distribution with the expressions obtained using eqs. (9), (11) and (14). For the MAH metric, such tests have been proposed by Mardia (1970) and have been extensively used in

many of the sciences dealing with multivariate data. The asymmetry of the p.d.f. about the origin, as measured by the deviation of the third moment, $E(R_{kl}^3)$, from zero, provides a powerful test of the null hypothesis about the multinormality of the anomalies. Wallace et al. (1991) examined the skewness of the ACC distribution and concluded using a univariate test that it was very skewed. In testing the significance, no account was taken of the spatial correlations which exist between the variables at different grid points. When the Mahalanobis metric is used, the third moment of the ACC becomes identically equal to the multivariate skewness defined by Mardia (1970): $E(R_{kl}^3)=b_{1,p}$. This statistic has the useful property of being invariant under non-singular linear transformations of the anomalies and is widely used as a powerful test of multinormality (Mardia et al. 1979). For $p=1$, it reproduces the usual univariate skewness test of normality: $b_{1,1}=m_3^2/m_2^3$. The third moment of the ACC provides a good test of multivariate skewness, if and only if, the Mahalanobis metric is used. Mardia (1970) also proposed a multinormality test making use of the multivariate kurtosis $b_{2,p}$ which is identical to $E(R_{kk}^2)$ when the MAH metric is used. Such powerful tests can be shown to guarantee certain robustness for normal theory tests on means and covariances which is not the case for more ad hoc tests such as that of phase-averaged multinormality (Toth 1991b, 1995). It is of interest to use these tests to quantify the deviation from multinormality of climate anomalies as will be demonstrated in the following example.

6. A bivariate example: the Southern Oscillation index

Although the methodology is applicable to systems with any number of variables, it will be illustrated here using the bivariate $p=2$ contracted description of ENSO provided by the observations of sea-level pressure at Darwin and Tahiti from 1866–1995.

6.1. *Description of the data*

The SLP data from January 1866 to June 1991 are used to define the University of East Anglia, Climate Research Unit, Southern Oscillation Index (CRU-SOI). The period from July 1991 to May 1995 was completed using SLP records obtained from the NOAA Climate Analysis Center published in the Climate Diagnostics Bulletin. The series and its extensions are discussed in the articles of Ropelewski and Jones (1987) and there is speculation about the homogeneity or non-stationarity of the Tahiti SLPs in the mid 1930s (Elliott and Angell, 1988). The data have been used in previous ENSO studies — Chen (1982), Trenberth (1984) (hereafter denoted as T84), Wright (1985), Trenberth and Shea (1987); and numerous other studies.

6.2. *Defining interannual anomalies*

Detrended interannual anomalies, having zero mean and unit variance, have been constructed by (a) taking the centered 12 month running mean of the original series, (b) detrending it by subtracting its 121-month running mean, and (c) standardising by subtracting the mean and dividing by the standard deviation. Although the sharp-edged windowing in the 12 month running mean can lead to some side-lobe leakage (T84), the procedure has the advantage that it guarantees elimination of the annual cycle and its higher harmonics. This eliminates some of the spurious minima and maxima which occur when using the 11-point filter proposed in T84, and gives a slightly higher correlation between Darwin and Tahiti: -0.80 compared to -0.76 quoted in T84, for the period 1935–1982. The detrending is necessary due to the decadal drifts that occur in the series — for example the 0.5–1 mb decrease/increase in the Tahiti/Darwin SLPs since the mid-1970s (Trenberth and Hoar, 1996 and references therein). The detrending has the disadvantage of shortening the series to 1870–1990 (1440 monthly values) since a reliable running mean cannot be defined for the first and last 5 years when using a 121 month filter. The anti-correlated oscillations in the standardised Darwin and Tahiti anomalies can be clearly seen in Fig. 1. For the period 1870–1990, the coefficient of correlation between the Darwin and Tahiti anomalies is -0.61, and is less than the value -0.80 obtained for the more recent period 1935–1982 perhaps due to inhomogeneities in the observational record. The SLP at Darwin is strongly correlated with major El Niños such

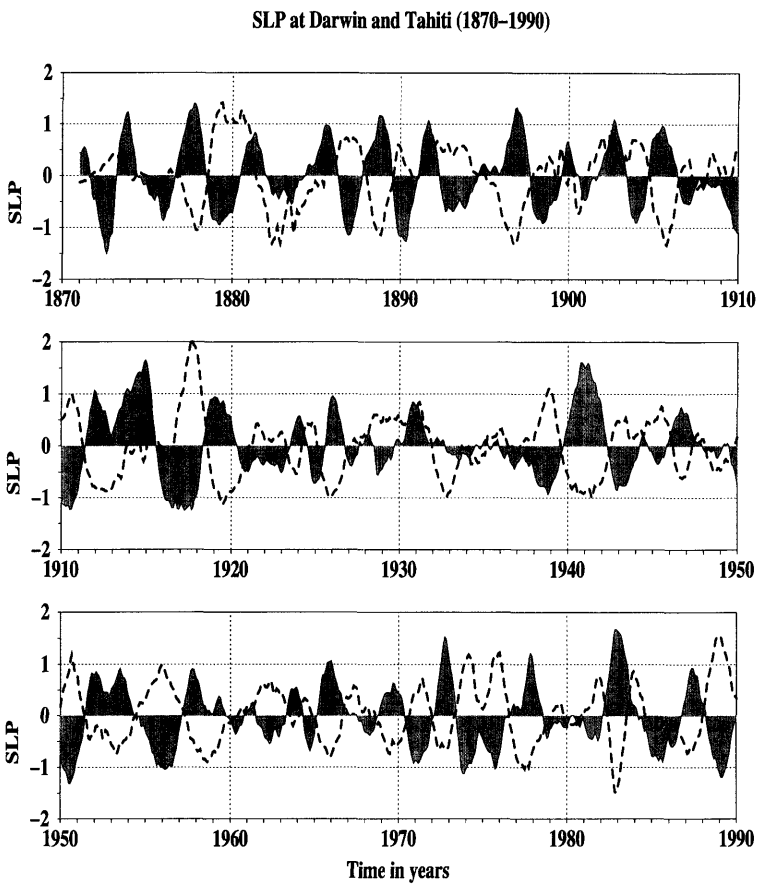

Fig. 1. Standardised interannual anomalies in detrended monthly mean sea-level pressures observed at Darwin (grey shaded) and Tahiti (dashed) for the period 1866–1995. Note the clear anticorrelations corresponding to longitudinal displacements in the convective activity associated with ENSO.

as 1982/83 and 1986/7 and anticorrelated with La Niñas such as 1988/89.

6.3. *Signal and noise*

Trenberth (1984) defined the SOI *signal* to be the linear combination of Darwin and Tahiti SLPs having maximum variance whereas the *noise* was defined to be the residual linear combination explaining the least variance. The square root of the ratio of the two variances was defined to be a measure of the signal-to-noise ratio (S/N). The signal and noise can be found by a bivariate principal component analysis and are the leading and trailing PCs of the Darwin/Tahiti SLP correlation matrix. By the spectral decomposition theorem, the real symmetric correlation matrix, $\boldsymbol{S}$, can be diagonalised as follows

$$\boldsymbol{S}=\mathbf{Q}\boldsymbol{\Lambda}\boldsymbol{Q}^{\mathrm{T}}=\frac{1}{\sqrt{2}}\begin{pmatrix}1 & 1\\ -1 & 1\end{pmatrix}\begin{pmatrix}1-r & 0\\ 0 & 1+r\end{pmatrix}\times\frac{1}{\sqrt{2}}\begin{pmatrix}1 & -1\\ 1 & 1\end{pmatrix}, \tag{17}$$

where r is the Darwin/Tahiti correlation (-0.61). The principal components, z', are rotated combinations of the SLPs at Darwin and Tahiti ($z'=\boldsymbol{Q}^{\mathrm{T}}z$) and their variances are given by the eigenvalues $(1-r)$ and $(1+r)$. Hence, the signal-to-noise ratio is given by $S/N=\sqrt{(1-r)/(1+r)}(=2.0)$. By defintion, the SOI

signal has a correlation of $\sqrt{(1-r)/2}(=0.90)$ with Darwin and $\sqrt{(1-r)/2}(=-0.90)$ with Tahiti — the equality in magnitude of correlation is a special consequence of having only two variables. Fig. 2 shows the signal (SOI) and noise time series, where the noise time series has a quarter of the variance of that of the signal time series, thereby giving a S/N ratio of 2. Furthermore, the noise time series has less coherence at low frequencies than does the signal time series which is more influenced by the slowly evolving sea surface temperature boundary conditions.

6.4. *The ACC distribution*

Fig. 3 shows the probability density function for the anomaly correlation R_{kl} calculated using three different metrics: $\boldsymbol{M}=\boldsymbol{I}$ (COR), $\boldsymbol{M}=\boldsymbol{S}^{-1}$ (MAH2), and $\boldsymbol{M}=\boldsymbol{S}^{-}$ (MAH1). The inverse covariance metric (MAH2) is given by $\boldsymbol{M}=\boldsymbol{S}^{-1}=\boldsymbol{Q}\ \mathrm{diag}\{(1-r)^{-1},\ (1+r)^{-1}\}\ \boldsymbol{Q}^{\mathrm{T}}$ and corresponds to calculating the correlation coefficient as the Euclidean scalar product in 2-dimensional PC space. The metric $\boldsymbol{M}=\boldsymbol{\Sigma}^{-}=\boldsymbol{Q}\mathrm{diag}\{(1-r)^{-1},\ 0\}\ \boldsymbol{Q}^{\mathrm{T}}$ is the rank one pseudo-inverse of the covariance metric and corresponds to calculating the correlations using only the signal PC (the SOI). This is reasonable since the SOI signal is considered to be the ENSO-related predictable component whereas the noise part is considered to be more due to internal atmospheric processes and observational uncertainties and hence is considered to be less predictable.

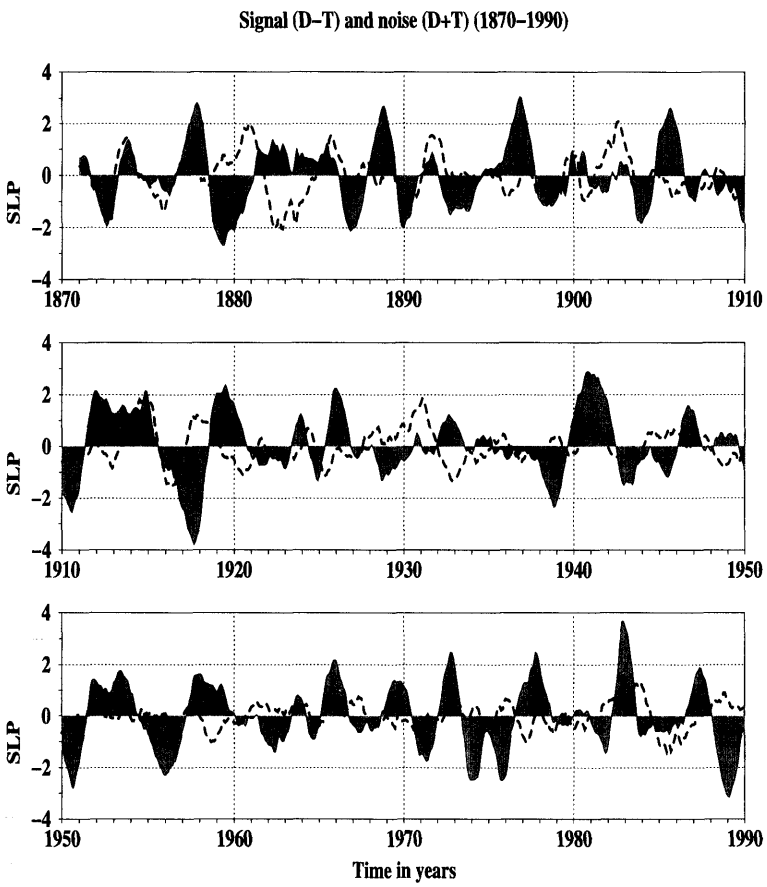

Fig. 2. The first (grey shaded) and second (dashed) principal components of the two time series shown in Fig. 1. The first PC is an optimal definition of the Southern Oscillation Index (SOI) and the second PC represents internal atmospheric and observational noise. The ratio of signal-to-noise standard deviations is 2.0.

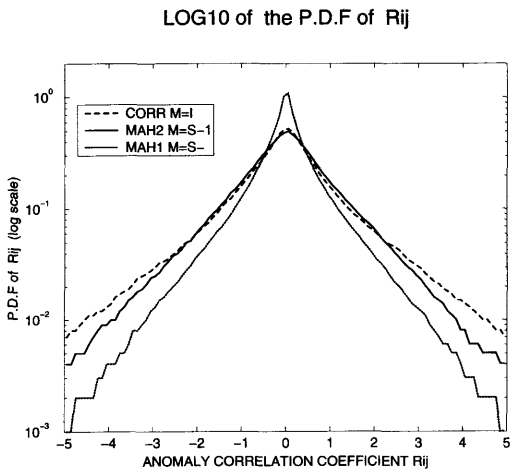

Fig. 3. The probability distributions of the anomaly correlation coefficient $R_{kl}=z_k^{\mathrm{T}}\boldsymbol{M}z_l$ (logarithmic scale) for 3 different metrics. The bin size was chosen to be 0.1 and no smoothing has been applied.

From the linearity in Fig. 3, it can be seen that the MAH2 p.d.f. decays exponentially whereas the MAH1 p.d.f. decays more rapidly, in agreement with the analytical expressions obtained in Subsection 5.2. Table 1 tabulates the low order moments of R_{kl} for the three metrics. The first-order moment is identically zero as a result of the mean of the anomalies being defined to be zero. The values expected assuming multinormally distributed anomalies have been calculated using the previously derived expressions with the coefficients given by $\beta_m=(1-r)^m+(1+r)^m$ for the COR metric and $\beta_m=p$ for the Mahalanobis metrics ($p=1, 2$). The values in Table 1 are close to the values expected assuming multinormality which suggests that the null hypothesis was reasonable. Examination of Table 1 reveals that the MAH2 metric gives the smallest kurtosis whereas MAH1 gives the largest kurtosis. The COR metric gives an intermediate kurtosis in agreement with its variance dimension $N_{\mathrm{D}}^{(1)}=\beta_1^2/\beta_2=2/(1+r^2)$ which has the value 1.5, and therefore lies between 1 for the MAH1 metric and 2 for the MAH2 metric.

6.5. *Are the anomalies multinormally distributed?*

The agreement with the values in parentheses in Table 1, suggests that the mulitnormal null hypothesis is valid. This can be tested rigorously by using the multinormal skewness test proposed by Mardia (1970) where the multinormal skewness is defined as $b_{1,2}=E(R_{kl}^3)$ calculated using the metric MAH2 (i.e., $b_{1,2}=0.122$). Critical values for this statistic are tabulated in Table 2 of Mardia (1974). Assuming 100 d.o.f. in the time series (i.e., about one independent d.o.f. per year), gives a 5% critical value of $b_{1,2}$ of 0.581. Hence at 5% confidence, we can accept the null hypothesis that the anomalies are multinormally distributed. Examination of the table of critical values, reveals that there would have to be at least 4000 d.o.f. in the time series (certainly a huge overestimate) before the null hypothesis could be rejected at 5%.

The multinormality of the anomalies in this bivariate case can be checked more directly by plotting their p.d.f.'s against that of a Gaussian (Fig. 4). It can be seen that the values from the Darwin, Tahiti, SOI signal and noise time series all lie close to the inverted parabola expected for the log of a Gaussian curve. The deviation from normality can be tested by using the univariate skewness, $b_1=m_3^2/m_2^3$, which takes the values of 0.065, 0.014, 0.002, and 0.069 for the time series of Darwin, Tahiti, signal (SOI) and noise, respectively. For one hundred degrees of freedom, the critical value at 5% of b_1 is 0.151 (Pearson and Hartley, 1972). Hence at 5% confidence, the SLP

Table 1. *Moments of $R_{kl}=z_k^{\mathrm{T}}\boldsymbol{M}z_l$ and kurtosis for Darwin and Tahiti SLP anomalies using three different metrics $\boldsymbol{M}=\boldsymbol{I}$, $\boldsymbol{S}^-$, and $\boldsymbol{S}^-$; values expected assuming multinormally distributed anomalies are shown in parentheses for comparison*

Moment/metric	COR	MAH2	MAH1
$E(R_{kl})$	0.000 (0.00)	0.000 (0.00)	0.000 (0.00)
$E(R_{kl}^2)$	2.735 (2.74)	1.994 (2.00)	0.998 (1.00)
$E(R_{kl}^3)$	0.044 (0.00)	0.122 (0.00)	−0.007 (0.00)
$E(R_{kl}^4)$	55.80 (62.9)	27.00 (24.0)	8.000 (9.00)
$E(R_{kl}^4)/E(R_{kl}^2)^2$	7.46 (8.38)	6.79 (6.0)	8.02 (9.0)

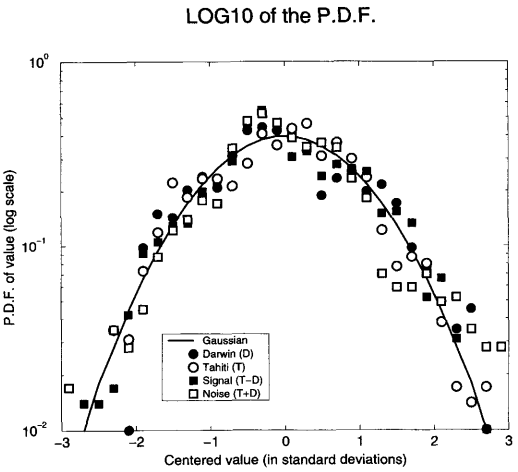

Fig. 4. The probability density functions of the time series depicted in Figs. 1, 2 (logarithmic scale). The solid curve is the parabola expected for the logarithm of a Gaussian distribution. The bin size was chosen to be 0.2 and no smoothing has been applied.

anomalies and their PCs are Gaussian distributed thereby confirming our null hypothesis and previous statements (Chu and Katz, 1985). The spatial and temporal averaging involved in defining monthly mean SLPs together with the central limit theorem, helps to explain why the SLP anomalies should be close to normally distributed.

7. Concluding remarks

The probability density function for a generalised anomaly correlation coefficient defined with an arbitrary weighting metric, has been derived analytically under the assumption that the anomaly patterns are multinormally distributed. The resulting distribution and its moments depend on the choice of metric used in defining the correlation coefficient. The moments can be used to estimate the number of spatial dimensions in the anomaly patterns and a generalised expression is presented. Such dimension estimates depend on the choice of metric and are generally not invariant under linear transformations of the anomalies. The kurtosis of the distribution depends inversely on the number of d.o.f. and hence reducing the dimensionality in order to increase the number of analogues will have the undesirable effect of increasing the number of outliers in the distribution. For example, correlating over a smaller geographical region for regional forecasts will reduce the number of d.o.f. but at the expense of giving correlation coefficients having more outliers. The null hypothesis of multinormal anomalies can be tested a posteriori by examining the moments of the correlation coefficient distribution.

As in ordinary least squares linear prediction, the (pseudo-)inverse of the covariance matrix plays a special role as the metric defining the matrix of weights used for correlating two patterns.

1. It gives an ACC that is invariant under linear transformations of the grid point variables ($z \rightarrow Az$). For example, invertible linear interpolation of the data onto another grid with the same number of points will not change the results.

2. The different principal modes of variability are given equal weight and hence spatial correlations (teleconnections) have been taken into account. Naively correlating using the same weight for all the grid points (COR metric) treats all grid points as equal and takes no account of the teleconnection structure between values at different grid points.

3. The use of a pseudo-inverse helps to eliminate the noisy part of the signal which is believed to be unpredictable. For example, in the Darwin/Tahiti case, it is the leading PC which is believed to be predictable (SOI) and not the trailing PC which represents observational error and internal atmospheric processes — hence the S^- pseudo-inverse is an appropriate metric for improving the signal-to-noise ratio in predictions.

4. The use of the pseudo-inverse rather than the inverse substantially lowers the dimensionality of the phase space from the number of grid points down to the rank of the covariance matrix. This has the advantage of making the search for analogues in a lower dimensional space more tractable (Van den Dool, 1994) whilst hopefully retaining the essential modes containing the signal.

5. The resulting ACC distribution is the closest to Gaussian, and has the least kurtosis. There are less outliers, which are difficult to predict and can therefore reduce the average predictive skill.

6. The moments of the ACC distribution give dimension estimates equal to the rank of the covariance matrix, and this is invariant under non-singular linear transformations, which is not

the case for dimension estimates made with other metrics.

7. The second and third moments of the ACC are equal to the multivariate kurtosis and skewness measures introduced by Mardia (1970) and provide powerful ways to test the multinormality null hypothesis.

For cases where a unique inverse of the covariance matrix exists, Bell (1982,1986) and Hasselmann (1993) have demonstrated that the inverse covariance matrix gives optimal fingerprints for detecting climate change and recommended its use in detecting such changes. Thacker (1996b) also noted that the use of such a metric has a simple geometric interpretation in terms of scalar products. In this study, it has been argued that the inverse of the covariance matrix does not in general exist, and therefore the best one can do is to use the pseudo-inverse of the covariance matrix which corresponds to the minimum length solution. The use of such a truncated Mahalanobis metric corresponds to taking the Euclidean scalar product of the first r principal components (scaled to have unit variance), where r is the estimated rank of the sample covariance matrix. In other words, instead of correlating spatial maps over all the grid points, correlations are made using the leading r principal components. The truncation is necessary since not all the principal components represent predictable signal and hence it makes no sense to include these in the correlation coefficient. In future studies of weather prediction and climate change fingerprinting, it will be of interest to define the correlation coefficients using the truncated Mahalanobis metric as a weight — in this way the multivariate nature of the data will be taken into account and the noisy unpredictable components will be discarded. For this approach to be effective, care should be taken to estimate the sample covariance matrix as robustly as possible using all the available data. It should be pointed out that in this study and in other studies such as that of Hasselmann (1993), the implicit assumption has been made that the predicted anomalies have come from the same population as the observed ones, which is only the case if the model is *perfect* and the observations are free from systematic measurement errors. In practice, systematic errors will cause the covariance matrix of the model generated anomalies to differ from the covariance matrix of the observed anomalies as may be witnessed by differences between the model and observed eigenvectors. In such cases, an optimal strategy would be to assess predictions by performing a canonical correlation analysis between modelled and observed anomalies in their appropriately truncated PC spaces (Barnett and Preisendorfer, 1987), and this approach can be considered as a statistical way of correcting for systematic errors and optimally reducing noise.

The approach adopted to derive the p.d.f. of the generalised covariance can also be applied to obtain the p.d.f. of the generalised root mean square measure, albeit at the expense of more complicated mathematical expressions. The above conclusions concerning the invariance, dimension, kurtosis, and noise reduction properties of the Mahalanobis metric remain the same. The root mean squared error in Mahalanobis space is equal to the square root of the sum of the squares of the leading principal components and hence can be interpreted geometrically as the distance of a point from the origin in PC-space. Such a simple interpretation is not possible if one uses the RMS in grid point space (as is often done in practise), since the variables are correlated and hence the distance is given by the cosine law rather than by the simple pythagorean sum. The simple geometric interpretation obtained with the Mahalanobis metric could be usefully employed in the hypersphere method used to define covering sets for estimating the local probability density of phase space (Toth, 1993).

The Mahalanobis metric can be used no matter what one chooses as anomalies, for example, one could use anomalies of geopotential height, streamfunction, vorticity etcetera. As previously mentioned, the resulting ACC is invariant under *non-singular* linear transformations and so different anomalies can sometimes give identical results. In general, however, the anomalies are not related by non-singular linear transformations. For example, vorticity is given by the Laplacian of the streamfunction and hence the streamfunction is only determined up to a constant — in this case, the transformation is linear yet singular and hence vorticity and streamfunction could give different results. The choice of anomaly field should be decided using dynamical arguments and physical conservation laws but then the truncated Mahalanobis metric should be used as the

weighting. Such an approach would represent the best of both dynamical and statistical considerations. In general, the eigenvectors of the sample covariance matrix do not correspond to Fourier modes and the truncated Mahalanobis metric is therefore not equivalent to energy or enstrophy norms which are local in wavenumber space (Palmer et al., 1997). In some cases, such as in predicting total rainfall amounts, the multinormal null hypothesis may be inappropriate, and it may be better to work with transformed quantities such as the square root of the rainfall which is more normally distributed. This corresponds to using a non-flat metric which gives more weight to anomalies nearer the origin. There are also other theoretical reasons why non-flat metrics may prove to be useful as ideal measures of real fluid flows (Pasmanter, 1996). The choice of optimal similarity measures for predictability studies and the underlying probability distributions of weather and climate anomalies remain challenges which require further attention.

8. Acknowledgements

It is a pleasure to acknowledge Dr. M. Allen, Dr. F. J. Doblas Reyes, Prof. K. V. Mardia, Dr. F. Molteni, Dr. T. Palmer, and Dr. R. Pasmanter, for useful discussions throughout this study. I also wish to thank the reviewers for their pertinent remarks and Dr. Z. Toth for sending me copies of his thought-provoking articles. The CRU-SOI historical data set was generously provided by Dr. K. Sperber of PCMDI from the NASA Climate Data System, and Professor D. Collett kindly supplied the Biometrika tables containing the confidence limits for $\sqrt{b_1}$ and b_2.

REFERENCES

Barnett, T. P. and Preisendorfer, R. W. 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.

Barnett, T. P. and Schlesinger, M. E. 1987. Detecting changes in global climate induced by greenhouse gases. *J. Geophys. Res.* **92**, 14772–14780.

Bell, T. L. 1982. Optimal weighting of data to detect climate change: application to the carbon dioxide problem. *J. Geophys. Res.* **87**, 11161–11170.

Bell, T. L. 1986. Theory of optimal weighting to detect climate change. *J. Atmos. Sci.* **43**, 1694–1710.

Besse, P. 1992. PCA stability and choice of dimensionality. *Statistics and Probability Letters* **13**, 405–410.

Bronshtein, I. N. and Semendyayev, K. A. 1985. *Handbook of mathematics.* Van Nostrand Reinhold Company, Inc., New York.

Chen, W. Y. 1982. Assessment of Southern Oscillation sea-level pressure indices. *Mon. Wea. Rev.* **110**, 800–807.

Chu, P. S. and Katz, R. W. 1985. Modeling and forecasting the Southern Oscillation: A time domain approach. *Mon. Wea. Rev.* **113**, 1876–1888.

Elliott, W. P. and Angell, J. K. 1988. Evidence for changes in Southern Oscillation relationships during the last 100 years. *J. Climate* **1**, 729–737.

Fraedrich, K., Ziehmann, C. and Sielmann, F. 1995. Estimates of spatial degrees of freedom. *J. Climate* **8**, 361–369.

Gutzler, D. S. and Shukla, J. 1984. Analogs in the wintertime 500 mb height field. *J. Atmos. Sci.* **41**, 177–189.

Hasselmann, K. 1976. Stochastic climate models. Part I: Theory. *Tellus* **28**, 473–485.

Hasselmann, K. 1993. Optimal Fingerprints for the Detection of Time-dependent Climate Change. *J. Climate* **6**, 1957–1971.

Horel, J. D. 1985. Persistence of the 500 mb height field during Northern Hemisphere Winter. *Mon. Wea. Rev.* **113**, 2030–2042.

Jolliffe, I. T. 1986. Principal component analysis. Springer Verlag, New York.

Karoly, D. J., Cohen, J. A., Meehl, G. A., Mitchell, J. F. B., Oort, A. H., Stouffer, R. J. and Wetherald, R. T. 1994. An example of fingerprint detection of greenhouse climate change. *Climate Dyn.* **10**, 97–105.

Kendall, M., Stuart, A. and Ord, J. K. 1983. *The advanced theory of statistics*, 4th edition. Griffin Press, High Wycombe, UK.

Lorenz, E. N., 1969. Atmospheric predictability as revealed by naturally occurring analogues. *Tellus* **26**, 636–646.

Mardia, K. V. 1970. Measures of multivariate skewness and kurtosis with applications. *Biometrika* **57**, 519–530.

Mardia, K. V. 1974. Applications of some measures of multivariate skewness and kurtosis in testing normality and robustness studies. *Sankhyā: The Indian Journal of Statistics* **36**, 115–128.

Mardia, K. V., Kent, J. T. and Bibby, J. M. 1979. *Mulitvariate analysis.* Academic Press Limited, London.

Marshall, J. and Molteni, F. 1993. Towards a dynamical understanding of planetary-scale flow regimes. *J. Atmos. Sci.* **50**, 1792–1818.

Menke, W. 1989. *Geophysical data analysis: discrete inverse theory.* Academic Press, Inc., San Diego.

Miyakoda, K., Hembree, G. D., Strickler, R. F. and Shulman, I. 1972. Cumulative results of extended forecast experiments (I). Model performance for winter cases. *Mon. Wea. Rev.* **100**, 836–855.

Palmer, T. N. and Anderson, D. L. T. 1994. The prospects for seasonal forecasting — A review paper. *Quart. J. Roy. Meteor. Soc.* **120**, 755–793.

Palmer, T. N., Gelaro, R., Barkmeijer, J. and Buizza, R. 1997. Singular vectors, metrics and adaptive observations. *J. Atmos. Sci.*, in press.

Pasmanter, R. A., 1996. Metric structures of inviscid flows. *Chaotic dynamics.*

Pearson, E. S. and Hartley, H. O. 1972. *Biometrika tables for statisticians*, vol. 2. Cambridge University Press.

Ropelewski, C. F. and Jones, P. D. 1987. An extension of the Tahiti-Darwin Southern Oscillation Index. *Mon. Wea. Rev.* **115**, 2161–2165.

Santer, B. D., Taylor, K. E., Wigley, T. M. L., Penner, J. E., Jones, P. D. and Cubasch, U. 1995. Towards the detection and attribution of an anthropogenic effect on climate. *Climate Dyn.* **12**, 77–100.

Simmons, A. J., Mureau, R. and Petroliagis, T. 1995. Error growth and estimates of predictability from the ECMWF forecasting system. *Quart. J. Roy. Meteor. Soc.* **121**, 1739–1771.

Tett, S. F. B., Mitchell, J. F. B., Parker, D. E. and Allen, M. 1996. Human influence on the atmospheric vertical temperature structure: detection and observations.*Science* **274**, 1170–1173.

Thacker, W. C. 1996. Climate fingerprints, patterns, and indices. *J. Climate* **9**, 2259–2261.

Thacker, W. C. and Lewandowicz, R. 1996. Climate indices, principal components, and the Gauss-Markov theorem. *J. Climate* **9**, 1942–1958.

Toth, Z. 1991a. Intercomparison of circulation similarity measures. *Mon. Wea. Rev.* **119**, 55–64.

Toth, Z. 1991b. Circulation patterns in phase space: a multinormal distribution? *Mon. Wea. Rev.* **119**, 1501–1511.

Toth, Z. 1993. Preferred and unpreferred circulation types in the northern hemisphere wintertime phase space. *J. Atmos. Sci.* **50**, 2868–2888.

Toth, Z. 1995. Degrees of freedom in northern hemisphere circulation data. *Tellus* **47A**, 457–472.

Trenberth, K. E., 1984. Signal versus noise in the Southern Oscillation. *Mon. Wea. Rev.* **112**, 326–332.

Trenberth, K. E. and Hoar, T. J. 1996. The 1990-95 El Niño-Southern Oscillation event: longest on record. *Geophys. Res. Lett.* **23**, 57–60.

Trenberth, K. E. and Shea, D. J. 1987. On the evolution of the Southern Oscillation. *Mon. Wea. Rev.* **115**, 3078–3096.

Van den Dool, H. M. 1994. Searching for analogues, how long must we wait? *Tellus* **46A**, 314–324.

Van den Dool, H. M. and Chervin, R. M. 1986. A comparison of month-to-month persistence of anomalies in a general circulation model and in the earth's atmosphere. *J. Atmos. Sci.*, **43**, 1454–1466.

Wallace, J. M., Cheng, X. and Sun, D. 1991. Does low-frequency atmospheric variability exhibit regime-like behavior? *Tellus* **43AB**, 16–26.

Wallace, J. M. and Gutzler, D. 1981. Teleconnections in the geopotential height field during the northern hemisphere winter. *Mon. Wea. Rev.* **109**, 784–812.

Wright, P. B. 1985. The Southern Oscillation: An ocean-atmosphere feedback system? *Bull. Amer. Meteor. Soc.* **66**, 398–412.

Zwiers, F. W. 1987. Statistical considerations for climate experiments. Part II: Multivariate tests. *J. Clim. App. Meteorology* **26**, 477–487.