# A Review of Methods for Statistical Climate Forecasting

**Edward Campbell**

**Technical Report 06/134**

**18 August 2006**

# Table of Contents

# 1 Introduction

In this document I detail the statistical methods commonly applied in the climatological literature, and note some possible research directions. In section 2 I provide an overview of the climatological literature. I have sketched the technical details of the methods described to act as a reference, and to provide clarity as to the mathematical details in section 3. I provide detailed comments on the statistical approaches identified in the mainstream climatological literature. Section 4 is devoted to a review and description of a number of potentially highly useful statistical methods that have received little or no attention in the climatological literature.

# 2 Overview

The thrust of the climatology literature seems to be in two major directions. First and foremost is the discovery of so-called “teleconnections”. That is, links between climatological events that whilst far apart in distance seem to display a causal relationship. Secondly, a good deal of effort is engaged in the analysis of low frequency climate variability, which modulate the higher frequency events that contribute to Australia’s uniquely diverse climate. There is good reason to be optimistic that the low frequency components of climate variability are open to understanding, since they are likely to be the result of large-scale ocean-atmosphere interactions.

Teleconnections are typically sought via statistical methods, and once discovered physical reasoning is brought to bear to seek understanding. The statistical methods applied fall overwhelmingly in the category of linear, and are typically multivariate in nature. Principal component analysis (PCA) and canonical correlation analysis (PCA) are ubiquitous. PCA is most commonly termed Empirical Orthogonal Function analysis. In climatology it is usual to think of variables as being measured over two- or three-dimensional fields, which can be difficult to analyse. An approach to this problem is to break the observed fields down into a sum of orthogonal contributions, hence the connection with PCA. See section 3.2 for further details. Discriminant analysis has been put forward as a nonlinear technique by some, which is hard to justify given the resulting discriminant functions fall into the framework of a linear statistical model. See section 3.7 for discussion of this point. I suspect the confusion is due to confusion over common versus separate within-groups covariance matrices.

There has been a little work on genuinely nonlinear methods using ideas from dynamical systems theory. I would add though that analog forecasting has strong parallels with such ideas. One notable example bringing these ideas together is provided by Drosdowsky (1994). My own view of analog forecasting is that the approach does not make best use of the available data, even if anti-analogs are used (Nicholls and Katz, 1991). This approach turns essentially continuous data into discrete historical events, from which we hope to understand the current weather situation and make forecasts based on the best available analog or combination of analogs- see Casey (1995). A more efficient approach is to try to estimate the underlying dynamics, and we discuss these ideas in section 3.9.

The step towards physical-statistical models is, I think, an important one if we are to derive better understanding of our climate and thence improved forecasting skill. This represents a fusing together of physical and empirical models to extract the best from both. To place this in context, we need to provide a broader physical basis for our statistical modelling. Physical models are typically expressed as systems of (partial) differential equations for a state vector describing the climate system. These models are typically developed at coarse spatial (in particular) and temporal resolutions. Empirical models however apply naturally at the scales of observations on the climate system.

There is a significant school of thought in meteorology and climatology that the physical world evolves according to essentially deterministic rules, which may or may not demonstrate chaotic behaviour. See, for example, Gleeson (1966) and Epstein (1969) where the only role for a stochastic mechanism is in representing uncertainty in initial values.

There is much evidence in the statistical literature (Smith, 1992) that simple deterministic chaos does not adequately explain the data arising from physical systems, and it is clear that a fusion of dynamical and statistical approaches is required. I would argue that the natural world is genuinely stochastic in nature, and a natural way to proceed is to form stochastic differential equations to model key physical processes. These would be used to gain qualitative insights, and be a complement to physical-statistical models.

# 3 Statistical Methods

## 3.1 Correlation and Regression Analysis

Extensive use has been made of correlation and regression methods in attempting to establish evidence of "teleconnections" (Nicholls, 1991)- climate events that are related whilst being far removed in space. Teleconnections are a feature of the Southern Oscillation because of the vast scale on which it takes place. Nicholls (1991) notes a characteristic of the El-Niño Southern-Oscillation (ENSO) "for rainfall anomalies to appear in many areas at the same time. Thus droughts in India, North China, Australia and parts of Africa and the Americas tend to occur approximately simultaneously …". In particular, lagged correlations have been used to detect time-lagged climate effects (Drosdowsky, 1993c). Linear regression methods have been used by many authors to model the relationship between the Southern Oscillation Index (SOI) and a variety of phenomena of interest in economic terms.

Nicholls (1986) used lagged correlations to show a statistically significant correlation between Australian sorghum yield and Darwin pressure, an indicator of the Southern Oscillation. Nicholls found that stronger relationships exist between yield and Darwin pressure *trends*, rather than absolute or scaled readings. After the removal of trends, due to improved technology and introduction of new cultivars for example, a linear regression model relating yield to the January-March to June-August trend explained about 50% of the yield variation. Given that the crop is planted between October and February this is a potentially very useful management tool. Indeed, a slightly improved model incorporating the trend up to October was rejected since a prediction at the time of planting was considered less useful from a management perspective.

Rimmington and Nicholls (1993) examined the use of the Southern Oscillation Index (SOI- standardised difference between sea-surface air pressure in Darwin and Tahiti) to forecast wheat yields in Australia. As found by Nicholls (1986), trends in the Southern Oscillation, as measured by SOI, had more predictive power than the observed value. Rimmington and Nicholls (1993) note that "A large proportion of the year-to-year variation of Australian wheat yield is due to variation in the available soil moisture which is determined by the balance of rainfall and evapotranspirative losses." They go on to state that "Much of the interannual variation in rainfall over the Australia [sic] wheat-belt is related to the … (ENSO) phenomenon … Variations in temperature, wind and therefore evapotranspiration may also be related to ENSO events." However, the pre-existing soil moisture profile is not used in model-building. The skill levels, as measured by $R^2$, peaks at 36% for Queensland and falls as low as 6% for South Australia. This suggests that whilst SOI alone accounts for a significant share of the variation in Australia wheat yield, a significant proportion is left unexplained.

An alternative to linear regression is provided by Russell et al. (1993), who use a methodology known as Alternating Conditional Expectations (ACE). This method identifies nonlinear transformations of the data that allow a linear regression model to be applied to the transformed data. Russell et al. (1993) note that variables besides SOI are potentially of use in building a predictor of rainfall, noting in particular seas-surface temperatures and that ACE can be applied when multiple predictors are available. They only apply ACE to SOI data however, but the level of complexity of including sea-surface temperatures is not as great as is claimed in this paper.

A number of interesting observations on the use of correlation and regression methods were made by Drosdowsky and Williams (1991), and are worth quoting here:

> '… these anomalies (geopotential height) are not symmetric, being more intense during the positive (SOI) phase. This may be due to the *nonlinear* nature of the latent heat forcing over northern Australia and Indonesia. During ENSO (negative phase) events convection is reduced, but not entirely absent over this region, while the increases in precipitation during the positive phase may be many times above the mean.
>
> … Locally, however, there are significant "nonlinear" deviations. The most significant deviations of these occur during summer in the geopotential height field over the Tasman sea, …
>
> … The effects of these nonlinearities on the evolution of extreme events need to be studied further, as do the implications for linear-regression forecasting schemes using the SOI as a predictor.'

This acknowledges a clear need to model inherently nonlinear phenomena.

*3.1.1 Research Opportunities*

- Methods assume all series examined are independent; an interesting approach might be to fit autoregressive models incorporating exogenous variables.
- Methods as applied tend not to look for evidence of nonlinear responses, which could be built in quite readily. Russell et al. (1993) examine the use of ACE, but many other nonparametric regression approaches exist. These include Additive and VAriance Stabilising transformations (AVAS) and projection pursuit regression. Note that projection pursuit regression is able to model interactions between predictors whereas ACE and AVAS cannot, and is also easily extendable to model multivariate responses.
- It seems that clearly meaningful explanatory variables are often not being incorporated, notably pre-exisiting soil moisture and sea-surface temperature.

## 3.2 Empirical Orthogonal Functions

It is well known that under very mild conditions a function $f(\cdot)$ may be expanded as a sum of functions that are mutually orthogonal to one another. These orthogonal functions may then be thought of as components of $f(\cdot)$. A random function $Y(t,\mathbf{x})$, where $t$ denotes time and $\mathbf{x} \in D$ location, may be decomposed into a doubly orthogonal expansion

$$Y(t,\mathbf{x}) = \sum_k Z_k(t)\varphi_k(\mathbf{x}) \ . \qquad (1)$$

This representation of a random function is known as the Karhunen-Loève expansion, although in the meteorological literature, where typically $D = \Re^2$, it is known as the Empirical Orthogonal Function (EOF) expansion.

The eigenfunctions $\{\varphi_k(\mathbf{x})\}$ are said to be *analytically* orthogonal since

$$\int_D \varphi_i(\mathbf{x})\varphi_j(\mathbf{x})d\mathbf{x} = \delta_{ij}\lambda_i,$$

where $\delta$ denotes the Kronecker delta function and $\lambda_i$ the corresponding eigenvalue. The random functions $\{Z_j(t)\}$ are said to be statistically orthogonal since

$$E\left[Z_i(t)Z_j(t)\right] = \delta_{ij}\lambda_i.$$

The coefficients $\{Z_k(t)\}$ are found by projecting each realisation onto the $k$th eigenfunction, thus:

$$Z_k(t) = \int_D Y(t,\mathbf{x})\varphi_k(\mathbf{x})d\mathbf{x}.$$

It is possible to show that the eigenfunctions are the solutions of the Fredholm integral equation

$$\int_D R(\mathbf{x},\mathbf{x}')\varphi_k(\mathbf{x}')d\mathbf{x}' = \lambda_k\varphi_k(\mathbf{x}), \qquad (2)$$

where $R(\mathbf{x},\mathbf{x}')$ is the autocorrelation of the process $Y(t,\mathbf{x})$. The functions $\{\varphi_k(\mathbf{x})\}$ are therefore the eigenfunctions of the symmetric semi-definite correlation kernel and the $\{\lambda_k\}$ are the associated eigenvalues. It can be shown that there is a countably infinite number of such eigenfunctions, so that the correlation kernel may be written as

$$R(\mathbf{x},\mathbf{x}') = \sum_{k=1}^{\infty} \lambda_k\varphi_k(\mathbf{x})\varphi_k(\mathbf{x}').$$

It can be shown that the truncated decomposition

$$y*(t,\mathbf{x}) = \sum_{k=1}^{K} z_k(t)\varphi_k(\mathbf{x})$$

minimises the mean integrated squared error

$$E\left\{\int_D [Y(t,\mathbf{x}) - y*(t,\mathbf{x})]^2 d\mathbf{x}\right\} = \sum_{k=K+1}^{\infty} \lambda_k . \qquad (3)$$

The spectral representation is optimal in the sense that this error is a minimum compared to $K$ terms of any orthonormal system (Cohen and Jones (1969)).

Note that the eigenfunctions may be thought of as orthogonal to the correlation kernel in the sense that

$$\int_{D^2} R(\mathbf{x},\mathbf{x}')\varphi_i(\mathbf{x})\varphi_j(\mathbf{x}')d\mathbf{x}d\mathbf{x}' = \delta_{ij}\lambda_i. \qquad (4)$$

According to Obled and Creutin (1986), the "first coherent numerical treatment of the problem (numerical solution of the Fredholm integral equation to give an eigenfunction decomposition) was proposed by Cohen and Jones (1969)," who formulated the problem as a regression model defined on a random field. The above description of EOFs is based on the aforementioned papers.

It is often said that EOFs is a form of principal component analysis, and the reason for this becomes clear when we consider the numerical problem we are faced with.. Given observations of the random function for a given time at $p$ locations, we have to estimate $R(\mathbf{x},\mathbf{x}')$ and then solve the Fredholm equation (2). The numerical approximation to this equation is

$$\sum_{j=1}^{p} R(\mathbf{x}_i,\mathbf{x}_j)\varphi_k(\mathbf{x}_j)\Delta_j = \lambda_k\varphi_k(\mathbf{x}_i), \qquad i = 1,\ldots,p, \qquad (5)$$

where $\Delta_j$ is the element of area associated with the observation at $\mathbf{x}_j$. Many other approximations are possible however, and some of these are referenced by Obled and Creutin (1986), who also examine some alternatives in detail, focusing on the choice of an orthogonal basis set for the expansion (1).

If we let

$$\Gamma_{ij} = R(\mathbf{x}_i,\mathbf{x}_j)\sqrt{\Delta_i\Delta_j}$$
$$\nu_j^{(k)} = \varphi_k(\mathbf{x}_i)\sqrt{\Delta_i}$$

then

$$\sum_{j=1}^{p} \Gamma_{ij}\nu_j^{(k)} = \lambda_k\nu_i^{(k)},$$

which represents the eigen-decomposition of the covariance matrix $\Gamma$ with eigenvalues $\{\lambda_k\}$. The eigenvectors $\nu_i^{(k)}$ estimate the eigenfunctions $\varphi_k(\mathbf{x}_i)$ at the observation points, multiplied by $\sqrt{\Delta_i}$ .

As a final theoretical note, Cohen and Jones (1969) used the Karhunen-Loève expansion to examine a regression model defined on a continuous random field. The example they use to illustrate the methodology involves a response variable being the temperature at National Airport, Washington, D.C. and the random field is the 700 mb height observed on a grid of points. Data were observed every 12 hours for 2 years.

There are many applications of EOFs published in the meteorology and climatology literature. Obled and Creutin (1986) cast the method in terms of optimal interpolation, which follows particularly from equation (3), and provide a case study on interpolation of rainfall fields.

Nicholls (1987) discusses two published applications of EOFs to study the nature of Southern Oscillation teleconnections. This paper also noted the use of extended or complex EOF analysis that incorporates temporal as well as spatial relationship. The theory was described in some detail by Barnett (1983), and is identical to the method termed 'complex PCA' by Horel (1984). This approach is discussed in detail below. A further recent example was provided by Burkhardt and James (1998), who used an extended EOF (EEOF) analysis to measure the intensity of a storm track.

### *3.2.1 Research Opportunities*

The most interesting opportunities focus on solution of the Fredholm integral equation in different contexts. Obled and Creutin (1986) conclude their paper by noting that the EOF method is still lacking some mathematical support, "especially in its relations to its connections with spectral analysis and stochastic differential equations." The stochastic differential equation approach could be used to build some appropriate Physics into a statistical model, to yield a realistic covariance model.

### 3.3 Principal Component Analysis

Principal Component Analysis (PCA) arises indirectly in the literature as a method of estimating EOFs, and directly as an exploratory statistical tool. We discussed EOFs above, and concentrate here on its use in exploratory analysis.

The data in meteorological applications typically comprise 3 dimensions: Stations (location), Time and the variables observed so the data matrix may be thought of as a cube. There are two distinct modes of analysis depending on whether we consider the observations to be stations for a fixed time (S-mode), or if we consider each individual time to be a variable and each station an observation (T-mode), thus revealing temporal associations. This is discussed by Drosdowsky (1993a), who also cites a number of such analyses in the literature.

Drosdowsky also discusses the rotation of principal components, given that principal components tend to provide a first component with "generally large loadings", with structures of interest in the second component that are often predictable, such as dipoles. This restricts the usefulness of the method, and one reason for this is the orthogonality constraint applied to successive principal components. Richman (1986) is referenced as providing an in-depth discussion of the merits of rotating principal components. Jolliffe (1989) suggests rotation be reserved for groups of components, not necessarily the first few with the largest variances, that have approximately equal variance (eigenvalues). Drosdowsky (1993a) compares the use of both an orthogonal (VARIMAX) and an oblique rotation (PROMAX).

Nicholls (1989) used PCA to simplify the pattern of Australian rainfall, and then examined correlations of the first 2 principal components with sea-surface temperature (SST). A comparison of Varimax and Promax rotations was made, with little difference found in this case. Smith (1994) examined the ability of PCA to predict Australian winter rainfall using Indian Ocean SSTs, employing principal components regression to find relationships between SST and rainfall principle components.

Paterson et al. (1978) used PCA in a very different application to classify regions of the south-west of Western Australia so that experimental locations could be chosen that represented the range of experimental locations in the state.

An alternative to PCA is Complex Principal Component Analysis (CPC), a good introduction to which is provided by Horel (1984). As noted above, this method is also termed 'Complex EOF analysis' (or Extended EOF, EEOF), seemingly interchangeably. A key advantage of CPC over real PCA is its ability to detect travelling waves, where PCA can only detect standing oscillations.

The first step in the analysis is to derive a complex data matrix where the real part is the original data, and the imaginary part is the original data shifted in phase by 90°. The next step is to determine the eigen-decomposition of the complex cross-correlation matrix. The final step is to rotate the complex principal components using the Varimax criterion to obtain a solution that emphasises regional relationships.

In mathematical terms, a scalar field $u_j(t)$ at location *j* and time *t* can be expanded in a Fourier series as

$$u_j(t) = \sum_{\omega} \left[ a_j(\omega)\cos\omega t + b_j(\omega)\sin\omega t \right],$$

for Fourier coefficients $a_j$ and $b_j$ at frequency $\omega$. In order to investigate propagating features it is natural to examine the frequency domain properties of the scalar field by the transformation

$$U_j(t) = \sum_{\omega} c_j(\omega) e^{-i\omega t}, \qquad (6)$$

where $c_j(\omega) = a_j(\omega) + ib_j(\omega)$. Expanding equation (6), we obtain

$$\begin{aligned} U_j(t) &= \sum_{\omega} \left\{ a_j(\omega)\cos\omega t + b_j(\omega)\sin\omega t + i\left[ b_j(\omega)\cos\omega t - a_j(\omega)\sin\omega t \right] \right\} \\ &= u_j(t) + i\hat{u}_j(t). \end{aligned}$$

We see that the real part is simply the original data field whilst the imaginary part is its Hilbert transform. This represents a filtering of $u_j(t)$ in which amplitude remains unchanged but each component's phase is advanced by 90°. Horel shows how the Hilbert transform can be estimated by the FFT.

An important property of CPC analysis is revealed by a number of examples considered by Horel. The Hilbert transform contains as much energy due to noise as the original data, since it does not act as a low-pass filter. This can be minimised by appropriate filtering in the calculation of the Hilbert transform by choosing suitable FFT weights.

The subsequent eigen-decomposition is based on a correlation matrix calculated using normalised data so that

$$r_{jk} = \left[U_j(t)^* U_k(t)\right]_t,$$

where * denotes complex conjugate and $[]_t$ a time average. Horel then examines the eigen-decomposition of this matrix and subsequent rotation of the principal components.

*3.3.1 Research Opportunities*

- There may be some interest in nonlinear PCA, which is a technique for incorporating data on different measurement scales, such as categorical. For further information on this see Kroonenberg et al. (1997) and references therein. This isn't the only formulation of nonlinear PCA however, since it is thought of by some as a PC optimisation using a criterion other than variance; I prefer to think of this as projection pursuit.
- CPC might be of use in space-time EDA to assess correlation structures.
- Rotation of PCs- see review article Richman (1986), J. Clim., 6, 295-335.

### 3.4 Singular Value Decomposition

The singular value decomposition (SVD) has often been viewed as competing with canonical correlation analysis (CCA) (Cherry, 1996). It is used in climatology to examine covariance relationships between two random fields, and we examine the reasoning for this following the development of Cherry (1996).

The random fields are referred to conventionally as left and right fields. Let $\mathbf{S} = \left(S_1, \ldots, S_{N_S}\right)'$ and $\mathbf{Z} = \left(Z_1, \ldots, Z_{N_Z}\right)'$ denote random vectors arising from the left and right fields respectively. It is assumed that these vectors each have mean 0, and are observed at times $t = 1, \ldots, T$. Assuming that observations at different times are independent, we may estimate the covariance matrix between the two fields to be

$$\hat{C}_{sz} = \frac{1}{T-1}\mathbf{s}'\mathbf{z},$$

where **s** and **z** represent the observed values of the random vectors.

We seek linear combinations of the form

$$X_i = \mathbf{a}_i'\mathbf{S}, \qquad Y_i = \mathbf{b}_i'\mathbf{Z}$$

such that $Cov(x_i, y_i)$ is maximised, subject to the constraints that **a** and **b** are unit vectors and that $\mathbf{a}_i'\mathbf{a}_j = \mathbf{b}_i'\mathbf{b}_j = 0 \ \forall i \neq j$.

SVD is advanced by Cherry (1996) as a means for solving this problem, referencing seemingly its first application in climatology by Prohaska (1976). To see why, we examine the process of maximising $Cov(x_i, y_i)$, dropping the subscript *i* for ease of exposition. Now,

$$\begin{aligned} Cov(X,Y) &= \mathbf{a}'Cov(\mathbf{S},\mathbf{M})\mathbf{b} \\ &= \mathbf{a}'C_{SZ}\mathbf{b}. \end{aligned}$$

Incorporating the unit vector constraints via Lagrange multipliers, we seek to maximise

$$C(\mathbf{a},\mathbf{b}) = \mathbf{a}'C_{SZ}\mathbf{b} + \tfrac{1}{2}\lambda_1(1-\mathbf{a}'\mathbf{a}) + \tfrac{1}{2}\lambda_2(1-\mathbf{b}'\mathbf{b}).$$

Thus,

$$\left.\begin{aligned}\frac{\partial C}{\partial \mathbf{a}} &= C_{sz}\mathbf{b} - \lambda_1 \mathbf{a} \\ \frac{\partial C}{\partial \mathbf{a}} &= C_{sz}\mathbf{b} - \lambda_1 \mathbf{a}\end{aligned}\right\} = 0,$$

for a maximum (in this case). Rearranging these equations we have the equivalent simultaneous system

$$\left.\begin{aligned}C'_{SZ}C_{SZ}\mathbf{b} &= \psi\mathbf{b} \\ C_{SZ}C'_{SZ}\mathbf{a} &= \psi\mathbf{a}\end{aligned}\right\},$$

where $\psi = \lambda_1 \lambda_2$ is a common eigenvalue.

These equations taken individually define an eigen-decomposition problem. Taken simultaneously requires a more general approach, and this is supplied by the SVD which decomposes the covariance matrix $C_{SZ}$ as

$$C_{SZ} = ADB'.$$

The columns of $A$ are formed from the orthonormal eigenvectors of $C_{SZ}C'_{SZ}$ corresponding to the largest $N_Z$ eigenvalues, and has dimensions $N_S$ x $N_Z$. The diagonal matrix $D$ has as its elements the positive square-root of the eigenvalues of $C'_{SZ}C_{SZ}$, and so has dimensions $N_Z$ x $N_Z$. The matrix $B$ is an orthogonal $N_Z$ x $N_Z$ matrix consisting of the orthonormal eigenvectors of the $N_Z$ x $N_Z$ matrix $C'_{SZ}C_{SZ}$. Note that the first $N_Z$ eigenvalues are common to $C'_{SZ}C_{SZ}$ and $C_{SZ}C'_{SZ}$. The remaining $N_S - N_Z$ eigenvalues of $C_{SZ}C'_{SZ}$ are equal to zero. See Seber (1984), pp 504 for more details. As a final note on the SVD, the eigenvectors as noted above are mutually orthogonal and so naturally satisfy the final requirement that $\mathbf{a}'_i\mathbf{a}_j = \mathbf{b}'_i\mathbf{b}_j = 0 \;\forall i \neq j$.

There are many applications of SVD in the literature, such as Wallace et al. (1992) who compare the method with PCA and then use the SVD to calculate canonical correlation vectors. They also use a technique known as combined PCA (CPCA), described in more detail by Bretherton et al. (1992), in which the data from two fields are combined. Bretherton et al. (1992) also describe a technique known as second field PCA (SFPCA), in which principal component amplitudes (loadings) are correlated with the second field.

### *3.4.1 Research Opportunities*

- Generalised SVD for more than 2 fields cf correspondence analysis.
- Projection pursuit

## 3.5 Cluster Analysis

Drosdowsky (1993a) used cluster analysis in an attempt to regionalise Australian rainfall anomalies, comparing it to a rotated principal component analysis. The results showed that for a small number of clusters the regional structure was somewhat different using the two techniques, whilst for a larger number of clusters the regional structure was essentially the same. Cluster analysis was also employed by Wolter (1987) in an exploratory data analysis mode.

### *3.5.1 Research Opportunities*

Rather than the exploratory emphasis of conventional cluster analysis, it would be of interest to explore model-based approaches. One approach is to view the population from which the data were drawn as being a mixture of components (clusters). The number of components could be assessed in a Frequentist framework using AIC. In a Bayesian framework, a reversible jump MCMC approach could be adopted (Green (1995), and this approach was explored by Richardson and Green (1997). A model-based approach was also explored by Banfield and Raftery (1993), who explore the use of multivariate normal distributions for individual clusters, but whose size, shape and orientation are allowed to vary in a number of ways. They also explore non-Gaussian extensions and an approximate Bayes factor for selecting the number of clusters. An alternative mixture formulation is provided by Cutler and Breiman (1994), who apply a procedure in which multivariate observations are expressed as mixtures of archetypal patterns that are chosen by nonlinear optimisation.

Jolliffe (1989) notes the mixture approach and suggests a less model-based approach and goes on to discuss a projection pursuit methodology. This has been explored by Nason (1995), who finds that the method is much more efficient than PCA in the presence of noise. We discuss this technique in the section on statistical methods of potential value (section 4.2).

### 3.6 Canonical Correlation Analysis

Canonical Correlation Analysis is a technique for exploring relationships between a collection of potential explanatory variables (**X**) and a set of response variables (**Y**). Nicholls (1987) applied CCA to a set of explanatory variables comprising bimonthly Darwin pressures over a 22 month period. The particular set of 22 months was chosen to encompass a typical SO cycle of about a year and to locate the period of strongest correlation (July to December) in the middle of the set. This will also enable lagged correlations to be detected. The response variables were Tahiti Pressure, Willis Island air temperature and south-east Australian rainfall.

Nicholls noted the usefulness of this procedure in detecting and exploring teleconnections. I would note though that it will only detect linear relationships, and may be thought of as a generalisation of multiple regression to the multi-response case.

#### *3.6.1 Research Opportunities*

- Examination of nonlinear extensions may be of interest.

### 3.7 Linear Discriminant Analysis

Drosdowsky and Chambers (1998) used linear discriminant analysis to classify rainfall categories in terms of predictors selected via multiple regression. The first stage in their analysis was to reduce the available sea-surface temperature anomaly (SSTa) data using a PCA. A step-wise multiple regression relating rainfall to principal components of SSTa and SOI was conducted, which gave reasonable predictions. It was however preferred to produce a probabilistic forecast, so the predictors identified in the regression modelling were used in a discriminant analysis.

Drosdowsky and Chambers (1998) noted an unexpected poor performance of the discriminant analysis procedure in this case, and stated two potential explanations for this. First, the use of too many predictors in the discrimination model and, secondly, “those selected by the multiple regression selection may not work best with the inherently nonlinear discriminant analysis procedure.” It is not at all clear that discriminant analysis is “inherently nonlinear”. The discriminant function is in fact linear when the rainfall category covariance matrices are assumed to be equal, and quadratic when they are allowed to be different and 2 rainfall categories are allowed. The procedure cannot really be thought of as nonlinear in a statistical sense or even in a physical sense, since the predictors enter the model in a linear fashion. A better approach would be to apply step-wise variable selection directly in the discriminant analysis procedure. Such procedures are now widely available, as Proc StepDisc in SAS® software for example.

#### *3.7.1 Research Opportunities*

- Rather than reducing the data via PCA, it would be of interest to use the SST and any other available data directly in the discrimination technique. This will inevitably lead to a situation where there are more variables than observations, which could be dealt with via Chemometrics techniques- partial least squares etc.

- Examine the use of minimal spanning trees for nonlinear discrimination, as shown by Friedman and Rafsky (1981).

### 3.8 Analogs

The idea behind analogue forecasting is a simple one: given current meteorological conditions, we search back through the data record for the closest match. Our forecast is then the outcome from this match, and an ensemble forecast may be obtained by taking the nearest neighbours from the past record. There are numerous applications of this idea in the literature.

Drosdowsky (1994) discusses analog forecasting as a nonlinear method by adopting a state space approach to the analysis of climatological series. Analogs are then constructed by examining the historical record for states close to the current state. The approach draws heavily on the work of Sugihara and May (1990). Nicholls and Katz. (1991) note a number of world meteorological agencies that use analog forecasting techniques. In addition they note the additional inclusion of anti-analogs, where the past pattern is opposite to the present. These are used with the evolution observed in the past reversed in order to make the forecast. Weighted averages of several analog forecasts have also been used, which was also noted by Drosdowsky (1994).

#### *3.8.1 Research Opportunities*

- It would be of interest to explore methods for developing prediction limits using this technique.

## 3.9 Time Series Analysis

Our main interest is in *nonlinear* approaches to time series analysis, inspired by the dynamical systems approach described by Tong (1990). The approach of Lall et al. (1996) to nonlinear forecasting has been identified as a method to be investigated for application in the IOCI (Ruprecht et al. (1996), pp14). Note however that McCaffrey et al. (1992) is an earlier reference to the use of nonparametric regression in nonlinear time series, and contains a more detailed statistical development.

This technique does not deal with the statistical properties of the series analysed, so this remains an open research question. The essence of the approach is to reconstruct the nonlinear mapping using nonparametric regression. The published approach uses Multivariate Adaptive Regression Splines (MARS), but there is also work using local polynomial fitting (Lall et al. (1995)). Generalised Cross Validation is used to assess the model fit, and thence selection of time lag and embedding dimension. Application of the method to modelling the level of the Great Salt Lake yields impressive results, particularly in revealing relatively low order dynamics (Sangoyomi et al. (1996)). This seems largely to be because the level of the lake is the result of many climate-driven processes which are spatially averaged, effectively filtering out many high frequency components. The remaining low frequency components provide a useful guide to climate change (Lall and Mann (1995)).

An alternative approach to system identification is provided by Abarbanel and Lall (1996). The time lag is selected as the first minimum of the *mutual information* (defined at the end of this section) plot, whilst the method of *global false nearest neighbours* is used to select the embedding dimension. The question of how many of these dimensions represent dynamical activity is explored using *local false neighbours*. Predictability of the system is explored by calculating both local and global Lyapunov exponents.

Dimension estimation for the Great Salt Lake data is further explored by Sangoyomi et al. (1996), who examine approaches based on scaling and nearest neighbours. The discussion of scaling leads to consideration of notions of generalised dimension measures, which include the well-known box counting dimension and correlation dimension, which is much used in the dynamical systems community.

To conclude this section we define some fundamental terms and concepts from this field, with references to additional material. Unless stated otherwise, the definitions given are based on Tong (1990).

### *3.9.1 Dynamical Systems Concepts*

The context for most of these definitions is the nonlinear iterated map

$$x(t+T) = f^T\big(x(t), x(t-\tau), \ldots, x\big(t-(m-1)\tau\big)\big), \qquad f^T : \mathbf{R}^m \to \mathbf{R}, \qquad (7)$$

which we refer to on occasion.

*Phase Space*

This concept is perhaps most readily introduced by consideration of the following set of nonlinear differential equations:

$$\dot{\mathbf{z}}(t) = F[\mathbf{z}(t)], \qquad (8)$$

where $\mathbf{z}$ is a $d$-dimensional process. This system can be written as a higher-order differential equation in terms of a single state variable $x$:

$$x^{(d)} = f\big(x,\ x',\ \ldots,\ x^{(d-1)}\big).$$

There are therefore many possible phase space representations, which are used to display the dynamics of the system. The most obvious is to plot the path of $\mathbf{z}(t)$ through time, but the *Packard-Takens Method (Fake Observables)* tells us that any $d$ observables will do. For example, if only one state variable, $x$ say, is available then we can use a pseudo-phase space $\mathbf{x}_t = \big[x(t), x(t-\tau), \ldots, x\big(t-(m-1)\tau\big)\big]$ instead. If the solution to (1) lies on an attractor of dimension $d_A < d$, then choosing the integer $m > 2d_A$ is a sufficient condition for unfolding the attractor from the scalar time series $x(t)$.

For further details see Packard et al. (1980) and Sangoyomi et al. (1996).

*Attractor*

Loosely, if it exists, the attractor of a dynamical system is what in the long term it settles down too. Thus, we neglect transient behaviour. Any point in phase space close to the attractor gets closer and closer to it.

More precisely, we give Tong (1990)'s (pp50) definition:

By an *attractor* for the mapping *f* we mean a compact set *A* such that the set

$$B = \left\{ \mathbf{x} : \lim_{n\to\infty} \inf_{\mathbf{y}\in A} \left\| f^{(n)}(\mathbf{x}) - \mathbf{y} \right\| = 0 \right\}$$

has positive Lebesgue measure and *A* is minimal w.r.t. this property. Here the notation $f^{(n)}$ denotes *n*-fold application of the mapping *f*.

*Limit Cycle/ Limit Point*

If the attractor is a set of *T* points $\{x_1, \ldots, x_T\}$ such that

$$f(\mathbf{x}_n) = f(\mathbf{x}_{n+1}), \qquad n = 1, \ldots, T-1$$

and

$$f(\mathbf{x}_T) = \mathbf{x}_1$$

then we call the attractor *A* a *limit cycle*. If $T = 1$ then we call it a *limit point*.

*Lyapunov Exponents*

A Lyapunov exponent measures the rate at which neighbouring points diverge or converge on repeated application of a mapping *f*. Let $x_0$ and $x_0'$ denote two different initial points. After *n* iterates of the mapping $x_{t+1} = f(x_t)$ we have that

$$\begin{aligned} x_n - x_n' &= f^{(n)}(x_0) - f^{(n)}(x_0') \\ &\approx \frac{d}{dx} f^{(n)}(x_0)(x_0 - x_0'). \end{aligned}$$

Applying the chain rule,

$$\frac{df^{(n)}(x_0)}{dx} = f'(x_0) f'(x_1) \ldots f'(x_{n-1}).$$

If the factors are all of comparable size then this derivative increases/decreases exponentially, which is the behaviour of the difference $(x_n - x_n')$. It therefore makes sense to consider the average rate of change

$$\lambda(x_0) = \lim_{n\to\infty} \ln \left| \frac{d}{dx} f^{(n)}(x_0) \right|^{1/n}.$$

Under some technical considerations $\lambda(x_0)$ turns out to be independent of $x_0$ almost surely. In this case $\lambda(x_0) = \lambda$ for almost all $x_0 \in \mathbf{R}$ and

$$|x_n - x_n'| \approx e^{\lambda n} |x_0 - x_0'|.$$

We call $\lambda$ the Lyapunov exponent of $f$.

For greater technical detail, see Tong (1990), pp 58-59. Of particular interest is the material on the probability measure implied by *f*, $\rho$ say, so that the Lyapunov exponent may be calculated as $E_\rho(\log|f'(X)|)$.

*State Space*

A fundamental concept of dynamical systems theory is that of a *state* (or *state variable or space*). Loosely speaking, a state variable represents the most efficient condensation of information contained in the past and present about the future. See Tong (1990) pp186-187 for more discussion, and links with *Predictor Space*.

A standard approach to studying differential equation systems is to make a change of variable to produce a system of equations, each of which is of order 1. This is illustrated by Tong (1990) of the Van der Pol equation:

$$\ddot{x} + \left(x^2 - 1\right)\dot{x} + x = 0.$$

If we set $y = \dot{x}$ then we may rewrite this equation in the standard *state space* form with $(x, y)'$ as the state vector:

$$\dot{x} = y \qquad\qquad \dot{y} = \left(1 - x^2\right)y - x.$$

*Time Lag*

This is the parameter $\tau$ in equation (2), the main role of which to reduce the correlation between observations.

*Embedding Dimension*

This is the parameter *m* in equation (2). Embedding dimension is used find the *dimension* of the system. If the embedding dimension chosen is too small, then the true dynamics of the system will not be revealed. For example, setting $\tau = 1$, if $x_t \approx x_{t-1}$ then selecting $m = 1$ will lead to a phase space representation of $\left(x_t, x_{t-1}\right)$ as a straight line, regardless of the true dynamics. Only when a larger value is chosen will the true dynamics be revealed (see *nearest neighbours* discussion below).

*Estimating Embedding Dimension*

The discussion here is based on Sangoyomi et al. (1996).

*1. Scaling Ideas*

The simplest notion of dimension is based on the observation that a *d*-dimensional object has volume of the form $V = r^d$, where $r$ is some characteristic length. This suggests that dimension may be estimated as $\log(V)/\log(r)$. This definition can be generalised to describe a class of dimension estimators, with important special cases. These are described in detail by Sangoyomi et al. (1996).

The generalisation arises, in essence, by considering the probability measure induced on the phase space by the mapping *f*, and partitioning the data to form an empirical density estimate. The size of the partition is measured by the characteristic length *r*.

An important special case is the so-called *correlation dimension*, denoted $D_2$ by Sangoyomi et al. (1996) (who also reference algorithms for its calculation):

$$D_2 = \lim_{r \to 0} \frac{\log\left(\sum p_i^2\right)}{\log r},$$

where $p_i$ denotes the proportion of the data falling into the *i*th partition. This is a popular technique for estimating embedding dimension. $D_2$ measures the probability of finding a pair of points in a partition, and so measures variation in probability mass.

*2. Geometry and Nearest Neighbours*

Abarbanel and Kennel (1992), as described by Sangoyomi et al. (1996), introduced an idea called *false neighbours* based on simple geometrical principles to estimate the embedding dimension. To illustrate, consider the 4 points on the unit circle at 0°, 90°, 180° and 270°. Projected onto the *x*-axis there will be only 3 distinct points, since the points at 90° and 180° are projected to the same point in 1 dimension, and these points are said to be *false neighbours*. We conclude from this that 1-dimension is not a suitable phase space representation in this case. Going from 2 to 3 dimensions however we end up with no false neighbours. Consequently at $m = 2$ the dynamics of the system have *unfolded*.

Note that this procedure estimates an integer dimension $d_E$ that is the minimum needed to unfold the dynamics, and is such that $\text{ceiling}(d_A) \leq d_E \leq 2d_A + 1$.

### *3.9.2 RSS Open Meeting on Chaos*

The Royal Statistical Society held an open meeting on chaos, comprising a number of read papers and discussion. the proceedings were published in 1992, volume 54 of the Journal of the Royal Statistical Society B. Our interest is less on chaos, since it is unlikely that truly chaotic dynamics are detectable in climatic systems where noise plays an influential role. Indeed, this general point was made by a number of contributors to this meeting. It is most likely that deterministic chaos will be detectable in carefully controlled laboratory experiments. Our emphasis is on the use of dynamical systems ideas applied within a statistical modelling framework. The results described by the participants are encouraging and highly interesting. We provide a brief overview now of some of the key findings.

Smith (1992) concentrates on the problem of estimating the dimension of a dynamical system, finding that the presence of noise in moderate to large signal-to-noise ratios tends to obscure the presence of deterministic chaos. This is a useful statistical perspective, since some researchers in other fields still seem to believe that seemingly random perturbations in time can be modelled as deterministic chaos. Much of this work has met with disappointing results, which is not surprising in the result of Smith's findings. Smith concludes that dimension estimation should be a precursor to estimating the nonlinear mapping $f$ of equation (7), which will indicate if this exercise is worthwhile and some indication of the likely embedding dimension required.

Nychka et al. (1992) examine the calculation of the largest Lyapunov exponent as an indicator of chaos, and estimation of $f$ using thin-plate splines and neural networks in the statistical model

$$X_t = f(x_{t-L}, x_{t-2L}, \ldots x_{t-dL}) + e_t.$$

In this formulation the series $\{e_t\}$ represents a random shock to the system. Like Smith (1992), Nychka et al. (1992) find moderate to large noise is sufficient to obscure and deterministic chaotic signal. However, this statistical approach is an interesting combination of dynamical systems and statistical thinking. They note in their conclusion that the Lyapunov exponent as defined in their paper is a joint property of the intrinsic nonlinear map $f$ and the extrinsic random shocks. In the absence of random shocks, the attractor for the system would be different, and hence the Lyapunov exponents would be different. Their strategy in estimating $f$ is to avoid estimating the full multivariate $d$-dimensional function. Using splines, they seek approximations flexible approximations in two dimensions, searching over possible pairs of lags. Their neural network approach represents $f$ via univariate functions taking linear combinations of the lagged values as arguments.

Broomhead and Huke (1992) show that system identification using embedding methods is unaffected, at least in theory, by data filtering, provided a finite-order non-recursive filter is used. The question of the impact on numerical *estimates* of dimension is not so clear however. They describe a situation where over-estimates can result even from non-recursive filtering, and such situations should be considered when selecting a suitable filter. The authors refer also to an alternative approach based on singular value decomposition that seems to provide good results in the present of noise.

A number of other papers were presented during this session, but they were either less relevant or repeated the same points. I reference them here for completeness: Casdagli (1992, Wolff (1992), Cheng and Tong (1992).

### *3.9.3 Research Opportunities*

- Examine the feasibility of incorporating exogenous variables.
- Is it possible to build a nonlinear time series into a spatial field?
- Statistical modelling for Lall et al. (1996).
    - Links with ideas of stochastic resonance to explain occasional very large fluctuations?
- Links with EEOF (Extended EOF) and random field regression?

### 3.10 Downscaling

Downscaling refers to the process by which output from relatively coarse spatial resolution general circulation/global climate models (GCMs) is refined to scales of most relevance to decision makers. GCMs produce output at a resolution that resolves only 4 points in south-west WA (CSIRO-DARLAM), whereas scales at individual catchment levels are required by water resource planners, for example. For an example application, see Barrow and Semenov (1995). GCMs also tend to have a coarse temporal resolution, which can be an issue in some applications (Wilby and Wigley (1997)).

In their review, Wilby and Wigley (1997) cite four key downscaling approaches. Amongst the earliest were regression based approaches which generally seek to establish a relationship between sub-grid scale response variables and grid-scale resolution predictors. Regression methods in all guises are relevant here, including neural network approaches. The second method cited is termed ‘weather pattern approaches’, in which observed weather station records are related to a given weather classification scheme. Statistical methods applied include principal component and canonical correlation analysis. A third approach is provided by stochastic weather generators, which use features of the climate to drive weather simulations. An illustration of this approach as well as regression methods is provided by Semenov and Barrow (1997). Wilby and Wigley (1997) note the use of Limited Area Models as a fourth approach.

A newer method described by Hughes and Guttorp (1994b) uses a stochastic model to relate broad-scale atmospheric circulation patterns to local rainfall. The model is driven by a set of latent weather states, which links the broad and fine scales. A recent application of this model, known as a non-homogeneous hidden Markov model, to south-west WA is given by Hughes et al. (1999). The results obtained thus far using this model are very encouraging, although the technique is highly computationally intensive. This seems to be mainly due to an expensive likelihood calculation, which may benefit from the application of a fully Bayesian or pseudo-likelihood ( Robert and Titterington (1998)) approach.

#### *3.10.1 Research Opportunities*

- Computational improvements- fully Bayesian/pseudo-likelihood.
- Examining climate variability using downscaling models.
- Use of explicitly Bayesian methods to represent parameter uncertainty.
- Joint modelling of precipitation and temperature.

# 4 Other Statistical Methods of Potential Value

In this section I detail some statistical methods that I believe are of particular relevance. However, the state of the art in climatology is clearly not making use of much in contemporary statistics. I would point in particular to generalized linear models (McCullagh and Nelder, 1995) and generalized additive models (Hastie and Tibshirani, 1986), the backbone of modern applied statistics. Generalized additive models place a strong emphasis on the fitting of flexible nonparametric functions to data, and there have been many developments in this field in recent times. A climatological example is provided by Qiu and Yandell (1998).

### 4.1 Functional Data Analysis

As defined by Meiring and Nychka (1998b), this is the study of curves represented by irregular samples from the curves, which may be contaminated by measurement error. There are two principle differences with multivariate analysis. First, the underlying continuity of the domain in functional data analysis and, secondly, the smoothness of the curves.

Meiring and Nychka (1998b) discuss an example where they apply functional data analysis (FDA) to stratospheric ozone at particular locations resolved in a vertical profile. They consider a class of varying coefficient functional data models. The coefficients of a basis function expansion depend on covariates, such as the quasi-biennial oscillation and season. These models are found to be sensitive to the complex interactive effects of the covariates on the shape of the vertical ozone profile. Further details of this work are described in Meiring and Nychka (1998a).

FDA offers great potential where data naturally arise as curves, as illustrated by the above example. A number of papers have appeared in the statistical literature in recent times, such as Pezzulli and Silverman (1993), Ramsay (1991), Silverman (1996) and references therein. A detailed introduction to FDA is provided by

Ramsay and Silverman (1997). The fundamental objective of FDA is to decompose the underlying theoretical functional response into orthogonal components, or basis functions. These can be parameterised for the particular application at hand to provide physical insight to the processes involved. For example, in the vertical ozone profile application of Meiring and Nychka described above, vertical ozone $Z(a, t)$ at altitude *a* and time *t* is decomposed as

$$Z(a,t) = \sum_{j=1}^{M} c_j(t, \mathbf{X}(t)) B_j(a) + \varepsilon(a,t) \qquad (9)$$

where

$$c_j(t, \mathbf{X}(t)) = \mu_j + \varphi_j(t)\alpha(t) + f_j(\mathbf{X}(t)) + \eta_j(t). \qquad (10)$$

Each basis function $B_j(a)$ is a continuous function of altitude; $\varepsilon(a,t)$ is a residual space-time process, $\mu_j$ is the mean of the *j*th coefficient; $\varphi_j(t)$ is a periodic function of time of year, used to model the quasi-biennial oscillation (QBO) in the equatorial wind direction that follows an approximate 28 month cycle and propagates down through the stratosphere; $\alpha(t)$ is a function of time *t* and $f_j(\mathbf{X}(t))$ is a non-parametric function of the covariates $\mathbf{X}(t)$; $\eta_j(t)$ is a residual process on the coefficient scale.

On substituting (10) into (9) we see that "this model allows for an overall vertical mean, a periodically and vertically varying trend which is linear if $\alpha(t) = t$, complex interactive and non-linear effects of covariates $\mathbf{X}(t)$ on the vertical profile shape, and an overall space-time residual process."

### 4.2 Projection Pursuit

Projection pursuit in its exploratory mode may be thought of as a generalisation of PCA and SVD, which seeks to maximise the variance of linear combinations of the form $Y = \mathbf{a}^T\mathbf{X}$ for a unit vector **a**. Thus $Var_X(\mathbf{a}) = \mathbf{a}^T S\mathbf{a}$, where *S* denotes the sample variance matrix of the data **X**. This leads to the optimisation problem

$$\arg\max\ Var_X(\mathbf{a}) \qquad \text{subject to } \mathbf{a}^T\mathbf{a} = 1.$$

In projection pursuit we generalise this to

$$\arg\max\ I_X(\mathbf{a}) \ \text{ subject to } \mathbf{a}^T\mathbf{a} = 1,$$

where the function $I_X(\cdot)$ is used to describe properties of the data that are *interesting*, rather than simply the variance. Clearly this can be generalised to compare more than one field. In this more general setting there typically is no analytical solution and numerical techniques must be applied.

I am not aware of any applications of projection pursuit in the climatological literature, but for an interesting illustration of its practical application see Walden (1994). In this paper a minimum entropy metric is applied to find distributions that are not uni-modal. The 'least interesting' distribution using this metric is a normal distribution, and we seek distributions that are far removed in entropy space. For a more theoretical and philosophical treatment of projection pursuit see Jones and Sibson (1987).

The central idea of searching for interesting projections has found other applications in the literature. Principal among these are projection pursuit regression and projection pursuit density estimation. Projection pursuit regression is very similar to the idea of generalized additive models, using sums of smooth functions of the predictors to represent a regression function. Such methods offer a useful generalisation of conventional linear methods to allow more realistic modelling. For a description of projection pursuit regression see Friedman and Stuetzle (1981), whilst for an account of generalized additive models see Hastie and Tibshirani (1986). I don't see such strong applications for projection pursuit density estimation, but this technique is of potential benefit if multivariate density estimates are required. The technique is due to Friedman et al. (1984).

### 4.3 Space-Time Covariance Modelling

This is a very important theme in contemporary statistics, rather than a single method. The need for genuine spatio-temporal models in climatology has been recognised for some time. See, for example, Zhang *et al*. (1998), Glaseby (1998), Drosdowsky (1993a, 1993b) and references therein. The typical approach has been to consider spatial and temporal behaviour separately.

This area is complicated by the intricate mathematical modelling so often involved, as exemplified by Jones and Zhang (1996) for example. Indeed, it is difficult to make progress without simplifying assumptions such as separable covariance structures. That is, if we observe a random variable $Z(\mathbf{x}, t)$, where typically $\mathbf{x} \in \Re^2$ denotes spatial location and $t \in \Re^+$ denotes time, then

$$Cov\left[Z(\mathbf{x}, t), Z\left(\mathbf{x}^*, t^*\right)\right] = C_1\left(x, x^*\right)C_2\left(t, t^*\right).$$

Thus the process covariance function may be found by modelling space and time separately, which implies that they are independent. In most practical cases this is an unrealistic assumption. Examples of this approach are provided by Stein (1986) and Jones and Zhang (1996).

An alternative approach is to abandon an explicitly mathematical modelling approach for a more statistical one. In doing so we lose touch to some degree with the physics of the processes under consideration, but we gain the potential for increased physical insight. There has been a growing trend in climatology to adopt genuinely spatio-temporal data analysis tools, as seen in the work of, for example, Burkhardt and James (1998), Glaseby (1998), Hutchinson (1995), Weare and Nasstrom (1982), and Barnett (1983).

Much of the work in meteorology involves extensions to the EOF technique, which I have discussed elsewhere. In most cases this analysis assumes temporal stationarity and estimates a spatial covariance structure. In recent times the approach has been extended to include temporal structure as well– see Weare and Nasstrom (1982) for an introduction. To be readily interpretable however the process analysed must be stationary in space and time, which is rarely the case.

An alternative approach is to develop statistical models that can be fitted using available data. Nott and Dunsmuir (1998) provide an interesting introduction to this subject, in the context of modelling wind fields. They develop a model for the observed process of the form

$$Z(\mathbf{x}, t) = \mu(\mathbf{x}) + \eta(\mathbf{x}, t) + \varepsilon(\mathbf{x}, t),$$

where $\mu(\cdot)$ denotes the spatial trend, $\eta(\cdot,\cdot)$ a zero-mean spatio-temporal process and $\varepsilon(\cdot,\cdot)$ a zero-mean measurement error. Nott and Dunsmuir (1998) describe approaches based on a spatial deformation technique (Sampson (1986), Sampson and Guttorp (1992)) and a direct kernel estimation technique due to Oehlert (1993). They note a number of severe computational and theoretical drawbacks with the spatial deformation technique. Their own method is based on assumptions of local spatial stationarity, but temporal stationarity is required. They report inferior performance of the kernel technique to their own approach.

I am particularly interested in ideas of spatio-temporal modelling within the framework of stochastic differential equations, which I view as a means to bring the physical and statistical modelling together. For support of this view, see Dawson's contribution to the discussion of Eynon and Switzer (1983), where a stochastic differential equation can be written down. In climate physics problems it may be possible to do the same, then the challenge would be to derive important statistical-physical properties from such an equation. For an introduction to stochastic (partial) differential equations, see Ikeda and Watanabe (1995).

### 4.4 Penalized Discriminant Analysis

Penalized Discriminant Analysis (PDA) is an extension of classical linear discriminant analysis (LDA), developed by Hastie et al. (1995). LDA is a highly useful technique when carefully applied, but suffers from a number of deficiencies. When presented with large numbers of highly correlated predictors, to paraphrase Hastie *et al*. (1995), LDA is too flexible and tends to overfit the data. An example of an application of interest to us is in relating SST fields to observed rainfall. In cases where the class boundaries in predictor space are complex and nonlinear LDA tends to be inflexible and underfits the data. For further discussion of approaches in this latter case, see Hastie *et al*. (1994).

The idea behind PDA is that these deficiencies can be overcome by appropriate 'regularization' of the within-groups covariance matrix, denoted $\Sigma_W$. There are two distinct motivations for this:

1. When the number of predictor variables is high relative to the number of observations, we cannot reliably estimate $\Sigma_W$;
2. Even when the sample size is sufficient to reliably estimate $\Sigma_W$, coefficients of spatially smooth variables tend to be spatially rough. Interpretation would be greatly aided by a smooth version, given that the fit is not compromised.

Hastie et al. (1995) propose that $\Sigma_W$ be replaced by $\Sigma_W + \lambda\Omega$, where $\Omega$ is a "roughness"-type penalty matrix. The LDA analysis then proceeds as usual. Note that Hastie et al. (1995) go on to show equivalence of PDA with penalised versions of canonical correlation analysis and optimal scoring. A key point here is that optimal scoring provides a link between nonparametric regression and discriminant analysis, providing a rich class of modelling tools in classification. For an alternative perspective on regularization in discriminant analysis using mixtures see Hastie and Tibshirani (1996).

As noted above, PDA is potentially of use in relating SST fields to observed rainfall. Rainfall could be classified as 'Low', 'Normal' and 'High' for example and these classifications related to the available SST data. A research problem here is to choose an appropriate roughness penalty. If a suitable discriminant function, or set of functions, could be found in this way then the potential for an SOI-like index to contribute to other procedures is obvious.

### 4.5 Bayesian Hierarchical Modelling

An exciting recent development is the use of Bayesian hierarchical methods to develop hybrid physical-statistical models, which provide for a sophisticated balance between physical and statistical modelling. The idea driving these methods is that there are many sources of information available to aid understanding of physical systems. We may make use of observations of various kinds, as well as models of various sub-systems. The Bayesian hierarchical approach allows us to integrate these sources of information, including the uncertainty in each component. For a general introduction see Wikle (2003). Some examples of applying this thinking to physical processes may be found in Berliner et al. (2003), Berliner (2003) and Berliner et al. (2000).

Suppose that we are studying a physical process $P$, which may be a collection of sub-processes, with physical parameters $\eta$. In observing the process $P$ we generate data $D$ and so statistical parameters $\theta$. We assume that all of these elements are subject to uncertainty, and seek to develop a model for the joint probability distribution denoted $[D, P, \eta, \theta]$. We may apply Bayes' theorem (Bernardo and Smith, 1994 pp 2) to factorise this joint probability model as

$$[D, P, \eta, \theta] = \left[D \middle| P, \eta, \theta\right]\left[P \middle| \eta, \theta\right][\eta, \theta]. \qquad (11)$$

We may now make some modelling assumptions. In the first term, conditional on $P$ and $\theta$ there is no further information in the physical parameters $\eta$ about the data $D$. Similarly for the second term, given $\eta$ there is no further information in the statistical parameters $\theta$ on the physical process $P$. We may therefore simplify (11) to

$$[D, P, \eta, \theta] = \left[D \middle| P, \theta\right]\left[P \middle| \eta\right][\eta, \theta]. \qquad (12)$$

We see that the joint probability model is the product of a model for the data, a process model and a prior parameters model. The prior parameters model captures available information on the parameters before the data are collected. For more details see Berliner (2003). A key point to note about this so-called physical-statistical model is the interconnection between the data and process models. The physical and statistical components are coupled by conditioning the data model on the physical process $P$.

It can be shown that the distribution of the process and parameters conditional on the data, the so-called posterior distribution, is such that

$$[\eta, P, \theta | D] \propto \left[D \middle| P, \theta\right]\left[P \middle| \eta\right][\eta, \theta]. \qquad (13)$$

In this way we can learn about the physical parameters through observation.

Algorithms for fitting physical-statistical models represent an active area of research. Campbell (2005) uses the importance sampling Monte Carlo approach of Berliner et al. (2003). This requires us to generate a relatively small ensemble from the prior parameters model, and pass each member of the ensemble through the physical process model. This physical process ensemble is then resampled so that a much larger sample drawn approximately from the posterior distribution (13) is obtained. This is done by assigning probabilities to each member of the ensemble, calculated using the observed data, and then sampling from them with replacement. Ensembles close to the observed data will be assigned a relatively high probability.

# References


Abarbanel, H. D. I. and Kennel, M. B. (1992). Local false nearest neighbours and dynamical dimensions from observed chaotic data. *Phys. Rev.*, **E47,** 3057-3068.

Abarbanel, H. D. I. and Lall, U. (1996). Nonlinear dynamics of the Great Salt Lake: system identification and prediction. *Clim. Dyn.,* **12,** 287-297.

Banfield, J. D. and Raftery, A. E. (1993). Model-based Gaussian and non-Gaussian clustering. *Biometrics,* **49,** 803-821.

Barnett, T. P. (1983). Interaction of the Monsoon and Pacific trade wind system at interannual time scales. Part I: The equatorial zone. *Mon. Wea. rev.,* **111,** 756-773.

Barrow, E. M. and Semenov, M. A. (1995). Climate change scenarios with high spatial and temporal resolution for agricultural applications. *Forestry,* **68,** 349-360.

Bates, B. C. and Charles, S. P. (1998). Downscaling climate model simulations: A hydrologic perspective. CSIRO Land and Water, Technical Report

Berliner, L. M. (1993). Hierarchical Bayesian time series models. *$XV^{th}$ Workshop on Maximum Entropy and Bayesian Methods*.

Berliner, L. M., Wikle, C. K. and Cressie, N. (2000). Long-lead prediction of Pacific SST via Bayesian dynamic modeling. *Journal of Climate*, **13,** 3953-3968.

Berliner, L. M. (2003). Physical-statistical modeling in geophysics. *Journal of Geophysical Research-Atmospheres,* 108 (D24).

Berliner, L. M., Milliff, R. F. and Wikle, C. K. (2003). Bayesian hierarchical modeling of air-sea interaction. *Journal of Geophysical Research,* **108,** 1-1:1-18.Bretherton, C. S., Smith, C. and Wallace, J. M. (1992). A intercomparison of methods for finding coupled patterns in climate data. *J. Climate,* **5,** 541-560.

Bernardo, J. M. and Smith, A. F. M. (1994) *Bayesian Theory,* John Wiley & Sons, Chichester.

Broomhead, D. S. and Huke, J. P. (1992). Linear filters and non-linear systems. *J. R. Statist. Soc. B,* **54,** 373-382.

Campbell, E. P. (2005). A note on statistical modelling in nonlinear systems. *Journal of Climate,* **18,** 3099-3110.

Casdagli, M. (1992). Chaos and deterministic versus stochastic non-linear modelling. *J. R. Statist. Soc. B,* **54,** 303-328.

Casey, T. (1995). Optimal linear combination of seasonal forecasts. *Aust. Met. Mag.,* **44,** 219-224.

Charles, S. P., Hughes, J. P., Bates, B. C. and Lyons, T. J. (1996). Assessing downscaling models for atmospheric circulation - local precipitation linkage. *Int. Conf. Water Resour. & Environ. Res.: Towards the 21st Century*. Water Resources Research Centre, Kyoto University, Kyoto, Japan,.

Cheng, B. and Tong, H. (1992). On consistent nonparametric order determination and chaos. *J. R. Statist. Soc. B,* **54,** 427-450.

Cherry, S. (1996). Singular value decomposition analysis and canonical correlation analysis. *J. Clim.,* **9,** 2003-2009.

Cohen, A. and Jones, R. H. (1969). Regression on a random field. *J. Am. Statist. Assoc.,* **64,** 1172-1182.

Cutler, A. and Breiman, L. (1994). Archetypal analysis. *Technometrics,* **36,** 338-347.

Drosdowsky, W. (1993a). An analysis of Australian seasonal rainfall anomalies: 1950-1987. I: Spatial patterns. *Int. J. Climatol.,* **13,** 1-30.

Drosdowsky, W. (1993b). An analysis of Australian seasonal rainfall anomalies: 1950-1987. II: Temporal variability and teleconnection patterns. *Int. J. Climatol.,* **13,** 111-149.

Drosdowsky, W. (1993c). Potential predictability of winter rainfall over southern and eastern Australia using Indian Ocean sea-surface temperature anomalies. *Aust. Met. Mag.,* **42,** 1-6.

Drosdowsky, W. (1994). Analog (nonlinear) forecasts of the Southern Oscillation index time series. *Weather and Forecasting,* **9,** 78-84.

Drosdowsky, W. and Chambers, L. (1998). Near global sea surface temperature anomalies as predictors of Australia seasonal rainfall. Bureau of Meteorology Research Centre, Research Report 65.

Drosdowsky, W. and Williams, M. (1991). The Southern Oscillation in the Australian region. Part I: Anomalies at the extremes of the Oscillation. *J. Climate,* **4,** 619-639.

Epstein, E. S. (1969). Stochastic dynamic prediction. *Tellus,* **21,** 739-759.

Eynon, B. P. and Switzer, P. (1983). The variability of rainfall acidity. *Can. J. Statist.,* **11,** 11-24.

Friedman, J. H. and Rafsky, L. C. (1981). Graphics for the Multivariate Two-Sample Problem. *J. Am. Statist. Assoc.,* **76,** 277-287.

Friedman, J. H. and Stuetzle, W. (1981). Projection pursuit regression. *J. Am. Statist. Assoc.,* **76,** 817-823.

Friedman, J. H., Stuetzle, W. and Schroeder, A. (1984). Projection pursuit density estimation. *J. Am. Statist. Assoc.,* **79,** 599-608.

Glaseby, C. A. (1998). Modelling multivariate spatio-temporal weather data using latent Gaussian processes. *7th International Meeting on Statistical Climatology*. Whistler, BC, Canada.

Gleeson, T. A. (1966). A causal relation for probabilities in synoptic meteorology. *J. Appl. Meteorol.,* **5,** 365-368.

Green, P. J. (1995). Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika,* **82,** 711-732.

Hastie, T., Buja, A. and Tibshirani, R. (1995). Penalized discriminant analysis. *Ann. Statist.,* **23,** 73-102.

Hastie, T. and Tibshirani, R. (1986). Generalized additive models. *Statist. Sci.,* **1,** 297-318.

Hastie, T. and Tibshirani, R. (1996). Discriminant analysis by Gaussian mixtures. *J. R. Statist. Soc. B,* **58,** 155-176.

Hastie, T., Tibshirani, R. and Buja, A. (1994). Flexible discriminant analysis by optimal scoring. *J. Am. Statist. Assoc.,* **89,** 1255-1270.

Horel, J. D. (1984). Complex principal component analysis: Theory and examples. *J. Clim. Appl. Meteorol.,* **23,** 1660-1673.

Huang, J., Higuchi, K. and Shabbar, A. (1998). Multiresoulution spectral analysis and its application to studying the relationship between NAO and ENSO. *7th International Meeting on Statistical Climatology*. Whistler, BC, Canada.

Hughes, J. P. and Guttorp, P. (1994b). A class of stochastic models for relating synoptic atmospheric patterns to regional hydrologic phenomena. *Water Resour. Res.,* **30,** 1535-1546.

Hughes, J. P. and Guttorp, P. (1994a). A class of stochastic models for relating synoptic atmospheric patterns to regional hydrologic phenomena. *Water Resour. Res.,* **30,** 1535-1546.

Hughes, J. P. and Guttorp, P. (1994c). Incorporating spatial dependence and atmospheric data in a model of precipitation. *J. Appl. Meteorol.,* **33,** 1503-1515.

Hughes, J. P., Guttorp, P. and Charles, S. P. (1999). A non-homeogeneous hidden Markov model for precipitation occurrence. *Appl. Statist.,* **48,** 15-30.

Hutchinson, M. F. (1995). Stochastic space-time weather models from ground-based data. *Agric. For. Meteorol.,* **73,** 237-264.

Ikeda, N. and Watanabe, S. *Stochastic Differential Equations and Diffusion Processes,* North-Holland/Kodansha, Tokyo.

Jolliffe, I. T. (1989). Rotation of ill-defined principal components. *Appl. Stats.,* **38,** 139-147.

Jones, M. C. and Sibson, R. (1987). What is projection pursuit? *J. R. Statist. Soc. A,* **150,** 1-36.

Jones, R. H. and Zhang, Y. (1996) Models for continuous stationary space-time processes , Unpublished Manuscript, .

Kroonenberg, P. M., Harch, B. D., Basford, K. E. and Cruickshank, A. (1997). Combined analysis of categorical and numerical descriptors of Australian groundnut accessions using nonlinear principal components analysis. *J. Ag. Biol. Env. Statist.,* **2,** 294-312.

Lall, U. and Mann, M. (1995). The Great Salt Lake: A barometer of low-frequency climatic variability. *Water Resour. Res.,* **31,** 2503-2515.

Lall, U., Moon, Y.-I. and Bosworth, K. (1995) Locally weighted polynomial regression: Parameter choice and application to forecasts of the Great Salt Lake , Draft manuscript, .

Lall, U., Sangoyomi, T. and Abarbanel, H. D. (1996). Nonlinear dynamics of the Great Salt Lake: Nonparameteric short-term forecasting. *Water Resour. Res.,* **32,** 975-985.

McCaffrey, D. F., Ellner, S., Gallant, A. R. and Nychka, D. W. (1992). Estimating the Lyapunov exponent of a chaotic system with nonparametric regression. *J. Am. Statist. Assoc.,* **87,** 682-695.

McCullagh, P. and Nelder, J. A. (1995) *Generalized Linear Models,* Chapman & Hall, London.

Meiring, W. and Nychka, D. W. (1998a). Functional data analysis for vertical profiles. *Interface 98*.

Meiring, W. and Nychka, D. W. (1998b). Functional data analysis of vertical Ozone profiles. *7th International Meeting on Statistical Climatology*. Whistler, BC, Canada.

Nason, G. (1995). Three-dimensional projection pursuit. *Appl. Statist.,* **44,** 411-430.

Nicholls, N. (1986). Use of the southern oscillation to predict Australian sorghum yield. *Agric. For. Meteorol.,* **38,** 9-15.

Nicholls, N. (1987). The use of canonical correlation to study teleconnections. *Mon. Wea. Rev.,* **115,** 393-399.

Nicholls, N. (1989). Sea surface temperatures and Australian winter rainfall. *J. Clim.,* **2,** 965-973.

Nicholls, N. (1991). The El Niño / Southern Oscillation and Australian vegetation. *Vegetatio 91: Vegetation and climate interactions in semi-arid areas*.

Nicholls, N. and Katz., R. W. (1991) In *Teleconnections linking worldwide climate anomalies*(Eds, Glantz, H. and Nicholls, N.) Cambridge University Press, New York, pp. 511-525.

Nott, D. and Dunsmuir, W. T. M. (1998). Analysis of spatial covariance structure from monitoring data. Department of Statistics, UNSW, Technical Report S98-6.

Nychka, D., Ellner, S., Gallant, A. R. and McCaffrey, D. (1992). Finding chaos in noisy systems. *J. R. Statist. Soc. B,* **54,** 399-426.

Obled, C. and Creutin, J. D. (1986). Some developments in the use of empirical orthogonal functions for mapping meteorological fields. *J. Clim. Appl. Meteorol.,* **25,** 1189-1204.

Oehlert, G. W. (1993). Regional trends in sulfate wet deposition. *J. Am. Statist. Assoc.,* **88,** 390-399.

Packard, N. H., Crutchfield, J. P., Farmer, J. D. and Shaw, R. S. (1980). Geometry from a time series. *Phys. Rev. Lett.,* **45,** 712-716.

Paterson, J. G., Goodchild, N. A. and Boyd, W. J. R. (1978). Classifying environments for sampling purposes using a principal component analysis of climatic data. *Agric. Meteorol,* **19,** 349-362.

Pezzulli, S. and Silverman, B. W. (1993). Some properties of smoothed principal components analysis for functional data. *Comp. Statist.,* **8,** 1-16.

Prohaska, J. (1976). A technique for analyzing the linear relationship between two meteorological fields. *Mon. Wea. Rev.,* **104,** 1345-1353.

Qiu, P. and Yandell, B. (1998). A local polynomial jump-detection algorithm in nonparametric regression. *Technometrics,* **40,** 141-152.

Ramsay, J. O. and Silverman, B. W. (1997) *Functional Data Analysis,* Springer-Verlag, New York.

Ramsay, J. O. D., C.J. (1991). Some tools for functional data analysis. *J. R. Statist. Soc. B,* **53,** 539-572.

Richardson, S. and Green, P. J. (1997). On Bayesian analysis of mixtures with an unknown number of components. *J. R. Statist. Soc. B,* **59,** 731-792.

Richman, M. B. (1986). Rotation of principal components. *J. Climatol.,* **6,** 293-335.

Rimmington, G. M. and Nicholls, G. M. (1993). Forecasting wheat yields in Australia with the southern oscillation index. *Aust. J. Agric. Res.,* **44,** 625-32.

Robert, C. P. and Titterington, D. M. (1998). Reparametrization strategies for hidden Markov models and Bayesian approaches to maximum likelihood estimation. *Statistics and Computing,* **8,** 145-158.

Ruprecht, J. K., Bates, B. C., Stokes, R. A. and eds. (1996). Climate Variability and Water Resources Workshop. Water and Rivers Commission, Technical Report series No WRT 5.

Russell, J. S., McLeod, I. M., Dale, M. B. and Valentine, T. R. (1993). The southern oscillation index as a predictor of seasonal rainfall in the arable areas of the inland Australian subtropics. *Aust. J. Agric. Research,* **44,** 1337-49.

Sampson, P. D. (1986). Spatial covariance estimation by scaled-metric scaling and biorthogonal grids. Department of Statistics, University of Washington, Technical Report 91.

Sampson, P. D. and Guttorp, P. (1992). Nonparametric estimation of nonstationary spatial covariance structure. *J. Am. Statist. Assoc.,* **87,** 108-119.

Sangoyomi, T. B., Lall, U. and Abarbanel, H. D. I. (1996). Nonlinear dynamics of the Great Salt Lake: Dimension estimation. *Water. Resour. Res.,* **32,** 149-159.

Seber, G. A. F. (1984) *Multivariate Observations,* John Wiley & Sons, Brisbane.

Semenov, M. A. and Barrow, E. A. (1997). Use of a stochastic weather generator in the development of climate change scenarios. *Clim. Change,* **35,** 397-414.

Silverman, B. W. (1996). Smoothed functional principal components analysis by choice of norm. *Ann. Statist.,* **24,** 1-24.

Smith, I. (1994). Indian Ocean sea-surface temperature patterns and Australian winter rainfall. *Int. J. Climatol.,* **14,** 287-305.

Smith, R. L. (1992). Estimating dimension in noisy chaotic time series. *J. R. statist. Soc. B,* **54,** 329-351.

Stein, M. (1986). A simple model for spatial-temporal processes. *Water Resour. Res.,* **22,** 2107-2110.

Sugihara, G. and May, R. M. (1990). Nonlinear forecasting as a way of distinguishing chaos from measurment error in time series. *Nature,* **344,** 734-741.

Tong, H. (1990) *Non-linear time series. A dynamical systems approach,* Oxford University Press, New York.

Walden, A. T. (1994). Spatial clustering: using simple summaries of seismic data to find the edge of an oil-field. *Appl. Statist.,* **43,** 385-398.

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

Weare, B. C. and Nasstrom, J. S. (1982). Examples of extended empirical orthogonal function analyses. *Monthly Weather Review,* **110,** 481-485.

Whitcher, B., Percival, D. B. and Guttorp, P. (1998). Bivariate wavelet analysis with an application to the Madden-Julian oscillation. *7th International Meeting on Statistical Climatology*. Whistler, BC, Canada.

Wikle, C. K. (2003). Hierarchical models in environmental science. *International Statistical Review,* **71,** 181-199.

Wilby, R. L. and Wigley, T. M. L. (1997). Downscaling general circulation model output: a review of methods and limitations. *Prog. Phys. Geog.,* **21,** 530-548.

Wolff, R. C. L. (1992). Local Lyapunov exponents: looking closely at chaos. *J. R. Statist. Soc. B,* **54,** 353-372.

Wolter, K. (1987). The Southern Oscillation in surface circulation and climate over the tropical Atlantic, eastern Pacific, and Indian Oceans as captured by cluster analysis. *J. Clim. Appl. Meteorol.,* **26,** 540-558.

Zhang, X., Sheng, J. and Shabbar, A. (1998). Modes of interannual and interddecadal variability of Pacific SST. *J. Climate*, **11,** 2556-2569.