IBC 94

XVIIth International Biometric Conference

HAMILTON, ONTARIO CANADA

8-12 AUGUST 1994

PROCEEDINGS

Volume 1:

INVITED PAPERS

Challenges in Multivariate Spatio-Temporal Modeling

Colin Goodall[1] and Kanti V. Mardia[2]
[1] Department of Statistics, Pennsylvania State University, University Park, PA 16802, USA
[2] Department of Statistics, University of Leeds, Leeds, LS2 9JT, UK

SUMMARY

We first review recent modeling approaches to spatio-temporal data. Next we introduce a model that combines effectively two well established approaches of spatial statistics and time series — one of Kriging in the field of spatial data and the other of the Kalman filter for time series data. We call our approach to this spatio-temporal state space model the spatio-temporal Kalman filter. The spatial field components in the spatio-temporal Kalman filter may take a variety of forms, including empirical orthogonal functions and so-called principal fields. We discuss some general properties of the model, and emphasize that the model allows prediction in space as well as in time, and using observations at sites that may differ from one time to the next. We emphasize the selection of, and diagnostics for, and time-evolution of the spatial fields. Multivariate observations are accommodated through a common set of spatial fields, and a matrix state. A simulated example is given, to provide insight into the spatio-temporal Kalman filter.

1 Introduction

Many examples exist of data that can be viewed as spatial and temporal processes. For example in meteorology sets of rainfall measurements are available from a number of sites over a period of time. Estimation of variables at unsampled times or unsampled spatial locations requires the extension of existing techniques into the space-time domain. Such work dates back at least to the 1860's when, as Haslett (1989) points out, Francis Galton (1863) collected data on weather from stations throughout Europe, from which he was able to conclude that the atmosphere is dominated by vast, coherent weather systems. We give here a brief review of the topic from the modeling point of view. For excellent recent overviews we refer the readers to Haslett (1989), to Rouhani and Myers (1990), and to Guttorp and Sampson (1994).

In some instances the extension into the space-time domain has been avoided. If the time component were integrated out then we are left with the problem of interpolation and prediction in space only, for which the method of kriging is appropriate. Similarly the spatial dependence could be removed, leading to the use of time series methods. Neither approach is entirely satisfactory. In particular it should be established whether spatial or temporal dependence is dominant before one is removed.

Another solution is to consider the spatio-temporal phenomenon as a realization of a random function in $d + 1$ dimensions (d dimensions in space plus one for time). Although in appearance straightforward, this presents both theoretical and practical problems as there are major differences between temporal and spatial phenomena (see, e.g., Rouhani and Myers (1990)):

1. One dimensional temporal data is ordered, contrary to spatial data. Also, space and time are measured on two different scales which cannot be compared.

2. The locations available are often sparse, whereas the time series associated with each are long. Or infrequent synoptic coverage (images) may be available. The accuracies of estimated temporal and spatial structures are likely to be quite different.

3. Space-time data can often exhibit temporal periodicity and spatial non-stationarity.

In response to these problems Rouhani and Wackernagel (1990) took the time series at each measurement point as separate, but correlated. Thus they focussed on the dimension richest in information, which allows backwards and forwards forecasting, but not spatial prediction. When considering environmental monitoring data, one can remove the effect of time and then view the data as repeated measurements in space (see Loader and Switzer (1992), Sampson and Guttorp (1992), and Mardia and Goodall (1993)). Random effects can also be included, as in the analysis of longitudinal data with either temporal correlation (Zeger and Diggle (1994) and Donnelly, Laird and Ware (1993)), or spatial correlation (Donnelly, Ware and Laird (1994)). For an interesting state space view of longitudinal modeling, see Jones (1993). The reader is also referred to related papers in Patil and Rao (1993), and also Haslett *et al* (1991).

A general approach is taken by Mardia and Goodall (1993) who write down an $(nTm \times nTm)$ covariance matrix Σ, where n is the number of sites, T the number of time points, and m the dimensionality of the observation at each site and time. This matrix is then simplified in stages, first to a Kronecker product that separates out the temporal part, $\Sigma = \Sigma_T \otimes \Sigma_{NM}$, and then to a Kronecker of three matrices $\Sigma = \Sigma_N \otimes \Sigma_T \otimes \Sigma_M$, where Σ_T is $T \times T$, Σ_{NM} is $nm \times nm$, Σ_N is $n \times n$, and Σ_M is $m \times m$. Oelhert (1993) also uses a factored covariance model. While the factorization appears very general, there is a possibly significant drawback associated with the factored structure, namely extra symmetries, e.g.

$$\Sigma_{ijk,i'j'k'} = \Sigma_{ij'k,i'jk'} . \tag{1.1}$$

Consider a spatio-temporal field $X(\mathrm{s}, t)$, where s is the location within the geographical domain of interest, D, and t indexes time. A variety of models for X have been considered. The simplest model, see for example Rouhani and Myers (1990), would be

$$X(\mathrm{s}, t) = Z(\mathrm{s}) + U(t) \ \text{ or } \ X(\mathrm{s}, t) = Z(\mathrm{s})U(t) \tag{1.2}$$

where $Z(\mathrm{s})$ and $U(t)$ are assumed independent, or at least uncorrelated. Rouhani and Myers (1990) highlight the special cases where one component or the other is deterministic, and give examples of such models from the literature. Haslett (1989) reviews various methods, including the method of optimal statistical objective analysis (OSOA), closely related to kriging, in which

$$X(\mathrm{s}, t) = P(\mathrm{s}, t) + \varepsilon(\mathrm{s}, t). \tag{1.3}$$

$P(\mathrm{s}, t)$ is a numerical prediction ('first guess') based on past values of $X(\mathrm{s}, t)$; $\varepsilon(\mathrm{s}, t)$ is the first guess error. Given the differences between the observations and the first guess predictions, together with a first guess at a new site s_0, the objective is to predict $\varepsilon(\mathrm{s}_0, t)$. The prediction is then used to obtain a refined estimate of $X(\mathrm{s}_0, t)$. Haslett's emphasis is on spatial, not temporal, structure.

Mardia and Goodall (1993), among others, consider models for a multivariate response in which

$$\mathbf{X}(\mathrm{s}, t) = \boldsymbol{\mu}(\mathrm{s}, t) + \boldsymbol{\varepsilon}(\mathrm{s}, t) \tag{1.4}$$

with $E[\mathbf{X}(\mathrm{s}, t)] = \boldsymbol{\mu}(\mathrm{s}, t)$ the set of trend surfaces, and $\boldsymbol{\varepsilon}$ a vector-valued mean zero second-order stationary temporal-spatial Gaussian process with covariance Σ. Together with the possible factorizations of Σ noted above, other semiparametric strategies are considered to remove the effect of time and thus view the data as repeated measurements in space.

Høst *et al* (1994) utilize the decomposition

$$X(\mathrm{s}, t) = \mu(\mathrm{s}, t) + S(\mathrm{s}, t)U(\mathrm{s}, t) \tag{1.5}$$

where μ, S and U are spatio-temporal random fields representing mean, standard deviation and residual respectively, taken to be mutually independent. μ and S are further decomposed as sum and product respectively, as in equation (1.4). The aim is to obtain more accurate representation of spatial interpolation errors.

Some of the ingredients in modeling spatial temporal multivariate monitoring and image data are the following.

1. A spatial model, which may be second-order stationary or intrinsic stationary.
2. A temporal model, which, unlike the spatial model, is not reversible.
3. A multivariate model.
4. A model for images.

These are interrelated in potentially complicated ways. For example, an expression for the cross-covariogram and cross variograms have been written down for multivariate data (Cressie (1993)), and simultaneous ARMA modeling in space and time (STARMA) has been investigated (e.g. Pfeifer and Deutsch 1980a,b). A general model should allow for the following, in addition to points 1-4 above.

5. Data sampled at different sites at different times, as with the interleaved hexagonal subgrids envisaged in the design of EMAP (Environmental Monitoring and Assessment Program) (Overton *et al*, 1990).
6. Synoptic coverage (images), most likely at infrequent intervals
7. Prediction at arbitrary spatial locations at arbitrary, possibly future times
8. The possibility that data occur at irregular intervals of time
9. Short series of data, e.g. annual samples over 5 years, and medium length sequences, e.g. monthly monitoring over 5 years, and long sequences, e.g. daily monitoring.

Here we begin to develop and to investigate a general structure for handling such data. Fundamentally, the approach is to model each data source as a linear combination of spatial fields, that are either fixed or are slowly changing in time. The linear combination is the state of the system, and its changes with time is modeled according to a general state space model. This resembles a random effects model (the linear combination is random), with the errors uncorrelated from time to time, but not, in general, uncorrelated in space. The overall broad objectives of this approach are as follows:

(a) Spatial prediction at a time for which data are available,
(b) Spatial-temporal prediction — typically at a future time,
(c) Network design, including design in both space and time, and
(d) Data description and summary through model fitting.

2 The Kalman filter for spatio-temporal data

Kriging has proved to be an effective tool in spatial statistics whereas the Kalman filter has proved itself as an established technique in time series analysis. We incorporate both these in our approach to spatio-temporal data and call it the spatial-temporal Kalman filter. Green and Titterington (1988) have introduced a Kalman filter for restoring sequences of images, but their model, discussed below, is complementary in the sense that their state vector is a spatial field unlike in our formulation. A method closely related to the spatio-temporal Kalman filter is used in numerical weather prediction, as pointed out by Haslett (1989).

We decompose the data into trend and correlated error components:

$$X(\mathrm{s},t) = \mu(\mathrm{s},t) + \varepsilon(\mathrm{s},t). \tag{2.1}$$

The covariance $\mathrm{cov}(\varepsilon(\mathrm{s},t),\ \varepsilon(\mathrm{s}',t'))$ might take one of several forms, involving all four arguments, $\Sigma(\mathrm{s},\mathrm{s}',t,t')$, or differences (error stationary in space and time), $\Sigma(\mathrm{s}'-\mathrm{s},\ \ t'-t)$. Here we assume that $\mu(\mathrm{s},t)$ contains all the temporal association — the error covariance is then $\Sigma(\mathrm{s},\mathrm{s}')$, or just $\Sigma(\mathrm{s}'-\mathrm{s})$ in the spatial stationary case of no time dependence. We might also assume that the $\varepsilon(\mathrm{s},t)$ are pure measurement error, and are uncorrelated.

The trend component must accommodate various types of changing pattern with time.

i. Periodic fluctuations.

ii. Steady change in a pattern.

iii. Innovation, transformation, migration, and removal of features.

Periodic fluctuation immediately points to a multivariate time series model, which can be expressed as an instance of the general state space model, and hence approached using the Kalman filter. Additional flexibility is usefully provided by the extended, or nonlinear, Kalman filter. We start with the extended setup, and write the trend surface as a sum of p products

$$\mu(\mathrm{s},t)=h_1(\mathrm{s},t)\alpha_1(t)+h_2(\mathrm{s},t)\alpha_2(t)+\ldots+h_p(\mathrm{s},t)\alpha_p(t)=\mathrm{h}(\mathrm{s},t)^T\boldsymbol{\alpha}(t), \tag{2.2}$$

say. Each component $h_i(s,t)\alpha_i(t)$ includes a slowly changing factor $h_i(s,t)$, and a more rapidly changing factor, $\alpha_i(t)$, which has *no* dependence on space.

For example, in monthly data, h may have two components, $h_1(\mathrm{s},t)$ containing typical winter pattern and $h_2(\mathrm{s},t)$ containing typical summer pattern. The quantities $\alpha_1(t)$ and $\alpha_2(t)$ follow annual cycles, 180° out of phase — from winter to summer $\alpha_1(t)$ decreases while $\alpha_2(t)$ increases. Changes in h_1 and h_2 reflect longer term evolution, spanning several years. If a pattern other than $h_1(\mathrm{s},t)$ and $h_2(\mathrm{s},t)$ were to occur but not persist, then that pattern would be subsumed in the noise $\varepsilon(\mathrm{s},t)$. Or, if the pattern recurs, perhaps sporadically, then a third field $h_3(\mathrm{s},t)$ may be introduced.

The formula

$$X(\mathrm{s},t)=\mathrm{h}(\mathrm{s},t)^T\boldsymbol{\alpha}(t)+\varepsilon(\mathrm{s},t) \tag{2.3}$$

provides a (linear) low-dimensional summary of the trend part of the random field $X(\mathrm{s},t)$. However, a finite linear combination of fields $h_i(s,t)$ cannot always be found. In the case of image sequences, a typical test problem (Hainsworth and Mardia (1992), Sutherland and Titterington (1994)) is the rotation of a black rectangle against a white background, with noise added. For that, a spatial mapping is used, e.g.

$$\mu(\mathrm{s},t)=h_0(\mathrm{s},t)+h_1(g_{\boldsymbol{\alpha}(t)}(\mathrm{s}),t) \tag{2.4}$$

where $\boldsymbol{\alpha}(t)$ contains the parameters of an affine transformation, $\mathrm{s}\mapsto g_{\boldsymbol{\alpha}(t)}(\mathrm{s})$, as in Hainsworth and Mardia. Or $\boldsymbol{\alpha}(t)$ may specify a realization of a deformable template $h_1(\boldsymbol{\alpha}(t);\ .\ )$:

$$\mu(s,t)=h_0(s,t)+h_1(\boldsymbol{\alpha}(t);s,t), \tag{2.5}$$

as in Sutherland and Titterington.

A second example of the use of equation (2.2) is in modeling deseasonalised data, where $h_1(s,t)$ might be a mean field, and $h_2(s,t)$ differences from the mean. The coefficients $\alpha_2(t)$ may fluctuate among both negative and positive values. A systematic trend in time in $X(\mathrm{s},t)$ might be captured in the slow evolution of $h_1(\mathrm{s},t)$, or in the sequence $\{(\alpha_1(t),\alpha_2(t)),\ t=1,\ldots,T\}$.

A Bayesian perspective on the Kalman filter is well known, see e.g. Meinhold and Singpurwalla (1983). The general probability model for the extended Kalman filter in equation (2.2) may be approached as follows. Let $\mathrm{x}(t)$ denote the vector of observations at time t. Let $\mathrm{x}(\leq t)$ denote the collection of such vectors of observations $\{\mathrm{x}(t') : t' = 1, \ldots, t\}$. Let $\mathcal{A}(t)$ denote the state space event at time t, $\mathcal{A}(t) = \{\mathrm{h}(\mathrm{s}, t), \alpha(t) : \mathrm{s} \in \mathcal{D}\}$ for spatial domain $\mathcal{D}$. Then

$$P[\, \mathcal{A}(t) \mid \mathrm{x}(\leq t), \mathcal{A}(0)\,] \;\propto\; P[\, \mathrm{x}(t) \mid \mathcal{A}(t)\,]\; P[\, \mathcal{A}(t) \mid \mathcal{A}(t-1)\,]\; P[\, \mathcal{A}(t-1) \mid \mathrm{x}(\leq t-1), \mathcal{A}(0)\,]. \quad (2.6)$$

This equation makes explicit the time dependent structure of the observed $\mathrm{x}(t)$. The entire p-vector of states $\alpha(t)$, and the entire p-vector of spatial fields enter into the equation for the probability model. Using a Markov random field structure for spatial dependence, there may be some simplification. A Monte Carlo Markov chain is one type of approach suitable also for image sequence problems, such as for deformable templates.

Alternatively, with distributional structure, the extended Kalman filter is a recursion for estimating the $\alpha(t)$ and $\mathrm{h}(\mathrm{s}, t)$ in this setting, and also for predicting $\alpha(t)$ and $\mathrm{h}(\mathrm{s}, t)$ at future times. The extended Kalman filter for this setup can be written with two sets of state equations (for a synopsis on the observation and state equations of the general state space model, see Section 2.2):

$$\begin{aligned} \alpha(t) &= P\,\alpha(t-1) \;+\; \eta(t) \\ \mathrm{h}(\mathrm{s}, t) &= Q\,\mathrm{h}(\mathrm{s}, t-1) \;+\; \zeta(\mathrm{s}, t), \end{aligned} \quad (2.7)$$

or these might be combined into a single state equation. The extended Kalman filter may include also state equations for the covariance, or the coefficients of a variogram or covariogram. More general forms of the (linear) Kalman filter, to include multivariate and image data are noted below.

A solution might be attempted by alternating between two (linear) Kalman filters, one for $\mathrm{h}(\mathrm{s}, t)$ and the other for $\alpha(t)$. There are two issues concerning the identification of $\mathrm{h}_i(\mathrm{s}, t)$ and $\alpha_i(t)$ in each product. The simpler issue is the relative magnitudes of $\mathrm{h}(\mathrm{s}, t)$ and $\alpha(t)$. We might constrain each field $h_i(\mathrm{s}, t)$ to have mean one integrated over the spatial domain $\mathcal{D}$ at each occasion t. The more complex issue is to establish the usefulness of $\alpha(t)$. The key is that $\mathrm{h}(\mathrm{s}, t)$ should be slowly varying. One possibility is to bound the coefficient of variation of $\mathrm{h}(\mathrm{s}, t)$. A second is to 'smooth' $\mathrm{h}(\mathrm{s}, t)$ in time more heavily — this is typically done by appending $\mathrm{h}(\mathrm{s}, t-1)$, $\mathrm{h}(\mathrm{s}, t-2)$, ..., etc. to the state vector; the full version is used in the state equations, but $\mathrm{h}(\mathrm{s}, t)$ appears in the observation equation. In some cases a third possibility, that $\alpha(t)$ has defined periodicity (12 months say), can be used: for a related technique in seasonal decomposition of time series see Cleveland and Devlin (1982).

2.1 The Kalman filter for $\mathrm{h}(\mathrm{s}, t)$ only

The ordinary, or linear, Kalman filter provides best linear unbiased predictions when either $\mathrm{h}(\mathrm{s}, t)$ or $\alpha(t)$ is treated as fixed, non-time varying. For image sequences, Green and Titterington (1988) set $p = 1$, $\alpha_t = 1$, and $h(\mathrm{s}, t)$ to be the true scene, with the state equations

$$h(\mathrm{s}, t) = \sum_{\mathrm{s}' \in \mathcal{I}} Q(\mathrm{s}, \mathrm{s}')h(\mathrm{s}', t-1) + \zeta(\mathrm{s}) \quad (2.8)$$

where the summation is over all pixels in the image $\mathcal{I}$, and Q is a transition matrix that, in the sth row, typically gives greatest weight to the neighbours s$'$ of s . Neighbourhood structure is initially introduced into $h(\mathrm{s}, t_0)$ through a Markov random field prior on $h(\mathrm{s}, t_0)$. As an alternative to maximum *a posteriori* (MAP) estimation, a series of simulations from the prior on $h(\mathrm{s}, t_0)$ can be used, and followed through the Kalman filter recursions.

A major concern in this approach — see also Titterington (1990) — is computational; Green and Titterington (1988) give some algorithms that appear computationally feasible, although still to be thoroughly explored in practical applications. Approaches that put greater structure into the images, e.g.

deformable templates (Sutherland and Titterington (1994), Phillips and Smith (1994)), are promising both computationally and from the modeling stand point.

Neighbourhood dependence is introduced in $h(\mathrm{s}, t_0)$ and (possibly) maintained through Q. Titterington (personal communication) has suggested an additional penalty term (in equation (2.6)) to reinforce neighbourhood dependence at each step. A different approach is that of Mardia *et al* (1992), Hainsworth & Mardia (1993) whose $k - ICM$ algorithm uses a space-time neighbourhood. For space-time *smoothing* of regular and irregular data, see Goodall and Phelan (1990).

2.2 The spatio-temporal Kalman filter

We turn now to the Kalman filter for $\alpha(t)$ only. In our setting of (multivariate) spatial-temporal data, it is most natural, and practicable, to fix the h_i to be functions of space only, $h_i = h_i(\mathrm{s})$, and then to model the state vector $\alpha(t)$ by a set of state equations. Measurement errors must be independent in time, but may be correlated in space. One major practical advantage is that there is no requirement that data is obtained at the same sites at all times. A challenge is to find a set of suitable $h_i(\mathrm{s})$. We address that challenge in Section 3, after first reviewing here the general state space model. Our notation is that of Janacek and Swift (1993).

In the general state space model the observation, or measurement, equation for a scalar is

$$X(t) = H\alpha(t) + \varepsilon(t), \tag{2.9}$$

where the measurement errors $\varepsilon(t)$ are NID$(0, \sigma^2)$, the elements of $\alpha(t): p \times 1$ are the state variables at time t, and H is a $1 \times p$ matrix (row vector) of parameters. The state, or system equation is

$$\alpha(t) = P\alpha(t-1) + K\eta(t), \tag{2.10}$$

where $P : p \times p$ is the transition matrix, $K : p \times q$ is the state noise coefficient matrix, and $\eta(t) : p \times 1$ is the state noise, or innovation vector, $\eta(t) \sim$ NID$(0, \Sigma_\eta)$. Generally, some but not all of the parameters H, σ^2, P, K, and Σ_η are known. For example, we might assume a specific structural or ARMA model — more often, we would do model selection in a class of such models — in which case P and K are given, and Σ_η simplifies. There is also a problem of identification, which may be resolved by selection of a canonical form (see, e.g., Janacek and Swift (1993)).

A vector version (vector $\mathrm{X}(t)$) of this model would be sufficient for monitoring station data, considered to be n correlated time series, but ignoring spatial relationships and in particular unmonitored locations. To remedy this, we replace the parameter vector H by a p-vector of spatial fields, $\mathrm{h}(\mathrm{s})$. The observation equation becomes

$$X(\mathrm{s}, t) = \mathrm{h}(\mathrm{s})^T\alpha(t) + \varepsilon(\mathrm{s}, t). \tag{2.11}$$

$X(\mathrm{s}, t)$ is infinite-dimensional, indexed by location s in the spatial domain $\mathcal{D}$, and $\varepsilon(\mathrm{s}, t)$ is a spatial-correlated error process,

$$\mathrm{cov}\,(\varepsilon(\mathrm{s}, t), \varepsilon(\mathrm{s}', t)) = \Sigma_\varepsilon(\mathrm{s}, \mathrm{s}'). \tag{2.12}$$

By definition, the state vector contains all the temporal covariance, $\mathrm{cov}\,(\varepsilon(\mathrm{s}, t), \varepsilon(\mathrm{s}', t')) = 0$ whenever $t \neq t'$.

The spatio-temporal Kalman filter *is a technique for obtaining predictions in this spatio-temporal state space model defined through equations (2.11), (2.12), and (2.10). The* extended spatio-temporal Kalman filter *is a technique for obtaining predictions in the nonlinear spatio-temporal state space model defined through equations (2.1), (2.2), and (2.7).*

Three methods for specifying the p spatial fields $h_i(s)$ of $\mathrm{h}(s)$ are discussed in the next section. (These choices resolve much of the identification issue in the general state space model.) A traditional, systematic, but infeasible, approach would be maximum likelihood estimation of unknown parameters, including the $h_i(\mathrm{s})$, $i = 1, \ldots, p$. To fix ideas, suppose observations are available at a fixed set of sites $\mathrm{s}_1, \ldots, \mathrm{s}_n$ at each time $t = 1, \ldots, T$. Maximum likelihood estimates of the states $\alpha(t)$ may be obtained using the Kalman filter, while further nonlinear minimization is used to estimate the unknown parameters. Each of the $n \times p$ entries in the matrix H, with ith row $\mathrm{h}(\mathrm{s})^T$, could be considered a parameter. Or the $h_i(\mathrm{s})$ may be specified by a small number of parameters, e.g. certain variogram parameters as in Section 3.3. We can estimate $\mathrm{h}(\mathrm{s})$, Σ_ϵ, and Σ_η (or sets of parameters that specify these respective quantities), following Mardia and Marshall (1984).

For a time series of data, prediction and updating follow from the usual Kalman filter equations. Let (as above) $\mathrm{x}(t)$ denote the vector of observations at time t. Let $\mathrm{a}(t)$ be an estimate of $\alpha(t)$ at time t, and let $C(t)$ be an estimate of its variance. Then the one-step predictions are

$$\begin{aligned}
\mathrm{a}(t|t-1) &= E[\alpha(t) \mid \mathrm{x}(1), \mathrm{x}(2), \ldots, \mathrm{x}(t-1)] = P\mathrm{a}(t-1) && (2.13)\\
C(t|t-1) &= \operatorname{cov}(\alpha(t) - \mathrm{a}(t|t-1)) = PC(t-1)P^T + K\Sigma K^T && (2.14)\\
\hat{\mathrm{X}}(\mathrm{s}, t|t-1) &= \mathrm{h}(\mathrm{s})^T P\mathrm{a}(t-1). && (2.15)
\end{aligned}$$

Suppose at time t the vector $\mathrm{x}(t)$ comprises observations at sites $\mathrm{s}_1, \ldots, \mathrm{s}_n$. Let H be the associated $n \times p$ matrix with ith row $\mathrm{h}(\mathrm{s}_i)$. The updating equations of the Kalman filter are

$$\begin{aligned}
\mathrm{a}(t) &= \mathrm{a}(t|t-1) + C(t|t-1)H^T F(t)^{-1}(\mathrm{x}(t) - H\mathrm{a}(t|t-1)) && (2.16)\\
C(t) &= C(t|t-1) - C(t|t-1)H^T F(t)^{-1} H C(t|t-1) && (2.17)\\
F(t) &= HC(t|t-1)H^T + \Sigma_\epsilon. && (2.18)
\end{aligned}$$

Spatial prediction at s_0 and time t given observations at time t is considered in Section 4.1.

A key feature of this spatio-temporal Kalman filter is that spatial locations are associated through the $h_i(\mathrm{s})$ and the $\alpha(t)$, common to all spatial locations. Thus the sites and quantity of data may differ from one time to the next, making the model particularly suited to data collected on interleaved (EMAP) grids, or to accommodate less frequent synoptic coverage — cf. the goals of Section 1. The model is very general, and while it is related to other approaches in the literature, it is envisaged that a prolonged, multi-year program of research is required to develop the various aspects.

3 Some choices of spatial fields

In the spatio-temporal Kalman filter, the state vector describes a linear combination of one or more trend components. That linear combination *is* the trend at each time. Thus, in choosing the number of spatial fields p, the $h_i(\mathrm{s})$, and the state space model (defined by P and K), the principal objective is to capture *all* the temporal correlation in the trend component. *A priori* there is then no reason to suppose that the errors $\epsilon(\mathrm{s}, t)$ are uncorrelated across space. However, in kriging a richer, or more carefully chosen (Cressie (1993)) model for trend can often remove spatial dependency; thus it is not unreasonable to analyze the spatial dependency, and to incorporate it *directly* into the trend.

Three different approaches are proposed here. The first is empirical orthogonal functions, which, through the singular value decomposition of the $T \times n$ matrix of observations X, are associated with invariant patterns of temporal variation (left singular vectors) and of spatial variation (right singular vectors).

The second approach, and the one we emphasize here, is via kriging. The data at each time t is regarded as observations from a surface $x(\mathrm{s},t)$ that is a realization of a Gaussian random field $X(\mathrm{s},t)$. These surfaces are correlated through the state equation; the spatial pattern of this correlation is contained in the spatial fields $h_i(\mathrm{s})$. One or more standard choices of $h_i(\mathrm{s})$ may be considered — e.g. the constant surface of ordinary kriging, or the linear, or quadratic surfaces common in universal kriging. Beyond that, the spatial pattern should be recovered from the data — that is, from the variogram after fitting an intrinsic model to the data.

The third approach is mentioned only briefly here, but appears very promising. Rodriguez-Iturbe *et al* (1986, 1988) and Cox and Isham (1988), and others, develop the use of a marked point process model for sequences of rainfall fields. Each point is random in space and time, and the mark specifies the spatial extent and amplitude of a Gaussian-shaped kernel representing a rain cell; the amplitude is multiplied by a factor that decays exponentially with time. Goodall and Phelan (1990) and Phelan and Goodall (1990) fit an explicit realization of a this model to a short sequence of images of rainfall intensity. Our suggestion here is to replace the exponential model for the amplitude factor by an entry in the state vector. A given kernel will appear and disappear from the state vector; the palette of possible kernels might be enhanced by introducing non-Gaussian shaped kernels.

3.1 Empirical orthogonal functions

Cohen and Jones (1969) and Preisendorfer (1988) discuss construction of random fields, or empirical orthogonal functions (EOF), using the Karhunen-Loève or singular value decomposition of a matrix $X\ :\ T \times n$ of T observations at n sites. For repeated measures data, the first p (out of n) EOFs provide a best summary of the available data. Thus one approach (example below) is to take $p = 2$, and $h_1 = h_{\mathrm{mean}}$, the mean field and $h_2 = h_{\mathrm{diff}}$, the first EOF of the centered X comprising 'typical' differences across time.

Cohen and Jones (1969) describe how the principal components can be interpolated between sites to give a spatial field — see also the excellent overview of Guttorp and Sampson (1994). One ingredient is the introduction of weights proportional to the area around each point. Note that the general state space model may require particular choices of h — for example, if we were using a basic structural model, then h would include an *intercept* (mean field), slope (trend field), and seasonal component.

3.2 Principal components ('warps') of Kriging

We call a vector of spatial fields h *principal fields* when they are obtained from an eigendecomposition of the 'kriging' matrix, as follows. Let X denote the random vector of values of a random field at n sites, s_1, $\mathrm{s}_2, \ldots, \mathrm{s}_n$. Define L to be the $(n+q) \times (n+q)$ 'kriging' matrix

$$L = \begin{pmatrix} E & F^T \\ F & 0 \end{pmatrix}, \tag{3.1}$$

where $F : n \times q$ has ith row f_i comprising polynomial terms in the s_i, e.g. $\mathrm{f}_i = (1|\mathrm{s}_i^T)$, or some other known functions of s_i and alternative covariates. The f_i specify the usual trend surface of kriging, which we call the kriging trend. The trend for our spatial temporal model may include the kriging trend, plus additional spatial fields (the principal fields) obtained from the residual variation after removing the kriging trend. $E : n \times n$ contains 'estimates' of the covariance of $X(\mathrm{s},t)$ and $X(\mathrm{s}',t)$, or 'estimates' of the variance of $X(\mathrm{s},t) - X(\mathrm{s}',t)$. Three different choices of 'estimates' are possible, reflecting progressively greater specificity in the model.

I. Empirical 'estimates' obtained after we remove the kriging trend, e.g. a linear or a quadratic surface is removed at each time. (Detrending and deseasonalising the data *does not* appear advisable here, as it is most important to include the temporal dependence in the spatial fields.)

II. A parametric covariogram or variogram fit to the empirical estimates.

III. A predetermined variogram (or covariogram) that does not depend on the data, for example the matrix loss function for thin-plate splines (Green and Silverman (1994), Kent and Mardia (1994)), where, for a pair of sites at distance r, the corresponding entry in E is $E(r) = r^2 \log r^2$ when there are two spatial dimensions, and $E(r) = r$ when three spatial dimensions.

Consider first the formulation in terms of covariances. We write

$$L = \begin{pmatrix} \Sigma & F^T \\ F & 0 \end{pmatrix}. \tag{3.2}$$

For an additional location, s, write $\sigma = \sigma(\mathrm{s})$ for $\mathrm{cov}(\mathbf{X}, X(\mathrm{s}))$, and $\mathrm{f} = \mathrm{f}(\mathrm{s})$ for the vector of carriers at s. For a realization x of $\mathbf{X}$, the *kriging predictor* at s is

$$\hat{X}(\mathrm{s}) = \begin{pmatrix} \sigma^T & \mathrm{f}^T \end{pmatrix} \begin{pmatrix} \Sigma & F^T \\ F & 0 \end{pmatrix}^{-1} \begin{pmatrix} \mathrm{x} \\ 0 \end{pmatrix}. \tag{3.3}$$

Note that if $\mathrm{s} = \mathrm{s}_i$, one of the given sites, then $\hat{X}(\mathrm{s}_i) = X(\mathrm{s}_i)$, so that the surface is interpolating.

The $n \times n$ upper left submatrix of L^{-1} may be written, assuming Σ is nonsingular,

$$K = \Sigma^{-1} - \Sigma^{-1}F(F^T\Sigma^{-1}F)^{-1}F^T\Sigma^{-1} \tag{3.4}$$

and the $p \times n$ lower left submatrix of L^{-1} may be written

$$G = (F^T\Sigma^{-1}F)^{-1}F^T\Sigma^{-1}. \tag{3.5}$$

Then

$$\hat{X}(\mathrm{s}) = \sigma(\mathrm{s})^T K\mathrm{x} + \mathrm{f}(\mathrm{s})^T G\mathrm{x}. \tag{3.6}$$

Note that $G\mathrm{x}$ is the coefficients in the generalized least squares regression, with covariance Σ, of x on F, and that $KF = 0$ and $GF = I$.

The quantities $K\mathrm{x}$ and $G\mathrm{x}$ depend only on the set of sites $\mathrm{s}_1, \ldots, \mathrm{s}_n$, and the vector x. Letting s vary determines a prediction surface, which is the sum of a linear combination $G\mathrm{x}$ of the trend functions $\mathrm{f}(\mathrm{s})$ and a linear combination $K\mathrm{x}$ of the covariance functions $\sigma(\mathrm{s})$. This is the general formulation for thin-plate splines (see, for example, Wahba (1990)).

The matrix K is the *partial information matrix*. K is not in general invertible. Consider the quadratic form, or weighted norm of x,

$$m(\mathrm{x}) = \mathrm{x}^T K\mathrm{x} = \mathrm{x}^T \left(\Sigma^{-1} - \Sigma^{-1}F(F^T\Sigma^{-1}F)^{-1}F^T\Sigma^{-1}\right)\mathrm{x}. \tag{3.7}$$

This is the norm of x defined using K or, in other words, the population Mahalanobis distance of x from 0 relative to the partial information K, or, equivalently, the Mahalanobis distance of $\mathrm{r} = (I - \mathcal{P})\mathrm{x}$ relative to the covariance Σ, where $\mathcal{P}$ is projection onto the trend surface defined by F. The quantity $\mathrm{x}^T K\mathrm{x}$ is the natural generalisation of the bending energy found in formulating thin-plate splines, as in Bookstein (1989). The following development of principal fields parallels Bookstein's (1989) description of principal warps, interpreted as deformations, for pairs of bivariate thin-plate splines.

Let the spectral decomposition of K be

$$K = UDU^T\,, \tag{3.8}$$

where D is diagonal with diagonal elements the eigenvalues $d_1 \geq d_2 \geq \ldots \geq d_{n-q} \geq d_{n-q+1} = \ldots = d_n = 0$. In general, q of the eigenvalues of K are zero, which follows from $KF = 0$. Also, $U^T U = UU^T = I_n$. Let u_i be the ith eigenvector of K, that is, the ith column of U. Setting $\mathrm{x} = \mathrm{u}_i$ yields Mahalanobis distance d_i. Bookstein (1989) associates the monotone sequence of eigenvalues with a progression from eigenvectors (of residuals) that tend to differ for closely spaced sites s_i — for such sites, the corresponding entries in Σ are large, to eigenvectors for which the largest differences in residuals are between widely separated sites.

Let $\mathrm{y} = U^T\mathrm{x}$, that is $\mathrm{x} = U\mathrm{y}$ so that x is the linear combination y of the eigenvectors u_i. The Mahalanobis distance of x is just

$$m(\mathrm{x}) = \sum_{i=1}^{n} d_i y_i^2\,. \tag{3.9}$$

There are two cases to consider: (1) the regression case, as above, in which the spatial random field has covariance Σ, and is typically second-order stationary with trend, and (2) the intrinsic random field case, in which the spatial random field typically possesses intrinsic stationarity, and polynomial trend of some order is annihilated. An analysis analogous to that given above applies for intrinsic random processes also. Bookstein's (1989) analysis is a special case of a self-similar intrinsic field, with $E(r) = r^2 \log r^2$ in his case (see Mardia *et al* (1994)).

3.3 Principal fields

If there is no trend — the mean $\mu(\mathrm{s}, t)$ of the spatial process is zero — then the best unbiased linear predictor is given, from first principles, by

$$\sigma(\mathrm{s})^T \Sigma^{-1}\mathrm{x}\,. \tag{3.10}$$

Eliminating trend implies replacing Σ^{-1} by K. Hence, heuristically, the 'best unbiased linear predictor', omitting a trend term, is

$$\sigma(\mathrm{s})^T K\mathrm{x}\,. \tag{3.11}$$

These expressions follow also from equations (3.4) and (3.6). Now for $\mathrm{x} = \mathrm{u}_i$, the ith eigenvector of K, we have the predictor

$$\sigma(\mathrm{s})^T K\mathrm{u}_i = \sigma(\mathrm{s})^T d_i \mathrm{u}_i = \sigma(\mathrm{s})^T \tilde{\mathrm{u}}_i \quad \text{say,} \tag{3.12}$$

where $\tilde{\mathrm{u}}_i$ is the normalized u_i.

We call $h_i(s)$ the ith principal field, where

$$h_i(\mathrm{s}) = \sigma(\mathrm{s})^T \mathrm{u}_i\,, \tag{3.13}$$

for $i = 1, \ldots, n$. We further define q kriging trend fields *using* $\mathrm{f}(\mathrm{s})$.

Since $KF = 0$, the eigenvectors $\mathrm{u}_{n-q+1}, \ldots, \mathrm{u}_n$ are linear combinations of the columns of F. Thus at $\mathrm{s}_1, \ldots, \mathrm{s}_n$ q of these spatial fields are redundant, and the total is just n. However, considering additional sites s, $\mathrm{f}(\mathrm{s})$ is prescribed in advance, while the $h_i(\mathrm{s})$ $(i > n - q)$ are defined by the second order structure. These sets of fields will span different subspaces in general.

4 Practical choices and multivariate data

4.1 Design sites and non-design sites

In practice we may start with the simplest case, when the trend surface is a constant ($q = 1$). We may also attempt to use only a subset of the principal fields. A different possibility is to use all the principal fields, but to construct the principal fields using a restricted set of sites, called the *design sites*, that are fewer in number than the sites where data are gathered. In fact, there need be no observations *at* the design sites.

These n design sites $\mathrm{s}_1, \ldots, \mathrm{s}_n$ are chosen at the start of model building. They, together with the choice of E and the kriging trend $\mathrm{f}(\mathrm{s})$, determine the principal fields $h_i(\mathrm{s})$. In the second part of a two-stage procedure, observations at time t comprising a set of $n^*(t)$ sites $\mathrm{s}_1^*(t), \ldots, \mathrm{s}_{n^*(t)}^*(t)$ are modeled using the set of $p = n + q$ principal fields and kriging trend fields that together are components of the trend surface. The $n^*(t)$ observations at time t are kriged with trend surface defined by the $n + q$ spatial fields and the appropriate second-order structure — which may not be the same as the E used to define the $h_i(\mathrm{s})$. In place of the usual generalized least-squares estimate of the trend surface coefficients, the estimated state vector $\mathrm{a}(t)$ is used. Elsewhere we consider in detail such a two stage approach, in which one spatial prediction problem (with n observations) is used to define a family of trend surfaces for a second spatial prediction problem (with n^* observations).

The choice of design sites is akin to knot selection in splining — the knots may be most dense where there is most rapid variation in the 'signal' component in the data. The adaptive function of an adaptive bandwidth smoother may prove useful in assessing the scale of an apparent 'signal', or trend.

We might take the number of design sites to be a small proportion of the number of sites for which observations are available. The design sites may be 'representative locations' from among the sites, or may be chosen at intermediate locations. For example, the Utility Acid Precipitation Study Program (UAPSP) monitoring data comprise measurements at $n = 18$ sites in the Eastern USA of three pollution variables, hydrogen, sulphate (SO_4) and nitrate (NO_3) deposition. Up to 79 observations are available, at 28-day intervals; some data are of poor quality and therefore missing. Given these 18 sites we might choose 4-5 design sites, using one of the site-selection criteria described by Mardia and Goodall (1993).

A detailed development of the Kalman filter model for the UAPSP monitoring data is in preparation for presentation. We investigate (1) the simple alternative of using mean and difference fields, h_{mean} and h_{diff} (Section 3.1), (2) the use of principal spatial fields with $q = 1$ and 4-5 design sites, (3) the use of principal spatial fields with $q = 1$ and $n = 18$ design sites at the monitoring station locations, and (4) principal spatial fields with more elaborate trend surfaces.

4.2 Diagnostics for the $h_i(\mathrm{s})$

The position and number of design sites may vary relatively slowly with time. A large number of design sites allows for complicated spatial fields. As time develops, and new patterns of pollution emerge, the set of design sites may be varied. Techniques used in determining monitoring station location, referred to above, may be useful.

We may also monitor the population Mahalanobis distance $m(\mathrm{x}(t))$ to assess the adequacy of the spatial fields. A marked increase in the criterion suggests the need for more design sites. A location-specific version of m helps indicate where these sites should be positioned. Let $b(r)$ be a kernel function that is a monotone decreasing function of distance r, and let $\mathrm{b}(\mathrm{s})$ be the n-vector with ith element

$b(\sqrt{(s_i - s)^T(s_i - s)})$. Define the spatial diagnostic field at time t

$$m(x(t), s) = (b(s) * x(t))^T K(b(s) * x(t)) \tag{4.1}$$

where '$*$' denotes elementwise multiplication of the two vectors. We may visually monitor the sequence of diagnostic fields $m(x(t), s)$, and either adjust the locations of individual design sites, or introduce new design sites. Special care is needed with these criteria when the number and positions of the sites vary with time.

Changes in the choice of design sites must be accommodated into the state vector. (The more general problem is identifiability in the extended Kalman filter when the mean field is $\mu(s,t) = h(s,t)^T\alpha(t)$, as in equation (2.2).) A heuristic approach is the following. Let $h(s,t)$ and $h(s,t+1)$ denote two (different) vectors of spatial fields at times t and $t+1$ respectively. The state vector $\alpha(t)$ may be updated to $\alpha(t)'$ according to the regression of $h(s,t)^T\alpha(t)$ on $h(s,t+1)$; specifically, the regression coefficients are $\alpha(t)'$. *Then* the data at time $t+1$ are used to update $\alpha(t)'$ to $\alpha(t+1)$. In practice, if the changes from $h(s,t)$ to $h(s,t+1)$ are slight, the regression may be almost trivial, $\alpha(t) \approx \alpha(t)'$.

A further concern (Mardia and Goodall (1993)) is non-stationarity in the spatial covariance structure. If after a single geographic deformation the data are stationary in the deformed space (Sampson (1986), Sampson & Guttorp (1992), Monestiez & Switzer (1991)), then the entire model can be constructed in the deformed space. Additional complications emerge when the geographic deformations vary with time also.

4.3 Modeling multivariate data and images

For multivariate data, the observation equation is

$$X(s,t) = A(t)h(s) + \varepsilon(s,t) \tag{4.2}$$

and the state equation is unchanged from equation (2.10), with $\alpha(t) = \text{vec}\, A(t)$. Images may be included also. Let $\tilde{s}$ index a pixel centered at s, and $\tilde{h}_i(\tilde{s})$ be the integral of $h_i(s)$ over the pixel $\tilde{s}$. For a vector of images $I(\tilde{s},t)$ at time t the observation equation is

$$I(\tilde{s},t) = B(t)\tilde{h}(\tilde{s}) + \xi(\tilde{s},t). \tag{4.3}$$

The state vector including multivariate and image data is

$$\alpha(t)^T = (\text{vec}\, A(t))^T \,\vdots\, (\text{vec}\, B(t))^T \tag{4.4}$$

Appropriate distributions for the measurement errors and innovation vectors may be formulated.

5 A simulation experiment

To gain insight into the model we are developing a graded series of examples using simulated and real data. We describe the aims of the various examples, but give the details of only one example here — the other examples are presented elsewhere. The example presented here is simulation of a sequence of vector-valued observations, using the specialization of the general state space model to a random walk plus noise (see Janacek and Swift (1993)) for which $p = 2$. This simple example illustrates how empirical orthogonal functions, *inter alia* the right singular vectors of the $T \times n$ data matrix X corresponding to the p largest singular values, can be used to recover a basis for the p columns of the matrix H, where H comprises the $\mathrm{h}(\mathrm{s}, t)$ at $\mathrm{s}_1, \ldots, \mathrm{s}_n$. We also show that the Kalman filter can recover the time sequence of states very well.

Given a set of general state space equations, with observational error covariance Σ_ϵ, and innovation covariance Σ_η, two key points must be made concerning estimating H from the EOFs of X. First, Σ_η must be large enough relative to Σ_ϵ, so that the signal is sufficiently strong — a precise statement must take into account the parameter matrix P and the *initial state* α_0. Second, if H is to be estimated unbiasedly, the subspace spanned by the columns of H must be an invariant subspace of Σ_ϵ (see Preisendorfer, 1988, p195). A series of modifications to the first example illustrates these points.

The use of spatial *fields* unifies observations at different sites into a single model, analyzed using the spatio-temporal Kalman filter. Our second example in the graded series illustrates this most important feature through a simulation in which the $h_i(\mathrm{s})$ comprise a basis for several realizations of a Gaussian random field with specified covariance. Observations at a set of n sites are used to estimate the matrix H and thus to predict the spatial field at additional sites and times. Two error models are considered: the first where errors are (uncorrelated) pure measurement error, the second where errors are correlated and satisfy the invariant subspace condition. This second alternative requires a two stage, 'conditional simulation' model, as the original realization of H determines some features of the subsequent observational error at each time step.

Our third example is an analysis of nitrate deposition in the eastern United States, with data from the UAPSP study.

5.1 Random walk plus noise

We set the parameters of the general state space model as follows. The matrix H has dimension $n \times p = 10 \times 2$. The columns are orthonormal, with first column (spatial field) proportional to the constant — each entry is $1/\sqrt{10}$ — and second column obtained by linear transformation of a vector of independent Gaussian random variates. The errors $\epsilon(t)$ in the observation equation are taken to be independent normal with variance 1/100 times the 10×10 identity matrix. The parameter matrix P in the state equation is the multiple k of the identity matrix, with $k = 0.5$ and $k = 1$. The system is stable with $k = 0.5$, but is not stable with $k = 1$. The error variance of the innovations $\eta(t)$ is 2×2 diagonal with non-zero entries $1/100$ (associated with the constant field), and $1/10$ (associated with the second field). This model was simulated for $T = 100$ time steps, beginning with state vector $\alpha(0) = (1, 0)^T$.

A typical simulation with $k = 0.5$ is shown in Figure 1. To avoid clutter, only representative time series of the simulated observations are shown the upper panel; the '+' symbols at the right hand side indicate the relative positions of all 10 entries in the second column of H (with no noise added). The time series of the simulated states are shown in the lower panel. The state associated with the constant field has the smaller variability of the two. The sequences of 'o's and '+'s show respectively the estimates of the states using the Kalman filter. The initial state vector for the simulation, $\alpha(0) = (1, 0)^T$, were also used to initialize the Kalman filter, but the sequences were insensitive to other choices of initial state vector, e.g. with $\alpha(0) = (100, -100)^T$, the $\alpha(t)$ converged in very few iterations.

The components in the $T \times n$ matrix X parallel to the columns of H (the spatial fields) are measured by the eigendecomposition of $H^t X^T X H$. In a typical simulation, the eigenvalues are 12.97 and 1.87 respectively, which sum to 14.84. The eigenvector is a rotation through $-87.62°$, that is, very close to $-90°$ which corresponds to the fact that the second component of $\alpha(t)$ is the larger of the two on the average. Because several conditions are met — the innovations have unequal variance, H is column-orthonormal, and P, Σ_η, and Σ_ϵ are all diagonal — the expectation of $H^t X^T X H$ is a diagonal matrix,

Fig. 1 (upper): 3 of 10 selected time series of observations

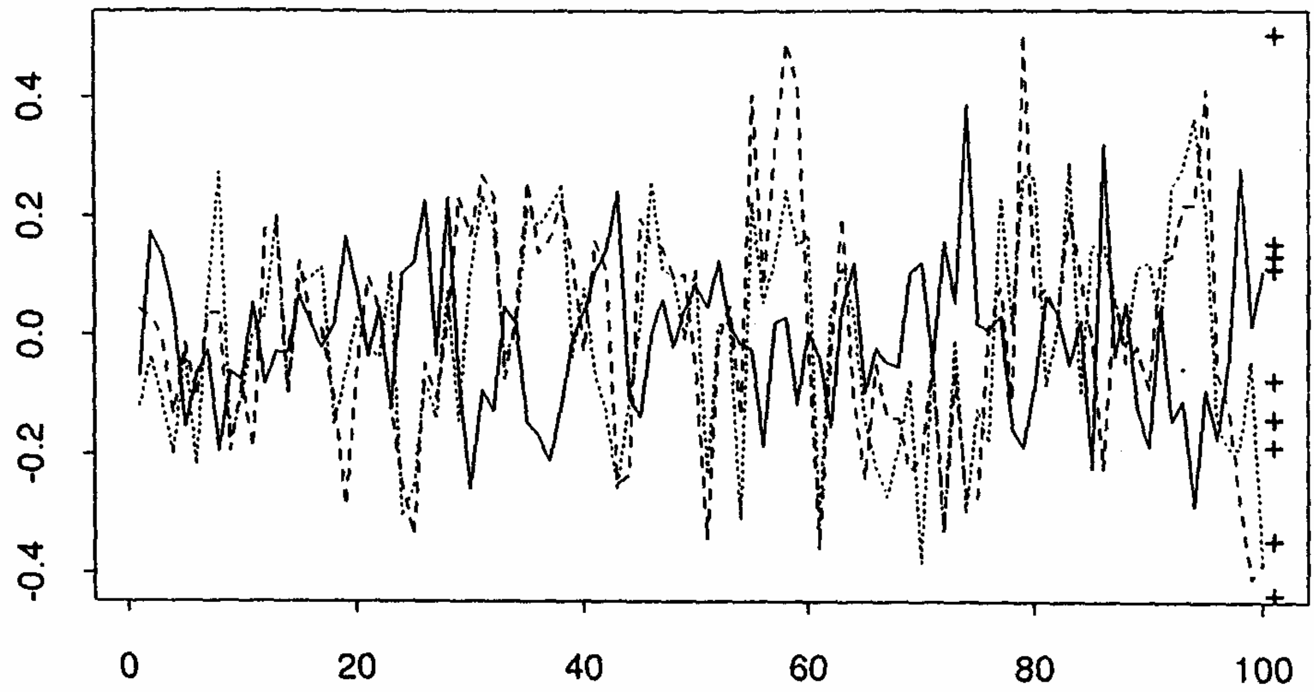
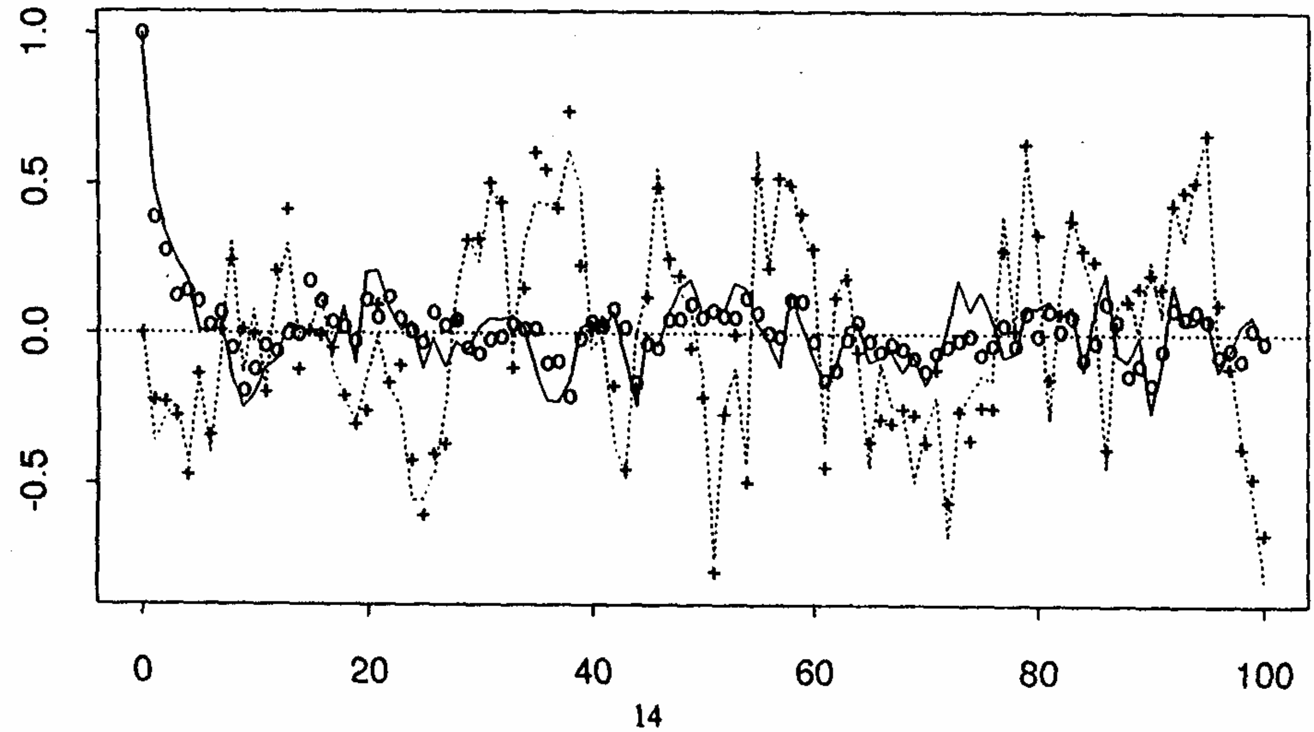


Fig. 1 (lower): time series of state and state estimates



with entries as follows. Ignoring initial conditions, for innovations with variance σ_η^2, and measurement error σ_ϵ^2, the corresponding entry equals

$$\frac{T - k^2(1-k^{2T})}{1-k^2}\sigma_\eta^2 + T\sigma_\epsilon^2 \approx \frac{T}{1-k^2}\sigma_\eta^2 + T\sigma_\epsilon^2. \tag{5.1}$$

With $k = 0.5$, and $T = 100$, the two diagonal entries are 2.33 (observed, 1.87) and 14.33 (observed, 12.97) respectively. The ratio of the two eigenvalues is 6.14 (observed, 6.05).

A second diagnostic is to measure the correspondence between the linear subspaces spanned by the two columns of H, $\mathbf{h}_1$ and $\mathbf{h}_2$ respectively, and by the first two right singular vectors, $\mathbf{v}_1$ and $\mathbf{v}_2$, of X respectively. We obtain related information by regressing each column of H in turn on the first two right singular vectors (with no constant term in the regression). The results are

$$\hat{\mathbf{h}}_1 = -0.0425\mathbf{v}_1 - 0.8852\mathbf{v}_2 \qquad R^2 = 78.5\% \tag{5.2}$$
$$\hat{\mathbf{h}}_2 = +0.9953\mathbf{v}_1 - 0.0459\mathbf{v}_2 \qquad R^2 = 99.3\%. \tag{5.3}$$

The second column of H is better estimated than the first column.

When $k = 1$ expectations can still be calculated, but there is considerably greater variability — and dependence on initial conditions. The matrix of eigenvectors is *not* diagonal.

Acknowledgements

This work was undertaken while Colin Goodall was visiting the University of Bristol. The work was supported in part by US Environmental Protection Agency Cooperative Agreement CR-821783 to Penn State University, Department of Statistics, by NSF grant DMS-9208656 to Penn State University (supporting both authors), and by funds of Leeds University. We are grateful to Mike Titterington for discussions of some of the modeling ideas, notably equation (2.6), to various authors for preprints of their work — especially to Paul Sampson for a preprint of Guttorp and Sampson (1994) — and to Alistair Walder for his helpful comments on the draft.

References

Bookstein, F. L. (1989). Principal warps: Thin-plate splines and the decomposition of deformations. *IEEE Trans. Pattern Anal. Machine Intell.*, 11:567–585.

Cleveland, W. S. and Devlin, S. J. (1982). Calendar effects in monthly time series: Modeling and adjustment. *Jour. Amer. Statist. Assoc.*, 77:520–528.

Cohen, A. and Jones, R. H. (1969). Regression on a random field. *Jour. Amer. Statist. Assoc.*, 64:1172–1182.

Cox, D. R. and Isham, V. S. (1988). A simple spatial-temporal model of rainfall. *Proc. Roy. Soc. Lond.*, A 415:317-328.

Cressie, N. A. C. (1993). *Statistics for Spatial Data*. Wiley, 2nd edition.

Donnelly, C. A., Laird, N. M., and Ware, J. H. (1993). Prediction and creation of smooth curves for temporally correlated longitudinal data. *To appear*.

Donnelly, C. A., Ware, J. H., and Laird, N. M. (1994). Regression analysis of spatially correlated data: the Kanawha County health study. In Patil, G. P. and Rao, C. R., editors, *Enviromental statistics*, pages 643–660. Elsevier, North Holland.

Galton, F. (1863). *Meteorographica, or Methods of mapping the Weather.* MacMillan, London.

Goodall, C. R. and Phelan, M. J. P. (1990). Edge-preservative smoothing and the assessment of point process models for GATE rainfall fields. In Prabhu, N. U. and Basawar, I. V., editors, *Statistical Inference in Stochastic processes*, pages 35–66.

Green, P. J. and Silverman, B. W. (1993). *Nonparametric regression and generalized linear models*, volume 58 of *Monographs on Statistics and Applied Probability*. Chapman and Hall, London.

Green, P. J. and Titterington, D. M. (1988). Recursive methods in image processing. *Bull. Int. Statist. Int.*, 52-4:51–67.

Guttorp, P. and Sampson, P. D. (1994). Methods for estimating heterogeneous spatial covariance functions with enviromental applications. *To appear.*

Hainsworth, T. J. and Mardia, K. V. (1993). A Markov random field restoration of image sequences. In Chellappa, R. and Jain, A. K., editors, *Markov random fields: Theory and applications*, pages 409–445. Academic Press.

Haslett, J. (1989). Space time modelling in meteorology - a review. *Bull. Int. Statist. Inst.*, 51:229–246. book 1.

Haslett, J., Bradley, R., Craig, P., Unwin, A., and Wills, G. (1991). Dynamic graphics for exploring spatial data with application to locating global and local anomalies. *Amer. Statist.*, 45:234–242.

Høst, G., Omre, H., and Switzer, P. (1994). Spatial interpolation errors for monitoring data. *To appear.*

Janacek, G. and Swift, L. (1993). *Time series: forecasting, simulations, applications.* Ellis Horwood, New York.

Jones, R. H. (1993). *Longitudinal data with serial correlation: a state space approach.* Chapman and Hall, London.

Kent, J. T. and Mardia, K. V. (1994). The link between kriging and thin-plate splines. To appear in Festschrift volume to Peter Whittle, Wiley.

Loader, C. and Switzer, P. (1992). Spatial covariance estimation for monitoring data. In Guttorp, P. and Walden, A., editors, *Statistics in enviromental and earth sciences*, pages 52–69. Charles W. Griffin, London.

Mardia, K. V. and Goodall, C. R. (1993). Spatial-temporal analysis of multivariate enviromental monitoring data. In Patil, G. P. and Rao, C. R., editors, *Multivariate Enviromental Statistics*. Elsevier.

Mardia, K. V., Hainsworth, T. J., and Haddon, J. F. (1992). Deformable templates in image sequences. In 11^{th} *IAPR Inter. Conf. on Pattern Recognition*, pages 132–135, The Hague, Netherlands. IEEE.

Mardia, K. V., Kent, J. T., Goodall, C. R., and Little, J. A. (1994). Kriging and splines with derivative information. Technical Report STAT 93/22/C, Dept. of Statistics, Leeds University.

Mardia, K. V. and Marshall, R. J. (1984). Maximum likelihood estimation of models for residual covariance in spatial regression. *Biometrika*, 71:135–146.

Meinhold, R. J. and Singpurwalla, N. D. (1983). Understanding the Kalman filter. *Amer. Statist.*, 37:123–127.

Monestiez, P. and Switzer, P. (1991). Semiparametric estimation of nonstationary spatial covariance models by metric multidimensional scaling. Technical Report 165, Dept. of Statistics, Stanford University.

Oehlert, G. W. (1993). Regional trends in sulfate wet deposition. *Jour. Amer. Statist. Assoc.*, 88:390–399.

Overton, W. S., White, D., and Stevens, D. L. (1990). *Design report for EMAP: Enviromental monitoring and assessment program.* US EPA.

Patil, G. P. and Rao, C. R., editors (1993). *Multivariate Enviromental Statistics.* Elsevier.

Pfeifer, P. E. and Deutsch, S. J. (1980a). Identification and interpretation of first order space-time ARMA models. *Technometrics*, 22:397–408.

Pfeifer, P. E. and Deutsch, S. J. (1980b). A three-stage iterative procedure for space-time modelling. *Technometrics*, 22:35–47.

Phelan, M. J. and Goodall, C. R. (1990). An assessment of a generalized WGR model for GATE rainfall. *Jour. Geophysical Research-Atmospheres*, 95:7603–7615.

Phillips, D. B. and Smith, A. F. M. (1994). Dynamic image analysis using Bayesian shape and texture models. In Mardia, K. V., editor, *Statistics and Images: Vol. 2*. Carfax, Oxford.

Preisendorfer, R. W. (1988). *Principal component analysis in meteorology and oceanography*. Elsevier, Amsterdam.

Rodriguez-Iturbe, I., Cox, D. R., and Eagleson, P. S. (1986). Spatial modelling of total storm rainfall. *Proc. Roy. Soc. Lond.*, A 403:27-50.

Rodriguez-Iturbe, I., Cox, D. R., and Isham, V. S. (1988). A point process model for rainfall: further developments. *Proc. Roy. Soc. Lond.*, A 417:283-298.

Rouhani, S. and Myers, D. E. (1990). Problems in space-time kriging of geohydrological data. *Math. Geology*, 22:611–623.

Rouhani, S. and Wackernagel, H. (1990). Multivariate geostatistical approach to space-time data analysis. *Water Resources Res.*, 26:585–591.

Sampson, P. D. (1986). Spatial covariance estimation by scaled-metric scaling and biorthogonal grids. Technical Report 102, Dept. of Statistics, Univ. of Washington, Seattle.

Sampson, P. D. and Guttorp, P. (1992). Nonparametric estimation of nonstationary spatial covariance structure. *Jour. Amer. Statist. Assoc.*, 87:108–119.

Sutherland, A. I. and Titterington, D. M. (1994). Bayesian analysis of image sequences. Technical report, Dept. of Statistics, Univ. of Glasgow.

Titterington, D. M. (1990). Modelling and restoration of image sequences. In McWhirter, J. G., editor, *Mathematics in signal processing II*. Clarendon Press, Oxford.

Wahba, G. (1990). *Spline models for observational data.* SIAM, Philadelphia.

Zeger, S. L. and Diggle, P. J. (1994). Semi-parametric models for longitudinal data with application to CD4 cell numbers in HIV seroconverters. *To appear.*