Journal of the Royal Statistical Society
SERIES C Applied Statistics

Appl. Statist. (2009)
58, *Part* 5, *pp.* 737–752

Spatiotemporal smoothing and sulphur dioxide trends over Europe

Adrian W. Bowman,

University of Glasgow, UK

Marco Giannitrapani

Shell Global Solutions International, Amsterdam, The Netherlands

and E. Marian Scott

University of Glasgow, UK

[Received October 2007. Final revision February 2009]

Summary. Spatiotemporal models for sulphur dioxide pollution over Europe are considered within an additive model framework. A suitable description of the spatiotemporal correlation structure of the data is constructed and incorporated in the analysis of the additive model, to ensure that standard errors and other forms of analysis reflect the form of variation that is exhibited by the data. To deal with the large sample size, an updating formula based on binning is derived to provide a computationally manageable implementation of the back-fitting algorithm. Interaction terms involving space, time and seasonal effects are also considered. This requires three-dimensional smoothing which is implemented by repeated application of lower dimensional marginal smoothing operations. The properties of this form of smoothing are examined and the estimators are shown to have first-order behaviour, inherited from the marginal operations, which is equivalent to the full multivariate versions. These models and methods are applied to the sulphur dioxide data, allowing detailed and informative descriptions of the spatiotemporal patterns to be created.

Keywords: Additive models; Correlated data; Non-parametric smoothing; Spatiotemporal data; Sulphur dioxide; Trend

1. Introduction

There is a very wide variety of settings in which data are collected with both a temporal and a spatial structure. A classic example is in environmental monitoring where measurement devices are positioned over a spatial region of interest and data are collected from each in a regular temporal pattern. A further example arises in the mapping of brain activity by using continuously recording probes which are positioned over a human skull. This type of data has generated very considerable interest in spatiotemporal models which can give flexible descriptions of changes in mean values, and if appropriate also in variance and covariance structures, over both space and time.

This paper considers spatiotemporal measurements on sulphur dioxide (SO_2) airborne pollution over Europe from 1990 to 2001. These data were collected through the 'European monitoring and evaluation programme' (EMEP) which co-ordinates the monitoring of airborne

Address for correspondence: Adrian W. Bowman, Department of Statistics, University of Glasgow, Glasgow, G12 8QQ, UK.
E-mail: adrian@stats.gla.ac.uk

 0035–9254/09/58737

pollution over Europe. Further information is available at the Web site `http://www.emep.int`, where the data can also be obtained. Berge *et al.* (1999) considered the long-term trends in these data and provided models of the flux between countries. Giannitrapani *et al.* (2006) considered the relationship between observed measurements and published emissions data, using information on wind direction and speed where this was available. The aim of the present paper is to develop statistical models which can describe the spatial and temporal patterns simultaneously.

Fig. 1(a) shows the data in time series form for a random sample of monitoring stations, with SO_2 levels on a log-scale to reduce skewness. A general downward trend is apparent in each series, although there is considerable variability and some irregularity, with gaps at the beginning, the end or even the middle of the time period of interest. Some of the variability is attributable to seasonal variation, as is common with environmental time series. This is illustrated in Fig. 1(b) where data for the same sample of stations are plotted by month. The boxplots show a cyclical pattern in the mean level of $\log(SO_2)$, with reduced values in the summer months.

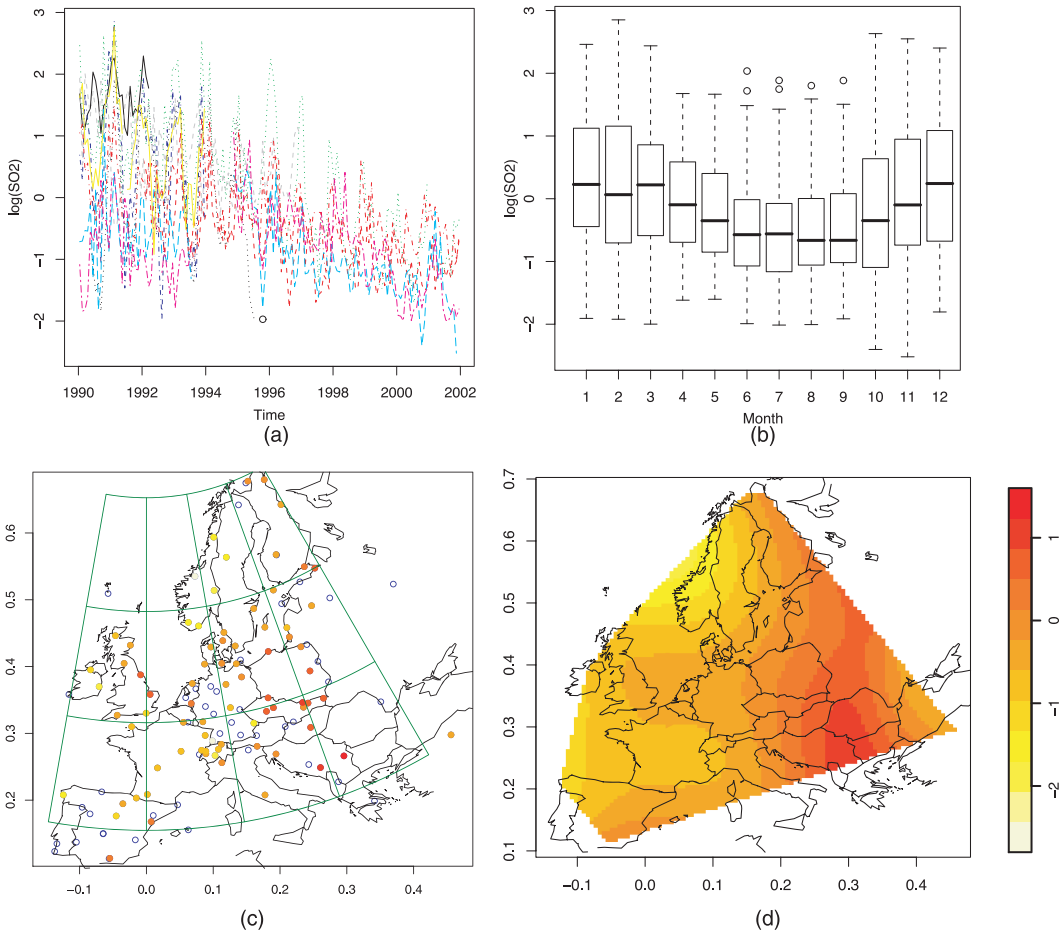

Fig. 1. $\log(SO_2)$ measurements over time for a sample of monitoring stations, (b) data for the same stations and time period but plotted against month, (c) location of the monitoring stations (the shading indicates the $\log(SO_2)$ values for those stations which have measurements available for April 1998; the grid shows the positions of lines of latitude and longitude at 10° intervals) and (d) estimate of spatial trend constructed by smoothing the data for April 1998

Fig. 1(c) shows the locations of the monitoring stations. Some of these, particularly those in the east, contributed data only in the latter part of the period of interest. As an illustration of the information that is available on spatial distribution, the shading of the points indicates the level of $\log(SO_2)$ in April 1998, for those stations for which data are available. Some broad patterns are apparent, with elevated levels in eastern Europe.

The map which has been superimposed highlights an issue about the way in which spatial location should be encoded for use in a statistical model. The use of simple longitude and latitude corresponds to the projection of the Earth's surface onto a cylinder which touches the Earth at the equator. Where the region of interest spans a substantial range of latitude, this causes distortion, with distances stretched horizontally as latitude increases. The Lambert (or Schmidt) projection addresses this by preserving area. This projection is defined by the transformation of longitude and latitude, expressed in radians as θ and ϕ, to the new co-ordinate system $x = r\sin(\theta)$, $y = r\cos(\theta)$, where $r = 2\sin\{\frac{1}{2}(\pi/2 - \phi)\}$. To illustrate the effect of the projection, grid lines which correspond to lines of latitude and longitude at 10° intervals have been superimposed on the plot. Fisher *et al.* (1987) gave further details of this and other available projections.

Fig. 1(d) shows the surface that is created by spatial smoothing of the plotted points for April 1998. There are many ways of implementing smoothing in this context and the details of the method that is used here are deferred to Section 2. The benefit of this step is a clearer visual representation of spatial trend. However, when considering the collection of surfaces that are created by smoothing the data spatially for each individual month in the 12-year period of interest, it is clear that, in addition to smooth systematic change over time, these surfaces will show considerable variation from month to month, in the same manner that time series data exhibit considerable variation from point to point. Simultaneous spatial and temporal smoothing is therefore a natural development to consider when interest lies in identifying both spatial and temporal trends.

Additive models, as described by Hastie and Tibshirani (1990), provide a very useful framework which allows smoothing to be incorporated in a flexible regression structure. Holland *et al.* (2000) analysed regional trends in SO_2 across the eastern USA by using an additive model to describe the patterns at individual monitoring stations and amalgamating these at a spatial level by using a Markov random field. This leads naturally to a Bayesian formulation and therefore connects with the large current literature on Bayesian spatial models. Banerjee *et al.* (2004) gave a good introduction to this general area, where models are often specified implicitly through conditional relationships at local neighbourhood level. Fahrmeir *et al.* (2004) described Bayesian methods of fitting explicitly additive structures. In a distinct but related line of development, Kammann and Wand (2003) showed how spatial prediction using kriging and additive terms for other covariates can be combined within a mixed model framework. This formulation was discussed more widely in Ruppert *et al.* (2003).

The various approaches to models which involve spatial data can broadly be distinguished by the assumptions that are made on which parts of the model are regarded as fixed underlying trend and which parts arise from spatial covariance, random effects or other random quantities. This general issue was discussed in Cressie (1993) and, specifically in the context of models for SO_2, by Hobert *et al.* (1997). The perspective that is taken in the present paper is to focus on systematic trend surfaces which provide 'low degree-of-freedom' descriptions of both spatial and temporal patterns by accommodating both of these within a standard additive regression setting. For example, Giannitrapani *et al.* (2007) used an additive model, which was similar to that of Holland *et al.* (2000), for an individual station in the European data and incorporated an appropriate temporal correlation structure in the inferential part of the analysis. The principal aim of the current paper is to develop methods which allow both spatial and temporal

components to be expressed in an additive model, along with suitable descriptions of the spatial and temporal variation.

With large data sets, the implementation of spatial modelling poses a significant challenge and several proposals for efficient computational methods have recently been made to address this. Fuentes (2007) considered likelihood methods for fitting Gaussian processes and adapted efficient spectral methods for lattice data to irregularly spaced observations through a gridded approximation. Cressie and Johannesson (2008) considered spatial prediction in the context of very large sample sizes and suggested the construction of non-stationary covariance functions from a fixed number of basis functions in an approach called 'fixed rank kriging'. Banerjee *et al.* (2008) tackled the same issue by projecting the underlying spatial process onto a lower dimensional subspace that was represented by a set of knots, to create a 'predictive process model' from which kriging can subsequently take place. Stein (2007) considered data on the surface of the sphere and constructed a parsimonious model for the covariance function by using spherical harmonics. However, although these ideas offer attractive methods for modelling and prediction of the underlying spatial process, they all assume that any systematic components, containing trend or covariate effects, are represented in standard linear model form. In contrast, the focus of the present paper is the development of methods which allow the systematic components to be estimated in a non-parametric manner, through computational methods which are feasible and efficient for large sample sizes, while incorporating an appropriate description of the spatial and temporal processes.

A simple additive model for the EMEP data is constructed in Section 2, where efficient computational methods and a model for the spatial and temporal variation are also described. Section 3 discusses methods of smoothing which allow interaction between the spatial and temporal effects, to provide more realistic models of spatiotemporal data. This leads in Section 4 to a model for the EMEP data which allows spatial, temporal, seasonal and interaction effects to be estimated and interpreted. Some further discussion is given in Section 5.

The data that are analysed in the paper can be obtained from

`http://www.blackwellpublishing.com/rss`

2. A simple additive model

A simple prescription for an additive model which relates response data $\{y_i; i = 1, \ldots, n\}$ to covariate information $\{x_{ij}; i = 1, \ldots, n; j = 1, \ldots, p\}$ is

$$y_i = \mu + \sum_{j=1}^{p} m_j(x_{ij}) + \varepsilon_i, \tag{1}$$

where the functions m_j represent smooth, but otherwise unconstrained, regression functions. This very flexible extension of standard linear models was discussed by Stone (1985), Hastie and Tibshirani (1986) and Buja *et al.* (1989), with a synthesis provided by Hastie and Tibshirani (1990) and a more recent exposition given by Wood (2006).

2.1. Smoothing components

Non-parametric regression provides the building blocks for fitting an additive model and there is a strong degree of similarity in the effects of the various types of smoothing which can be adopted. Local linear, and where appropriate local mean, methods have been adopted here, by constructing estimators as the solution $\hat{\alpha}$ from the least squares criteria

$$\min_{\alpha}\left\{\sum_{i=1}^{n}(y_i-\alpha)^2\, w(x_i-x;h)\right\},$$

$$\min_{\alpha,\beta}\left[\sum_{i=1}^{n}\{y_i-\alpha-\beta(x_i-x)\}^2\, w(x_i-x;h)\right],$$

$$\min_{\alpha,\beta,\gamma}\left[\sum_{i=1}^{n}\{y_i-\alpha-\beta_1(x_{i1}-x_1)-\beta_2(x_{i2}-x_2)\}^2\, w(x_{i1}-x_1;h_1)\, w(x_{i2}-x_2;h_2)\right]$$

for the case of local mean and local linear methods with one and two covariates respectively. (This last case allows interaction terms for the combined effects of two explanatory variables.) The weight functions $w(x_i-x;h)$, which by default are taken here to be normal density functions with standard deviation h, ensure that observations near the point of interest x have most influence on the resulting estimate. The local linear form is generally preferred over the local mean because of its superior properties, as described by Fan and Gijbels (1996). However, the local mean remains useful in sample spaces where linear regression is not appropriate. The cyclical nature of seasonal effects is one example of this, where the weight function

$$w(x_i-x;h)=\exp\left[\frac{r}{h}\cos\left\{\frac{2\pi(x_i-x)}{r}\right\}\right]$$

creates an estimate with period r.

The vector of fitted values from each of these methods of non-parametric regression can be expressed as Sy, where the rows of the smoothing matrix S contain the weights which arise from the solution of the least squares criteria that are defined above. This allows the approximate degrees of freedom of the smoothing procedure to be defined as $\mathrm{tr}(S)$, by analogy with the hat matrix in a linear model. This provides a convenient scale on which the degree of smoothing can be expressed. Specification of the approximate degrees of freedom determines the value of the bandwidth h to be used in the weight functions.

2.2. Computational feasibility

The back-fitting algorithm is a well-established method of fitting additive models. This involves repeated application of non-parametric regression on the partial residuals that are created by subtracting the current component estimates, with the exception of the one of current interest, from the observed data. Specifically, the iterations are defined by

$$\hat{m}_j^{(l)}=S_j\left(y-\hat{\mu}-\sum_{k<j}\hat{m}_k^{(l)}-\sum_{k>j}\hat{m}_k^{(l-1)}\right),\qquad j=1,\ldots,p, \tag{2}$$

where S_j denotes the smoothing matrix for covariate j and $\hat{\mu}$ denotes a vector filled with the sample mean $\bar{y}$. The columns of the matrix S_j are centred to have mean 0, to ensure identifiability, corresponding to the constraints $\Sigma_i\, m_j(x_{ij})=0$, as a result of the 'grand mean' parameter μ in the model. Standard errors and other inferential tools are most easily constructed from the projection matrices which generate the component estimates after the convergence of the back-fitting algorithm. Giannitrapani *et al.* (2010) have pointed out that these can be identified by reformulating the iterations (2) in matrix form as

$$P_j^{(l)}=S_j\left(I_n-P_0-\sum_{k<j}P_k^{(l)}-\sum_{k>j}P_k^{(l-1)}\right). \tag{3}$$

The matrix P_0 is filled with the value $1/n$ to estimate μ by $\bar{y}$. Successive iterations from the

starting point which sets $P_j^{(0)}$, for $j = 1, \ldots, p$, to a matrix filled with 0s converge to a set of projection matrices which construct the fitted values as $\hat{m} = (\Sigma_{j=0}^{p} P_j) y$.

From a computational viewpoint, formulation (2) is very straightforward to implement whereas formulation (3) is very demanding for large data sets because of the need to store the matrices P_j, each of which has dimension $n \times n$. For example, the SO_2 data involve 11654 observations, which make working with the standard projection matrices impractical. In alternative approaches to smoothing such as the penalized splines that were described by Marx and Eilers (1998), a 'low rank' solution to this problem is often adopted by restricting the knots, or basis elements, to a modest number. Kammann and Wand (2003) and Ruppert *et al.* (2003) have described the details. Analogous thinking with local linear methods leads to the use of binning which, when working with the covariate x_j, replaces the full response vector y by a vector of mean values $\tilde{y}$ within the 'bins' that are defined by a fine grid of values on the covariate axis. This simple and very effective technique has a long history in smoothing, as indicated by the early work of Härdle and Scott (1992) and Wand and Jones (1995). It is straightforward to show that smoothing these means, weighted by their associated sample sizes, produces the same result as smoothing the entire n-vector of data adjusted to the grid locations. The effect of using the grid, rather than original, locations is very small but brings a very large reduction in dimensionality. Bowman and Azzalini (2003) gave efficient computational formulae for local linear and other smoothing techniques in vector matrix and binned form.

The process of binning can be expressed in incidence matrices B_j whose ith column contains a 1 in the row corresponding to the bin containing observation x_{ij} and 0s elsewhere. The means of the response data within each bin are then constructed as $\Lambda^{-1} B_j y$, where Λ is a diagonal matrix constructed from the numbers of observations in each bin. When required, any quantities corresponding to the bin locations can be expanded to quantities corresponding to the original data locations simply by premultiplying by B_j^{T}. If the centred smoothing matrix $\tilde{S}_j$ uses the bin locations as covariate values and the bin frequencies as additional weights, then the projection matrix $\tilde{P}_j$ which constructs the estimates at the bin locations has the updating formula

$$\tilde{P}_j^{(l)} = \tilde{S}_j \Lambda_j^{-1} \left\{ B_j (I - P_0) - \sum_{k<j} B_j B_k^{\mathrm{T}} \tilde{P}_k^{(l)} - \sum_{k>j} B_j B_k^{\mathrm{T}} \tilde{P}_k^{(l-1)} \right\}. \tag{4}$$

This formulation uses projection matrices whose row dimensions are limited by the numbers of grid points that are used in the binning operations. The combination of incidence matrices $B_j B_k^{\mathrm{T}}$ is a matrix of low dimension in both rows and columns. Indeed, with careful programming these matrices do not need to be evaluated at all as the necessary assignments can be performed by careful use of indicator variables.

2.3. A model for the sulphur dioxide data

The simplest additive model which can be considered for the $\log(SO_2)$ data (y) involves components for space ((x_1, x_2) in Lambert co-ordinates), time (t, using years and months on a decimal year scale) and month (z) in the form

$$y = \mu + m_s(x_1, x_2) + m_t(t) + m_z(z) + \varepsilon. \tag{5}$$

The decomposition of the time effect into a seasonal cycle plus a longer-term trend follows the seasonal–trend decomposition approach of Cleveland *et al.* (1990) and has been used in the context of additive models and varying-coefficient models by Ferguson *et al.* (2007) and others.

Although there are methods for choosing the degree of smoothing which is employed for each component, these generally assume independent errors and can lead to inappropriate choices when that assumption is invalid, as is highly likely for the SO_2 data. A simple and very stable approach is to nominate the degrees of freedom that are associated with each component of the model. Since interest lies in the large-scale pattern of SO_2, modest numbers of degrees of freedom are appropriate and these were therefore set to 6 for the univariate terms and 12 for the bivariate spatial term. The sensitivity of the results to these choices will be investigated later in the paper. For the bivariate spatial term, the equal area construction of the Lambert co-ordinates leads to the adoption of a common smoothing parameter h in each of the spatial dimensions. In terms of binning, the 124 unique site locations and the 12 months of the year provide natural bin locations. For year on a decimal scale, 100 equal width bins were used to cover the scale.

The results of fitting model (5) are shown in Figs 2(a)–2(c). The spatial pattern shows higher levels of SO_2 in the south-east and over the Iberian peninsula. There is a steady decline over the years and a modest seasonal effect. Of course, this is an extremely simple model and, although it provides a useful benchmark for the broad features of the data, it is not credible that spatial effects should remain constant over time or throughout the year. Interactions therefore need to be investigated and this will be the subject of Section 3.

2.4. A model for correlation

Before embarking on the construction of more complex models it is helpful to consider the provision of standard error information to accompany the estimates and to supply information on their precision. The first step must be the construction of a model for the spatial and temporal correlations, using the residuals from fitted model (5). Giannitrapani *et al.* (2007) considered variograms of the SO_2 data over time and reached the conclusion that a Gaussian correlation function, $\exp\{-(d/\nu)^2\}$ over distance d, was appropriate and that its range parameter ν was broadly stable over time, which is consistent with a separable model for spatial and temporal correlation. The underlying trend surface was linear and some potential difficulties were noted when a non-parametric trend surface was used. However, with the assumption that the spatial correlations are stable over time the empirical variograms for each month can be pooled before the Gaussian model is fitted. This stabilizes the fitting process and produces the estimate $\hat{\nu} = 0.101$ for the range of the spatial correlation, with the interpretation that observations from locations which are separated by a distance of more than 0.2 in Lambert co-ordinates are effectively uncorrelated.

A simple strategy for temporal correlation involves a time series model to be fitted to the residuals from each site. An auto-regressive AR(1) model was found to be adequate, by inspection of the auto-correlation and partial auto-correlation functions. The median value of the estimated correlation parameters ρ can then be taken as a simple global estimate. This yielded the value $\hat{\rho} = 0.612$, indicating a strong degree of temporal correlation across the months. However, in interpreting both temporal and spatial correlations it should be noted that any bias from the estimation of model (5) will be exhibited in the residuals and may lead to some inflation of the estimates of correlation. This leads in turn to the conservative effect of increasing standard errors.

From these fitted parameters, estimates of the spatial and temporal correlation matrices $\hat{V}_s$ and $\hat{V}_t$ can be constructed. Under the assumption of separability, an estimate of the global correlation matrix $\hat{V}$ can be constructed as the Kronecker product of $\hat{V}_s$ and $\hat{V}_t$, with rows and columns omitted where there are gaps in the observed data. Since each component of the

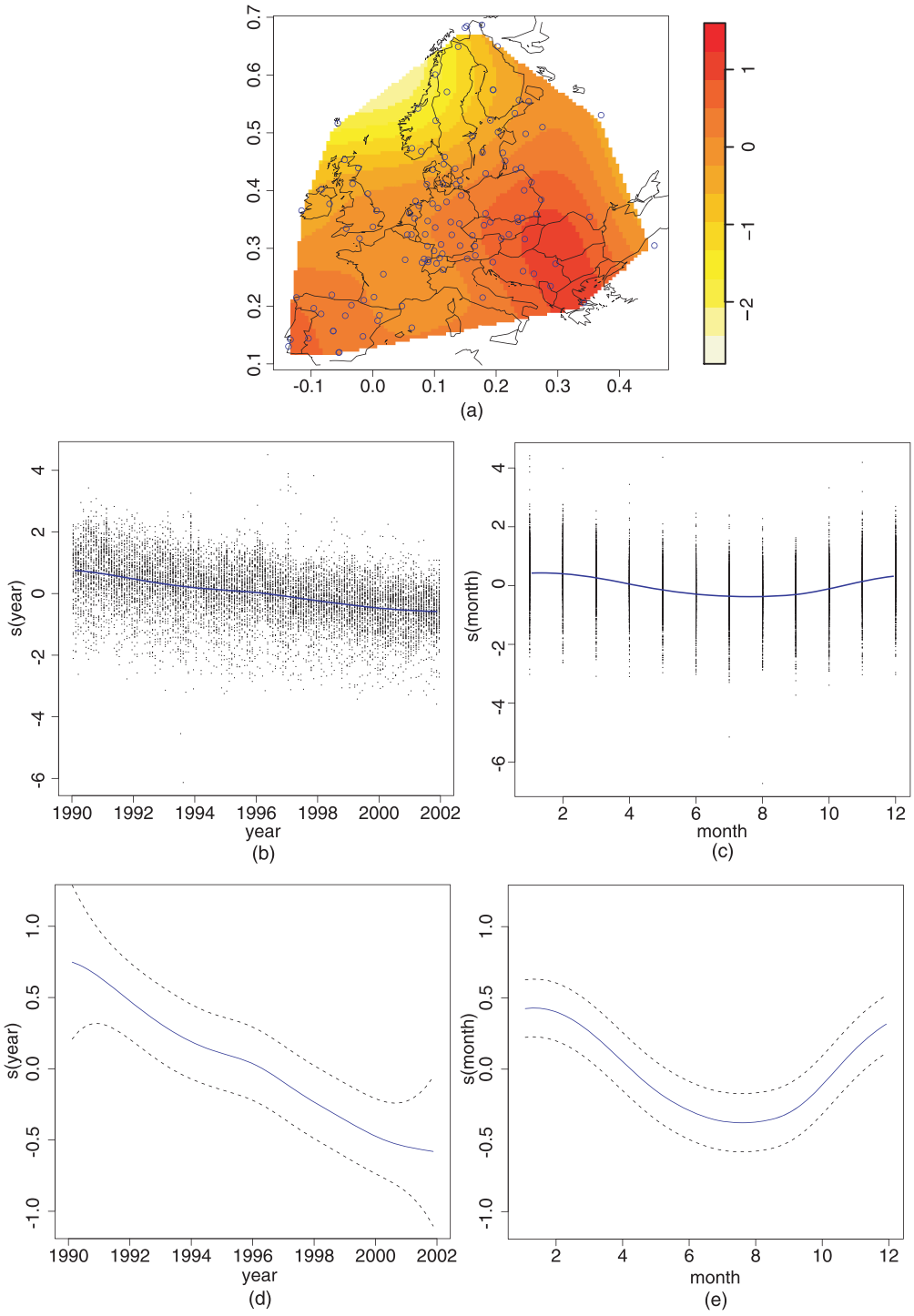

Fig. 2. Estimates of (a) the spatial, (b) year and (c) month components of additive model (5) (·, partial residuals), and estimates of the components for (d) year and (e) month (- - - - -, magnitude of 2 standard errors at each location)

additive model has an estimate of the form $\hat{m}_j = P_j y = B_j^{\mathrm{T}} \tilde{P}_j y$, the standard errors are readily available as the square roots of the diagonal entries of $P_j \hat{V} P_j^{\mathrm{T}} \sigma^2$. The error variance σ^2 can be estimated from the residual sum of squares $\mathrm{RSS} = y^{\mathrm{T}}(I-P)^{\mathrm{T}}(I-P)y$ and the approximate degrees of freedom associated with error. This latter quantity is $\mathrm{tr}\{(I-P)^{\mathrm{T}}(I-P)V\}$ which can be calculated from the reduced form of the projection matrices $\tilde{P}_j$, with careful programming. This results in the estimate $\hat{\sigma} = 0.893$. Lines corresponding to distances of 2 standard errors from the year and month estimates of additive model (5) have been added to Figs 2(d) and 2(e).

3. Spatiotemporal smoothing

The simple additive model of Section 2 provides a useful starting point for the construction of more realistic models, in particular where the shape of spatial effects changes with month and year. It would be very surprising if the spatial pattern of SO_2 was identical throughout the year or the decline in SO_2 had the same shape across the spatial locations of all monitoring stations. Spatiotemporal interaction terms such as $m_{s,t}(x_1, x_2, t)$ and $m_{s,z}(x_1, x_2, z)$ therefore need to be added to the model.

It is relatively unusual to consider smoothing across three variables simultaneously and the 'curse of dimensionality' which applies to multivariate smoothing techniques provides discouragement on the grounds of possible deterioration in the statistical properties of the resulting estimators. However, there are two features of the SO_2 data, which are shared by many spatiotemporal data sets, which suggest that a three-dimensional approach is feasible. The first is the large sample size with, in this case, 11 654 observations providing considerable information in terms of the ability to estimate complex features. The second is the grid structure of the data, with spatial location, year and month defining the margins of a three-dimensional structure where observations are present in a substantial proportion of the resulting cells. Currie *et al.* (2006) developed an elegant approach known as generalized linear array models, which is based on Kronecker products of spline bases, for data of this type where observations are available across all cells. The substantial number of missing cells in the SO_2 data complicates this particular approach here but the general principle of performing smoothing by combinations of marginal operations, which is exemplified in the generalized linear array models, offers an attractive way forward.

To illustrate the ideas, consider a simple problem of bivariate smoothing over the co-ordinates x_1 and x_2. As discussed in Section 2, this can be performed by fitting a planar surface in a local manner through weighted least squares. However, if the data lie on a regular grid, where the observations y_{ij} correspond to all the combinations of the marginal grid points $\{x_{i1}; i = 1, \ldots, n_1\}$ and $\{x_{j2}; j = 1, \ldots, n_2\}$, then an alternative approach is to smooth the data along x_{i1} for each fixed x_{j2} and then to smooth the resulting estimates along x_{j2} for each fixed x_{i1}. This operation can be expressed as $S_1 Y S_2^{\mathrm{T}}$, where S_1 and S_2 denote the smoothing matrices constructed from the marginal grids in x_1 and x_2 respectively and Y denotes the matrix whose (i, j)th element is y_{ij}. This formulation lies at the heart of the generalized linear array model approach of Currie *et al.* (2006). The dimensional efficiency is made apparent by rewriting $S_1 Y S_2^{\mathrm{T}}$ as $(S_2 \otimes S_1)y$, where '$\otimes$' denotes the Kronecker product and the vector y is created by successively appending the columns of Y. This gives a very compact and efficient bivariate smoothing procedure which produces an estimate of genuinely bivariate form but which requires only smoothing matrices whose dimensionalities are determined by the sizes of the marginal grids. Note also that the estimator is unaffected by the order in which the smoothing operations are applied.

With local linear estimators, it is possible to investigate the properties of this procedure. Following on from the notation of the previous paragraph, the estimate can be written as

$$\hat{m}(x_1, x_2) = \left(\sum_{j=1}^{n_2} s_j(x_2) \left\{ \sum_{i=1}^{n_1} s_i(x_1) y_{ij} \right\} \right) \tag{6}$$

where the $s_i(x_1)$ denote the values in the row of S_1 corresponding to estimation at x_1 and the $s_j(x_2)$ denote the values in the row of S_2 corresponding to estimation at x_2. When local linear smoothing is used to estimate a regression function $g(x)$ with a single covariate x whose observed values are distributed uniformly, it is well known that the mean and variance of the estimator have the asymptotic expressions

$$\mathbb{E}\{\hat{g}(x)\} = \sum_{i=1}^{n} s_i(x)\, g(x) = g(x) + \frac{h^2}{2} g''(x) + o(h^2),$$
$$\operatorname{var}\{\hat{g}(x)\} = \sum_{i=1}^{n} s_i^2(x)\sigma^2 = \frac{\sigma^2}{nh} \int w(z)^2 \, \mathrm{d}z + o\left(\frac{1}{nh}\right).$$

Fan and Gijbels (1996) discussed these and a variety of other theoretical results. Repeated application of these results allows the mean value of the estimator (6) to be written as

$$\begin{aligned}
\mathbb{E}\{\hat{m}(x_1, x_2)\} &= \left(\sum_{j=1}^{n_2} s_j(x_2) \left\{ \sum_{i=1}^{n_1} s_i(x_1)\, m(x_1, x_2) \right\} \right) \\
&= \left(\sum_{j=1}^{n_2} s_j(x_2) \left\{ m(x_1, x_2) + \frac{h_1^2}{2} \frac{\partial^2}{\partial x_1^2} m(x_1, x_2) + o(h_2^2) \right\} \right) \\
&= m(x_1, x_2) + \frac{h_1^2}{2} \frac{\partial^2}{\partial x_1^2} m(x_1, x_2) + \frac{h_2^2}{2} \frac{\partial^2}{\partial x_2^2} m(x_1, x_2) + o(h_1^2 + h_2^2).
\end{aligned}$$

Similarly, the variance of estimator (6) has the form

$$\begin{aligned}
\operatorname{var}\{\hat{m}(x_1, x_2)\} &= \sum_{j=1}^{n_2} s_j^2(x_2) \left\{ \sum_{i=1}^{n_1} s_i^2(x_1)\sigma^2 \right\} \\
&= \sum_{j=1}^{n_2} s_j^2(x_2) \left\{ \frac{\sigma^2}{n_2 h_2} \int w(z)^2 \, \mathrm{d}z + o\left(\frac{1}{n_2 h_2}\right) \right\} \\
&= \frac{\sigma^2}{n h_1 h_2} \left\{ \int w(z)^2 \, \mathrm{d}z \right\}^2 + o\left(\frac{1}{n h_1 h_2}\right),
\end{aligned}$$

where $n = n_1 n_2$. These expressions for the mean and variance are identical to those constructed by full bivariate smoothing, indicating that the marginal and bivariate approaches enjoy the same first-order statistical properties.

For bivariate smoothing, the marginal approach offers little benefit because very efficient methods of computation are available for the standard formulation, as described by Bowman and Azzalini (2003). However, efficient methods are not available for smoothing with three covariates and so the marginal approach is attractive. For example, smoothing over space and month can be performed by marginal smoothing of the data matrix whose rows correspond to sampling locations and whose columns correspond to months. Spatial smoothing is already a two-dimensional operation and the result is a smooth estimate of three simultaneous covariates. If the first-order properties of the estimator are required, then these can be derived by a simple

extension of the analysis that was given above. Specifically, if the first-order biases and variances of the marginal smoothing operations are denoted by $b_s(x_1, x_2)$, $v_s(x_1, x_2)\sigma^2$, and $b_z(z)$ and $v_z(z)\sigma^2$ then the first-order bias and variance of the resulting estimator are $b_s(x_1, x_2) + b_z(z)$ and $v_s(x_1, x_2)\, v_z(z)\sigma^2$ respectively.

As soon as there are irregularities in the data through missing cells, the simple Kronecker product form of the smoothing matrix $S_2 \otimes S_1$ must be replaced by a more complex matrix $\tilde{S}_{12}$ which contains appropriate combinations of the weights that are generated by the specific pattern of time points at different sites. Nonetheless, the marginal smoothing operations allow these weights to be constructed and combined in a very efficient manner. The dimensionalities of the smoothing matrices are given by the products of the numbers of bins that are involved in each margin. As discussed above, these are 124 for space and 12 for month. To constrain the dimensionality of interactions involving time, year was represented in integer form to span the 12 values from 1990 to 2001. There was very little loss of information in using this more restricted scale.

When interaction terms are included in a model, it is useful to be able to separate these from main effects in the usual way. A simple way to achieve this in the specification of the model is to decompose the smoothing operations into the components S_1, S_2 and $S_{12} = \tilde{S}_{12} - S_1 - S_2$. The component estimates that are generated by these terms can therefore be interpreted as main effects and interaction.

4. An interaction model for spatiotemporal trend

Model (5) was expanded to incorporate space–year, space–month and, for completeness, month–year interactions. The new model can be expressed as

$$\begin{aligned} y = \mu &+ m_s(x_1, x_2) + m_t(t) + m_z(z) \\ &+ m_s(x_1, x_2) : m_t(t) + m_s(x_1, x_2) : m_z(z) + m_t(t) : m_z(z) + \varepsilon, \end{aligned} \tag{7}$$

where the ':' notation indicates interaction terms. These terms can be defined more precisely as smooth surfaces $m_{s,t}$, $m_{s,z}$ and $m_{t,z}$ in the corresponding variables, with the constraints that the marginal sums over the observed values of each contributing variable are 0. This corresponds to the usual constraints in standard analysis of variance and it can be implemented by the smoothing matrix decomposition that was indicated at the end of Section 3.

The residuals from the fitted model were used to estimate the temporal and spatial correlation parameters ($\hat{\rho} = 0.551$ and $\hat{\nu} = 0.098$; see Section 2.4), for the construction of the spatiotemporal correlation matrix. The estimates of the main effect terms are displayed in Figs 3(a), 3(c) and 3(d). These main effects should now be interpreted only as a baseline to which the interaction terms are to be added. The space, year and month effects are similar in shape to those from the simple additive model (5), although the year and month terms show some dampening, presumably as a consequence of the variation explained by the interaction terms.

The estimate of the month–year interaction is displayed as a shaded surface in Fig. 3(b). The perturbations that are represented by the interactions are relatively small but show higher values in winter in earlier years, and lower values in summer in later years, than an additive structure would suggest. To quantify the evidence that an interaction is required, contour lines are drawn to show the regions where the estimate is more (or less) than 2 standard errors away from 0. This suggests that the evidence is modest.

The interaction terms for space–year and space–month each consist of 12 surfaces showing the deviation in spatial pattern that is associated with each year or month. These are most effectively

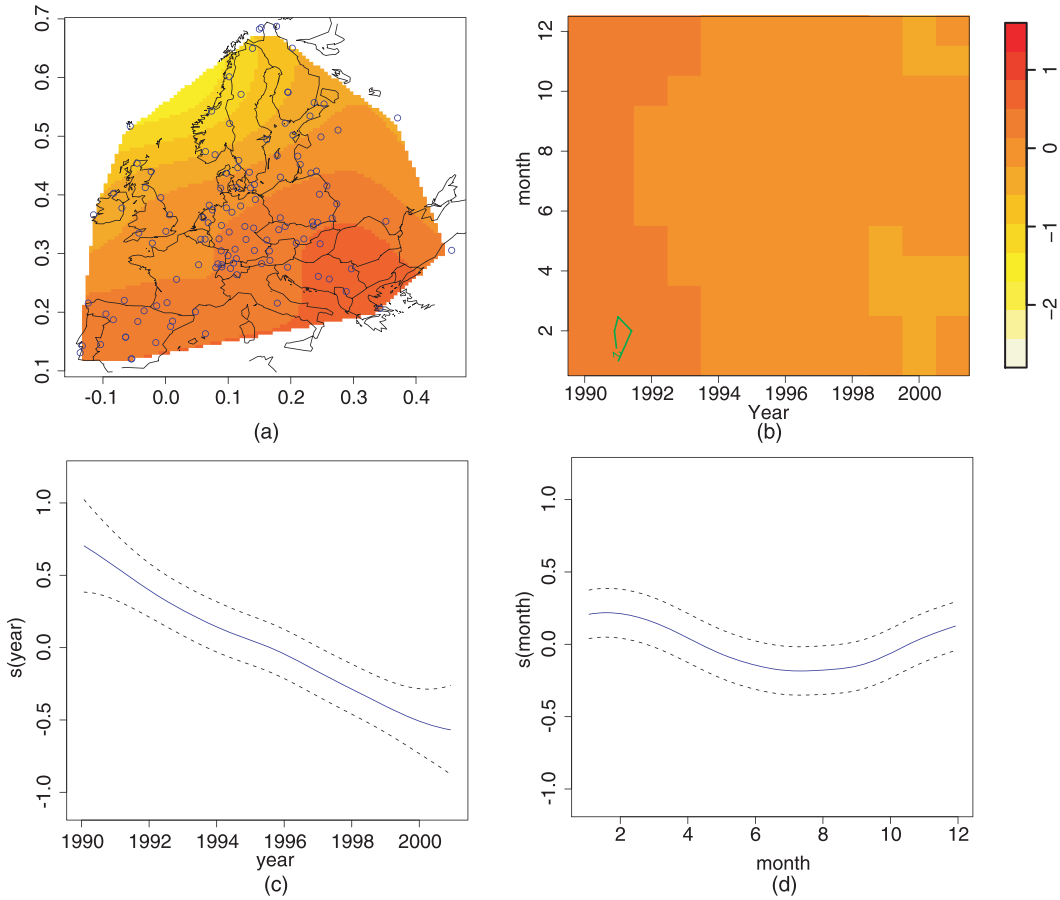

Fig. 3. (a) Estimates of the main effect of space, (b) interaction term for year and month, and estimates of main effects of (c) time and (d) month: the contour lines show the distance of the surface from 0 in units of standard errors, where this is greater than 2

viewed as an animation which displays the smoothly changing patterns across time and throughout the year. The collection of plots is available for viewing at `http://www.stats.gla.ac.uk/~adrian/spatial-smoothing.pdf`.

Fig. 4 shows a selection of these, from which the broad patterns are apparent. Northern regions are lower in earlier years, and south-eastern regions are higher in later years, than an additive structure suggests. In the spatial–seasonal interaction, northern regions are higher in winter, southern regions lower in summer and eastern regions lower in autumn than an additive structure suggests. Again, contours have been added to display the regions where the estimate is more than two standard errors away from 0.

The aim of the analysis is to investigate the nature of the interaction terms. However, it is also useful to confirm in a global manner that there is convincing evidence that each interaction term is indeed required. McMullan *et al.* (2007) discussed the use of correlation-adjusted residual sums of squares $y^{\mathrm{T}}(I-P)^{\mathrm{T}}V^{-1}(I-P)$ and subsequent approximate F-tests, following the linear model analogy that was proposed by Hastie and Tibshirani (1990). Giannitrapani *et al.* (2010) also considered more careful quadratic form calculations, based on an approach that was described in Bowman and Azzalini (1997) but incorporating the estimated correlation matrix.

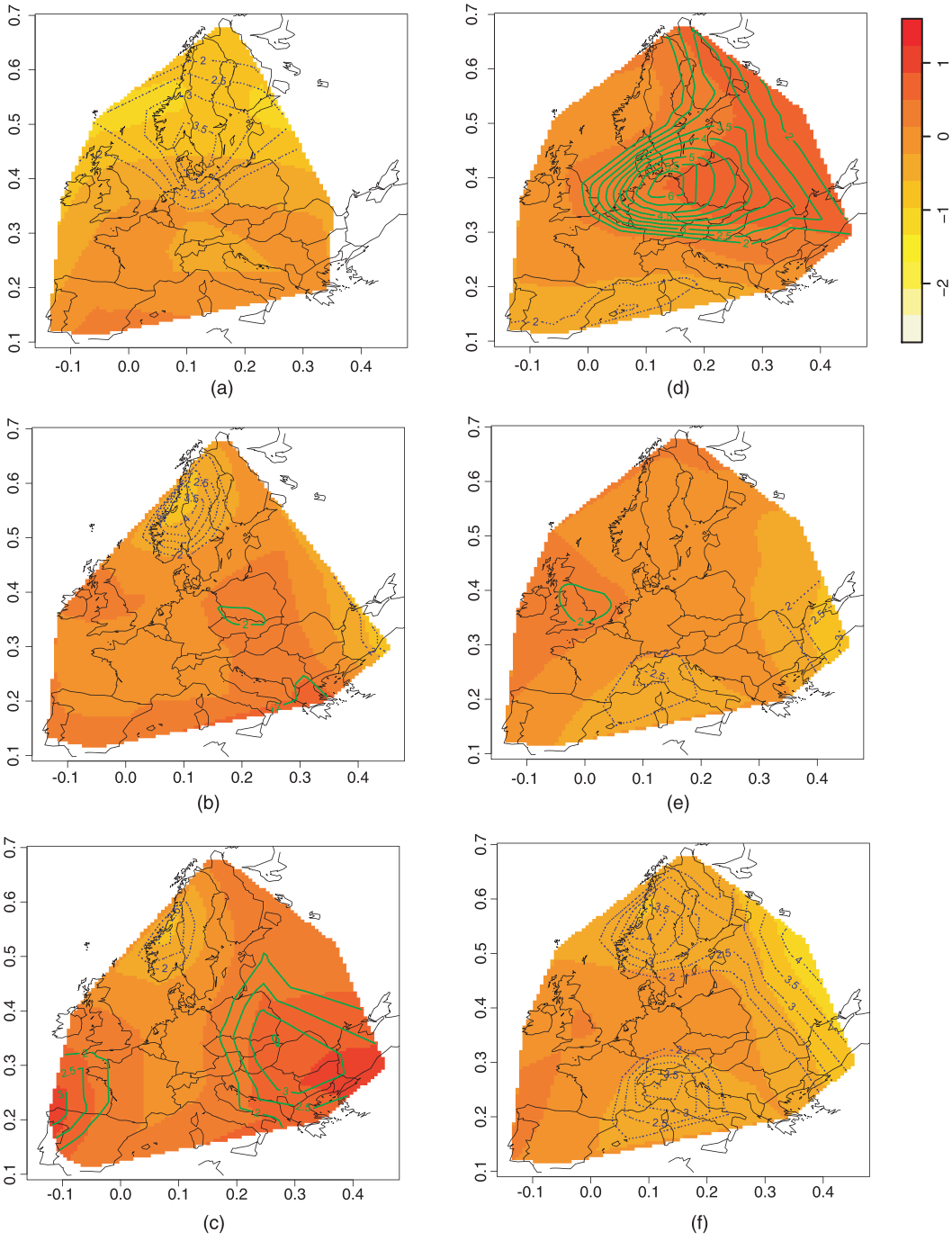

Fig. 4. Interaction terms between space and year for years (a) 1990, (b) 1995 and (c) 2000, and interaction terms between space and month for months (d) 1, (e) 5 and (f) 9: the contour lines show the distance of the surface from 0 in units of standard errors, where this is greater than 2

It is therefore feasible in principle to compare competing models and to identify the significance of individual model components.

However, the size of the computational task with large data sets needs to be considered in light of the expectation that significant effects are generally to be expected with large sample sizes. A simple strategy to confirm the presence of interaction effects is therefore to be preferred, with greater emphasis then given to the interpretation and description of these effects, as discussed above. A convenient form of hypothesis test uses as test statistic a measure of the size of the interaction estimate $\hat{m}_{\text{int}}$, namely $\hat{m}_{\text{int}}^{\mathrm{T}} \hat{m}_{\text{int}} / \hat{\sigma}^2$. The numerator of this test statistic can be written as $y^{\mathrm{T}} P_j^{\mathrm{T}} P_j y$, where j identifies the projection matrix for the interaction term, and so, ignoring bias and the variability that are associated with the estimation of σ^2, the mean value of the test statistic is $\nu = \mathrm{tr}(P_j^{\mathrm{T}} P_j V) = \mathrm{tr}(P_j V P_j^{\mathrm{T}})$. The test can then be completed by referring the test statistic to a χ^2_ν-distribution. For each of the three interaction terms this produced an extremely small p-value, indicating clear evidence that interaction is present. The largest p-value is 0.003 for the interaction between year and month. The weaker nature of this interaction is evident from the interaction plots in Figs 3 and 4.

In Section 2.3, a specific choice of the degree of smoothing that was applied to the data, in terms of modest degrees of freedom, was proposed. This makes an effective decision on the smoothness of the underlying trend surface, with remaining variability attributed to spatial and temporal covariances. It is clearly important to investigate the sensitivity of the results to the particular choices of degrees of freedom. Model (7) was refitted using smaller degrees of freedom (4 for univariate and 8 for spatial smoothing) and larger degrees of freedom (8 for univariate and 20 for spatial smoothing). The results, which are available for inspection at the Web site mentioned above, change only in minor details. This offers considerable reassurance that the conclusions are relatively insensitive to the particular degree of smoothing that was applied.

5. Discussion

The aim of this paper has been to extend the additive models framework to accommodate spatiotemporal data. This required smoothing over three covariates, to allow spatiotemporal interactions in the description of the mean part of the model, as well as the incorporation of spatiotemporal correlation to ensure that standard errors and global tests are constructed appropriately. The presence of missing data complicated the calculations by preventing the use of simple Kronecker product forms. However, similar principles were involved even if their computational implementation required considerable effort.

One of the main benefits of the fitted models is the ability to visualize both main effects and interactions in appropriately constructed plots. The structure of the data is sufficiently complex that exploratory plots of raw data cannot display the underlying patterns clearly. An additive model, with interactions and enhanced by information on standard errors, can display these structures in a very effective manner. The use of estimates and standard errors is quite a straightforward approach. However, the construction of these estimates and standard errors is not; nor is the nature of the data to which they are applied. We therefore believe that the value of the methods proposed lies in providing tools for an essentially descriptive, but nonetheless informative, analysis of spatiotemporal data.

As with any model, care must be taken in interpreting the results. The analysis here has aimed to describe the mean pattern of SO_2 at broad scales, both spatially and temporally. There may be individual stations whose measurements are consistently out of line with the mean level for the surrounding region but this behaviour is absorbed into a larger-scale representation. It should

also be noted that the largest concentrations are indicated over Bulgaria and Romania, where no monitoring stations are located. The high values here are induced by the high values that are observed from stations immediately to the west. Similarly, the reduction in the height of the mean level surface to the east of these countries is dictated by only two stations. However, the influence of these features will be represented in the standard errors, which it will clearly be important to construct and consider.

Acknowledgements

Marco Giannitrapani acknowledges support from the Chancellor's fund of the University of Glasgow, and the Centre for Ecology and Hydrology, Edinburgh. The assistance of Rognvald Smith, Centre for Ecology and Hydrology, in obtaining access to the data and giving advice on their context is also gratefully acknowledged.

References

Banerjee, S., Carlin, B. and Gelfand, A. (2004) *Hierarchical Modeling and Analysis for Spatial Data*. Boca Raton: Chapman and Hall–CRC.

Banerjee, S., Gelfand, A. E., Finley, A. O. and Sang, H. (2008) Gaussian predictive process models for large spatial data sets. *J. R. Statist. Soc.* B, **70**, 825–848.

Berge, E., Bartnicki, J., Olendrzynsky, K. and Tsyro, S. (1999) Long-term trends in emissions and transboundary transport of acidifying air pollution in Europe. *J. Environ. Mangmnt*, **57**, 31–50.

Bowman, A. and Azzalini, A. (1997) *Applied Smoothing Techniques for Data Analysis*. Oxford: Oxford University Press.

Bowman, A. and Azzalini, A. (2003) Computational aspects of nonparametric smoothing with illustrations from the sm library. *Computnl Statist. Data Anal.*, **42**, 545–560.

Buja, A., Hastie, T. and Tibshirani, R. (1989) Linear smoothers and additive models. *Ann. Statist.*, **17**, 453–555.

Cleveland, R., Cleveland, W., McRae, J. and Terpenning, I. (1990) STL: a seasonal-trend decomposition procedure based on loess. *J. Off. Statist.*, **6**, 3–73.

Cressie, N. A. (1993) *Statistics for Spatial Data*. New York: Wiley.

Cressie, N. and Johannesson, G. (2008) Fixed rank kriging for very large spatial data sets. *J. R. Statist. Soc.* B, **70**, 209–226.

Currie, I. D., Durban, M. and Eilers, P. H. C. (2006) Generalized linear array models with applications to multidimensional smoothing. *J. R. Statist. Soc.* B, **68**, 259–280.

Fahrmeir, L., Kneib, T. and Lang, S. (2004) Penalized structured additive regression for space-time data: a Bayesian perspective. *Statist. Sin.*, **14**, 731–761.

Fan, J. and Gijbels, I. (1996) *Local Polynomial Modelling and Its Applications*. London: Chapman and Hall.

Ferguson, C. A., Bowman, A. W., Scott, E. M. and Carvalho, L. (2007) Model comparison for a complex ecological system. *J. R. Statist. Soc.* A, **170**, 691–711.

Fisher, N., Lewis, T. and Embleton, B. (1987) *Statistical Analysis of Spherical Data*. Cambridge: Cambridge University Press.

Fuentes, M. (2007) Approximate likelihood for large irregularly spaced spatial data. *J. Am. Statist. Ass.*, **102**, 321–331.

Giannitrapani, M., Bowman, A. and Scott, E. (2010) Additive models for correlated data with applications to air pollution monitoring. In *Statistical Methods for Trend Detection and Analysis in the Environmental Sciences* (eds R. Chandler and E. Scott). London: Wiley. To be published.

Giannitrapani, M., Bowman, A., Scott, E. and Smith, R. (2006) Sulphur dioxide in Europe: statistical relationships between emissions and measured concentrations. *Atmos. Environ.*, **40**, 2524–2532.

Giannitrapani, M., Bowman, A., Scott, E. and Smith, R. (2007) Temporal analysis of spatial covariance of SO_2 in Europe from 1990 to 2001. *Environmetrics*, **1**, 1–12.

Härdle, W. and Scott, D. (1992) Smoothing by weighted average of rounded points. *Computnl Statist.*, **7**, 97–128.

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

Hastie, T. and Tibshirani, R. (1990) *Generalized Additive Models*. London: Chapman and Hall.

Hobert, J., Altman, N. and Schofield, C. (1997) Analyses of fish species richness with spatial covariate. *J. Am. Statist. Ass.*, **92**, 846–854.

Holland, D., De Oliveira, V., Cox, L. and Smith, R. (2000) Estimation of regional trends in sulfur dioxide over the eastern United States. *Environmetrics*, **11**, 373–393.

Kammann, E. E. and Wand, M. P. (2003) Geoadditive models. *Appl. Statist.*, **52**, 1–18.

Marx, B. and Eilers, P. (1998) Direct generalized additive modeling with penalized likelihood. *Computnl Statist. Data Anal.*, **28**, 193–209.
McMullan, A., Bowman, A. W. and Scott, E. M. (2007) Water quality in the River Clyde: a case study of additive and interaction models. *Environmetrics*, **18**, 527–539.
Ruppert, D., Wand, M. P. and Carroll, R. (2003) *Semiparametric Regression*. London: Cambridge University Press.
Stein, M. (2007) Spatial variation of total column ozone on a global scale. *Ann. Appl. Statist.*, **1**, 191–210.
Stone, C. (1985) Additive regression and other nonparametric models. *Ann. Statist.*, **13**, 689–705.
Wand, M. P. and Jones, M. C. (1995) *Kernel Smoothing*. London: Chapman and Hall.
Wood, S. (2006) *Generalized Additive Models: an Introduction with R*. London: Chapman and Hall–CRC.