16
From spatio-temporal smoothing to functional spatial regression: a penalized approach

Maria Durban, Dae-Jin Lee, M. Carmen Aguilera-Morillo, Ana M. Aguilera

16.1
Introduction

When a response variable is observed several times and in different locations, we talk about space-time data. A different approach based on Functional Data Analysis (FDA) can be considered when the *data* are seen as functions of time sampled at different locations. The aim of this work is to show that both perspectives share many features when a roughness penalty approach is used to ensure the smoothness of the estimated functions. Furthermore, this common framework will facilitate the improvement of existing methods in both perspectives.

In the context of spatial data, the ideas of smoothing has been present for many years. For example, [1] combined kriging with low rank smoothing for continuous covariates. Most recently, [2] built up on this idea and developed a spatial random effects model and fixed rank kriging, based on predicting the coefficients of a set of basis functions. In a similar context, [3] proposed the use of two-dimensional B-spline basis with random coefficients whose covariance structure was given by a penalty matrix controlled by separate smoothing parameters for each dimension.

In the last 20 years, several approaches have been developed in the context of spatio-temporal data. Some of them are based on the idea of Kalman filters [4, 5]. Others have a Bayesian perspective, in particular, hierarchical models based of Markov Random fields have become very popular [6, 7]. Here, we adopt the method introduced by [8] based on penalized smooth mixed models. These models use B-spline basis and discrete penalties [9] in a multidimensional setting [9, 10], but the methodology can be applied for any basis and quadratic penalty. Penalized splines do not rely on stochastic processes, and so, there is no need to model the covariance function. Although the number and position of the knots for the basis functions have to be chosen, the use of penalties relaxes the importance of the knots placement [11]. The mixed model representation of P-splines [12] solves the problems of the selection of the smoothing parameter (since they become a ratio of variance parameters), and the non-identifiability common in additive models (see [13] for further details).

Smoothing is also a key tool in the context of FDA where sample functions are usually observed with error and need to be pre-smoothed. An interesting review of different ways of including smoothing in FDA methodologies can be seen in [14]. In this context P-splines were used for smoothing the sample curves and estimating different FDA models such as PCA, functional logit regression and functional PLS, among others ([15, 16, 17, 18]).

Alternative approaches for predicting spatio-temporal data are based on using different FDA methodologies for modelling a set of continuous-time curves with spatial dependence. On the one hand, classical geostatistical tools such as kriging were extended for this purpose in [19, 20, 21, 22, 23]. On the other hand, functional regression models with a functional response have been recently applied in [24]. The spatial information is introduced in terms of scalar covariates and considering a three dimensional P-spline penalty that combines the two-dimensional p-spline discrete penalty used for spatial regression ([25, 3]) with the continuous penalty (based on the second order squared derivatives of the parameter functions) used for functional regression ([14]).

The rest of the Chapter is organized as follows: Section 2 introduces a penalized approach for smoothing spatial data, and the reparametrization of this approach into a mixed model is presented in Section 3. Section 4 gives a general framework for smoothing spatio-temporal data using an ANOVA-type decomposition. The benefits of this approach are shown in a small simulation study. P-spline functional spatial regression is introduced in Section 5. Finally, in Section 6, we analyze an air pollution dataset in Spain using both the functional spatial regression and spatio-temporal smoothing approach.

16.2
Smoothing spatial data via penalized regression

Suppose, for simplicity, that we observe a response variable, y_i, at a finite set of spatial locations $s_i = (u_i, v_i)$, $i = 1, \ldots, n$, and y_i is normally distributed. There are many different approaches to smoothing and predicting spatial data: geostatistical models [26], Bayesian hierarchical models [27] or penalized regression [28], among others. Although they approach the smoothing problem from different perspectives, they are intimately related, for example, penalized splines can be interpreted as Bayes estimates with a suitable Gaussian process prior [29, 30], and spline fitting is well known to be a special case of kriging [31]. In this section we focus on the use of penalized regression splines (this will help us to see immediately the links with spatial functional regression).

Penalized regression splines or P-splines [9, 32] are based on the use of a rich basis for regression and a penalty on the coefficients to control the smoothness of the fit. There are many possibilities for the choice of basis (B-splines, thin plate regression splines, etc.) and penalties (differenced or derivative based penalties), we will illustrate (without loss of generality) the methodology using B-spline basis and second order difference penalties. The model proposed to capture the spatial dependence is

$$y_i = f(u_i, v_i) + \epsilon_i \quad \epsilon_i \sim N(0, \sigma^2), \tag{16.1}$$

where the smooth function f can be expressed in terms of a number of basis functions. Some authors suggest the use of radial basis functions or thin plate splines, or a more computationally efficient version of thin plate regression splines proposed by [33]. These bases have the limitation of being isotropic smoothers, and the selection of knots to construct the basis is not trivial. We follow the approach of [34], and use tensor product of B-spline basis with equally spaced knots. Although the e domain of the tensor product smooth is a rectangle or cuboid, it is often the case that the covariates only occupy part of the domain; in that case, a simple solution is to drop the basis functions that are to be evaluated at zero, and the corresponding components of the penalty. In the case of scattered data, the basis is constructed from the tensor product of marginal B-spline basis defined in [35] so that

$$f(u, v) = \sum_{k=1}^{q} \sum_{l=1}^{r} a_{kl} B_k^U(u) B_l^V(v) \tag{16.2}$$

where $\left\{B_k^U(u) : k = 1, \ldots, q\right\}$ and $\left\{B_l^V(v) : l = 1, \ldots, r\right\}$ are the marginal B-spline basis for each spatial coordinate. Let us denote by $\mathbf{B}^U$ the $n \times q$ matrix of values of the B-spline basis along u evaluated at the sample spatial locations u_i and by $\mathbf{B}^V$ the $n \times r$ matrix of values of the B-spline basis along v evaluated at the sample spatial coordinates v_i.

If unpenalized regression was used, then, the coefficients a_{kl} could be chosen by minimizing the least squared problem

$$S = \sum_{i}^{n} (y_i - f(u_i, v_i))^2 = \sum_{i} \left(y_i - \sum_{k=1}^{q} \sum_{l=1}^{r} a_{kl} B_k^U(u_i) B_l^V(v_i) \right)^2. \tag{16.3}$$

In this case, the smoothness of the spatial surface is controlled by the the number of B-spline basis in each dimension. As an alternative approach, it is possible to introduce a penalty that constraints coefficients that are next to each other to be similar. By construction, the domain of the tensor product smooth is a rectangle, and the coefficients a_{kl} are arranged in a matrix $\mathbf{A}$ of size $q \times r$, and so, we penalized the coefficients along the rows and columns of that matrix, i.e.

$$Pen(\mathbf{A}) = \lambda_u \sum_{l=1}^{r} \|\mathbf{P_u a_{.l}}\|^2 + \lambda_v \sum_{k=1}^{q} \|\mathbf{P_v a_{.k}}\|^2, \tag{16.4}$$

where $\mathbf{P_u}$ imposes a penalty on the columns of $\mathbf{A}$ ($\mathbf{a_{.l}}$ corresponds to column l) and $\mathbf{P_v}$ imposes a penalty on each row of $\mathbf{A}$ ($\mathbf{a_{.k}}$). One important feature of (16.4) is the fact that λ_u and λ_v can be different. This allows different amounts of smoothing along the two dimensions.

Using expression (16.2, model (16.1) can be expressed in matrix form as

$$\mathbf{y} = f(\mathbf{u}, \mathbf{v}) + \boldsymbol{\epsilon} = \mathbf{Ba} + \boldsymbol{\epsilon}, \quad \boldsymbol{\epsilon} \sim N(\mathbf{0}, \mathbf{I}\sigma^2),$$

where the "row-Tensor" product of two matrices denoted by symbol $\square$ is defined as

$$\mathbf{B} = \mathbf{B}^U \square\, \mathbf{B}^V = (\mathbf{B}^V \otimes \mathbf{1}_q') \odot (\mathbf{1}_r' \otimes \mathbf{B}^U)\,. \tag{16.5}$$

The basis $\mathbf{B}$ is of dimension $n \times qr$, the operator $\odot$ is the *Hadamard or "element-wise" matrix product*, and $\mathbf{1}_q$ and $\mathbf{1}_r$ are column vectors of ones of length q and r respectively. In the case of data on a regular grid, the basis is calculated as the Kronecker product of the marginal basis, $\mathbf{B} = \mathbf{B}^U \otimes \mathbf{B}^V$. Figure ~~1~~ plots a portion of the basis functions in this case.

Figure 16.1

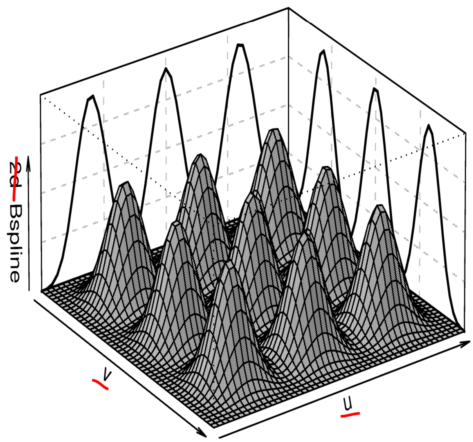

Figure 16.1 Portion of the B-spline basis (tensor product of nine cubic splines) in the case of data in a regular grid.

Then, the penalized least squares problem is written as

$$\mathbf{S}(\mathbf{a};\mathbf{y},\lambda_u,\lambda_v) = (\mathbf{y} - \mathbf{Ba})'(\mathbf{y} - \mathbf{Ba}) + \mathbf{a}'\mathbf{Pa}, \tag{16.6}$$

where $\mathbf{P}$ is expressed as

$$\mathbf{P} = \lambda_u \mathbf{I}_r \otimes \mathbf{P}_u + \lambda_v \mathbf{P}_v \otimes \mathbf{I}_q. \tag{16.7}$$

In particular, when $\mathbf{P}_u$ and $\mathbf{P}_v$ are based on second order differences, i.e., $\mathbf{P}_u = \Delta'\Delta$ (and Δ a matrix that forms differences of order 2), the structure imposed by this penalty is such that each coefficient a_{kl} depends on the eight next neighboring coefficients along the coordinate axes (in the Bayesian approach, this would be the covariance structured of the coefficients). Of course, the dependence can be easily modified if differences are necessary along other directions.

Conditional on the values of λ_u and λ_v,

$$\hat{\mathbf{y}} = \hat{f}(\mathbf{u},\mathbf{v}) = \mathbf{B}(\mathbf{B}'\mathbf{B} + \mathbf{P})^{-1}\mathbf{B}'\mathbf{y} = \mathbf{H}\mathbf{y}.$$

The matrix $\mathbf{H}$ is called the hat matrix and it is very useful tool. It shows that the smoother is linear, and its trace gives a measure of the effective dimension of the model [36].

Optimization of smoothing parameters can be done using leave-one-out cross-validation, information criteria, etc. We choose an approach that takes advantage of the connections between penalized splines and mixed models. Details are given in the next section

16.3
Penalized smooth mixed models

The connection between nonparametric regression and mixed models was established many years ago [37, 38], but it became popular much later [12, 25]. This approach has many advantages: i) the smoothing parameter is estimated via maximum likelihood, and ii) it can deal easily with with the identifiability problems that appear in models with more than one smooth term.

Smoothing penalties can also be viewed as resulting from improper Gaussian prior distributions on the spline coefficients, i.e., model (16.1) can be expressed as

$$\mathbf{y} = \mathbf{B}\mathbf{a} + \boldsymbol{\epsilon}, \quad \mathbf{a} \sim N(\mathbf{0}, \mathbf{P}^{-}) \quad \boldsymbol{\epsilon} \sim N(\mathbf{0}, \mathbf{I}\sigma^2), \tag{16.8}$$

since $\mathbf{P}^{-}$ is the Moore-Penrose pseudo-inverse of $\mathbf{P}$ (since penalties based on differences or derivatives are semi-definite positives, within the number of zero eigenvalues equal to the order of the difference/derivative). To avoid the improper distribution we propose a re-parametrization of the model in which we separate the penalized and unpenalized coefficients yielding a mixed model. Our aim will be to reformulate model (16.1) (and therefore model (16.8) as

$$f(\mathbf{u}, \mathbf{v}) = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\alpha, \text{ with } \alpha \sim N(\mathbf{0}, \mathbf{G}),$$

where the basis and coefficients are re-parameterized as

$$\mathbf{B} \rightarrow [\mathbf{X} : \mathbf{Z}] \quad \text{and} \quad \mathbf{a} \rightarrow (\boldsymbol{\beta}, \alpha).$$

The transformation is based on the singular value decomposition (SVD) of the penalty matrix given in (16.7) (which is a function of the SVD of the marginal penalties $\mathbf{P}_u$ and $\mathbf{P}_v$). Using a similar approach to [13], we find that the transformation $\mathbf{T}$ needed to re-parameterized the model bases is

$$\mathbf{T} = [\mathbf{U}_{un}u \otimes \mathbf{U}_{vn} : \mathbf{U}_{us} \otimes \mathbf{U}_{vn} : \mathbf{U}_{un} \otimes \mathbf{U}_{vs} : \mathbf{U}_{us} \otimes \mathbf{U}_{vs}], \tag{16.9}$$

where $\mathbf{U}_{un}$ and $\mathbf{U}_{us}$ are the eigenvectors corresponding to the zero and non-zero eigenvalues of $\mathbf{P}_u$ (and similarly for $\mathbf{P}_v$). Then, the matrices of fixed and random effects are $\mathbf{BT} = [\mathbf{X} : \mathbf{Z}]$,

$$\mathbf{X} = \left(\mathbf{X}^U \square \mathbf{X}^V\right) \tag{16.10}$$

$$\mathbf{Z} = \left(\mathbf{Z}^U \square \mathbf{X}^V : \mathbf{X}^U \square\, \mathbf{Z}^V : \mathbf{Z}^U \square \mathbf{Z}^V\right), \tag{16.11}$$

where $\mathbf{Z}^U = \mathbf{B}^U \mathbf{U}_{su}$. Then, columns of $\mathbf{X}$ span the polynomial null space of $\mathbf{P}$ and the columns of $\mathbf{Z}$ span its complement. The covariance matrix of the random effects α is diagonal with elements

$$\mathbf{G} = \begin{pmatrix} \lambda_u \widetilde{\mathbf{\Sigma}}_U \otimes \mathbf{I}_d & & \\ & \lambda_v \mathbf{I}_d \otimes \widetilde{\mathbf{\Sigma}}_V & \\ & & \lambda_u \widetilde{\mathbf{\Sigma}}_U \otimes \mathbf{I}_{r-d} + \lambda_v \mathbf{I}_{q-d} \otimes \widetilde{\mathbf{\Sigma}}_V, \end{pmatrix}^{-1}, \tag{16.12}$$

$\widetilde{\mathbf{\Sigma}}_U$ and $\widetilde{\mathbf{\Sigma}}_V$ are the non-zero eigenvalues of the marginal penalty matrices, $\mathbf{I}$ is an identity matrix, and d is the dimension of the null space of $\mathbf{P}$. This partition allows the representation of the fitted surface in terms of the sum of three components: one for $\mathbf{u}$ (latitude), one for $\mathbf{v}$ (longitude), and an interaction component which depends on both geographical components simultaneously.

The estimates of the coefficients $\boldsymbol{\beta}$ and α follow from standard mixed model theory [39],

$$\widehat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{y} \tag{16.13}$$

$$\widehat{\alpha} = \mathbf{G}\mathbf{Z}'\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\widehat{\boldsymbol{\beta}}). \tag{16.14}$$

where $\mathbf{V} = \sigma^2\mathbf{I} + \mathbf{ZGZ}'$. In the mixed model setting, smoothing parameters λ_u and λ_u become the ratio of variances, therefore, they may be estimated by maximizing the residual log-likelihood (REML) of [40]

$$\begin{aligned} \ell(\lambda_1, \lambda_2, \sigma^2) &= -\frac{1}{2}\log|\mathbf{V}| - \frac{1}{2}\log|\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}| - \\ &\quad -\frac{1}{2}\mathbf{y}'(\mathbf{V}^{-1} - \mathbf{V}^{-1}\mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1})\mathbf{y}. \end{aligned} \tag{16.15}$$

Recently, [41] presented a fast algorithm for the estimation of smoothing parameter in the context of multidimensional smooth mixed models.

The definition of matrix $\mathbf{Z}$ and covariance matrix $\mathbf{G}$ suggests that the smooth function $f(\mathbf{u}, \mathbf{v}))$ accounting for the spatial structure in the data can be decomposed as the sum of three components: one latitude, on for longitude and another for the interaction between them. This decomposition has inspired a new class of smooth models called *P-spline ANOVA models* that have an immediate application in the case of spatial and spatio-temporal data.

16.4 P-spline smooth ANOVA models for spatial and spatio-temporal data

Sometimes, fitting a multidimensional smooth model of the form $f(\mathbf{x}_1, \ldots, \mathbf{x}_k)$ can be restrictive. For example, in the case of spatio-temporal data, using the model

$$\mathbf{y} = f_{st}(\mathbf{u}, \mathbf{v}, \mathbf{t}) + \boldsymbol{\epsilon}\,, \quad \boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I}) \tag{16.16}$$

can lead to a poor fitting. For example, if there is a strong additive effect of $time$, but the interaction with the geographical location is relatively small, fitting model (16.16) using multidimensional P-splines will impose an interaction model and not an additive one. In order to accommodate all possible settings we propose the use of the following model as a general approach for the smoothing of space-time data:

$$E[\mathbf{y}] = \gamma + \sum_{i=1}^{k} f_i(\mathbf{x}_i) + \sum_{i<j} f_{ij}(\mathbf{x}_i, \mathbf{x}_j) + \cdots + f_{1,\ldots,k}(\mathbf{x}_1, ..., \mathbf{x}_k), \tag{16.17}$$

where γ is a constant term, f_i are additive univariate functions of the i^{th} covariate, f_{ij} a two-dimensional interaction smooth function of the pair of covariates $(\mathbf{x}_i, \mathbf{x}_j)$, and so on until a k^{th} order interaction. These type of models can be seen as a functional version of *Analysis-Of-Variance* (ANOVA). Using this terminology, model (16.17) is the sum of smooth functions of *main effects* and *two-way interactions*, *three-way interactions*, and so on. These models have been considered in the literature in the context of Smoothing Splines, as SS-ANOVA models [42, 43]. Since, main effects are *contained* in the higher order interactions, it is necessary to impose constraints in order to make the model identifiable. This may be complicated and computationally expensive when there are higher order interactions. As an alternative, we propose a Low-rank S-ANOVA model. P-splines use low-rank bases functions, and so, they are computationally less demanding than other approaches.

In the case of spatio-temporal data, a model such as (16.17) might not be realistic. In general, we will be interested in the spatial effect, the temporal effect, and the interaction between them. Expressing the model in this way we are, implicitly, giving more flexibility to the spatio-spatial structure in the model, and we can gain insight on process behind our data (for example, we can test whether space and time are separable). Therefore, we will consider a *reduced* version of model (16.17) (see [13] for a full description)

$$\mathbf{y} = \gamma + f_s(\mathbf{u}, \mathbf{v}) + f_t(\mathbf{t}) + f_{st}(\mathbf{u}, \mathbf{v}, \mathbf{t}) + \boldsymbol{\epsilon}\,, \quad \boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I}). \tag{16.18}$$

f_s represents the spatial structure common along time, f_t is the common temporal pattern shared by all locations, and f_{st} would account for departures from this overall functions across space and along time. Each of this functions is expressed again in terms of basis functions and coefficients. The B-spline regression basis for this model would be

$$\mathbf{B} = [\mathbf{1}_{nt} : \mathbf{B}_s \otimes \mathbf{1}_t : \mathbf{1}_n \otimes \mathbf{B}_t : \mathbf{B}_s \otimes \mathbf{B}_t], \tag{16.19}$$

where $\mathbf{B}_s$ is the two-dimensional basis defined in (16.5), $\mathbf{B}_t$ is the B-spline basis for the time effect and $\mathbf{B}_s \otimes \mathbf{B}_t$ is the basis for the interaction. If data are collected at the same time points for all locations, the 3d-basis is constructed using the Kronecker product, if time points are different, the box-product is used instead.

The vector of regression coefficients is $\mathbf{a} = (\gamma, \mathbf{a}^{(s)\prime}, \mathbf{a}^{(t)\prime}, \mathbf{a}^{(st)\prime})'$ and the penalty matrix is block-diagonal with penalties over $\mathbf{a}$ of the form

$$\mathbf{P} = \text{blockdiag}(0, \mathbf{P}_{(s)}, \mathbf{P}_{(t)}, \mathbf{P}_{(st)}), \tag{16.20}$$

where $\mathbf{P}_{(s)}$ is the two-dimensional penalty matrix for the spatial component, with smoothing parameters λ_u and λ_v as in (16.7), i.e.

$$\mathbf{P}_{(s)} = \lambda_u \mathbf{P}_u \otimes \mathbf{I}_r + \lambda_v \mathbf{I}_q \otimes \mathbf{P}_v, \tag{16.21}$$

$\mathbf{P}_{(t)}$ is the one-dimensional penalty matrix for the time component, with smoothing parameter λ_t, and $\mathbf{P}_{(st)}$ is the three-dimensional penalty matrix for the spatio-temporal component with smoothing parameters τ_u, τ_v and τ_t

$$\mathbf{P}_{(st)} = \tau_u \mathbf{I}_t \otimes \mathbf{P}_u \otimes \mathbf{I}_r + \tau_v u \mathbf{I}_t \otimes \mathbf{I}_q \otimes \mathbf{P}_v + \tau_t \mathbf{P}_t \otimes \mathbf{I}_r \otimes \mathbf{I}_q. \tag{16.22}$$

The B-spline model matrix for this model is not of full rank since the space spanned by $\mathbf{B}_t$ is contained in the space spanned by $\mathbf{B}_s \otimes \mathbf{B}_t$, and therefore we encounter the identifiability problems mentioned above. Several approaches have been taken to overcome this problem: (i) add a ridge penalty [44], or (ii) identify and impose the constraints numerically [10]. However, the first alternative may induce to numerical problems, and the second method is difficult to extend in the case of more than 2-way interactions. We use here a simpler and more efficient approach based on removing the linearly dependent columns of the basis (identifying the columns to be removed is immediate when the mixed model representation is used). We adapt the transformation given (16.9) to the spatio-temporal model and find that the mixed model matrices are

$$\begin{array}{rcl}
\mathbf{X} & = & [\overbrace{\mathbf{1}_t \otimes \mathbf{x}_s :}^{f_s(\mathbf{u},\mathbf{v})} \quad \overbrace{\mathbf{t} \otimes \mathbf{1}_n :}^{f_t(\mathbf{t})} \quad \overbrace{\mathbf{t} \otimes \breve{\mathbf{x}}\,] \qquad\qquad\qquad\qquad}^{f_{s,t}(\mathbf{u},\mathbf{v},\mathbf{t})} \\
\mathbf{Z} & = & [\,\mathbf{1}_t \otimes \mathbf{Z}_s : \quad \mathbf{Z}_t \otimes \mathbf{1}_n : \quad \mathbf{t} \otimes \mathbf{Z}_s : \mathbf{Z}_t \otimes \breve{\mathbf{X}}_s : \mathbf{Z}_t \otimes \mathbf{Z}_s],
\end{array} \tag{16.23}$$

where $\breve{\mathbf{x}} = (\mathbf{u} : \mathbf{v} : \mathbf{x}_s)$, $\mathbf{x}_s = \mathbf{v} \Box \mathbf{u}$, and covariance of the random effects is given by

$$\mathbf{G} = \text{blockdiag}(\mathbf{F}_{(s)}, \mathbf{F}_{(t)}, \mathbf{F}_{(s,t)})^{-1} \tag{16.24}$$

with blocks

$$\begin{array}{rcl}
\mathbf{F}_{(s)} & = & \begin{pmatrix} \lambda_u \widetilde{\mathbf{\Sigma}}_u \otimes \mathbf{I}_d & & \\ & \lambda_v \mathbf{I}_d \otimes \widetilde{\mathbf{\Sigma}}_v & \\ & & \lambda_u \widetilde{\mathbf{\Sigma}}_u \otimes \mathbf{I}_{r-d} + \lambda_v \mathbf{I}_{q-d} \otimes \widetilde{\mathbf{\Sigma}}_v \end{pmatrix}, \\
 & & \\
\mathbf{F}_{(t)} & = & \lambda_t \widetilde{\mathbf{\Sigma}}_t \ , \\
\mathbf{F}_{(s,t)} & = & \text{blockdiag}\left(\mathbf{F}^{(1)}_{(s,t)}, \mathbf{F}^{(2)}_{(s,t)}, \mathbf{F}^{(3)}_{(s,t)}\right).
\end{array}$$

where

$$\mathbf{F}^{(1)}_{(s,t)} = \begin{pmatrix} \tau_u \widetilde{\boldsymbol{\Sigma}}_u \otimes \mathbf{I}_d & & \\ & \tau_v \mathbf{I}_d \otimes \widetilde{\boldsymbol{\Sigma}}_v & \\ & & \tau_u \widetilde{\boldsymbol{\Sigma}}_u \otimes \mathbf{I}_{r-2} + \tau_2 \mathbf{I}_{q-2} \otimes \widetilde{\boldsymbol{\Sigma}}_v \end{pmatrix},$$

$$\mathbf{F}^{(2)}_{(s,t)} = \begin{pmatrix} \tau_t \widetilde{\boldsymbol{\Sigma}}_t \otimes \mathbf{I_d} & & \\ & \tau_u \mathbf{I}_{p-d} \otimes \widetilde{\boldsymbol{\Sigma}}_u + \tau_t \widetilde{\boldsymbol{\Sigma}}_t \otimes \mathbf{I}_{q-d} & \\ & & \tau_v \mathbf{I}_{p-d} \otimes \widetilde{\boldsymbol{\Sigma}}_v + \tau_t \widetilde{\boldsymbol{\Sigma}}_t \otimes \mathbf{I}_{r-d} \end{pmatrix},$$

$$\mathbf{F}^{(3)}_{(s,t)} = \tau_u \mathbf{I}_{p-d} \otimes \widetilde{\boldsymbol{\Sigma}}_u \otimes \mathbf{I}_{r-d} + \tau_v \mathbf{I}_{p-d} \otimes \mathbf{I}_{q-d} \otimes \widetilde{\boldsymbol{\Sigma}}_2 + \tau_t \widetilde{\boldsymbol{\Sigma}}_t \otimes \mathbf{I}_{q-d} \otimes \mathbf{I}_{r-d}\,.$$

Again, estimation of fixed effects coefficients, prediction of random effects, and estimation of smoothing parameters can be done by using standard mixed models methodology. However, the size of the data sets in the spatio-temporal context, makes difficult the use of standard software. We overcome this problem by using the Generalized Linear Array Models (GLAM) algorithms developed by [34] to calculate (16.13), (16.14) and (16.15) (see [13] for details).

16.4.1 Simulation study

We undertake a small simulation study to show that an ANOVA-type model is preferable to an additive model, or a pure interaction model (which is a common spatio-temporal model with non-separable covariance structure). For simplicity we restrict our simulation to the 2d-case, and generate data from 3 possible models:

$$\eta^{(1)} = f_1(\mathbf{x}_1) + f_1(\mathbf{x}_2), \qquad \text{(\textit{``Two main effects model''})}$$
$$\eta^{(2)} = f_{1,2}(\mathbf{x}_1, \mathbf{x}_2), \text{ and} \qquad \text{(\textit{``Interaction model''})}$$
$$\eta^{(3)} = f_1(\mathbf{x}_1) + f_2(\mathbf{x}_2) + f_{1,2}(\mathbf{x}_1, \mathbf{x}_2). \qquad \text{(\textit{``Two main effects and interaction''})}$$

and,

$$\begin{aligned} f_1(\mathbf{x}_1) &= \sin(2\pi \mathbf{x}_1), \\ f_2(\mathbf{x}_2) &= \cos(3\pi \mathbf{x}_2), \text{ and} \\ f_{1,2}(\mathbf{x}_1, \mathbf{x}_2) &= 3\ \sin(2\pi \mathbf{x}_1)\ (2\mathbf{x}_2 - 1). \end{aligned}$$

We consider the case of data on a regular grid, the covariates $\mathbf{x}_1$ and $\mathbf{x}_2$ chosen in $[0, 1]$ with dimensions $n_1 = 30$ and $n_2 = 20$, respectively. Figure 16.2 shows the simulated true smooth functions and true surfaces for the proposed scenarios.

Two hundred replicates of three smooth mixed models (*additive* , *anova* and *interaction* models) were fitted for each scenario, with a combination of $\sigma = 0.25$, $\sigma = 0.5$, and $\sigma = 1$. Marginal B-splines bases $\mathbf{B}_1$ and $\mathbf{B}_2$ were calculated with 8 and 6 knots respectively, with cubic splines. Second order marginal penalties were

Figure 16.2

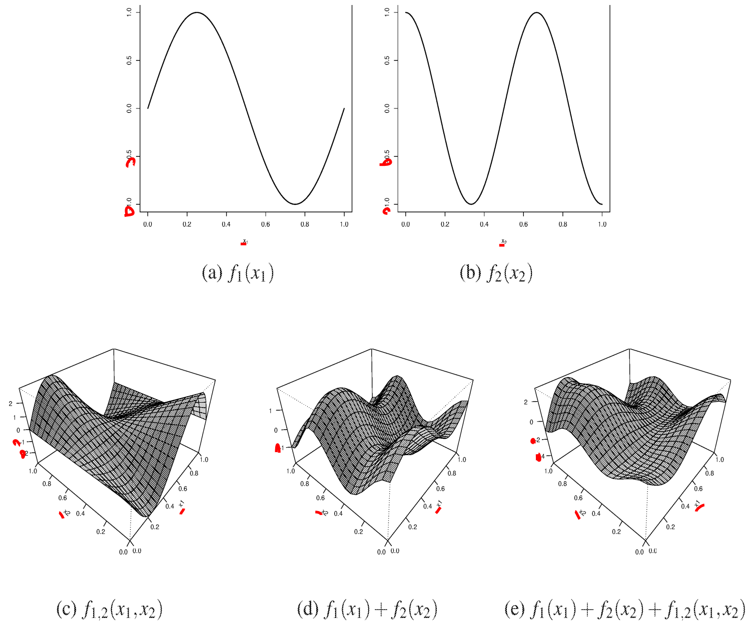

Figure 16.2 Simulated functions: (a) and (b) are the nonlinear main effects of $\mathbf{x}_1$ and $\mathbf{x}_2$; (c) is the additive surface of main effects; (d) is interaction surface and (e) is the sum of the main effects and the interaction surfaces.

used in the fitting procedure, and smoothing parameters were chosen by minimizing by REML. To check each model's performance we computed the mean square error (MSE) for each replicate. Figure 16.3 shows the box-plots of the $\log(\text{MSE})$ values for fitted smooth models. The gray shaded box-plot corresponds to the model from which we have simulated each scenario (i.e., in scenario 1, we consider $\eta^{(1)}$ as a function of two main effects, and thus the *additive* model is the favored model). S-ANOVA model clearly gave better results in scenarios 2 and 3 (interaction model, and additive plus interaction model). Additive model performed slightly better than the S-ANOVA model in scenario 1. In this case, the S-ANOVA model reduces to an additive model when the smoothing parameters in the interaction $(\tau_1, \tau_2) \to \infty$. The poor performance of the S-ANOVA model in some replicates might be due to numerical problems, since we have considered an upper bound for the smoothing parameters equal to 10^6.

Figure 16.3

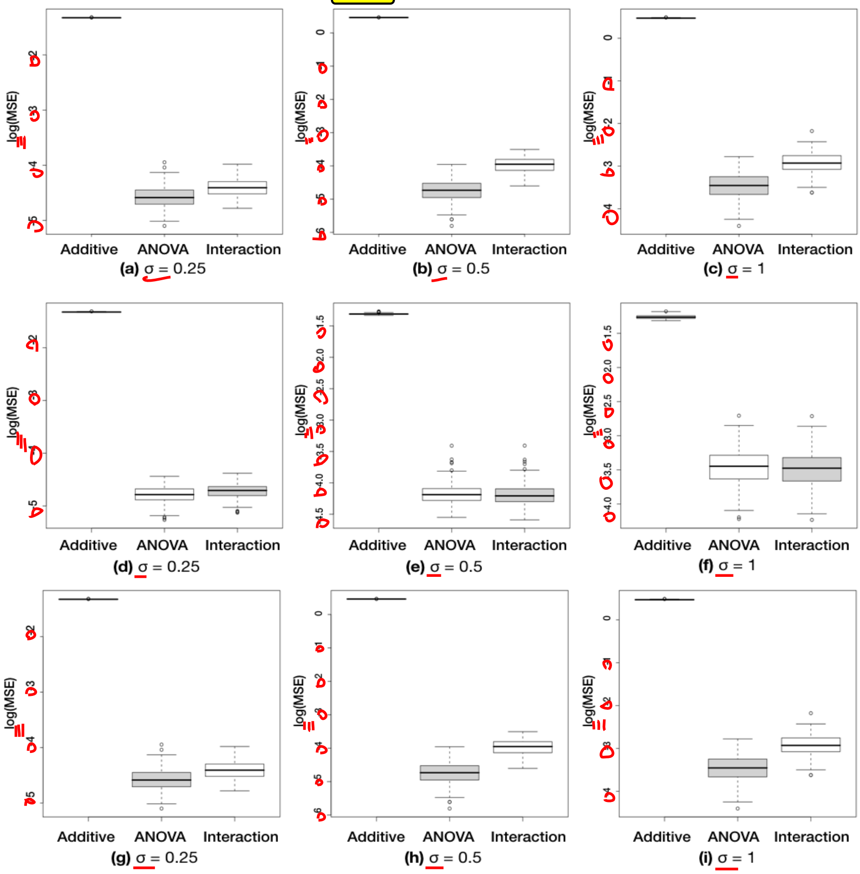

Figure 16.3 $\log$(MSE) of fitted smooth model for $R = 200$: scenario 1 (top), scenario 2 (middle), and scenario 3 (bottom).

16.5
P-spline functional spatial regression

An alternative method for modelling and predicting spatio-temporal data is using a functional data analysis based approach as the one developed in [24]. In this case, we have a sample of spatially correlated sample curves $\{y_i(t) : t \in T, i = 1, \ldots, n\}$ which have been observed with error at a finite set of time points $\{t_j : j = 1, \ldots, m\}$ for each geographical location s_i. Then, the data can be seen as realizations of a spatial functional variable (spatio-temporal stochastic process)

$$\{X(s,t) : s \in S \subseteq \mathbb{R}^2, t \in T \subseteq \mathbb{R}\}, \tag{16.25}$$

where $s = (u, v)$ is a generic data location in the spatial domain $S = U \times V$ and U, V and T are real intervals.

Let us assume that the realizations of the functional variable X are square integrable functions in the spatio-temporal domain and belong to the $pqr-$dimensional tensor function space generated by the three univariate basis of B-splines, so that

$$x(s,t) = \sum_{k=1}^{q} \sum_{l=1}^{r} \sum_{h=1}^{p} a_{klh} \mathbf{B}_k^U(u) \mathbf{B}_l^V(v) \mathbf{B}_h^T(t). \tag{16.26}$$

This means that for all spatial locations, the associated sample curves belong to the finite-dimension space generated by the basis $\left\{\mathbf{B}_h^T : h = 1, \ldots, p\right\}$, so that they admit the basis expansion

$$x(s,t) = \sum_{h=1}^{p} a_h(s) \mathbf{B}_h^T,$$

where the basis coefficients are realizations of a multivariate spatial process given by

$$a_h(s) = \sum_{k=1}^{q} \sum_{l=1}^{r} a_{klh} \mathbf{B}_k^U(u) \mathbf{B}_l^V(v).$$

For each time point, the associated sample surfaces belong to the tensor function space generated by the basis $\left\{\mathbf{B}_k^U \mathbf{B}_l^V : k = 1, \ldots, q; l = 1, \ldots, r\right\}$ so that can be expressed as

$$x(.,t) = \sum_{k=1}^{q} \sum_{l=1}^{r} a_{kl}(t) \mathbf{B}_k^U \mathbf{B}_l^V,$$

where the basis coefficients are realizations of a multivariate stochastic process given by

$$a_{kl}(t) = \sum_{h=1}^{p} a_{klh} \mathbf{B}_h^T(t).$$

Once the basis coefficients in equation (16.26) are estimated from the discrete observation y_{ij}, the spatio-temporal functional variable can be estimated at unobserved locations and times (s_0, t_0) by replacing in such equation. This way we can predict the curve of temporal evolution of the variable across the temporal domain for not sampled geographical locations and the surface of spatial evolution of the variable across the spatial domain for any time point in the temporal domain.

The basis coefficients in Equation (16.26) can be estimated by introducing the spatial variability through the following functional spatial regression model [24]

$$\mathbf{y}(t) = \mathbf{B}_s \alpha(t) + \epsilon(t),\ \forall t \in T, \tag{16.27}$$

where $\mathbf{y}(t) = (y_1(t), \ldots, y_n(t))'$ is the vector of response functions, $\mathbf{B}_s$ is the two dimensional B-spline basis for the geographical position described in the previous section, $\alpha(t) = (\alpha_1(t), \ldots, \alpha_{qr}(t))'$ is the vector of parameter functions to be estimated and $\epsilon(t) = (\epsilon_1(t), \ldots, \epsilon_n(t))'$ the vector of error terms.

Let us assume a basis representation for the functional response $\mathbf{y}(t) = \mathbf{C}B^T(t)$, and a basis representation for the the functional parameter $\alpha(t) = \mathbf{A}B^T(t)$ with $C = (c_{ih})_{n\times p}$ and $A = (a_{(kl)h})_{qr\times p}$ being the corresponding matrices of basis coefficients and $B^T(t) = (B_1^T(t), \ldots, B_p^T(t))'$ being the vector of basis functions. Then, the model becomes

$$CB^T(t) = \mathbf{B}_s\mathbf{A}B^T(t) + \epsilon(t), \ \forall t \in T.$$

The matrix of parameters is estimated by penalized sum of squares, where we have separated regularization for space and time, furthermore, we use a non-isotropic penalty term for space to allow more flexibility, i.e.

$$\begin{aligned} PSSE(y,\alpha) &= \int (\mathbf{C}B^T(t) - \mathbf{B}_s\mathbf{A}\mathbf{B}^T(t))'(\mathbf{C}B^T(t) - \mathbf{B}_s\mathbf{A}\mathbf{B}^T(t))dt + \\ &+ \ vec(A)'[PEN^{U,V,T}]vec(A), \end{aligned} \tag{16.28}$$

where $PEN^{U,V,T}$ is defined as in (16.22). Interchanging the integration and summation operations implied by the matrix products, calculating the derivatives with respect to $\mathbf{A}$, and using some properties of the Kronecker product we obtain

$$vec(\mathbf{A}) = \left[\Psi \otimes (\mathbf{B}_s'\mathbf{B}_s) + Pen^{U,V,T}\right]^{-1} vec(\mathbf{B}_s'\mathbf{C}\Psi'),$$

where $\Psi = \int B^T B^T$ is the inner product matrix between the basis functions in the temporal domain.

16.6 Application to air pollution data

In this section we illustrate both penalized approaches (spatio-temporal smoothing and functional regression) with an application of air pollution data. The data set consists medians over the years 2005 to 2012 of daily ozone levels (O_3) at the 55 monitoring stations in Spain and Portugal. The raw data set together with the map with the geographical locations are shown in 16.4.

The data can be obtained from the `R` package `openair` available at CRAN. The Openair project is an initiative of the Natural Environment Research Council (NERC) that aims to provide a collection of open-source tools for the analysis of air pollution data (more details can be found at `http://www.openair-project.org/`.

16.6.1 Spatio-temporal smoothing

The P-spline ANOVA model in (16.18) was fitted with 10 basis functions for each longitude and latitude covariates to construct the spatial main effect basis $\mathbf{B}_s$, 18 basis functions for the temporal main effect basis $\mathbf{B}_t$, and the Kronecker product of both matrices for the basis of the interaction effect.

Figure 16.4

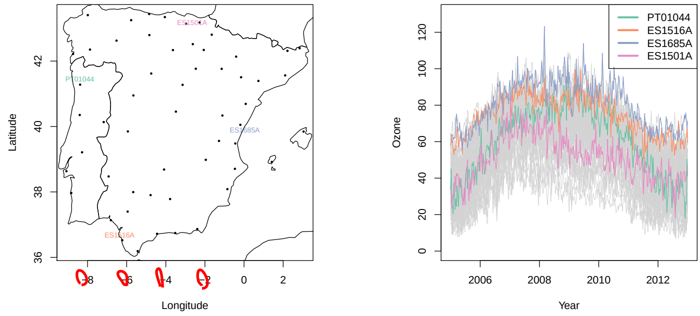

Figure 16.4 Medians of daily ozone curves (from 2002 to 2015) observed at 55 site in Spain and Portugal. Four locations are highlighted.

Figure 16.5

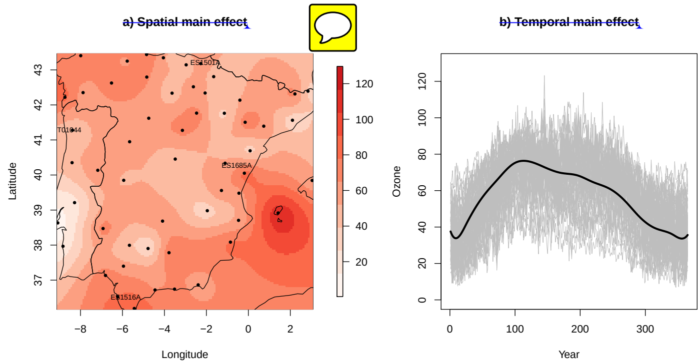

Figure 16.5 Smoothed spatial and temporal main effects for the ANOVA model.

The model is fitted using the mixed model formulation in (16.23) and REML for the estimation of the variance components. The smooth effects of space and time (i.e. $\widehat{f}_s$ and $\widehat{f}_t$ including the constant terms $\hat{\gamma}$) are shown in Figure 16.5 and represents the main spatial and temporal effects of the ANOVA decomposition. Figure 16.6 shows the space-time interaction estimated by the ANOVA model for four selected days in a year . We can clearly see that the spatial trend is not constant along the days, and it is quite different from the overall spatial trend shown in Figure 16.5, showing the need for a space-time interaction. Approximate F-test also concluded that the interaction term was significant in the model. Finally, Figure 16.7 shows the fitted curves at four selected locations (we will compare them later with the results from the spatial functional approach).

Figure 16.6

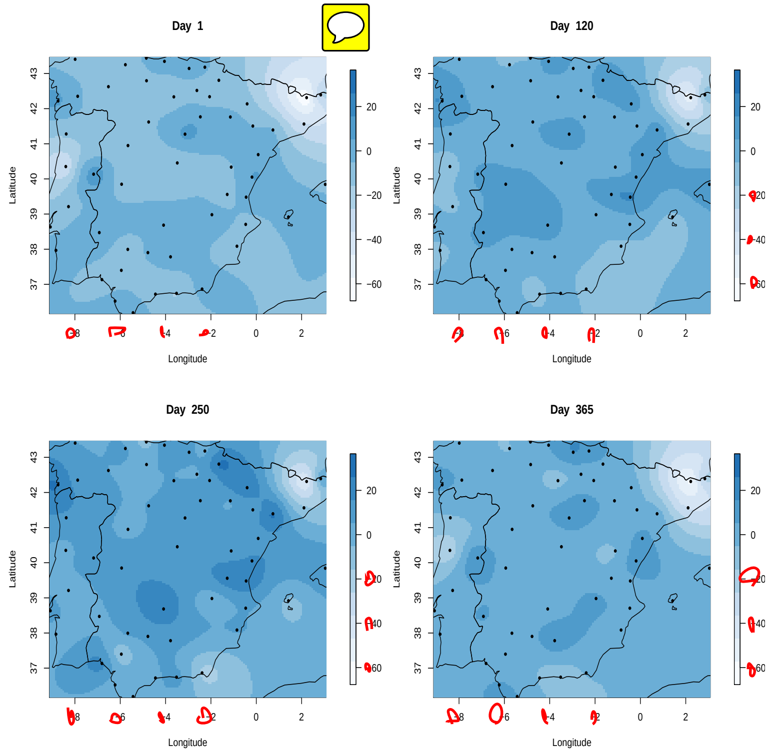

Figure 16.6 Smoothed spatio-temporal interaction for ANOVA model at four selected days.

16.6.2
Spatial functional regression

To apply the penalized functional spatial regression model 16.27 (PFSRM), we start by constructing a cubic B-spline representation of the curves in terms of 18 basis functions. The regression splines fitted this way can be seen in Figure 16.8. The PFSRM model is then estimated by using the Kronecker sum of three second-order P-spline penalties (two for space and one for time) and marginal temporal and spatial basis of dimension 365×18 and 55×100 respectively. The smoothing parameters were selected by generalized cross-validation (see [24] for details).

The predicted curves for each of the considered sites were computed by leave-one-out cross-validation. Figure 16.10 displays the predicted curves provided by PFSRM next to their mean curve and point wise confidence bands (computed as the mean $\pm$ two times the standard deviation). An example of predicted curve for one site superposed with its raw data is displayed in Figure 16.9.

Finally, we compare the fit of both approaches for four selected locations in Fig-

Figure 16.7

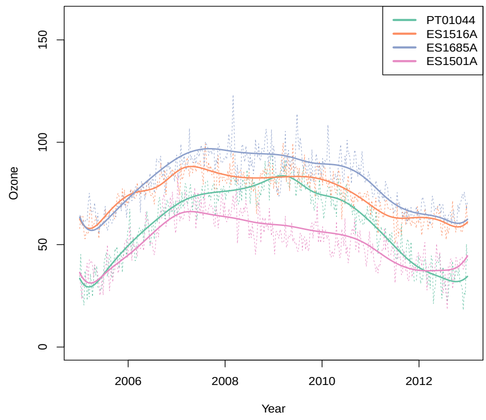

Figure 16.7 Smoothed spatio-temporal fit for ANOVA model at four selected locations.

(a) Figure 16.8 (b)

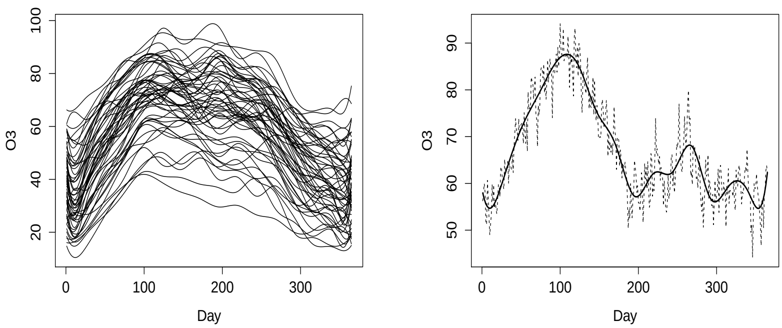

Figure 16.8 Regression splines fitted from the ozone raw data by using a cubic B-spline basis with dimension 18 (left). A sample path (dashed line) together with its basis representation (solid line) using 18 B-spline basis functions (right).

ure 16.11. The curves are quite similar although some discrepancies appear at the beginning and the end of the year.

Inspired by model (16.17), a more flexible functional approach would be

$$\mathbf{y}(t) = \gamma + f(\mathbf{u}, \mathbf{v}) + \boldsymbol{\alpha}(t) + \mathbf{B}_s\beta(t) + \epsilon(t), \ \forall t \in T, \tag{16.29}$$

and so model (16.27) becomes:

$$\mathbf{C}\boldsymbol{\theta}(t) = \mathbf{B}\mathbf{a} + \boldsymbol{\epsilon}(t), \ \forall t \in T,$$

Figure 16.9

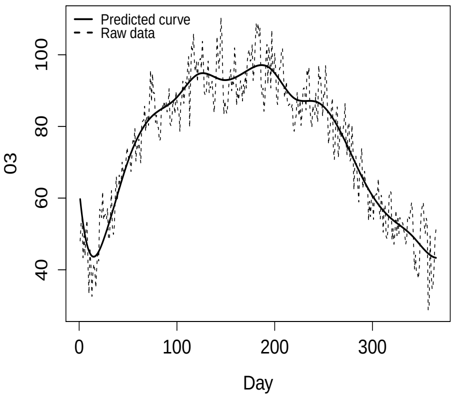

Figure 16.9 Predicted curve from the regression splines of the ozone raw data using 18 cubic B-spline basis functions (solid line) and the observed raw data (solid line) in one site.

Figure 16.10

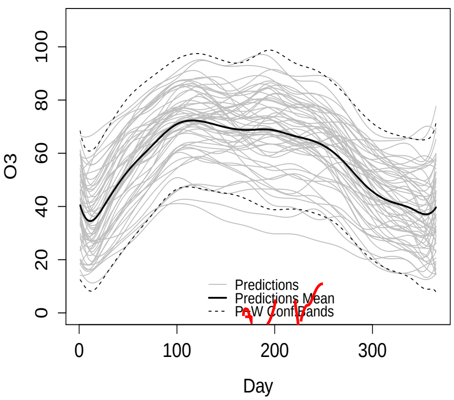

Figure 16.10 Predicted curves (grey) from the regression splines of the ozone raw data using 18 cubic B-spline basis functions) join to its mean curve (black and solid line) and the point wise confidence bands according to the mean $\pm$ two times the standard deviation (black and dashed line).

where $\mathbf{B}$ is given in (16.19, and the vector of regression coefficients is $\mathbf{a} = (\gamma, \mathbf{a}^{(s)\prime}, \mathbf{a}^{(t)\prime}, \mathbf{a}^{(st)\prime})'$. Each component of the vector would be penalized separately as described in (16.20). In order to properly identify the terms in the model, constraints need to be imposed. A possible approach is to constrain the coefficients in the model as follows (see [13]) for details)

$$\sum_h \mathbf{a}_h^{(t)} = \sum_k \mathbf{a}_{kl}^{(s)} = \sum_l \mathbf{a}_{kl}^{(s)} = 0$$
$$\sum_h \mathbf{a}_{klh}^{(st)} = \sum_k \mathbf{a}_{klh}^{(st)} = \sum_l \mathbf{a}_{klh}^{(st)} = 0.$$

Figure 16.11

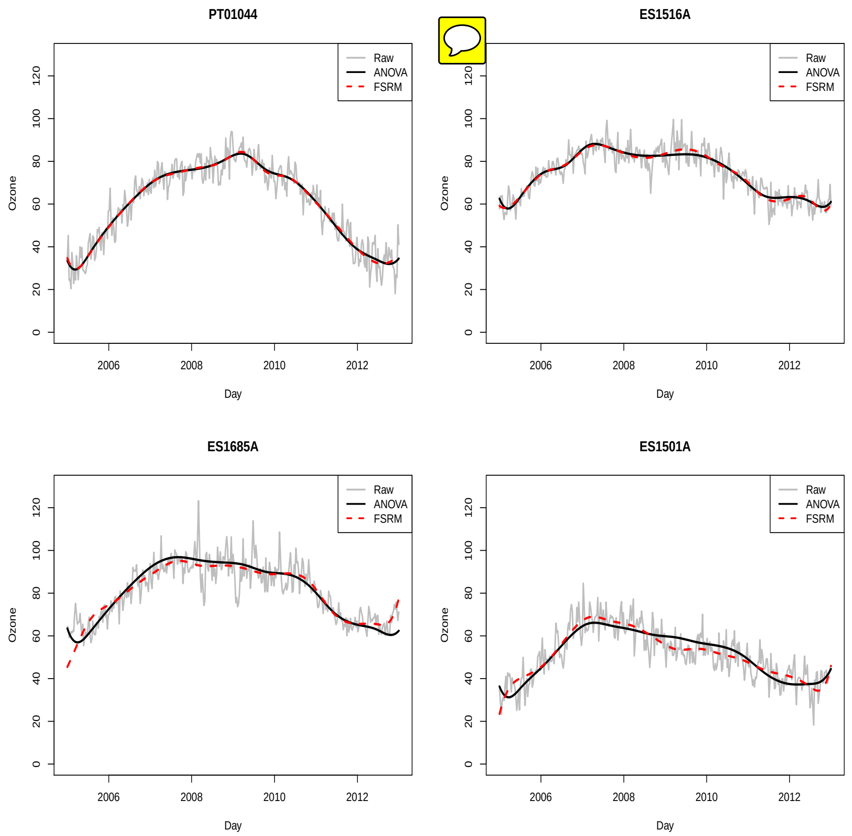

Figure 16.11 Smoothed spatio-temporal fit for ANOVA model at four selected locations.

~~16.7~~ Acknowledgments

This research has been funded by projects MTM2013-47929-P and MTM2014-52184-P from Secretaría de Estado Investigación, Desarrollo e Innovación, Ministerio de Economía y Competitividad, Spain, co-financed by European Regional Development Fund (ERDF).

References

1 Kammann, E. and Wand, M. (2003) Geoadditive Models. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, **52**, 1–18.
2 Cressie, N. and Johannesson, G. (2008) Fixed rank kriging for very large spatial data sets. *Journal of the Royals Statistical Society, B*, **70**, 209–226.
3 Lee, D.J. and Durbán, M. (2009) Smooth-car mixed models for spatial count data. *Computational Statistics & Data Analysis*, **53** (8), 2968–2979.
4 Huang, H.C. and Cressie, N. (1996) Spatio-temporal prediction of snow water equivalent using kalman filter. *Computational Statistics and Data Analysis*, **22**, 159–175.
5 Wikle, C.K. and Cressie, N. (1999) A dimension-reduced approach to space-time kalman filtering. *Biometrika*, **86**, 815–829.
6 Gössl, C., Auer, D.P., and Fahrmeir, L. (2001) Bayesian spatiotemporal inference in functional magnetic resonance imaging. *Biometrics*, **57**, 554–562.
7 Banerjee, S., Carlin, B.P., and Gelfand, A.E. (2004) *Hierarchical Modeling and Analysis for Spatial Data*, Monographs on Statistics and Applied Probability, Chapman and Hall.
8 Lee, D.J. and Durbán, M. (2011) P-spline ANOVA type interaction models for spatio-temporal smoothing. *Statistical Modelling*, **11**, 49–69.
9 Eilers, P. and Marx, B. (1996) Flexible smoothing with B-splines and penalties. *Statistical Science*, **11**, 89–121.
10 Wood, S. (2006) Low-rank scale-invariant tensor product smooths for generalized additive mixed models. *Biometrics*, **62**, 1025–1036.
11 Ruppert, D. (2002) Selecting the number of knots for penalized splines. *Journal of Computational and Graphical Statistics*, **11**, 735–757.
12 Currie, I.D. and Durbán, M. (2002) Flexible smoothing with p-splines: a unified approach. *Statistical Modelling*, **2**, 333–349.
13 Lee, D.J. (2010) *Smoothing mixed models for spatial and spatio-temporal data*, Ph.D. thesis, Universidad Carlos III de Madrid.
14 Ramsay, J.O. and Silverman, B.W. (2005) *Functional data analysis (Second Edition)*, Springer-Verlag.
15 Aguilera, A.M. and Aguilera-Morillo, M.C. (2013) Comparative study of different B-spline approaches for functional data. *Mathematical and Computer Modelling*, **58** (7-8), 1568–1579.
16 Aguilera, A.M. and Aguilera-Morillo, M.C. (2013) Penalized PCA approaches for B-spline expansions of smooth functional data. *Applied Mathematics and Computation*, **219** (14), 7805–7819.
17 Aguilera-Morillo, M.C., Aguilera, A.M., Escabias, M., and Valderrama, M.J. (2013) Penalized spline approaches for functional logit regression. *TEST*, **22** (2), 251–277.
18 Aguilera, A.M., Aguilera-Morillo, M.C., and Preda, C. (2016) Penalized versions of functional pls regression. *Chemometrics and Intelligent Laboratory Systems*, **154**, 80–92.
19 Delicado, P., Giraldo, R., Comas, C., and Mateu, J. (2009) Statistics for spatial functional data: some recent contributions. *Environmetrics*, **21**, 224–239.
20 Giraldo, R., Delicado, P., and Mateu, J.

(2010) Continuous time-varying kriging for spatial prediction of functional data: An environmental application. *Journal of Agricultural, Biological, and Environmental Statistics*, **15** (1), 66–82.

21 Giraldo, R., Delicado, P., and Mateu, J. (2011) Ordinary kriging for function-valued spatial data. *Environmental and Ecological Statistics*, **18** (3), 411–426.

22 Giraldo, R., Mateu, J., and Delicado, P. (2012) geofd: An R Package for Function-Valued Geostatistical Prediction. *Revista Colombiana de Estadística*, **35** (3), 385–407.

23 Menafoglio, A., Secchi, P., and Dalla Rosa, M. (2013) A universal kriging predictor for spatially dependent functional data of a hilbert space. *Electronic Journal of Statistics*, **7**, 2209–2240.

24 Aguilera-Morillo, M. and Durban, M. amd Aguilera, A. (2017) Prediction of functional data with spatial dependence: a penalized approach. *Stochastic Environmental Research and Risk Assessment*, **31**, 7–22.

25 Durbán, M. and Currie, I. (2003) A note on P-spline additive models with correlated errors. *Computational Statistics*, **18**, 251–262.

26 Diggle, P. and Ribeiro, P. (2007) *Model-Based Geostatistics*, Springer Series in Statistics, Springer, New York.

27 Banerjee, S., Carlin, B., and Gelfand, A. (2005) *Hierarchical Modeling and Analysis for Spatial Data*, Monographs on Statistics and Applied Probability, Chapman & Hall, London.

28 Farhmeir, L. and Kneib, T. (2011) *Bayesian Smoothing and Regression for Longitudinal, Spatial and Event History Data*, Oxford Statistical Science Series, Oxford University Press, New York.

29 Green, P. and Silverman, B. (1994) *Nonparametric Regression and Generalized Linear Models*, Monographs on Statistics and Applied Probability, Chapman & Hall, London.

30 Wahba, G. (1978) Improper priors, spline smoothing and the problem of guarding against model errors in regression. *Journal of the Royal Statistical Society B*, **40**, 364–372.

31 Dubrule, . (1994) Comparing splines and kriging. *Computers & Geosciences*, **10** (2-3), 327–338.

32 Eilers, P., Marx, B., and Durban, M. (2015) Twenty years of p-splines. *SORT (Statistics and Operations Research Transactions)*, **39**, 149–186.

33 Wood, N. (2003) Thin plate splines regression. *Journal of the Royal Statistical Society*, **65**, 95–114.

34 Currie, I.D., Durban, M., and Eilers, P.H.C. (2006) Generalized linear array models with applications to multidimensional smoothing. *Journal of the Royal Statististical Society-Series B*, **68**, 1–22.

35 Eilers, P., Currie, I., and Durban, M. (2006) Fast and compact smoothing on large multidimensional grids. *Computational Statistics & Data Analysis*, **50**, 61–76.

36 Hastie, T. and Tibshirani, R. (1990) *Generalized Additive Models*, Monographs on Statistics and Applied Probability, Chapman and Hall, London.

37 Green, P.J. (1987) Penalized likelihood for general semi-parametric regression models. *International Statistical Review/Revue Internationale de Statistique*, **55**, 245–259.

38 Speed, T. (1991) Comment on: that BLUP is a good thing: The estimation of random effects. *Statistical Science*, **6**, 15–51.

39 Searle, S., Casella, G., and McCulloch, C. (1992) *Variance components*, Wiley Series in Probability and Mathematical Statistics, Wiley.

40 Patterson, H. and Thompson, R. (1971) Recovery of inter-block information when block sizes are unequal. *Biometrika*, **58**, 545–554.

41 Rodriguez-Alvarez, M.X., Lee, D., Kneib, T., Durban, M., and Eilers, P. (2015) Fast smoothing parameter separation in multidimensional generalized P-splines: the SAP algorithm. *Statistics and Computing*, **25**, 941–957.

42 Chen, Z. (1993) Fitting multivariate regression functions by interaction spline models. *Journal of the Royal Statistical Society Series B-Statistical Methodology*, **55**, 473–491.

43 Cu, C. (2013) *Smoothing spline Anova Models*, Springer Series in Statistics, Springer, New York.

44 Eilers, P. and Marx, B. (1998) Direct generalized additive modeling with penalized likelihood. *Computational Statistics & Data Analysis*, **28**, 193–209.