P-spline ANOVA-type interaction models for spatio-temporal smoothing

Dae-Jin Lee and María Durbán

Statistical Modelling 2011 11: 49
DOI: 10.1177/1471082X1001100104

Statistical Modelling 2011; **11**(**1**): 49–69

P-spline ANOVA-type interaction models for spatio-temporal smoothing

Dae-Jin Lee and María Durbán
Department of Statistics, Universidad Carlos III de Madrid, Spain

Abstract: In recent years, spatial and spatio-temporal modelling have become an important area of research in many fields (epidemiology, environmental studies, disease mapping, etc.). However, most of the models developed are constrained by the large amounts of data available. We propose the use of penalized splines (*P*-splines) in a mixed model framework for smoothing spatio-temporal data. Our approach allows the consideration of interaction terms which can be decomposed as a sum of smooth functions similarly as an analysis of variance decomposition. The properties of the bases used for regression allow the use of algorithms that can handle large amount of data. We show that on imposing the same constraints as in a factorial design it is possible to avoid identifiability problems. We illustrate the methodology for Europe ozone levels in the period 1999–2005.

Key words: ANOVA decomposition; mixed models; Penalized splines; spatio-temporal data

Received May 2009; **revised** November 2009; **accepted** January 2010

1 Introduction

In recent years, there has been an enormous growth of data with spatio-temporal structure. These types of data arise in many contexts such as meteorology, environmental sciences, epidemiology or demography, among others. This wide variety of settings has generated a considerable interest in the development of spatio-temporal models. However, the complexity of the models needed and the size of the datasets have made this a challenging task. Our methodological development is motivated by the analysis of ozone levels collected at several monitoring stations in Europe between 1990 and 2005. Figure 1 presents the locations of the monitoring stations and the seasonal pattern in ozone levels in four different countries (Spain, Sweden, Austria and the UK). The plots show that the stations cover a large area where spatial trends are likely to appear (mostly due to climate conditions), and a clear seasonal pattern is present along the years. Therefore, a smoothing spatio-temporal model seems suitable to estimate simultaneously the spatial and temporal trends.

Address for correspondence: Dae-Jin Lee Departamento de Estadística, Universidad Carlos III de Madrid, Escuela Politécnica Superior, Av. Universidad, 30, 28911 Leganés, Madrid, Spain. E-mail: dae-jin.lee@uc3m.es

 10.1177/1471082X1001100104

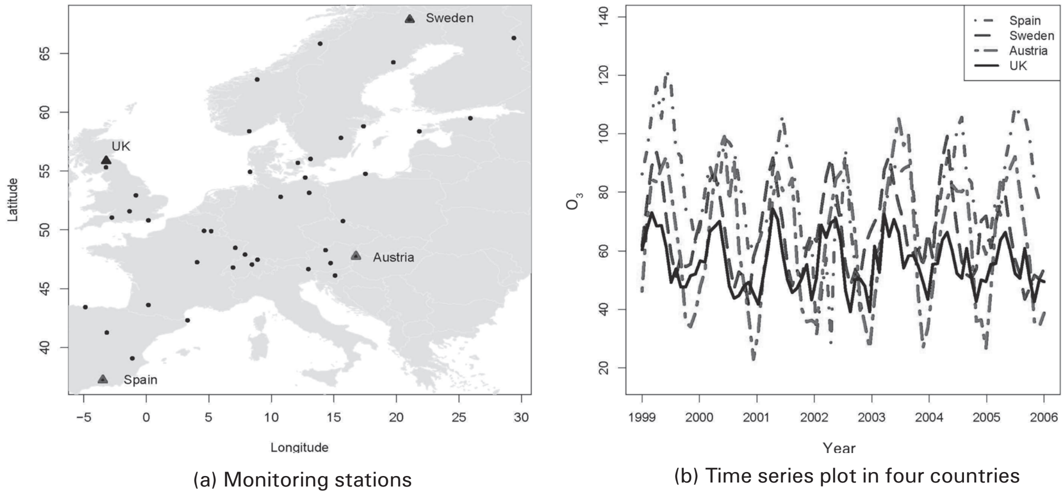


Figure 1 (a) Sample of 43 monitoring stations over Europe (b) O_3 levels in four selected countries

Here, we consider the use of penalized splines (Eilers and Marx, 1996) for smoothing spatio-temporal data. Recent papers outline the use of these methods in several applications for Gaussian and non-Gaussian responses in one or more dimensions (see, e.g., Currie *et al.*, 2004). In the multidimensional case, the common extension is the use of tensor product of B-spline bases (Currie *et al.*, 2006; Eilers *et al.*, 2006; Wood, 2006b). The work by Currie *et al.* (2006) introduced a methodology based on the development of generalized linear array methods, or GLAM, with a compact notation in which the data are arranged in an array structure or regular grid. The GLAM algorithms take advantages of the structure of the data, avoiding computational issues in storage and allow managing huge amount of data also with high speed and efficient computations in model estimation (see Currie *et al.*, 2006, Section 3). When data are scattered (as is the case of spatial data), Eilers *et al.* (2006) proposed the use of the 'row-wise' Kronecker or box product of individual B-spline basis.

Most of the common approaches in spatio-temporal smoothing are considered in the additive models framework. They extend the geoadditive models proposed by Kammann and Wand (2003) or assume a smooth function to model non-linear time effects (MacNab and Dean, 2001; Fahrmeir *et al.*, 2004; Kneib and Fahrmeir, 2006). This formulation implies that the response variable y is modelled as the sum of spatial and temporal effects of the form

$$f(\text{space}) + f(\text{time}).$$

This additive model does not account for the space–time interaction effect, and therefore, cannot reflect important features in the data. In general, this assumption implies a spatio-temporal correlation structure given by separable covariance terms for a spatial and temporal components, respectively. This approach is computationally very attractive but results too simplistic in real situations. In a very recent work, Bowman *et al.* (2009) considered spatio-temporal models within the additive framework for sulphur dioxide pollution over Europe. The space–time structure is constructed and incorporated in the additive model, the interaction terms involving time and seasonal effects are also considered in their study.

From a Bayesian perspective, recent works (Gössl *et al.*, 2001; Banerjee *et al.*, 2004) present non-separable hierarchical models based on Markov random fields in which both dependence structures are incorporated through the prior. In these models, the interaction is modelled by Kronecker products of precision matrices. This approach assumes isotropic processes, i.e., the same correlation in any spatial direction (which is unrealistic in many cases) and can be computationally intensive for large datasets.

In contrast, we propose more realistic models which allow for the consideration of the $3d$ interaction effect. We describe non-separable models for smoothing across spatial and temporal dimension simultaneously, which explicitly consider the interaction between space and time, and may easily be set into GLAM framework. Models with functional form are given by

$$f(\text{space}, \text{time}). \tag{1.1}$$

We allow for different amount of smoothing for spatial coordinates, and also for temporal dimension, and extend model (1.1) to explicitly consider different smooth additive terms for space and time and space–time interaction.

The paper is organized as follows. In Section 2, we introduce the P-spline methodology, and the representation as a mixed model in one and more dimensions. In Section 3, we propose a general methodology to represent interaction models using Smooth-analysis of variance (ANOVA) decompositions based on the mixed model reparameterization and the new bases. We also show that this is equivalent to impose linear constraints in the coefficients. In Section 4, the methods of previous sections are applied to the case of spatio-temporal data to consider models which incorporate the space–time interaction. We illustrate the use of the models with data from the analysis of ozone levels in Europe from 1990 to 2005 in Section 5. We conclude with a discussion and some concluding remarks. Some technical details on the models are deferred to the Appendix.

2 Smoothing with *P*-splines

Given a response $\boldsymbol{y}$ and covariate $\boldsymbol{x}$, a non-parametric model for the data would be given by

$$\boldsymbol{y} = f(\boldsymbol{x}) + \boldsymbol{\epsilon}, \quad \boldsymbol{\epsilon} \sim \mathcal{N}(0, \sigma^2 \boldsymbol{I}),$$

where $f(\cdot)$ is a smooth function and $\boldsymbol{\epsilon}$ is a Gaussian error term with variance $\sigma^2 \boldsymbol{I}$. The method proposed by Eilers and Marx (1996) considers f as a sum of local basis functions, i.e., $\boldsymbol{B\theta}$, where $\boldsymbol{B} = (\boldsymbol{B}_1(\boldsymbol{x}), \boldsymbol{B}_2(\boldsymbol{x}), ..., \boldsymbol{B}_c(\boldsymbol{x}))$ is an $n \times c$ matrix of B-splines (c depends on the degree and number of knots of the B-spline) constructed from the covariate $\boldsymbol{x}$, and $\boldsymbol{\theta}$ is the vector of regression coefficients. Although other bases could be considered, we choose the use of B-splines because they have better numerical properties and allow for an easy representation as mixed models and multidimensional smoothing.

The P-spline approach minimizes the penalized sum of squares

$$S(\boldsymbol{\theta}; \boldsymbol{y}, \lambda) = (\boldsymbol{y} - \boldsymbol{B\theta})'(\boldsymbol{y} - \boldsymbol{B\theta}) + \boldsymbol{\theta}'\boldsymbol{P\theta}, \tag{2.1}$$

where $\boldsymbol{P}$ is a discrete penalty matrix which depends on a smoothing parameter λ. This penalty term controls the smoothness of the fit applying penalties over adjacent coefficients. We define the $c \times c$ matrix $\boldsymbol{P}$, as

$$\boldsymbol{P} = \lambda \boldsymbol{D}'\boldsymbol{D}, \tag{2.2}$$

where $\boldsymbol{D}$ is a difference matrix, applied directly to the regression coefficients. For the rest of the paper, we consider a second-order difference as

$$(\boldsymbol{\theta}_1 - 2\boldsymbol{\theta}_2 + \boldsymbol{\theta}_3)^2 + \cdots + (\boldsymbol{\theta}_{c-2} - 2\boldsymbol{\theta}_{c-1} + \boldsymbol{\theta}_c)^2 = \boldsymbol{\theta}'\boldsymbol{D}'\boldsymbol{D\theta},$$

where $\boldsymbol{D}'\boldsymbol{D}$, in this case, defines a second-order penalty; but also first-or third-order penalties can also be used.

For a given value of λ, the minimization of (2.1) yields

$$\widehat{\boldsymbol{\theta}} = \left(\boldsymbol{B}'\boldsymbol{B} + \boldsymbol{P}\right)^{-1} \boldsymbol{B}'\boldsymbol{y}. \tag{2.3}$$

The choice of λ is, in general, subject to a certain criterion. Three possibilities are the Akaike information criterion (AIC) (Akaike, 1973), the Bayesian information criterion (Schwarz, 1978) and generalized cross-validation (Craven and Wahba, 1979). These criteria are straightforward to calculate P-splines, but calculation of effective dimension of the model is needed. Hastie and Tibshirani (1990) suggest to approximate the effective dimension or degrees of freedom (**edf**) with the hat-matrix of the smoother. The hat-matrix of the P-spline is defined as

$$\boldsymbol{H} = \boldsymbol{B}\left(\boldsymbol{B}'\boldsymbol{B} + \boldsymbol{P}\right)^{-1} \boldsymbol{B}'. \tag{2.4}$$

The properties of the trace of a matrix allow us to use a more computationally efficient expression for the edf:

$$\mathbf{edf} = \text{trace}(\boldsymbol{H}) = \text{trace}(\left(\boldsymbol{B}'\boldsymbol{B} + \boldsymbol{P}\right)^{-1} \boldsymbol{B}'\boldsymbol{B}), \tag{2.5}$$

which involves only $c \times c$ matrices, instead of the calculation of the $\boldsymbol{H}$ of $n \times n$.

2.1 Mixed model representation of *P*-splines

In recent years, smoothing techniques based on splines have become very popular, mostly due to their inclusion in the linear mixed model framework. The interest on this representation is due to the possibility of including smoothing in a large set of models (random effects models, correlated data or longitudinal studies, survival analysis), and the use of the methodology already developed for mixed models for the estimation and inference (see Brumback and Rice, 1998; Verbyla *et al.*, 1999; Wand, 2003; Welham *et al.*, 2007, among others).

The smoothing parameter, λ, becomes the ratio between the variance of the residuals and the variance of the random effect. The variance components can be estimated by residual or restricted maximum likelihood (see Patterson and Thompson, 1971; Schall, 1991). The goal here is to set a new basis which allows the representation of a P-spline model and its associated penalty as a mixed model:

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{Z\alpha} + \boldsymbol{\epsilon}, \quad \boldsymbol{\alpha} \sim \mathcal{N}(0, \boldsymbol{G}),\ \boldsymbol{\epsilon} \sim \mathcal{N}(0, \sigma^2 \boldsymbol{I}), \tag{2.6}$$

where $\boldsymbol{G} = \sigma^2_{\boldsymbol{\alpha}} \Lambda$ is the variance components matrix for the random effects $\boldsymbol{\alpha}$, for some definite positive matrix Λ and $\boldsymbol{X}$ and $\boldsymbol{Z}$ are the fixed and random effects matrices. This representation decomposes the fitted values as the sum of a polynomial/unpenalized part ($\boldsymbol{X\beta}$) and a non-linear/penalized ($\boldsymbol{Z\alpha}$) smooth term. The formulation as a mixed model is based on a reparameterization of the original non-parametric model. There are several alternatives depending on the bases and the penalty used, e.g, Wand (2003) described the representation with truncated power functions as bases, and Currie and Durbán (2002) and Currie *et al.* (2006) described the representation using B-splines bases.

Since this transformation is not unique, in this paper we follow a approach similar to Lee and Durbán (2009). We need to find a one-to-one transformation, $\boldsymbol{T}$, such that $\boldsymbol{BT} = [X : Z]$, $\boldsymbol{B\theta} = \boldsymbol{X\beta} + \boldsymbol{Z\alpha}$ and $[X : Z]$ has full rank, i.e.,

$$\boldsymbol{BTT'\theta} = [X : Z]\underbrace{\boldsymbol{T'\theta}}_{\omega}, \quad \omega' = (\boldsymbol{\beta}', \boldsymbol{\alpha}').$$

Under these conditions, the penalty $\boldsymbol{\theta}'\boldsymbol{P\theta}$ becomes $\boldsymbol{\alpha}'\boldsymbol{F\alpha}$, for some block-diagonal matrix $\boldsymbol{F}$.

The transformation we propose is based on the singular value decomposition (SVD) of the penalty matrix (2.2). Let $\boldsymbol{U\Sigma U}'$ be the SVD of $\boldsymbol{D}'\boldsymbol{D}$, assuming a second-order penalty, the diagonal matrix of eigenvalues has two zeroes and $c - 2$ positive eigenvalues, then $\boldsymbol{\Sigma} = \mathrm{diag}(0, 0, \widetilde{\boldsymbol{\Sigma}})$. We take $\boldsymbol{U}_n = [\boldsymbol{1}^* : \boldsymbol{u}^*]$ as the eigenvectors corresponding to the null space of the SVD, where $\boldsymbol{1}^*$ is $(1, ..., 1)'/\sqrt{c}$ and $\boldsymbol{u}^*$ is the vector $(1, ..., c)$ centered and scaled to have unit length. The matrix $\boldsymbol{U}_s$ is the sub-matrix which corresponds to the positive eigenvalues $\widetilde{\boldsymbol{\Sigma}}$. For the 1$d$ case, the transformation is defined as $\boldsymbol{T} = [\boldsymbol{U}_n : \boldsymbol{U}_s]$, with $\boldsymbol{T}$ orthogonal and the new coefficients are given by

$$\boldsymbol{\beta} = \boldsymbol{U}_n'\boldsymbol{\theta} \quad \text{and} \quad \boldsymbol{\alpha} = \boldsymbol{U}_s'\boldsymbol{\theta},$$

and the mixed model basis $\boldsymbol{X} = \boldsymbol{B}\boldsymbol{U}_n$ and $\boldsymbol{Z} = \boldsymbol{B}\boldsymbol{U}_s$. Furthermore, since $\boldsymbol{\beta}$ is unpenalized, we may replace $\boldsymbol{X} = \boldsymbol{B}[\mathbf{1}^* : \boldsymbol{u}^*]$ by $[\mathbf{1} : \boldsymbol{x}]$, where $\mathbf{1}$ is a vector of ones and $\boldsymbol{x}$ is the covariate. With this transformation, the penalty $\boldsymbol{\theta}'\boldsymbol{P}\boldsymbol{\theta}$ becomes $\boldsymbol{\alpha}'\boldsymbol{F}\boldsymbol{\alpha}$, for some $\boldsymbol{F}$. Given this reparameterization, we have that $\boldsymbol{F} = \boldsymbol{T}'\boldsymbol{P}\boldsymbol{T}$; in the $1d$ case, the penalty is the diagonal matrix $\boldsymbol{F} = \lambda\widetilde{\boldsymbol{\Sigma}}$.

Given the new basis and the new penalty, the penalized sum of squares (2.1) becomes

$$S(\boldsymbol{\beta}, \boldsymbol{\alpha}; \lambda) = (\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{Z}\boldsymbol{\alpha})'(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{Z}\boldsymbol{\alpha}) + \boldsymbol{\alpha}'\boldsymbol{F}\boldsymbol{\alpha}. \tag{2.7}$$

Taking derivatives on (2.7) with respect to the parameters, it is straightforward to obtain the standard mixed model equations for $\widehat{\boldsymbol{\beta}}$ and $\widehat{\boldsymbol{\alpha}}$, with variance components matrix defined as $\boldsymbol{G} = \sigma^2\boldsymbol{F}^{-1}$. Given the new transformed basis and penalty, it is possible to obtain the hat-matrix and its trace as in (2.4) and (2.5).

2.2 Multidimensional smoothing with tensor products

The natural extension of the P-spline methodology to more than one dimension is the use of tensor products of B-splines. Let us suppose for simplicity a general two dimensional case. A smooth model for the response y_{ij} would be given by

$$\boldsymbol{y} = f(\boldsymbol{x}_1, \boldsymbol{x}_2) + \boldsymbol{\epsilon} = \boldsymbol{B}\boldsymbol{\theta} + \boldsymbol{\epsilon}, \quad \boldsymbol{\epsilon} \sim \mathcal{N}(0, \sigma^2\boldsymbol{I}), \tag{2.8}$$

with covariates $\boldsymbol{x}_1' = (x_{11}, ..., x_{1n_1})$ and $\boldsymbol{x}_2' = (x_{21}, ..., x_{2n_2})$, where $\boldsymbol{B}$ is the regression basis constructed from the covariates $(\boldsymbol{x}_1, \boldsymbol{x}_2)$. As we pointed out before, we use B-spline bases since their extension to two or more dimensions can be easily done by using tensor products. For data in a regular grid (as mortality life tables or images), the regression matrix $\boldsymbol{B}$ is constructed from the Kronecker product of marginal bases:

$$\boldsymbol{B} = \boldsymbol{B}_2 \otimes \boldsymbol{B}_1, \ n_1 n_2 \times c_1 c_2 \tag{2.9}$$

(see Currie *et al.*, 2004, 2006). In the spatial smoothing context, where the $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ are, respectively, the geographic longitude and latitude, we have data as (x_{1i}, x_{2i}, y_i). Lee and Durbán (2009) proposed the use of the row-wise Kronecker product, or box product, defined in Eilers *et al.* (2006) (denoted by $\Box$ symbol):

$$\boldsymbol{B} = \boldsymbol{B}_2 \Box \boldsymbol{B}_1 = (\boldsymbol{B}_2 \otimes \mathbf{1}_{c_1}') \odot (\mathbf{1}_{c_2}' \otimes \boldsymbol{B}_1), n \times c_1 c_2, \tag{2.10}$$

where $\odot$ is the 'element-wise' matrix product and $\mathbf{1}_{c_1}$ and $\mathbf{1}_{c_2}$, are column vectors of ones of length c_1 and c_2, respectively. The operations in (2.10) are such that each row of $\boldsymbol{B}$ is the Kronecker product of the corresponding rows of $\boldsymbol{B}_2$ and $\boldsymbol{B}_1$. In the spatial case, where we have n spatial locations, both bases $\boldsymbol{B}_2$ and $\boldsymbol{B}_1$ have the same number of n rows.

It is worth mentioning that the use of the tensor products (Kronecker or box) depends on the data structure (grid or scattered), and it only affects the model basis, but the penalty used is the same in both cases. This penalty is applied over the regression coefficients which can be always set into array form. Let $\boldsymbol{\Theta}$, $c_1 \times c_2$, be the matrix of regression coefficients, where $\boldsymbol{\theta} = \text{vec}(\boldsymbol{\Theta})$, $c_1c_2 \times 1$. The penalty matrix for a $2d$ P-spline model is

$$\boldsymbol{P} = \lambda_1 \boldsymbol{I}_{c_2} \otimes \boldsymbol{D}_1'\boldsymbol{D}_1 + \lambda_2 \boldsymbol{D}_2'\boldsymbol{D}_2 \otimes \boldsymbol{I}_{c_1}, \tag{2.11}$$

where $\boldsymbol{D}_q$, with $q = 1, 2$, is again a second-order difference matrix. Then, (2.11) applies the discrete penalties over the rows and columns of $\boldsymbol{\Theta}$ and allows for anisotropic smoothing ($\lambda_1 \neq \lambda_2$), since the amount of smoothing can be different in each dimension (longitude and latitude).

The mixed model reparameterization presented in Section 2.1 can be extended to the multidimensional case. For example, using the SVD of the penalty (2.11), we obtain the mixed model bases $\boldsymbol{X}$ and $\boldsymbol{Z}$ as the tensor product of the marginal basis. Depending on the data structure, we use the Kronecker product ($\otimes$) or the box product ($\square$). For the spatial case, we obtain

$$\boldsymbol{X} = \boldsymbol{X}_2 \square \boldsymbol{X}_1 \quad \text{and} \tag{2.12}$$
$$\boldsymbol{Z} = [\boldsymbol{X}_2 \square \boldsymbol{Z}_1 : \boldsymbol{Z}_2 \square \boldsymbol{X}_1 : \boldsymbol{Z}_2 \square \boldsymbol{Z}_1], \tag{2.13}$$

with $\boldsymbol{X}_q = [\boldsymbol{1}_q : \boldsymbol{x}_q]$ and $\boldsymbol{Z}_q = \boldsymbol{B}_q \boldsymbol{U}_{qs}$ as the mixed model bases for $q = 1, 2$.

It can be shown that the penalty is a block-diagonal matrix defined by

$$\boldsymbol{F} = \begin{pmatrix} \lambda_1 \boldsymbol{I}_2 \otimes \widetilde{\boldsymbol{\Sigma}}_1 & & \\ & \lambda_2 \widetilde{\boldsymbol{\Sigma}}_2 \otimes \boldsymbol{I}_2 & \\ & & \lambda_2 \widetilde{\boldsymbol{\Sigma}}_2 \otimes \boldsymbol{I}_{c_1-2} + \lambda_1 \boldsymbol{I}_{c_2-2} \otimes \widetilde{\boldsymbol{\Sigma}}_1 \end{pmatrix}. \tag{2.14}$$

The decomposition of the bases (2.12) and (2.13) and the penalty (2.14) facilitates a key result for the rest of the paper; it allows us to represent the fitted $2d$ surface in terms of three components: a term for $\boldsymbol{x}_1$, a term for $\boldsymbol{x}_2$ and an interaction term $(\boldsymbol{x}_1, \boldsymbol{x}_2)$, which is not possible to obtain using the original tensor product bases in (2.9) or (2.10). In order to demonstrate this decomposition, we can expand the expressions for $\boldsymbol{X}$ and $\boldsymbol{Z}$ as follows:

$$\boldsymbol{X} \equiv [\boldsymbol{1}_n : \boldsymbol{x}_1 : \boldsymbol{x}_2 : \boldsymbol{x}_2 \square \boldsymbol{x}_1] \quad \text{and} \tag{2.15}$$
$$\boldsymbol{Z} \equiv [\boldsymbol{Z}_1 : \boldsymbol{Z}_2 : \boldsymbol{Z}_2 \square \boldsymbol{x}_1 : \boldsymbol{x}_2 \square \boldsymbol{Z}_1 : \boldsymbol{Z}_2 \square \boldsymbol{Z}_1], \tag{2.16}$$

the symbol $\equiv$ denoting the right-hand side of (2.15) and (2.16) have the same elements as (2.12) and (2.13), but in different order. We can identify the elements of the mixed model bases corresponding to each of the terms of the decomposition and obtain the overall $2d$ fit $f(\boldsymbol{x}_1, \boldsymbol{x}_2)$ in terms of the sum of $f(\boldsymbol{x}_1) + f(\boldsymbol{x}_2) + f(\boldsymbol{x}_1, \boldsymbol{x}_2)$. This decomposition can be viewed as an additive model with an interaction or as we detail in next section as an ANOVA model.

3 Smooth-ANOVA decomposition models

In the context of multidimensional smoothing, sometimes the interest lies in fitting complex models with functional form given by

$$\mathbb{E}[\boldsymbol{y}] = f_0 + \sum_{i=1}^{d} f_i(\boldsymbol{x}_i) + \sum_{i<j} f_{ij}(\boldsymbol{x}_i, \boldsymbol{x}_j) + \cdots + f_{1,\ldots,d}(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_d), \qquad (3.1)$$

where f_0 is a constant term and $f(\cdot)$ are smooth functions of the covariates. The decomposition in (3.1) can be viewed as a classical ANOVA (additive models (Hastie and Tibshirani, 1990) are a special case of model (3.1) when only main effects are included). These models have been considered in the literature in the context of Smoothing Splines, as SS-ANOVA models (see Chen, 1993; Gu, 2002). However, their use has been limited, mostly due to issues related to identifiability constraints and computational cost. As an alternative, we propose the use of penalized splines in models of the form (3.1). *P*-splines are based on low-rank bases functions, and so, they present an advantage over the SS-ANOVA approach. We also develop a method to construct identifiable models based on the mixed model representation of *P*-splines introduced in the previous section, as a Smooth-ANOVA (S-ANOVA) model. There are other alternatives to avoid the identifiability problems: *(i)* add a ridge penalty on the system of equations in (2.3) as in Marx and Eilers (1998) or *(ii)* identify and impose the constraints numerically (see Wood, 2006a). However, for higher order interactions, the first alternative requires a correct definition of the ridge penalty, and the second method may run into numerical problems.

For simplicity, we follow the spatial case illustrated in the previous section to introduce our S-ANOVA model. Suppose we now want to fit a model of the form

$$\boldsymbol{y} = \gamma + f_1(\boldsymbol{x}_1) + f_2(\boldsymbol{x}_2) + f_s(\boldsymbol{x}_1, \boldsymbol{x}_2) + \boldsymbol{\epsilon}, \qquad (3.2)$$

the model is defined in terms of two main effects for geographic coordinates (i.e., f_1 and f_2) and a spatial (two-way) interaction, f_s.

The *P*-spline methodology requires a basis and a penalty. The *B*-spline basis for this model is

$$\boldsymbol{B} = [\boldsymbol{1}_n : \boldsymbol{B}_1 : \boldsymbol{B}_2 : \boldsymbol{B}_s], \qquad (3.3)$$

where $\boldsymbol{1}_n$ is a column of ones of length n, for the intercept term, $\boldsymbol{B}_1$ and $\boldsymbol{B}_2$ the *B*-spline basis for the coordinates and $\boldsymbol{B}_s$ is the spatial, $n \times c_1c_2$, basis defined in (2.10). Then, model (3.2) can be written as

$$\boldsymbol{y} = \boldsymbol{B\theta} + \boldsymbol{\epsilon} = \gamma\boldsymbol{1}_n + \boldsymbol{B}_1\boldsymbol{\theta}_1 + \boldsymbol{B}_2\boldsymbol{\theta}_2 + \boldsymbol{B}_s\boldsymbol{\theta}_s + \boldsymbol{\epsilon},$$

where $\boldsymbol{\theta} = (\gamma, \boldsymbol{\theta}_1', \boldsymbol{\theta}_2', \boldsymbol{\theta}_s')'$ is the vector of regression coefficients. The penalty is of the form

$$\boldsymbol{P} = \begin{pmatrix} 0 & \cdots & & \\ \vdots & \lambda_1 \boldsymbol{D}_1'\boldsymbol{D}_1 & & \\ & & \lambda_2 \boldsymbol{D}_2'\boldsymbol{D}_2 & \\ & & & \tau_2 \boldsymbol{D}_2'\boldsymbol{D}_2 \otimes \boldsymbol{I}_{c_1} + \tau_1 \boldsymbol{I}_{c_2} \otimes \boldsymbol{D}_1'\boldsymbol{D}_1 \end{pmatrix} \tag{3.4}$$

and has a block-diagonal structure, where each block corresponds to the penalty over the coefficients of the model (note that the constant term is not penalized, and therefore the first entry is zero).

However, the regression matrix (3.3) (of dimension $n \times (1 + c_1 + c_2 + c_1c_2)$) is not full rank, (rank($\boldsymbol{B}$) = c_1c_2), so there are $(1 + c_1 + c_2)$ linearly dependent columns, and model (3.2) should be carefully modified in order to preserve the identifiability. Wood (2006b) pointed the need to construct appropiate model bases and penalties and to impose constraints to maintain the model identifiability and Wood (2006a, Chapter 4), suggested the use of the QR decomposition in order to identify numerically any linear dependent columns of model bases and remove them.

In contrast, we propose a more elegant way to construct identifiable model bases and penalties based on the reparameterization shown in Section 2.1. The mixed model representation of model (3.2) allows us to find that some terms are repeated, and we can avoid the problem by removing the column vector of $\mathbf{1}_n$'s in the fixed effects matrices.

It can also be demonstrated that this systematic elimination process is equivalent to impose the usual linear constraints on the model coefficients. This can be easily proved by recovering the penalty of the original parametrization. In model (3.2), these constraints are exactly equivalent to those applied in a factorial design with two main effects and a two-way interaction, i.e.,

$$\sum_i^{c_1} \boldsymbol{\theta}_{1i} = \sum_j^{c_2} \boldsymbol{\theta}_{2j} = 0\,, \qquad \text{for main effects and} \tag{3.5}$$

$$\sum_i^{c_1} \boldsymbol{\Theta}_{ij} = \sum_j^{c_2} \boldsymbol{\Theta}_{ij} = 0\,, \text{ for two-way interactions.} \tag{3.6}$$

Durbán and Currie (2003) used a similar approach in additive models context. To achieve identifiability, they proposed centering the B-spline basis matrices $\boldsymbol{B}_q$ leading to $\boldsymbol{B}_q^* = (\boldsymbol{I}_n - \mathbf{1}\mathbf{1}'/n)\boldsymbol{B}_q$, for $q = 1, 2$. This result premultiplies the basis by an $n \times n$ centering matrix, which centers the B-spline basis by rows. The approach presented in this section is in essence equivalent, since we constraint the coefficients which results equivalent as centering the B-spline basis by columns, i.e., $\boldsymbol{B}_q(\boldsymbol{I}_c - \mathbf{1}\mathbf{1}'/c)$, where c is the number of columns of the basis (much smaller than n).

Table 1 Set of regression coefficient constraints in a full 3*d* *P*-spline ANOVA-type model in (3.7)

	Constraints
Main effects	$\sum_i^{c_1} \boldsymbol{\theta}_i^{(1)} = \sum_j^{c_2} \boldsymbol{\theta}_j^{(2)} = \sum_k^{c_3} \boldsymbol{\theta}_k^{(3)} = 0$
Two-way interaction	$\sum_i^{c_1} \boldsymbol{\theta}_{ij}^{(1,2)} = \sum_j^{c_2} \boldsymbol{\theta}_{ij}^{(1,2)} = \sum_i^{c_1} \boldsymbol{\theta}_{ik}^{(1,3)} = \sum_k^{c_3} \boldsymbol{\theta}_{ik}^{(1,3)} = \sum_j^{c_2} \boldsymbol{\theta}_{jk}^{(2,3)} = \sum_k^{c_3} \boldsymbol{\theta}_{jk}^{(2,3)} = 0$
Three-way interaction	$\sum_i^{c_1} \boldsymbol{\theta}_{ijk}^{(1,2,3)} = \sum_j^{c_2} \boldsymbol{\theta}_{ijk}^{(1,2,3)} = \sum_k^{c_3} \boldsymbol{\theta}_{ijk}^{(1,2,3)} = 0$

These results can be extended for more dimensions. For example, in the 3d case, a full ANOVA-type decomposition with terms:

$$\begin{aligned} \boldsymbol{y} = \gamma &+ f_1(\boldsymbol{x}_1) + f_2(\boldsymbol{x}_2) + f_3(\boldsymbol{x}_3) + \\ &+ f_{1,2}(\boldsymbol{x}_1, \boldsymbol{x}_2) + f_{1,3}(\boldsymbol{x}_1, \boldsymbol{x}_3) + f_{2,3}(\boldsymbol{x}_2, \boldsymbol{x}_3) + \\ &+ f_{1,2,3}(\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{x}_3) + \boldsymbol{\epsilon}, \end{aligned} \tag{3.7}$$

where all main effects, two-way and three-way interaction are included in the model. For model (3.7), these constraints are shown in Table 1. In practice, the methodology proposed in this paper does not require to apply the identifiability contraints. The new reparameterization allows us to construct the mixed model bases $\boldsymbol{X}$ and $\boldsymbol{Z}$, removing the corresponding vector of $\mathbf{1}_n$'s, and then construct the associated penalty $\boldsymbol{F}$ with its smoothing parameters (see Appendix A.1).

4 Spatio-temporal smoothing with *P*-splines

In this section, we propose a novel way to look at spatio-temporal data based on *P*-spline methodology. We start by using non-separable models of the form

$$\mathbb{E}[\boldsymbol{y}] = f_{st}(\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{x}_t), \tag{4.1}$$

where $f_{st}(\cdot)$ is a smooth three-dimensional function across space and time, which explicitly considers the space–time interaction. The regression basis for a 3d interaction model (4.1) is

$$\boldsymbol{B} = \boldsymbol{B}_s \otimes \boldsymbol{B}_t, \; nt \times c_s c_t, \tag{4.2}$$

where $\boldsymbol{B}_s$ is the spatial B-spline basis defined in (2.10), of dimension $n \times c_s$, $c_s = c_1 c_2$, and $\boldsymbol{B}_t$ is $t \times c_t$ marginal B-spline basis for time.

Model (4.1) and basis given by (4.2) can easily be set into GLAM framework. We replace the $nt \times 1$ response vector $\boldsymbol{y}$ by the matrix $\boldsymbol{Y}$ of dimension $t \times n$ and the coefficient vector $\boldsymbol{\theta}$ of length $c_s c_t \times 1$ by an array of coefficients $\boldsymbol{\Theta}$ of dimension $c_t \times c_s$. In matrix notation, the model can be written as

$$\mathbb{E}[\boldsymbol{Y}] = \boldsymbol{B}_t \boldsymbol{\Theta} \boldsymbol{B}_s' \, . \tag{4.3}$$

Smoothness is imposed via a penalty matrix $\boldsymbol{P}$ based on second-order difference matrices $\boldsymbol{D}_1$, $\boldsymbol{D}_2$ and $\boldsymbol{D}_t$. Then, the penalty term in three-dimensions is

$$\boldsymbol{P} = \lambda_1 \; \boldsymbol{D}_1'\boldsymbol{D}_1 \otimes \boldsymbol{I}_{c_2} \otimes \boldsymbol{I}_{c_t} + \lambda_2 \; \boldsymbol{I}_{c_1} \otimes \boldsymbol{D}_2'\boldsymbol{D}_2 \otimes \boldsymbol{I}_{c_t} + \lambda_t \; \boldsymbol{I}_{c_1} \otimes \boldsymbol{I}_{c_2} \otimes \boldsymbol{D}_t'\boldsymbol{D}_t, \qquad (4.4)$$

which implies placing penalties over each dimension of the $3d$ array $\boldsymbol{\Theta}$. The penalty (4.4) allows spatial anisotropy and also a smoothing parameter λ_t, for the temporal component. Following the same procedure as in Section 2.1, we apply the SVD over the penalty (4.4). The new bases for model (4.1) can be written in a compact notation as

$$\boldsymbol{X} = \boldsymbol{X}_s \otimes \boldsymbol{X}_t \quad \text{and} \qquad (4.5)$$

$$\boldsymbol{Z} = [\boldsymbol{Z}_s \otimes \boldsymbol{X}_t : \boldsymbol{X}_s \otimes \boldsymbol{Z}_t : \boldsymbol{Z}_s \otimes \boldsymbol{Z}_t], \qquad (4.6)$$

where $\boldsymbol{X}_s$ and $\boldsymbol{Z}_s$ are the matrices defined in (2.12) and (2.13) for the spatial case. And $\boldsymbol{X}_t$ and $\boldsymbol{Z}_t$ are the similar matrices for time dimension. Finally, using the reparameterization presented earlier, it is straightforward to obtain the block-diagonal mixed model penalty $\boldsymbol{F}$ (see details in Appendix A.1).

As we showed in Section 2.2, the construction of the new bases (4.5) and (4.6) allows us to represent the fitted values in terms of the sum of additive components plus interactions (two-way and three-way interactions). For spatio-temporal data, this decomposition may be very useful in terms of the interpretability of the model fit, since we can decompose the overall fit not only as main effects of latitude and longitude (or other covariates) but also the spatial effects (two-way interaction) and specially the interaction between space and time (three-way interactions). However, in terms of model formulation, it does not account for independent and separate penalties since we have three smoothing parameters λ_1, λ_2 and λ_t for each of the dimensions of the model. That is, the amount of smoothing used for the additive terms is also used for the interactions. In some cases (as we will show in the analysis of the ozone data), this is not realistic, and so, we will apply the P-spline ANOVA methodology to the spatio-temporal setting.

4.1 Spatio-temporal S-ANOVA model

In many cases, the main interest when considering spatio-temporal data is to check for the presence of a space–time interaction. The S-ANOVA model formulation presented in Section 3 allows us to construct a more realistic model for spatio-temporal smoothing:

$$\boldsymbol{y} = \gamma + f_s(\boldsymbol{x}_1, \boldsymbol{x}_2) + f_t(\boldsymbol{x}_t) + f_{st}(\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{x}_t) + \boldsymbol{\epsilon}\,, \quad \boldsymbol{\epsilon} \sim \mathcal{N}(0, \sigma^2), \qquad (4.7)$$

where we explicitly consider a smooth term for the spatial surface, for temporal smooth trend and a smooth term for space–time interaction. A P-spline basis for model (4.7) would be

$$\boldsymbol{B} = [\boldsymbol{1}_{nt} : \boldsymbol{B}_s \otimes \boldsymbol{1}_t : \boldsymbol{1}_n \otimes \boldsymbol{B}_t : \boldsymbol{B}_s \otimes \boldsymbol{B}_t] \qquad (4.8)$$

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

Transforming the model bases (4.8) with the reparameterization of Section 2.1, we can easily remove the repeated terms and obtain new bases

$$\begin{aligned} \boldsymbol{X} &= [\overbrace{\boldsymbol{X}_s \otimes \boldsymbol{1}_t}^{f_s(\boldsymbol{x}_1,\boldsymbol{x}_2)} : \overbrace{\boldsymbol{1}_n \otimes \boldsymbol{x}_t}^{f_t(\boldsymbol{x}_t)} : \overbrace{\boldsymbol{x}_s \otimes \boldsymbol{x}_t \;]\qquad\qquad\qquad\qquad\qquad\qquad}^{f_{st}(\boldsymbol{x}_1,\boldsymbol{x}_2,\boldsymbol{x}_t)} \\ \boldsymbol{Z} &= [\;\boldsymbol{Z}_s \otimes \boldsymbol{1}_t : \boldsymbol{1}_n \otimes \boldsymbol{Z}_t : \boldsymbol{Z}_s \otimes \boldsymbol{x}_t : (\boldsymbol{x}_1 : \boldsymbol{x}_2 : \boldsymbol{x}_s) \otimes \boldsymbol{Z}_t : \boldsymbol{Z}_s \otimes \boldsymbol{Z}_t], \end{aligned} \tag{4.9}$$

where $\boldsymbol{x}_s = \boldsymbol{x}_2 \Box \boldsymbol{x}_1$. Both matrices are exactly the same as those for the $3d$ model (4.5) and (4.6) but in a different order.

Finally, the mixed model penalty for the S-ANOVA model is of the form

$$\boldsymbol{F} = \text{blockdiag}(\boldsymbol{F}_{(s)}, \boldsymbol{F}_{(t)}, \boldsymbol{F}_{(st)}). \tag{4.10}$$

The blocks $\boldsymbol{F}_{(s)}$ and $\boldsymbol{F}_{(t)}$ correspond, respectively, to the spatial and temporal mixed model penalty terms, with smoothing parameters λ_1, λ_2 and λ_t. And the last block $\boldsymbol{F}_{(st)}$ is the penalty term for the spatio-temporal interaction with smoothing parameters τ_1, τ_2 and τ_t. This term is constructed once we have removed the linear dependency (see the details of the construction of (4.10) in Appendix A.2).

For the S-ANOVA spatio-temporal model (4.7), the resultant mixed model reparameterization is equivalent to apply only a subset of constraints in Table 1. This is a linear constraint over the temporal main effect coefficient, i.e., $\sum_{t=1}^{c_t} \boldsymbol{\theta}_t^{(\mathrm{t})} = 0$, and constraints over the spatio-temporal array of coefficients, $\boldsymbol{\Theta}^{(\mathrm{st})}$, of dimensions $c_t \times c_1 \times c_2$:

$$\sum_i^{c_1} \boldsymbol{\theta}_{t,ij}^{(\mathrm{st})} = \sum_j^{c_2} \boldsymbol{\theta}_{t,ij}^{(\mathrm{st})} = \sum_i^{c_1} \sum_j^{c_2} \boldsymbol{\theta}_{t,ij}^{(\mathrm{st})} = 0. \tag{4.11}$$

5 Application to ozone levels in Europe: period 1999–2005

A repeated exposure to ozone pollution ground level may cause important damages to human health (including asthma, reduced lung capacity or susceptibility to respiratory illnesses), ecosystems and agricultural crops. The formation of ozone is increased by hot weather and in urban industrial areas, and the concentrations over Europe also present a wide variation and large differences due to climate conditions over the continent. Therefore, it is expected that ozone concentrations around Europe present a spatio-temporal pattern.

The harmful effects of ozone have become an important issue for the development of new policies. The European Environment Agency (EEA) has established a programme to monitor changes in ozone levels in the last decade. The EEA presents annual evaluation reports of ground-level ozone pollution in Europe from

April to September, based on the information submitted to the European Commission on ozone in ambient air (further information is available at the website http://www.eea.europa.eu/). According to this annual report, although emissions of ozone precursors have been reduced over the last decade, ozone pollution levels have not changed significantly in the period 1999–2005. The analysis of the data will confirm this statement, and it will show that different countries reach the largest values of ozone at different time points.

We analysed monthly averages of air pollution by ground-level ozone (in $\mu g/m^3$ units) over Europe from January 1999 to December 2005. The data were collected in 43 monitoring stations in 15 European countries. Following the methodology described in previous sections, we fitted three models to the data: *(i) spatio-temporal S-ANOVA model*; *(ii) 3d interaction model* and *(iii) space–time additive model*. The three model formulations are then:

$$\begin{array}{rl} \text{1. } \textbf{S-ANOVA:} & f_s(x_1, x_2) + f_t(x_t) + f_{s,t}(x_1, x_2, x_t), \\ \text{2. } \textbf{Interaction:} & f_{s,t}(x_1, x_2, x_t) \quad \text{and} \\ \text{3. } \textbf{Additive:} & f_s(x_1, x_2) + f_t(x_t). \end{array} \tag{5.1}$$

In order to fit the models, we set up the B-spline bases using the following parameters: *(i)* the number of (equidistant) internal knots, ndx; *(ii)* the degree of the P-spline, $bdeg$, and *(iii)* the order of the penalty, $pord$. We selected one knot for every four or five observations. The parameters were: $bdeg = 3$ (cubic B-splines), $pord = 2$ (second-order penalty) and $ndx_{(s)} = (10, 10)$ for both spatial dimensions and $ndx_{(t)} = 21$ for time, in order to have enough flexibility to capture the seasonal time trend. Then, the spatial bases $\boldsymbol{B}_1$ and $\boldsymbol{B}_2$ are of dimension 43×13 and $\boldsymbol{B}_t$ has dimension 84×24.

The mixed model formulation is straightforward following the methodology proposed in the paper: we construct matrices $\boldsymbol{X}$ and $\boldsymbol{Z}$ and the block-diagonal penalty $\boldsymbol{F}$ for each model. We compared the performance of the models in terms of the **AIC** calculated as

$$\mathbf{AIC} = \mathbf{Dev} + 2\ \mathbf{edf},$$

where **Dev** is the deviance calculated as $\mathbf{Dev}=\sum_i^n (y_i - \widehat{y}_i)^2$ and **edf** are the effective degrees of freedom of the model (d.f.), measured as the trace of the hat-matrix.

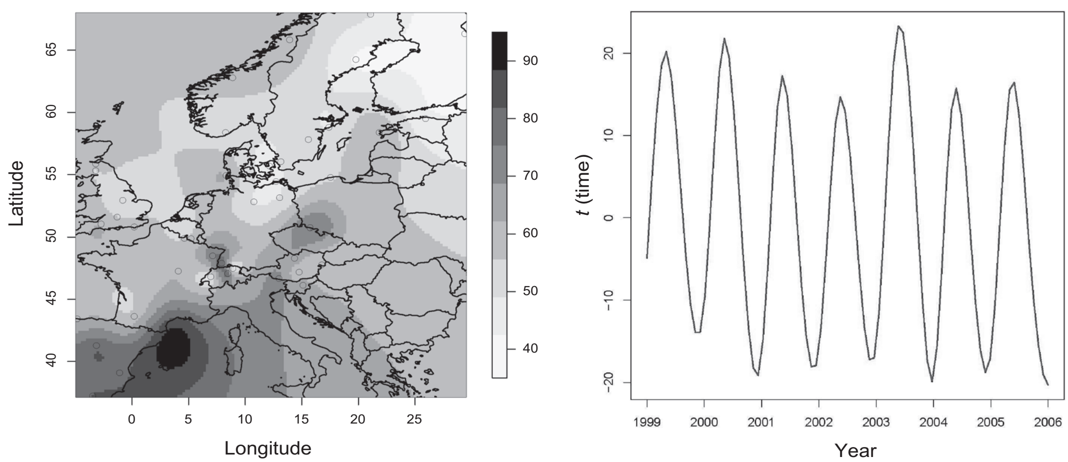
The results are summarized in Table 2. There is a superior performance of S-ANOVA and interaction models with respect to the additive model. This could be expected since it is unrealistic to force the spatial pattern of ozone concentrations to increase and decrease similarly in all locations. The interaction model, although giving a better fit, uses a large amount of edf. This is due to the fact that model has a single smoothing parameter for the temporal component. Then, the strong seasonal trend forces the model to use a small smoothing parameter (large d.f.). The S-ANOVA model performs better. It uses less d.f. because the model allows a different amount of smoothing in the additive temporal term and the spatio-temporal component, and, as we could expect, the temporal smoothing in the interaction does not need to be so strong. This results in a more parsimonious model.

Table 2 AIC and estimated degrees of freedom of fitted models

Model	AIC	Dev	d.f.
S-ANOVA	14280.73	13548.67	366.03
Interaction	14537.22	13007.12	765.05
Additive	16506.28	16374.32	65.98

Figure 2(a) shows the smoothed spatial surface for the ozone levels of the S-ANOVA model. The estimated spatial trend surface reflects a non-uniform picture across Europe, since the highest concentrations are observed in southern Europe in Mediterranean countries such as Spain, France and Italy, and the lowest levels are in north-west Europe and the UK. The seasonal cycle of ozone levels is captured by the temporal trend shown in Figure 2(b), where the highest levels are recorded during spring and summer months (April to September). The highest peak corresponds to the heat wave occurred in Europe during summer 2003. The spatio-temporal S-ANOVA model also allows the explicit modelling of the space–time interaction in addition to the spatial and temporal trends. Figure 3 shows this interaction from March to August 2002. As it can be seen from the sequence of figures, there are differences between north-west and southern and Mediterranean countries throughout the summer period.

The differences between additive and S-ANOVA models can be seen in Figure 4. We plotted the fitted values for four different monitoring stations against the raw ozone levels data. The additive model ignores the interaction and assumes a spatial



(a) Spatial surface: $f_s(x_1, x_2)$ (b) Time trend: $f_t(x_t)$

Figure 2 Spatial and temporal smooth terms for S-ANOVA model

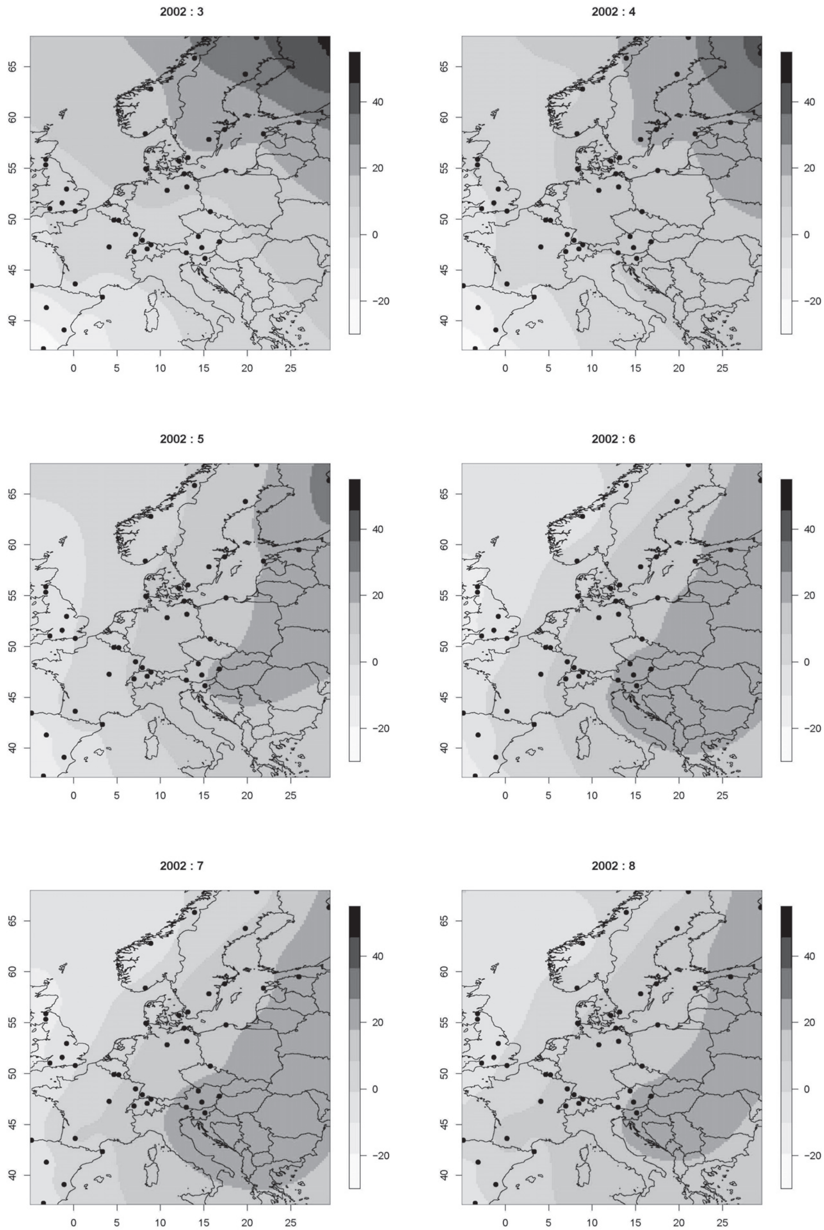
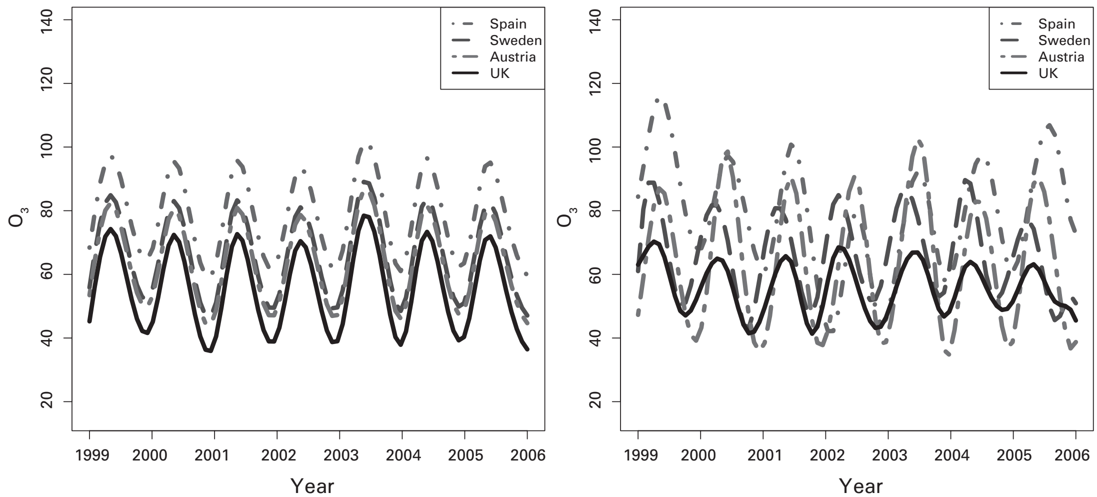


Figure 3 Spatio-temporal interaction fit for the spatio-temporal S-ANOVA model, from March to August 2002



(a) Additive model fit

(b) Spatio-temporal S-ANOVA model fit

Figure 4 Comparison of fitted values for monitoring stations in Spain, Sweden, Austria and UK

smooth surface over all monitoring stations that remains constant over time. The fitted values vary smoothly according to a seasonal pattern, but maintain the same differences among locations (Figure 4(a)). In contrast, the spatio-temporal S-ANOVA model fit is able to capture the individual characteristics of the stations throughout time. Figure 4(b) shows the particular phase and amplitude given the geographic and seasonal inter-annual variations of four monitoring stations. The high and low season for ozone concentrations are different, depending on the location and the cycle changes over time.

6 Discussion

We have presented a flexible modelling methodology for spatio-temporal data smoothing. We extend the P-spline approach to consider the smoothing over spatial and temporal dimensions by the construction of the model basis with the appropriate B-spline low-rank bases products and proposed an easy and direct procedure to avoid the identifiability problems based on the mixed model reparameterization of the model. This methodology allowed us to construct ANOVA-type models. This procedure is equivalent to apply constraints over the P-spline regression coefficients, and therefore, the connection with the classical ANOVA decomposition is straightforward. The array formulation of multidimensional P-spline models (Currie *et al.*, 2006) yields a unified framework for d-dimensional smoothing. It is possible to represent a d-dimensional $c_1 \times c_2 \times \cdots \times c_d$ array of coefficients by $\boldsymbol{\Theta}$ and apply the corresponding constraints. The interpretation of the constraints is also easier using the array form, since they are applied over each of the dimensions of the coefficients array. The array $\boldsymbol{\Theta}$ is flattened onto the dimension in which the constraints are applied, and reinstated in vector form (see Currie *et al.*, 2006; Eilers *et al.*, 2006, for software considerations).

In practice, it is also easy to extend the model by the incorporation of other relevant covariates as smooth additive terms or as interactions. One of the main benefits of the spatio-temporal S-ANOVA model proposed is the interpretation of the smoothing and the ability to visualize each of the terms of the decomposition in descriptive plots. The S-ANOVA model also gives a direct interpretation in terms of their smoothing parameters and regression coefficients, since we set independent and separate penalties and coefficients for each smooth term.

With large datasets, the computational implementation of the analyses of spatio-temporal data is very intensive and requires efficient computational methods. In the P-spline approach, the dimension of the bases involved in the smoothing depends basically on the number of knots, and therefore the dimensionality of the problem is reduced by setting a moderate number for each covariate dimension. However, when data often present a strong seasonal trend (which is very common in environmental problems), the size of the basis $\boldsymbol{B}_t$ has to be large (between 20 and 40 equidistant knots) in order to have enough degrees of freedom to capture the temporal structure. In this paper, we have found adequate number of four knots for each of the 7 years considered. If a larger sample of monitoring stations would have been considered in the study during a larger time period, the number of parameters in the interaction $\boldsymbol{B}_s \otimes \boldsymbol{B}_t$ could easily be of the order of thousands, and the computational burden prohibitive. Nevertheless, the GLAM also have an important role in the algorithms implementation, since they allow us to store the data and model matrices more efficiently and speed up the calculations. This computational aspect is a topic of current research.

A possible approach to capture the seasonal/cyclic patterns in environmental data, and also reduce the computational burden, would be to include specific structures for this setting as, e.g.: smooth components for the trend and the use of periodic (co)sine functions for the seasonal component (see Eilers *et al.*, 2008; Marx *et al.*, 2010), or periodic smoothing with specialized harmonic basis and penalties. However, the use of this type of models when the space–time interaction is present in the data would not avoid the identifiability problem addressed in this paper when an ANOVA-type model is considered. The methodology proposed, based on the mixed model representation, makes possible the incorporation of these alternatives in a more complex model.

Acknowledgements

The research of Dae-Jin Lee and María Durbán was supported by the Spanish Ministry of Science and Innovation (project MTM 2008-02901). The first author also acknowledges the financial support of Comunidad de Madrid (project CCG07-UC3M/HUM-3260) and Spanish Ministry of Education and Science (project SEJ 2005-06454). We also thank the referees for the valuable comments on our manuscript. These comments certainly helped to improve the readibility and the organization of the paper.

Appendix

A Mixed model penalties spatio-temporal *P*-splines models

This appendix presents the penalty terms for the reparameterization of 3d and spatio-temporal S-ANOVA models. We show how to construct the block-diagonal mixed model representation penalty, $\boldsymbol{F}$, associated to the new bases. It is clear to see that in both models, matrices $\boldsymbol{X}$ and $\boldsymbol{Z}$ are equivalent but their columns/blocks are in a different order. This reordering needs a proper specification of $\boldsymbol{F}$ and the smoothing parameters.

A.1 Penalty $\boldsymbol{F}$ in the 3d case

Given the 3d model (4.1), and the mixed model matrices shown in Section 4, i.e.,

$$\boldsymbol{X} = \boldsymbol{X}_s \otimes \boldsymbol{X}_t \text{ and}$$
$$\boldsymbol{Z} = [\boldsymbol{Z}_s \otimes \boldsymbol{X}_t : \boldsymbol{X}_s \otimes \boldsymbol{Z}_t : \boldsymbol{Z}_s \otimes \boldsymbol{Z}_t].$$

The block-diagonal penalty $\boldsymbol{F}$ has seven blocks:

$$\boldsymbol{F} = \text{blockdiag}(\boldsymbol{F}_{(1)}, \boldsymbol{F}_{(2)}, \boldsymbol{F}_{(1,2)}, \boldsymbol{F}_{(t)}, \boldsymbol{F}_{(1,t)}, \boldsymbol{F}_{(2,t)}, \boldsymbol{F}_{(1,2,t)}),$$

where

$$\begin{aligned}
\boldsymbol{F}_{(1)} &= \lambda_1 \boldsymbol{I}_2 \otimes \widetilde{\boldsymbol{\Sigma}}_1 \otimes \boldsymbol{I}_2,\\
\boldsymbol{F}_{(2)} &= \lambda_2 \widetilde{\boldsymbol{\Sigma}}_2 \otimes \boldsymbol{I}_2 \otimes \boldsymbol{I}_2,\\
\boldsymbol{F}_{(1,2)} &= \lambda_1 \boldsymbol{I}_{c_2-2} \otimes \widetilde{\boldsymbol{\Sigma}}_1 \otimes \boldsymbol{I}_2 + \lambda_2 \widetilde{\boldsymbol{\Sigma}}_2 \otimes \boldsymbol{I}_{c_1-2} \otimes \boldsymbol{I}_2,\\
\boldsymbol{F}_{(t)} &= \lambda_t \boldsymbol{I}_2 \otimes \boldsymbol{I}_2 \otimes \widetilde{\boldsymbol{\Sigma}}_t,\\
\boldsymbol{F}_{(1,t)} &= \lambda_1 \boldsymbol{I}_2 \otimes \widetilde{\boldsymbol{\Sigma}}_1 \otimes \boldsymbol{I}_{c_t-2} + \lambda_t \boldsymbol{I}_{c_2} \otimes \boldsymbol{I}_{c_1-2} \otimes \widetilde{\boldsymbol{\Sigma}}_t,\\
\boldsymbol{F}_{(2,t)} &= \lambda_2 \widetilde{\boldsymbol{\Sigma}}_2 \otimes \boldsymbol{I}_2 \otimes \boldsymbol{I}_{c_t-2} + \lambda_t \boldsymbol{I}_{c_2-2} \otimes \boldsymbol{I}_2 \otimes \widetilde{\boldsymbol{\Sigma}}_t,\\
\boldsymbol{F}_{(1,2,t)} &= \lambda_1 \boldsymbol{I}_{c_2-2} \otimes \widetilde{\boldsymbol{\Sigma}}_1 \otimes \boldsymbol{I}_{c_t-2} + \lambda_2 \widetilde{\boldsymbol{\Sigma}}_2 \otimes \boldsymbol{I}_{c_1-2} \otimes \boldsymbol{I}_{c_t-2} + \lambda_t \boldsymbol{I}_{c_2-2} \otimes \boldsymbol{I}_{c_1-2} \otimes \widetilde{\boldsymbol{\Sigma}}_t.
\end{aligned}$$

A.2 Penalty F in the S-ANOVA case

The spatio-temporal model (4.7) has mixed model bases as defined in Section 4.1

$$\begin{aligned} X &= [X_s \otimes \mathbf{1}_t \,:\, \mathbf{1}_n \otimes x_t \,:\, x_s \otimes x_t\,] \quad \text{and} \\ Z &= [\underbrace{Z_s \otimes \mathbf{1}_t}_{f_s(x_1,x_2)} : \underbrace{\mathbf{1}_n \otimes Z_t}_{f_t(x_t)} : \underbrace{Z_s \otimes x_t : (x_1 : x_2 : x_s) \otimes Z_t : Z_s \otimes Z_t}_{f_{st}(x_1,x_2,x_t)}]. \end{aligned}$$

The penalty matrix F is now of the block-diagonal form:

$$F = \text{blockdiag}(F_{(s)}, F_{(t)}, F_{(st)}),$$

where each block corresponds to the penalty over the smooth terms in the spatio-temporal S-ANOVA model. The first two blocks are exactly the same terms as in the $2d$ and $1d$ cases, with smoothing parameters λ_1, λ_2 and λ_t, i.e.,

$$F_{(s)} = \begin{pmatrix} \lambda_1 I_2 \otimes \widetilde{\Sigma}_1 & & \\ & \lambda_2 \widetilde{\Sigma}_2 \otimes I_2 & \\ & & \lambda_1 I_{c_2-2} \otimes \widetilde{\Sigma}_1 + \lambda_2 \widetilde{\Sigma}_2 \otimes I_{c_1-2} \end{pmatrix} \text{ and } \; F_{(t)} = \lambda_t \widetilde{\Sigma}_t.$$

The last block, $F_{(s,t)}$, is the penalty of the spatio-temporal interaction term, with smoothing parameters τ_1, τ_2 and τ_t. Since to build this block some columns in the bases have to be removed, it requires a more detailed presentation. We split $F_{(s,t)}$ into three sub-blocks: $F_{(s,t)} = \text{blockdiag}(F_{(1)}, F_{(2)}, F_{(3)})$, where

$$F_{(1)} = \begin{pmatrix} \tau_1 I_2 \otimes \widetilde{\Sigma}_1 & & \\ & \tau_2 \widetilde{\Sigma}_2 \otimes I_2 & \\ & & \tau_1 I_{c_2-2} \otimes \widetilde{\Sigma}_1 + \tau_2 \widetilde{\Sigma}_2 \otimes I_{c_1-2} \end{pmatrix}, \; F_{(2)} = \tau_t I_3 \otimes \widetilde{\Sigma}_t \text{ and}$$

$$\begin{aligned} F_{(3)} = {} & \tau_1 I_2 \otimes \widetilde{\Sigma}_1 \otimes I_{c_t-2} + \tau_t I_2 \otimes I_{c_2-2} \otimes \widetilde{\Sigma}_t + \tau_2 \widetilde{\Sigma}_2 \otimes I_2 \otimes I_{c_3-2} + \tau_t I_{c_2-2} \otimes I_2 \otimes \widetilde{\Sigma}_t \\ & + \tau_1 I_{c_2-2} \otimes \widetilde{\Sigma}_1 \otimes I_{c_t-2} + \tau_2 \widetilde{\Sigma}_2 \otimes I_{c_1-2} \otimes I_{c_3-2} + \tau_t I_{c_2-2} \otimes I_{c_1-2} \otimes \widetilde{\Sigma}_t. \end{aligned}$$

References

Akaike H (1973) Maximum likelihood identification of Gaussian autoregressive moving average models. *Biometrika*, **60**, 255–65.

Banerjee S, Carlin BP and Gelfand AE (2004) *Hierarchical modeling and analysis for spatial data*. Monographs on Statistics and Applied Probability 101. Chapman & Hall/CRC.

Bowman AW, Giannitrapani M and Scott EM (2009) Spatiotemporal smoothing and sulphur dioxide trends over Europe. *Applied Statistics*, **58**, 737–52.

Brumback B and Rice J (1998) Smoothing spline models for the analysis of nested and crossed samples of curves. *Journal of American Statistical Association*, **93**, 961–94.

Chen Z (1993) Fitting multivariate regression functions by interaction spline models. *Journal of the Royal Statistical Society, Series B*, **55**, 473–91.

Craven P and Wahba G (1979) Smoothing noisy data with spline functions: estimating the correct degree of smoothing by the method of cross-validation. *Numerical Mathematics*, **31**, 377–403.

Currie ID and Durbán M (2002) Flexible smoothing with *P*-splines: a unified approach. *Statistical Modelling*, **4**, 333–49.

Currie ID, Durbán M and Eilers PHC (2004) Smoothing and forecasting mortality rates. *Statistical Modelling*, **4**, 279–98.

Currie ID, Durbán M and Eilers PHC (2006) Generalized linear array models with applications to multidimensional smoothing. *Journal of the Royal Statistical Society, Series B*, **68**, 1–22.

Durbán M and Currie ID (2003) A note on *P*-spline additive models with correlated errors. *Computational Statistics*, **18**, 251–62.

Eilers PHC, Currie ID and Durbán M (2006) Fast and compact smoothing on large multidimensional grids. *Computational Statistics and Data Analysis*, **50**, 61–76.

Eilers PHC and Marx BD (1996) Flexible smoothing with *B*-splines and penalties. *Statistical Science*, **11**, 89–121.

Eilers PHC, Gampe J Marx BD and Rau R (2008) Modulation models for seasonal time series and incidence tables. *Statistics in Medicine*, **27**, 3430–41.

Fahrmeir L, Kneib T and Lang S (2004) Penalized structured additive regression for space-time data: a Bayesian perspective. *Statistica Sinica*, **14**, 715–45.

Gössl C, Auer DP and Fahrmeir L (2001) Bayesian spatiotemporal inference in functional magnetic resonance imaging. *Biometrics*, **57**, 554–62.

Gu C (2002) *Smoothing spline ANOVA models*. Springer Series in Statistics. Springer.

Hastie T and Tibshirani R (1990) *Generalized additive models*. Monographs on Statistics and Applied Probability. London: Chapman & Hall.

Kammann EE and Wand MP (2003) Geoadditive models. *Applied Statistics*, **52**, 1–18.

Kneib T and Fahrmeir L (2006) Structured additive regression for categorical space-time data: a mixed model approach. *Biometrics*, **62**, 109–18.

Lee D-J and Durbán M (2009) Smooth-CAR mixed models for spatial count data. *Computational Statistics and Data Analysis*, **53**, 2958–79.

MacNab Y and Dean CB (2001) Autoregressive spatial smoothing and temporal spline smoothing for mapping rates. *Biometrics*, **57**, 949–56.

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

Marx BD, Eilers PHC, Gampe J and Rau R (2010) Bilinear modulation models for seasonal tables of counts. *Statistics and Computing*.

Patterson H and Thompson R (1971) Recovery of inter-block information when block sizes are unequal. *Biometrika*, **58**, 545–54.

Schall R (1991) Estimation in generalized linear models with random effects. *Biometrika*, **78**, 719–21.

Schwarz G (1978) Estimating the dimension of a model. *Annals of Statistics*, **6**, 461–64.

Verbyla A, Cullis B, Kenward M and Welham S (1999) The analysis of designed experiments and longitudinal data using smoothing splines. *Journal of the Royal Statistical Society, Series C*, **48**, 269–312.

Wand MP (2003) Smoothing and mixed models. *Computational Statistics*, **18**, 223–49.

Welham SJ, Cullis BR, Kenward MG and Thompson R (2007) A comparison of mixed model splines for curve fitting. *Australian New Zealand Journal of Statistics*, **49**, 1–23.

Wood SN (2006a) *Generalized additive models—an introduction with R*. Texts in Statistical Science. Chapman & Hall.

Wood SN (2006b) Low-rank scale-invariant tensor product smooths for generalized additive mixed models. *Biometrics*, **62**, 1025–36.