UNIVERSIDAD CARLOS III DE MADRID working papers

Working Paper 09-33
Statistics and Econometrics Series 12
June 2009

Departamento de Estadística
Universidad Carlos III de Madrid
Calle Madrid, 126
28903 Getafe (Spain)
Fax (34) 91 624-98-49

P-SPLINE ANOVA-TYPE INTERACTION MODELS FOR SPATIO-TEMPORAL SMOOTHING

Dae-Jin Lee and María Durbán

Abstract

In recent years, spatial and spatio-temporal modelling have become an important area of research in many fields (epidemiology, environmental studies, disease mapping, ...). 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 ANOVA decomposition. The properties of the bases used for regression allow the use of algorithms that can handle large amount of data. We show that 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.

Keywords: ANOVA decomposition, Mixed models, Penalized Splines, Spatiotemporal data.

Universidad Carlos III de Madrid, Department of Statistics, Escuela Politécnica Superior, Campus de Leganés, Madrid, Spain. e-mail addresses: dae-jin.lee@uc3m.es (Dae-Jin Lee) and mdurban@est-econ.uc3m.es (María Durbán). **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).

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

May 28, 2009

Abstract

In recent years, spatial and spatio-temporal modelling have become an important area of research in many fields (epidemiology, environmental studies, disease mapping, ...). 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 ANOVA decomposition. The properties of the bases used for regression allow the use of algorithms that can handle large amount of data. We show that 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.

Keywords: ANOVA decomposition, Mixed models, Penalized Splines, Spatio-temporal data.

*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. Tel: +34 91 624 9175. Fax: +34 91 624 9177

†e-mail: mdurban@est-econ.uc3m.es

1 Introduction

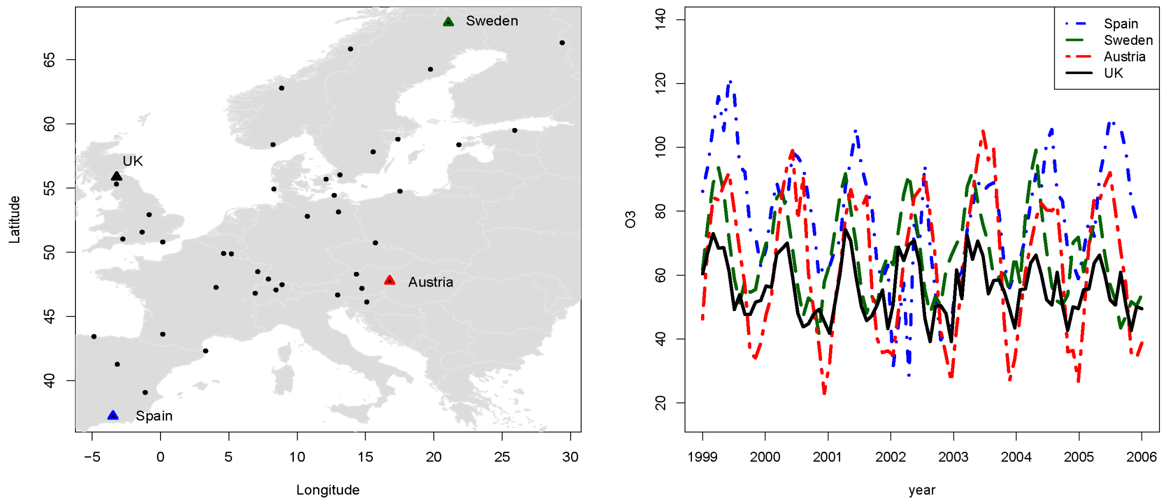
In recent years, there has been an enormous growth of data with spatio-temporal structure. This type of data arise in many contexts such as, meteorology, environmental sciences, epidemiology or demography, among others. This wide variety of settings has generated a considerable interest in the development of spatio-temporal models. However, the complexity of the models needed and the size of the data sets has 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 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 models seems suitable to estimate simultaneously the spatial and temporal trends.

Here, we consider the use of penalized splines (Eilers and Marx, 1996) for smoothing spatio-temporal data (Lee and Durbán, 2008). Recent papers outline the use of these methods in several applications for Gaussian and non-Gaussian responses in one or more dimensions (see for example Currie et al., 2004). In the multidimensional case, the common extension is the use of tensor product of B-splines 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 $\boldsymbol{y}$ is modelled as the sum of spatial and temporal effects of the form:

$$\mathbb{E}[\boldsymbol{y}] = f(\boldsymbol{x}_1, \boldsymbol{x}_2) + f(\boldsymbol{x}_t) \, .$$

This additive model, does not account for the space-time interaction effect, and therefore, can not 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. 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.



(a) Monitoring stations.

(b) Time series plot in four countries.

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

the same correlation in any spatial direction (which is unrealistic in many cases), and can be computationally intensive for large data sets.

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 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 the mixed model representation of multidimensional P-splines is described. Section 3 develops a general methodology to represent interaction models using ANOVA decompositions. In section 4, the methods of previous sections are applied to the case of spatio-temporal data, and section 5 illustrates the techniques with the analysis of ozone levels in Europe from 1990 to 2005. Concluding remarks are given in section 6. We defer to the Appendix some technical details of ANOVA decompositions.

2 P-splines for spatial data

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 the smooth function and $\boldsymbol{\epsilon}$ is a Gaussian error term with variance $\sigma^2\boldsymbol{I}$. The method proposed by Eilers and Marx (1996) consider 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

$$\boldsymbol{S}(\boldsymbol{\theta}; \boldsymbol{y}, \lambda) = \|\boldsymbol{y} - \boldsymbol{B\theta}\|^2 + \boldsymbol{\theta}'\boldsymbol{P\theta}, \tag{2.1}$$

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

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

where $\boldsymbol{D}$ is a difference matrix applied directly to the regression coefficients (a common choice is to consider a second order difference, which defines a quadratic penalty). Then, for a given value of λ, the minimization of (2.1), yields:

$$\widehat{\boldsymbol{\theta}} = (\boldsymbol{B}'\boldsymbol{B} + \boldsymbol{P})^{-1} \boldsymbol{B}'\boldsymbol{y} \ , \tag{2.2}$$

The choice of λ is, in general, subject to a certain criteria (AIC, BIC, CV or GCV).

In the multidimensional case, let's suppose that we have spatial data $(\boldsymbol{x}_{1i}, \boldsymbol{x}_{2j}, \boldsymbol{y}_{ij})$, where $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ are respectively geographic longitude and latitude and $\boldsymbol{y}$ is the response variable. A smooth model for the data would be given by:

$$\boldsymbol{y} = f(\boldsymbol{x}_1, \boldsymbol{x}_2) + \boldsymbol{\epsilon} = \boldsymbol{B\theta} + \boldsymbol{\epsilon} \ , \quad \boldsymbol{\epsilon} \sim \mathcal{N}(0, \sigma^2 \boldsymbol{I}) \ , \tag{2.3}$$

where $\boldsymbol{B}$ is the regression basis constructed from the spatial covariates $(\boldsymbol{x}_1, \boldsymbol{x}_2)$. As we pointed out before, we use B-splines 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:

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

(see Currie et al. (2004, 2006)). In the spatial smoothing context, 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):

$$\boldsymbol{B} = \boldsymbol{B}_2 \Box \boldsymbol{B}_1 = (\boldsymbol{B}_2 \otimes \boldsymbol{1}'_{c_1}) \odot (\boldsymbol{1}'_{c_2} \otimes \boldsymbol{B}_1) \ , n \times c_1 c_2 \ , \tag{2.4}$$

where $\odot$ is the "element-wise" matrix product and $\mathbf{1}_{c_1}$ and $\mathbf{1}_{c_2}$ are vectors of ones of length c_1 and c_2 respectively.

It is worth mentioning that the use of the products depends on the data structure (grid or scattered), and it affects the model basis and its product (Kronecker or Box), 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 $\mathbf{\Theta}$, $c_1 \times c_2$, be the matrix of regression coefficients, where $\boldsymbol{\theta} = \text{vec}(\mathbf{\Theta})$, $c_1 c_2 \times 1$. The penalty matrix for a $2d$ P-spline model is:

$$\boldsymbol{P} = \lambda_1 \boldsymbol{I}_{c_2} \otimes \boldsymbol{D}_1' \boldsymbol{D}_1 + \lambda_2 \boldsymbol{D}_2' \boldsymbol{D}_2 \otimes \boldsymbol{I}_{c_1} \ , \tag{2.5}$$

where $\boldsymbol{D}_q$, $q = 1, 2$, is again a difference matrix. Then, (2.5) applies the discrete penalties over the rows and columns of $\mathbf{\Theta}$, and allows for anisotropic smoothing ($\lambda_1 \neq \lambda_2$), since the amount of smoothing can be different in each dimension (longitude and latitude).

2.1 Mixed models representation of multidimensional 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 goal here is to set a new basis which allow the representation of a P-spline model and its associated penalty as a mixed model:

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

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 ($\boldsymbol{X}\boldsymbol{\beta}$) and a non-linear/penalized ($\boldsymbol{Z}\boldsymbol{\alpha}$) 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, for example, Wand (2003) described the representation with truncated power functions as bases, and Currie and Durbán (2002) described the representation using B-splines bases.

Since this transformation is not unique, in this paper we choose a similar approach to Currie et al. (2006) and Durbán et al. (2006). The smoothing parameter becomes the ratio $\lambda = \sigma^2 / \sigma^2_{\boldsymbol{\alpha}}$, and variance components can be estimated by residual or restricted maximum likelihood (REML) of Patterson and Thompson (1971).

We use here the spatial model (2.3) to illustrate the construction of the transformation, but the methodology will extend to any number of covariates. We will find a *one-to-one* (orthogonal) transformation, $\boldsymbol{T}$, such that, $\boldsymbol{B}\boldsymbol{T} = [\boldsymbol{X} : \boldsymbol{Z}]$, $\boldsymbol{B}\boldsymbol{\theta} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{Z}\boldsymbol{\alpha}$, and $[\boldsymbol{X} : \boldsymbol{Z}]$ has full rank, i.e.:

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

Under this 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 (2.5), and the simultaneous diagonalization of $\boldsymbol{D}_1'\boldsymbol{D}_1$ and $\boldsymbol{D}_2'\boldsymbol{D}_2$. Let $\boldsymbol{U}_1\boldsymbol{\Sigma}_1\boldsymbol{U}_1'$ be the SVD of $\boldsymbol{D}_1'\boldsymbol{D}_1$, assuming a second order penalty, the diagonal matrix of eigenvalues has two zeroes and $c_1 - 2$ positive eigenvalues, then $\boldsymbol{\Sigma}_1 = \text{diag}(0, 0, \widetilde{\boldsymbol{\Sigma}}_1)$.

We take $\boldsymbol{U}_{1n} = [\boldsymbol{1}_1^* : \boldsymbol{u}_1^*]$ as the eigenvectors corresponding to the null space of the SVD, where $\boldsymbol{1}_1^*$ is $(1, ..., 1)'/\sqrt{c_1}$ and $\boldsymbol{u}_1^*$ is the vector $(1, ..., c_1)$ centered and scaled to have unit length. The matrix, $\boldsymbol{U}_{1s}$ is the sub-matrix which corresponds to the positive eigenvalues $\widetilde{\boldsymbol{\Sigma}}_1$ (the definitions of $\boldsymbol{U}_{2n}$, $\boldsymbol{U}_{2s}$ and $\widetilde{\boldsymbol{\Sigma}}_2$ are similar).

Then, the transformation is defined as $\boldsymbol{T} = [\boldsymbol{T}_n : \boldsymbol{T}_s]$ as :

$$\begin{aligned} \boldsymbol{T}_n &= [\boldsymbol{U}_{2n} \otimes \boldsymbol{U}_{1n}] \text{ and} \\ \boldsymbol{T}_s &= [\boldsymbol{U}_{2n} \otimes \boldsymbol{U}_{1s} : \boldsymbol{U}_{2s} \otimes \boldsymbol{U}_{1n} : \boldsymbol{U}_{2s} \otimes \boldsymbol{U}_{1s}] \,, \end{aligned}$$

$\boldsymbol{T}$ is orthogonal, and the new coefficients are given by:

$$\boldsymbol{\beta} = \boldsymbol{T}_n'\boldsymbol{\theta} \quad \text{and} \quad \boldsymbol{\alpha} = \boldsymbol{T}_s'\boldsymbol{\theta}.$$

Using some matrix algebra results (Liu, 1999) it can be shown that the mixed model matrices $\boldsymbol{X}$ and $\boldsymbol{Z}$ calculated as $\boldsymbol{BT}$, are in fact, the tensor product of the marginal basis:

$$\boldsymbol{X} = \boldsymbol{X}_2 \square \boldsymbol{X}_1 \quad \text{and} \tag{2.7}$$

$$\boldsymbol{Z} = [\boldsymbol{X}_2 \square \boldsymbol{Z}_1 : \boldsymbol{Z}_2 \square \boldsymbol{X}_1 : \boldsymbol{Z}_2 \square \boldsymbol{Z}_1] \,, \tag{2.8}$$

where $\boldsymbol{X}_q = \boldsymbol{B}_q\boldsymbol{U}_{qn}$ and $\boldsymbol{Z}_q = \boldsymbol{B}_q\boldsymbol{U}_{qs}$, $q = 1, 2$ (this is an important result because we can avoid, in practice, the calculation of matrix $\boldsymbol{T}$). Furthermore, since $\boldsymbol{\beta}$ is unpenalized, we may replace $\boldsymbol{X}_q = \boldsymbol{B}_q[\boldsymbol{1}_q^* : \boldsymbol{u}_q^*]$ by $[\boldsymbol{1} : \boldsymbol{x}_q]$, where $\boldsymbol{1}$ is a vector of ones and $\boldsymbol{x}_q$, is the q^{th} covariate.

The expressions for $\boldsymbol{X}$ and $\boldsymbol{Z}$ given above can be expanded as:

$$\begin{aligned} \boldsymbol{X} &\equiv [\boldsymbol{1}_n : \boldsymbol{x}_1 : \boldsymbol{x}_2 : \boldsymbol{x}_2 \square \boldsymbol{x}_1] \quad \text{and} \\ \boldsymbol{Z} &\equiv [\boldsymbol{Z}_1 : \boldsymbol{Z}_2 : \boldsymbol{Z}_2 \square \boldsymbol{x}_1 : \boldsymbol{x}_2 \square \boldsymbol{Z}_1 : \boldsymbol{Z}_2 \square \boldsymbol{Z}_1] \,, \end{aligned}$$

This is a key result for the rest of the paper, since it allows the representation of the fitted surface in terms of three components: a term for $\boldsymbol{x}_1$, a term for $\boldsymbol{x}_2$, and an interaction term. It can be shown that the penalty is a block-diagonal matrix, $\boldsymbol{F} = \boldsymbol{T}'\boldsymbol{PT}$, defined by:

$$\boldsymbol{F} = \begin{pmatrix} \boldsymbol{0}_4 & & & \\ & \lambda_1 \boldsymbol{I}_{c_2} \otimes \widetilde{\boldsymbol{\Sigma}}_1 & & \\ & & \lambda_2 \widetilde{\boldsymbol{\Sigma}}_2 \otimes \boldsymbol{I}_{c_1} & \\ & & & \lambda_2 \widetilde{\boldsymbol{\Sigma}}_2 \otimes \boldsymbol{I}_{c_1 - 2} + \lambda_1 \boldsymbol{I}_{c_2 - 2} \otimes \widetilde{\boldsymbol{\Sigma}}_1 \end{pmatrix},$$

where $\mathbf{0}_4$ is a 4×4 matrix of zeros corresponding to the unpenalized fixed part, and with blocks corresponding to each smooth term, and the variance components matrix in mixed model (2.6) is given by $\boldsymbol{G} = \sigma^2 \boldsymbol{F}^{-1}$.

3 Smooth-ANOVA decomposition models

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

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

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 analysis-of-variance (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; Wahba and Luo, 1995; 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 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 introduced in the previous Section. There are other alternatives to avoid the identifiability problems: (i) add a ridge penalty on the system of equations in (2.2) as in Marx and Eilers (1998), or (ii) identify and impose the constraints numerically (see Wood, 2006a). However, the first alternative implies a correct definition of the penalty matrix, and the second method is difficult to extend in the case of more than 2-way interactions.

For simplicity, we will illustrate the procedure in the $2d$ spatial case:

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

the model is defined in terms of two main effects for geographic coordinates (i.e f_1 and f_2), and a spatial (2-way) interaction, f_s. The B-spline basis for this model is:

$$\boldsymbol{B} = [\mathbf{1}_n : \boldsymbol{B}_1 : \boldsymbol{B}_2 : \boldsymbol{B}_s]\ , \tag{3.3}$$

where $\mathbf{1}_n$ 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 (2.4). Then, model (3.2) can be written as:

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

where $\boldsymbol{\theta} = (\gamma, \boldsymbol{\theta}_1', \boldsymbol{\theta}_2', \boldsymbol{\theta}_s')'$ is the vector of regression coefficients. We impose smoothness on the coefficients by a block-diagonal penalty of the form:

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

However, the regression matrix (3.3) (of dimension $n \times (c_1 + c_2 + c_1c_2)$) 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 considered in order to preserve the identifiability. Wood (2006b) pointed the need to construct appropiate model bases and penalties, and impose constraints to maintain the model identifiability, and Wood (2006a, chap. 4), suggested the use of the QR decomposition in order to numerically identify 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 formulation as a mixed model.

Following the results in Section 2.1, we apply the SVD over the penalty matrix $\boldsymbol{P}$, and obtain the mixed model reparameterization. For the additive terms corresponding to covariates $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$, we have the mixed model matrices:

$$f_1(\boldsymbol{x}_1) \equiv [\boldsymbol{X}_1 : \boldsymbol{Z}_1] = [\mathbf{1}_n : \boldsymbol{x}_1 : \boldsymbol{Z}_1] \ , \text{ and} \tag{3.5}$$
$$f_2(\boldsymbol{x}_2) \equiv [\boldsymbol{X}_2 : \boldsymbol{Z}_2] = [\mathbf{1}_n : \boldsymbol{x}_2 : \boldsymbol{Z}_2] \ , \tag{3.6}$$

and for the interaction term:

$$\begin{aligned} f_s(\boldsymbol{x}_1, \boldsymbol{x}_2) &\equiv \boldsymbol{X}_2 \square \boldsymbol{X}_1 : \boldsymbol{X}_2 \square \boldsymbol{Z}_1 : \boldsymbol{Z}_2 \square \boldsymbol{X}_1 : \boldsymbol{Z}_2 \square \boldsymbol{Z}_1 \\ &\equiv [\mathbf{1}_n : \boldsymbol{x}_2] \square [\mathbf{1}_n : \boldsymbol{x}_1] : [\mathbf{1}_n : \boldsymbol{x}_2] \square \boldsymbol{Z}_1 : \boldsymbol{Z}_2 \square [\mathbf{1}_n : \boldsymbol{x}_1] : \boldsymbol{Z}_2 \square \boldsymbol{Z}_1 \ . \end{aligned} \tag{3.7}$$

The reparameterization used allow us to identify the linearly dependent columns in the bases (the columns of (3.5) and (3.6) are already contained in (3.7)). Therefore, we can solve the identifiability problems by simply removing the vector $\mathbf{1}_n$ in (3.7). In the main effects, we also remove one $\mathbf{1}_n$ vector. Then, the fixed and random effects matrices are:

$$\boldsymbol{X} = [\mathbf{1}_n : \boldsymbol{x}_1 : \boldsymbol{x}_2 : \boldsymbol{x}_2 \square \boldsymbol{x}_1], \ \text{ and} \tag{3.8}$$
$$\boldsymbol{Z} = [\boldsymbol{Z}_1 : \boldsymbol{Z}_2 : \boldsymbol{x}_2 \square \boldsymbol{Z}_1 : \boldsymbol{Z}_2 \square \boldsymbol{x}_1 : \boldsymbol{Z}_2 \square \boldsymbol{Z}_1]. \tag{3.9}$$

The mixed model representation will not be complete unless we give an expression for the variance-covariance matrix $\boldsymbol{G} = \sigma^2 \boldsymbol{F}^{-1}$. In order to obtain $\boldsymbol{F}$ we use the fact that $\boldsymbol{F} = \boldsymbol{T}'\boldsymbol{P}\boldsymbol{T}$, and so, we need a transformation $\boldsymbol{T}$ that takes into account the reduction in the dimension of matrices $\boldsymbol{X}$ and $\boldsymbol{Z}$ (details on the construction of $\boldsymbol{T}$ are given in Appendix A).

The new transformation matrix $\boldsymbol{T}$ has dimension $(1 + c_1 + c_2 + c_1c_2) \times c_1c_2$, and leads to the mixed model penalty:

$$\boldsymbol{F} = \text{blockdiag}(\mathbf{0}_4, \boldsymbol{F}_1, \boldsymbol{F}_2, \boldsymbol{F}_s) \ , \tag{3.10}$$

where

$$\begin{aligned} \boldsymbol{F}_1 &= \lambda_1 \widetilde{\boldsymbol{\Sigma}}_1 \ , \\ \boldsymbol{F}_2 &= \lambda_2 \widetilde{\boldsymbol{\Sigma}}_2 \ , \ \text{ and} \\ \boldsymbol{F}_s &= \text{blockdiag}(\tau_1 \widetilde{\boldsymbol{\Sigma}}_1, \tau_2 \widetilde{\boldsymbol{\Sigma}}_2, \tau_1 \boldsymbol{I}_{c_2-2} \otimes \widetilde{\boldsymbol{\Sigma}}_1 + \tau_2 \widetilde{\boldsymbol{\Sigma}}_2 \otimes \boldsymbol{I}_{c_1-2}) \ . \end{aligned}$$

Our aim is to show that the effect of removing the column of $\mathbf{1}_n$'s in the fixed effects matrices is equivalent to impose the usual constraints on the model coefficients, i.e., solving the identifiability problems in the mixed model results in transforming the original penalty (or the original coefficients). This can be easily proved by recovering the penalty of the original parametrization using the fact that

$$\breve{\boldsymbol{P}} = \boldsymbol{T}\boldsymbol{F}\boldsymbol{T}' = \boldsymbol{T}\boldsymbol{T}'\boldsymbol{P}\boldsymbol{T}\boldsymbol{T}',$$

for $\boldsymbol{P}$ and $\boldsymbol{F}$ given in (3.4) and (3.10). However, the matrix $\boldsymbol{T}$ in the ANOVA case is no longer orthogonal. Then, if we define $\boldsymbol{K} = \boldsymbol{T}\boldsymbol{T}'$, it is easy to show that $\boldsymbol{K}$ is a centering matrix:

$$\boldsymbol{K} = \begin{pmatrix} 1 & \cdots & & 0 \\ \vdots & \boldsymbol{K}_1 & & \\ & & \boldsymbol{K}_2 & \\ 0 & & & \boldsymbol{K}_2 \otimes \boldsymbol{K}_1 \end{pmatrix}, \tag{3.11}$$

where $\boldsymbol{K}_q = (\boldsymbol{I}_{c_q} - \mathbf{1}\mathbf{1}'/c_q)$, for $q = 1, 2$ (see Appendix A for details).

Then, the penalty that takes into acount the identifiability constraints is:

$$\breve{\boldsymbol{P}} = \boldsymbol{K}\boldsymbol{P}\boldsymbol{K}\ , \tag{3.12}$$

with rank$(\breve{\boldsymbol{P}}) = c_1c_2 - 4$.

In terms of the regression parameters, the matrix $\boldsymbol{K}$ apply constraints over the $\boldsymbol{\theta}$ coefficients, from (3.12), we have

$$\boldsymbol{\theta}'\breve{\boldsymbol{P}}\boldsymbol{\theta} = \widetilde{\boldsymbol{\theta}}'\boldsymbol{P}\widetilde{\boldsymbol{\theta}}\ ,$$

where

$$\widetilde{\boldsymbol{\theta}} = \boldsymbol{K}\boldsymbol{\theta} = \begin{pmatrix} \gamma & \cdots & & \\ \vdots & \boldsymbol{K}_1\boldsymbol{\theta}_1 & & \\ & & \boldsymbol{K}_2\boldsymbol{\theta}_2 & \\ & & & (\boldsymbol{K}_2 \otimes \boldsymbol{K}_1)\boldsymbol{\theta}_s \end{pmatrix}. \tag{3.13}$$

The coefficients of the 2-way interaction $(\boldsymbol{K}_2 \otimes \boldsymbol{K}_1)\boldsymbol{\theta}_s$ can be written in array form, $\boldsymbol{K}_1\boldsymbol{\Theta}\boldsymbol{K}_2$, and so, we are centering the coefficients matrix $\boldsymbol{\Theta}$ by rows and columns. These constraints are exactly equivalent to those applied in a factorial design with two main effects and a 2-way interaction, i.e.:

$$\sum_i^{c_1} \boldsymbol{\theta}_{1i} = \sum_j^{c_2} \boldsymbol{\theta}_{2j} = 0\ , \qquad \text{for main effects and} \tag{3.14}$$

$$\sum_i^{c_1} \boldsymbol{\Theta}_{ij} = \sum_j^{c_2} \boldsymbol{\Theta}_{ij} = 0\ , \quad \text{for 2-way interactions}\ . \tag{3.15}$$

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 a $n \times n$ centering matrix, which centers the B-spline basis by rows. The approach presented in this Section, improves their results, we have that $\boldsymbol{B}\widetilde{\boldsymbol{\theta}}$ is equal to $\widetilde{\boldsymbol{B}}\boldsymbol{\theta}$, where $\widetilde{\boldsymbol{B}}_q = \boldsymbol{B}_q\boldsymbol{K}_q$, is centered by columns, which is computationally more efficient.

Table 1: Set of regression coefficient constraints in a full $3d$ P-spline ANOVA-type model in (3.16).

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

Several nested models can also be considered from this approach, for instance if we consider only one additive term and the interaction, e.g. $f_1(\boldsymbol{x}_1) + f_s(\boldsymbol{x}_1, \boldsymbol{x}_2)$, following the methodology proposed in this Section, it is easy to show that we have to remove the first column of $\boldsymbol{X}_2$ in order to avoid identifiability problems. Then the mixed models bases are:

$$[\boldsymbol{X} : \boldsymbol{Z}] = [\mathbf{1}_n : \boldsymbol{x}_1 : \boldsymbol{x}_2 \Box \boldsymbol{X}_1 : \boldsymbol{Z}_1 : \boldsymbol{x}_2 \Box \boldsymbol{Z}_1 : \boldsymbol{Z}_2 \Box \boldsymbol{X}_1 : \boldsymbol{Z}_2 \Box \boldsymbol{Z}_1] \ ,$$

and penalty:

$$\boldsymbol{F} = \text{blockdiag}(\mathbf{0}_4, \lambda_1 \widetilde{\boldsymbol{\Sigma}}_1, \tau_1 \widetilde{\boldsymbol{\Sigma}}_1, \tau_2 \widetilde{\boldsymbol{\Sigma}}_2 \otimes \boldsymbol{I}_2, \tau_1 \boldsymbol{I}_{c_2-2} \otimes \widetilde{\boldsymbol{\Sigma}}_1 + \tau_2 \widetilde{\boldsymbol{\Sigma}}_2 \otimes \boldsymbol{I}_{c_1-2}) \ .$$

This model is equivalent to apply constraints over the main effect, $\sum_i^{c_1} \boldsymbol{\theta}_i = 0$, and over the columns of matrix $\boldsymbol{\Theta}$ for the 2-way interaction, i.e. $\sum_j^{c_2} \boldsymbol{\Theta}_{ij} = 0$.

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

$$\begin{aligned} \boldsymbol{y} =& \gamma + f_1(\boldsymbol{x}_1) + f_2(\boldsymbol{x}_2) + f_3(\boldsymbol{x}_3) + \\ &+ f_{1,2}(\boldsymbol{x}_1, \boldsymbol{x}_2) + f_{1,3}(\boldsymbol{x}_1, \boldsymbol{x}_3) + f_{2,3}(\boldsymbol{x}_2, \boldsymbol{x}_3) + \\ &+ f_{1,2,3}(\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{x}_3) + \boldsymbol{\epsilon} \ , \end{aligned} \tag{3.16}$$

where all main effects, 2-way and 3-way interaction are included in the model. For model (3.16) these constraints are shown in Table 1.

4 Spatio-temporal P-splines models

We start by proposing non-separable models of the form:

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

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 (2.4), of dimension $n \times c_s$, where $c_s = c_1 c_2$, 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, since we can express the data in a compact notation. 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 3-dimensions is:

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

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.

For the mixed model representation, we proceed as in Section 2.1 and use the SVD over the penalty (4.4). Then, we obtain the mixed model model matrices, which in compact notation can be written as:

$$\boldsymbol{X} = \boldsymbol{X}_s \otimes \boldsymbol{X}_t \, , \text{ and} \tag{4.5}$$

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

where $\boldsymbol{X}_s$ and $\boldsymbol{Z}_s$ are the fixed and random effects matrices for the spatial part defined in (2.7) and (2.8). For time dimension, we have the fixed and random matrices defined as $\boldsymbol{X}_t = [\boldsymbol{1}_t : \boldsymbol{x}_t]$ and $\boldsymbol{Z}_t = \boldsymbol{B}_t \boldsymbol{U}_{ts}$. The block-diagonal penalty in the mixed model is then:

$$\boldsymbol{F} = \text{blockdiag}(\boldsymbol{0}_8, \boldsymbol{F}_{(1)}, \boldsymbol{F}_{(2)}, \boldsymbol{F}_{(1,2)}, \boldsymbol{F}_{(t)}, \boldsymbol{F}_{(1,t)}, \boldsymbol{F}_{(2,t)}, \boldsymbol{F}_{(1,2,t)}) \, , \tag{4.7}$$

where $\boldsymbol{0}_8$ is a diagonal matrix of zeroes corresponding to the unpenalized fixed part and the remaining seven blocks are the penalty terms for the random part:

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

The construction of the mixed model bases (4.5) and (4.6) allows us to represent the fitted values in terms of the sum of additive components plus interactions (2-way and 3-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 effects) but also the spatial effects (2-way interaction) and specially the interaction between space and time (3-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 P-spline ANOVA model for spatio-temporal smoothing

The Smooth-ANOVA model formulation presented in previous Section allows us to consider more realistic models for spatio-temporal smoothing:

$$\boldsymbol{y} = \gamma + f_s(\boldsymbol{x}_1, \boldsymbol{x}_2) + f_t(\boldsymbol{x}_t) + f_{st}(\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{x}_t) + \boldsymbol{\epsilon}\ ,\ \ \boldsymbol{\epsilon} \sim \mathcal{N}(0, \sigma^2)\ , \tag{4.8}$$

where we explicitly consider a smooth term for the spatial surface, for temporal smooth trend, and a smooth term for space-time interaction.

The B-spline basis for model (4.8) is:

$$\boldsymbol{B} = [\mathbf{1}_{nt} : \boldsymbol{B}_s \otimes \mathbf{1}_t : \mathbf{1}_n \otimes \boldsymbol{B}_t : \boldsymbol{B}_s \otimes \boldsymbol{B}_t]\ , \tag{4.9}$$

and vector of regression coefficients $\boldsymbol{\theta} = (\gamma, \boldsymbol{\theta}^{(s)\prime}, \boldsymbol{\theta}^{(t)\prime}, \boldsymbol{\theta}^{(st)\prime})'$.

For each smooth term in model (4.8), the mixed model bases are:

$$f_s(\boldsymbol{x}_1, \boldsymbol{x}_2) \equiv\ [\mathbf{1}_n : \boldsymbol{x}_1 : \boldsymbol{x}_2 : \boldsymbol{x}_s : \boldsymbol{Z}_s] \otimes \mathbf{1}_t \tag{4.10}$$

$$f_t(\boldsymbol{x}_t) \equiv\ \mathbf{1}_n \otimes [\mathbf{1}_t : \boldsymbol{x}_t : \boldsymbol{Z}_t]\ \ \text{and} \tag{4.11}$$

$$f_{st}(\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{x}_t) \equiv\ [\mathbf{1}_n : \boldsymbol{x}_1 : \boldsymbol{x}_2 : \boldsymbol{x}_s] \otimes [\mathbf{1}_t : \boldsymbol{x}_t] : \boldsymbol{Z}_s \otimes [\mathbf{1}_t : \boldsymbol{x}_t] : [\mathbf{1}_n : \boldsymbol{x}_1 : \boldsymbol{x}_2 : \boldsymbol{x}_s] \otimes \boldsymbol{Z}_t : \boldsymbol{Z}_s \otimes \boldsymbol{Z}_t\ , \tag{4.12}$$

where $\boldsymbol{x}_s = \boldsymbol{x}_2 \Box \boldsymbol{x}_1$. It is easy to see that terms (4.10) and (4.11) also appear in (4.12). To avoid this linear dependency, we remove the columns of (4.12) that are already included in the previous smooth terms, as we did in the spatial case. Again, this can be easily done by removing the column vectors $\mathbf{1}_n$ and $\mathbf{1}_t$ from (4.12), thus we define for the spatio-temporal interaction:

$$f_{st}(\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{x}_t) \equiv\ [\boldsymbol{x}_1 : \boldsymbol{x}_2 : \boldsymbol{x}_s : \boldsymbol{Z}_s] \otimes [\boldsymbol{x}_t : \boldsymbol{Z}_t] \tag{4.13}$$

We also need to remove the vector $\mathbf{1}_t$ in the temporal smooth term (4.11), which lead us to the time main effect constraint. We rewrite the full mixed model bases as:

$$\boldsymbol{X} = [\mathbf{1}_n : \boldsymbol{x}_1 : \boldsymbol{x}_2 : \boldsymbol{x}_s] \otimes \mathbf{1}_t : \mathbf{1}_n \otimes \boldsymbol{x}_t : [\boldsymbol{x}_1 : \boldsymbol{x}_2 : \boldsymbol{x}_s] \otimes \boldsymbol{x}_t\ , \tag{4.14}$$

$$\boldsymbol{Z} = \boldsymbol{Z}_s \otimes \mathbf{1}_t : \mathbf{1}_n \otimes \boldsymbol{Z}_t : \boldsymbol{Z}_s \otimes \boldsymbol{x}_t : [\boldsymbol{x}_1 : \boldsymbol{x}_2 : \boldsymbol{x}_s : \boldsymbol{Z}_s] \otimes \boldsymbol{Z}_t\ . \tag{4.15}$$

Finally, the mixed model penalty for the ANOVA model is of the form:

$$\boldsymbol{F} = \text{blockdiag}(\mathbf{0}_8, \boldsymbol{F}_{(s)}, \boldsymbol{F}_{(t)}, \boldsymbol{F}_{(st)})\ . \tag{4.16}$$

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, which is constructed once we remove the linear dependency (see the details of the construction of (4.16) in Appendix (B.1)).

For the ANOVA spatio-temporal model (4.8), the resultant mixed model reparameterization is equivalent to apply constraints 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 $c_t \times c_1 \times c_2$.

$$\sum_i^{c_1} \boldsymbol{\theta}_{t,ij}^{(\mathrm{st})} = \sum_j^{c_2} \boldsymbol{\theta}_{t,ij}^{(\mathrm{st})} = \sum_i^{c_1} \sum_j^{c_2} \boldsymbol{\theta}_{t,ij}^{(\mathrm{st})} = 0. \tag{4.17}$$

See the details in Appendix (B.2).

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 the development of new policies. The European Environment Agency (EEA) has established a program 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-September, based on information submitted to the European Commission on ozone in ambient air. According to this annual reports, although emissions of ozone precursors have been reduced over the last decade, ozone pollution levels has not changed significantly in the period 1999-2005. The analysis of the data will confirm this statement, but it will show that different countries reach the largest values of ozone at different time points.

We analyzed monthly averages of air pollution by ground-level ozone (in $\mu g/m^3$ units) over Europe from January 1990 to December 2005. The data were collected in 43 monitoring stations in 15 EU countries. Following the methodology described in previous sections, we fitted 3 models to the data: (i) *spatio-temporal ANOVA model*; (ii) $3d$ *interaction model* and (iii) *space-time additive model*. The three models formulation are then:

i.	**ANOVA:**	$f_s(\boldsymbol{x}_1, \boldsymbol{x}_2) + f_t(\boldsymbol{x}_t) + f_{s,t}(\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{x}_t)$
ii.	**Interaction:**	$f_{s,t}(\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{x}_t)$, and
iii.	**Additive:**	$f_s(\boldsymbol{x}_1, \boldsymbol{x}_2) + f_t(\boldsymbol{x}_t)$

In order to fit the models, we set up the B-splines bases, using the following parameters: (1) the number of (equidistant) internal knots, ndx; (2) the degree of the P-spline,

Table 2: AIC and estimated degrees of freedom of fitted models.

Model	AIC	d.f.
ANOVA	14280.73	366.03
Interaction	14537.22	765.05
Additive	16506.28	65.98

$bdeg$; and 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.

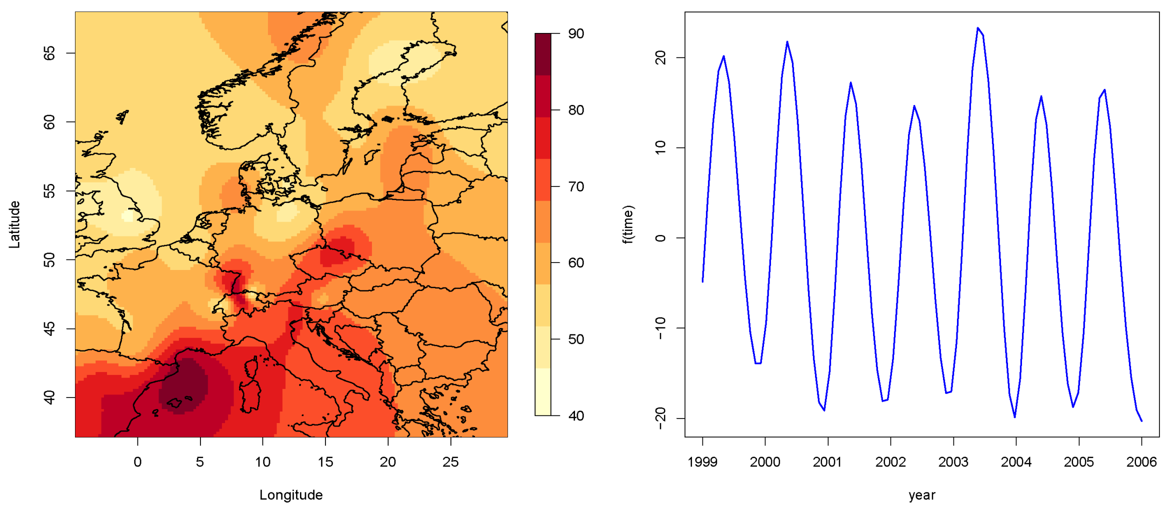
The mixed model formulation is straightforward following the methodology proposed in the paper: we construct matrices $\boldsymbol{X}$ and $\boldsymbol{Z}$, and the block-diagonal penalty $\boldsymbol{F}$ for each model. We compared the performance of the models in terms of the Akaike Information Criteria (**AIC**), and the effective degrees of freedom (**d.f.**) of the model, measured as $\mathrm{d.f.} = \text{trace}(\boldsymbol{H})$ (i.e. the trace of the hat-matrix). The results are summarized in Table 2. There is a superior performance of 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 ANOVA model performs better. It uses less d.f. because the model allows a different the 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.

Figure 2a shows the smoothed spatial surface for the ozone levels of the 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 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 2b, where the highest levels are recorded during spring and summer months (April-September). The highest peak corresponds to the heat-wave occurred in Europe during summer 2003. The spatio-temporal 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 April to September 2001. 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 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 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 4a). In contrast, the spatio-temporal ANOVA model fit, is able to capture the individual characteristics of the stations throughout time. Figure 4b 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 represent a d-dimensional $c_1 \times c_2 \times \cdots \times c_d$ array of coefficients by $\boldsymbol{\Theta}$, an 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).



(a) Spatial surface: $f_s(\boldsymbol{x}_1, \boldsymbol{x}_2)$ (b) Time trend: $f_t(\boldsymbol{x}_t)$

Figure 2: Spatial and temporal smooth terms for ANOVA model.

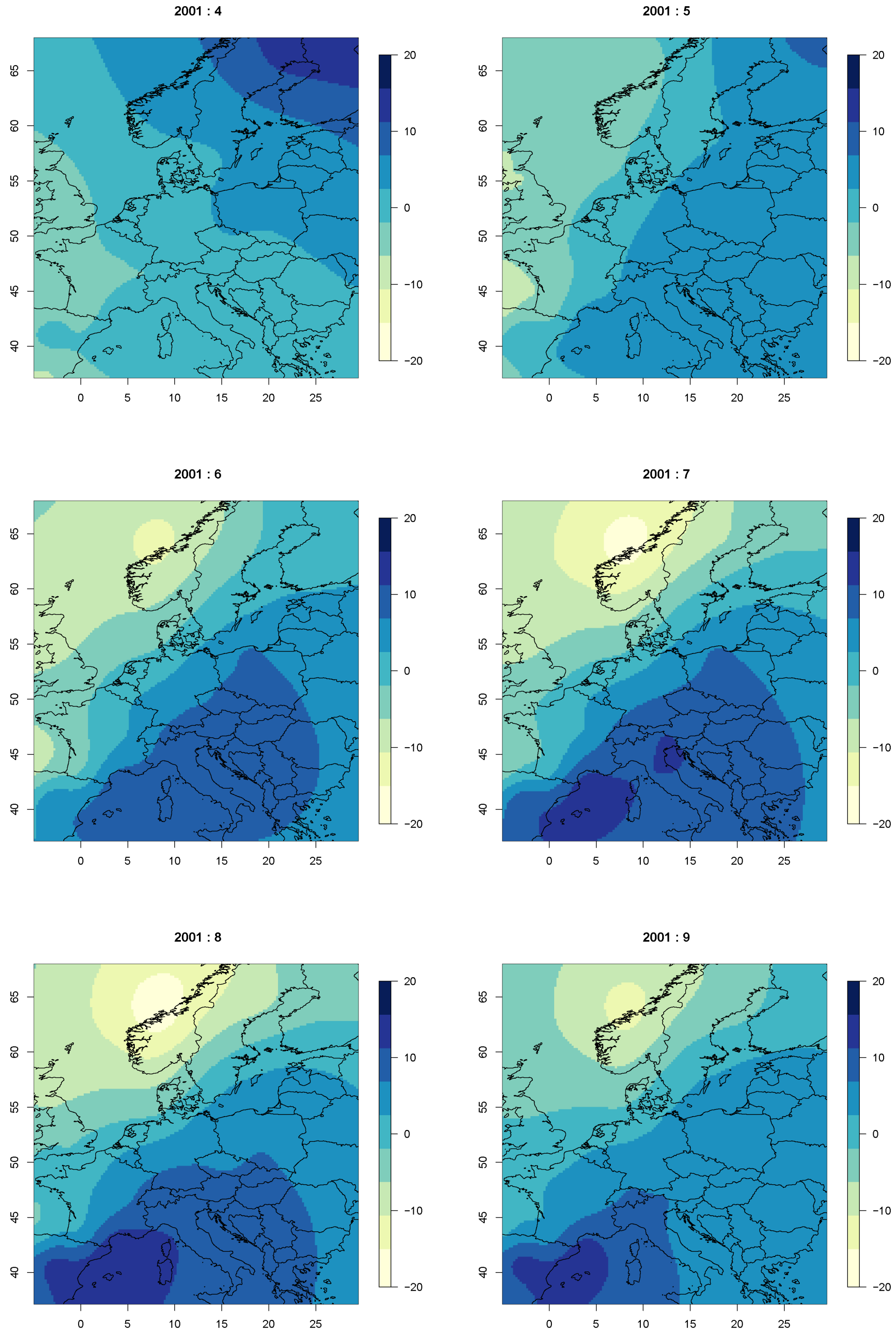
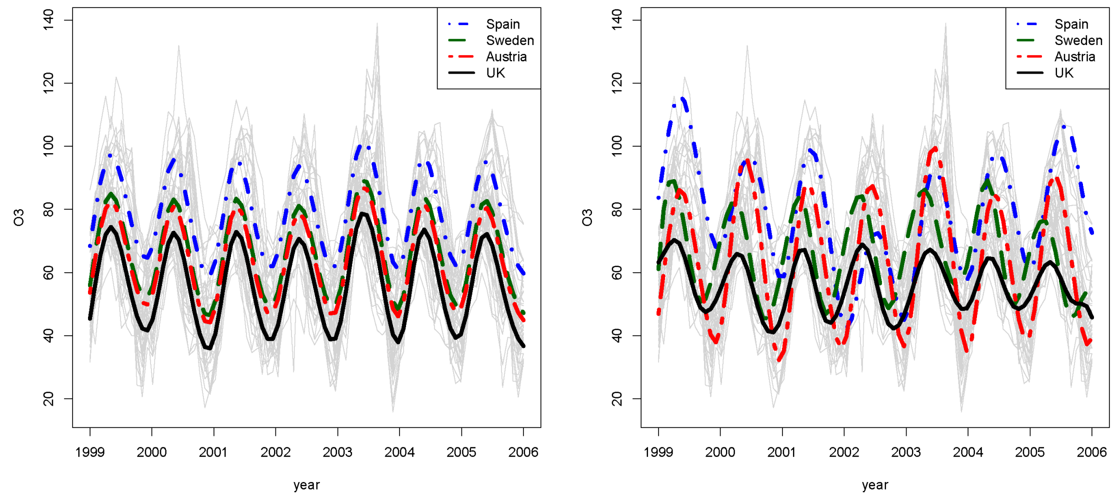


Figure 3: Spatio-temporal interaction fit for the spatio-temporal ANOVA model, from April to September 2001.

In practice, our approach does not require the construction of the transformation matrix $\boldsymbol{T}$. Since, as shown in Section 3, given the smooth-ANOVA model it is much more easier to construct the mixed model matrices $\boldsymbol{X}$ and $\boldsymbol{Z}$, removing the corresponding vector column of $\mathbf{1}_n$'s, and then construct the associated block-diagonal penalty $\boldsymbol{F}$



(a) additive model fit (b) spatio-temporal ANOVA model fit

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

with its smoothing parameters. 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 ANOVA model proposed is the interpretation of the smoothing and the ability of visualize each of the terms of the decomposition in descriptive plots. The 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 are very intensive and require efficient computational methods. In the P-spline approach, the dimension of the bases involved in the smoothing depends basicly 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 found adequate a number of 4 knots for each of the seven 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 methods also have an important role in the algorithms implementation, since 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.

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).

References

Banerjee, S., Carlin, B. P., and Gelfand, A. E. (2004). *Hierarchical Modeling and Analysis for Spatial Data*. Monographs on Statistics and Applied Probability 101. Chapman & Hall/CRC.

Brumback, B. and Rice, J. (1998). Smoothing Spline models for the analysis of nested and crossed samples of curves. *J. Am. Statist. Assoc.*, 93(443):961–994.

Chen, Z. (1993). Fitting multivariate regression functions by interaction spline models. *J. R. Statist. Soc. B*, 55:473–491.

Currie, I. D. and Durbán, M. (2002). Flexible smoothing with P-splines: A unified approach. *Statistical Modelling*, 2:333–349.

Currie, I. D., Durbán, M., and Eilers, P. H. C. (2004). Smoothing and forecasting mortality rates. *Statistical Modelling*, 4(4):279–298.

Currie, I. D., Durbán, M., and Eilers, P. H. C. (2006). Generalized linear array models with applications to multidimensional smoothing. *J. R. Statist. Soc. B*, 68:1–22.

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

Durbán, M., Currie, I. D., and Eilers, P. H. C. (2006). Mixed models, array methods and multidimensional density estimation. In *Proceedings of the 21st International Workshop on Statistical Modelling*, pages 143–150, Galway, Ireland.

EEA (2009). Air pollution by ozone across Europe during summer 2008. Technical report, European Environmental Agency. No 2/2009.

Eilers, P. H. C., Currie, I. D., and Durbán, M. (2006). Fast and compact smoothing on large multidimensional grids. *Computational Statistics and Data Analysis*, 50(1):61–76.

Eilers, P. H. C. and Marx, B. D. (1996). Flexible smoothing with B-splines and penalties. *Stat. Sci.*, 11:89–121.

Fahrmeir, L., Kneib, T., and Lang, S. (2004). Penalized structured additive regression for space-time data: A bayesian perspective. *Statistica Sinica*, 14:715–745.

Gössl, C., Auer, D. P., and Fahrmeir, L. (2001). Bayesian spatiotemporal inference in functional magnetic resonance imaging. *Biometrics*, 57(2):554–562.

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. Chapman and Hall, London.

Kammann, E. E. and Wand, M. P. (2003). Geoadditive models. *Journal of the Royal Statistical Society, C - Applied Statistics*, 52:1–18.

Kneib, T. and Fahrmeir, L. (2006). Structured additive regression for categorical space-time data: A mixed model approach. *Biometrics*, 62:109–118.

Lee, D.-J. and Durbán, M. (2008). P-spline ANOVA-Type interaction models for spatio-temporal smoothing. In Eilers, P. H. C., editor, *Proceedings in the 23nd International Workshop on Statistical Modelling*, pages 315–320, Utrecht, The Netherlands.

Lee, D.-J. and Durbán, M. (2009). Smooth-CAR mixed models for spatial count data. *Computational Statistics and Data Analysis*, 53(8):2968–2979.

Liu, S. (1999). Matrix results on the Khatri-Rao and Tracy-Singh products. *Linear Algebra and its Applications*, 289:267–277.

MacNab, Y. and Dean, C. B. (2001). Autoregressive spatial smoothing and temporal spline smoothing for mapping rates. *Biometrics*, 57:949–956.

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

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

Verbyla, A., Cullis, B., Kenward, M., and Welham, S. (1999). The analysis of designed experiments and longitudinal data using smoothing splines. *J. Roy. Stat. Soc. C*, 48:269–312.

Wahba, G. and Luo, Z. (1995). Smoothing spline ANOVA fits for very large nearly regular data sets, with application to historical global climate data. Technical report, Dept. of Statistics, Univ. of Wisconsin.

Wand, M. P. (2003). Smoothing and mixed models. *Computational Statistics*, 18:223–249.

Welham, S. J., Cullis, B. R., Kenward, M. G., and Thompson, R. (2007). A comparison of mixed model splines for curve fitting. *Aust. N. Z. J. Stat.*, 49(1):1–23.

Wood, S. N. (2006a). *Generalized Additive Models - An introduction with R*. Texts in Statistical Science. Chapman & Hall.

Wood, S. N. (2006b). Low-rank scale-invariant tensor product smooths for generalized additive mixed models. *Biometrics*, 62(4):1025–1036.

Appendix

A Transformation matrix $\boldsymbol{T}$ to impose identifiability constraints

For model (3.2), the transformation matrix $\boldsymbol{T}$ is defined by blocks. The procedure of removing the vector column $\mathbf{1}_n$ in the mixed model bases, is equivalent to remove the first column of the null space eigenvectors $\boldsymbol{U}_{1n}$ and $\boldsymbol{U}_{2n}$, i.e. $\mathbf{1}_1^*$ and $\mathbf{1}_2^*$ in the transformation matrix $\boldsymbol{T}$. Then, the model is reparameterized by considering the transformation $\boldsymbol{T} = [\boldsymbol{T}_n : \boldsymbol{T}_s]$, with:

$$\boldsymbol{T}_n = \begin{bmatrix} 1 & \cdots & & 0 \\ \vdots & \boldsymbol{u}_1^* & & \\ & & \boldsymbol{u}_2^* & \\ 0 & & & \boldsymbol{u}_2^* \otimes \boldsymbol{u}_1^* \end{bmatrix} \text{ and } \boldsymbol{T}_s = \begin{bmatrix} 0 & \cdots & \\ \boldsymbol{U}_{1s} & & \\ \vdots & \boldsymbol{U}_{2s} & \\ & & \boldsymbol{u}_2^* \otimes \boldsymbol{U}_{1s} : \boldsymbol{U}_{2s} \otimes \boldsymbol{u}_1^* : \boldsymbol{U}_{2s} \otimes \boldsymbol{U}_{1s} \end{bmatrix}, \quad \text{(A.1)}$$

The definition of $\boldsymbol{T}$ given above is based on the SVD decomposition of the penalty matrix of second order, $\boldsymbol{D}'\boldsymbol{D} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{U}'$, where $\boldsymbol{U} = [\boldsymbol{U}_n : \boldsymbol{U}_s]$. We consider for the null space: $\boldsymbol{U}_n = [\mathbf{1}^* : \boldsymbol{u}^*]$, the $c \times 2$, matrix of eigenvectors, where $\mathbf{1}^* = \mathbf{1}_c/\sqrt{c}$, with $\mathbf{1}$ a vector of ones of length $c \times 1$ and $\boldsymbol{u}^*$ is the vector $(1, 2, ...c)$ centered and scaled to have unit length.

The identity $c \times c$ matrix $\boldsymbol{I}_c$ can be decomposed as the sum:

$$\boldsymbol{I}_c = \boldsymbol{U}_n\boldsymbol{U}_n' + \boldsymbol{U}_s\boldsymbol{U}_s' , \quad \text{(A.2)}$$

where $\boldsymbol{U}_n\boldsymbol{U}_n' = \mathbf{1}^*\mathbf{1}^{*\prime} + \boldsymbol{u}^*\boldsymbol{u}^{*\prime}$. Then,

$$\boldsymbol{U}_s\boldsymbol{U}_s' = \boldsymbol{I}_c - \mathbf{1}^*\mathbf{1}^{*\prime} - \boldsymbol{u}^*\boldsymbol{u}^{*\prime} \quad \text{(A.3)}$$

where $\mathbf{1}^*\mathbf{1}^{*\prime} = \mathbf{1}\mathbf{1}'/c$. From (A.3), we have

$$\boldsymbol{U}_s\boldsymbol{U}_s' + \boldsymbol{u}^*\boldsymbol{u}^{*\prime} = \boldsymbol{I}_c - \mathbf{1}\mathbf{1}'/c \quad \text{(A.4)}$$

Using equation (A.4) and the definition of $\boldsymbol{T}$ given in (A.1) it is inmediate to prove that $\boldsymbol{T}\boldsymbol{T}' = \boldsymbol{K}$ in (3.11).

B Spatio-temporal Smooth-ANOVA

B.1 Mixed model penalty

The penalty matrix corresponding to basis (4.9) would be:

$$\boldsymbol{P} = \text{blockdiag}(0, \boldsymbol{P}_{(s)}, \boldsymbol{P}_{(t)}, \boldsymbol{P}_{(s,t)}) , \quad \text{(B.1)}$$

where $\boldsymbol{P}_{(s)}$ is the penalty matrix for the spatial $2d$-smooth term as in (2.5), with smoothing parameters λ_1 and λ_2. The penalty matrix for the temporal dimension is $\boldsymbol{P}_{(t)} = \lambda_t \boldsymbol{D}_t' \boldsymbol{D}_t$, and for the space-time interaction the penalty $\boldsymbol{P}_{(s,t)}$ is similar to (4.4) with smoothing parameters τ_1, τ_2, and τ_t. However, basis (4.9) is not of full rank, so we proceed as in Section 3, and define a matrix $\boldsymbol{T}$ that will enable us to calculate the mixed model penalty as $\boldsymbol{F} = \boldsymbol{T}'\boldsymbol{P}\boldsymbol{T}$. In this case: $\boldsymbol{T} = [\boldsymbol{T}_n : \boldsymbol{T}_s]$, and removing the appropiate null space eigenvectors, we have:

$$\boldsymbol{T}_n = \begin{bmatrix} \text{blockdiag}(1\ ,\ [\boldsymbol{1}_2^* \otimes \boldsymbol{u}_1^* : \boldsymbol{u}_2^* \otimes \boldsymbol{1}_1^* : \boldsymbol{u}_2^* \otimes \boldsymbol{u}_1^*]\ ,\ \boldsymbol{u}_t^*) \\ [\boldsymbol{1}_2^* \otimes \boldsymbol{u}_1^* : \boldsymbol{u}_2^* \otimes \boldsymbol{1}_1^* : \boldsymbol{u}_2^* \otimes \boldsymbol{u}_1^*] \otimes \boldsymbol{u}_t^* \end{bmatrix} \quad \text{and} \tag{B.2}$$

$$\boldsymbol{T}_s = \begin{bmatrix} \text{blockdiag}(0\ , [\boldsymbol{U}_{2n} \otimes \boldsymbol{U}_{1s} : \boldsymbol{U}_{2s} \otimes \boldsymbol{U}_{1n} : \boldsymbol{U}_{2s} \otimes \boldsymbol{U}_{1s}], \boldsymbol{U}_{ts}) \\ [\boldsymbol{U}_{2n} \otimes \boldsymbol{U}_{1s} : \boldsymbol{U}_{2s} \otimes \boldsymbol{U}_{1n} : \boldsymbol{U}_{2s} \otimes \boldsymbol{U}_{1s}] \otimes \boldsymbol{u}_t^* : [\boldsymbol{1}_2^* \otimes \boldsymbol{u}_1^* : \boldsymbol{u}_2^* \otimes \boldsymbol{1}_1^* : \boldsymbol{u}_2^* \otimes \boldsymbol{u}_1^*] \otimes \boldsymbol{U}_{ts} \end{bmatrix}. \tag{B.3}$$

Then, $\boldsymbol{F}$ is the block-diagonal matrix:

$$\boldsymbol{F} = \text{blockdiag}(\boldsymbol{0}_8, \boldsymbol{F}_{(s)}, \boldsymbol{F}_{(t)}, \boldsymbol{F}_{(st)})\ , \tag{B.4}$$

with blocks:

$$\begin{aligned} \boldsymbol{F}_{(s)} &= \text{blockdiag}(\lambda_1 \boldsymbol{I}_2 \otimes \widetilde{\boldsymbol{\Sigma}}_1, \lambda_2 \widetilde{\boldsymbol{\Sigma}}_2 \otimes \boldsymbol{I}_2, \lambda_1 \boldsymbol{I}_{c_2-2} \otimes \widetilde{\boldsymbol{\Sigma}}_1 + \lambda_2 \widetilde{\boldsymbol{\Sigma}}_2 \otimes \boldsymbol{I}_{c_1-2})\ , \\ \boldsymbol{F}_{(t)} &= \lambda_t \widetilde{\boldsymbol{\Sigma}}_t\ , \\ \boldsymbol{F}_{(s,t)} &= \begin{bmatrix} \text{blockdiag}(\tau_1 \boldsymbol{I}_2 \otimes \widetilde{\boldsymbol{\Sigma}}_1, \tau_2 \widetilde{\boldsymbol{\Sigma}}_2 \otimes \boldsymbol{I}_2, \tau_2 \widetilde{\boldsymbol{\Sigma}}_2 \otimes \boldsymbol{I}_{c_1-2} + \tau_1 \boldsymbol{I}_{c_2-2} \otimes \widetilde{\boldsymbol{\Sigma}}_1) \\ \text{blockdiag}(\tau_3 \boldsymbol{I}_3 \otimes \widetilde{\boldsymbol{\Sigma}}_3, \tau_1 \widetilde{\boldsymbol{\Sigma}}_1 \otimes \boldsymbol{I}_{c_t-2} + \tau_t \boldsymbol{I}_{c_1-2} \otimes \widetilde{\boldsymbol{\Sigma}}_t, \tau_2 \widetilde{\boldsymbol{\Sigma}}_2 \otimes \boldsymbol{I}_{c_t-2} + \tau_t \boldsymbol{I}_{c_2-2} \otimes \widetilde{\boldsymbol{\Sigma}}_t) \\ \tau_1 \boldsymbol{I}_{c_2-2} \otimes \widetilde{\boldsymbol{\Sigma}}_1 \otimes \boldsymbol{I}_{c_t-2} + \tau_2 \widetilde{\boldsymbol{\Sigma}}_2 \otimes \boldsymbol{I}_{c_1-2} \otimes \boldsymbol{I}_{c_t-2} + \tau_t \boldsymbol{I}_{c_2-2} \otimes \boldsymbol{I}_{c_1-2} \otimes \widetilde{\boldsymbol{\Sigma}}_t \end{bmatrix}. \end{aligned}$$

B.2 Linear constraints over coefficients in the spatio-temporal case

To demonstrate how to obtain the constraints for the space-time interaction in (4.17), we use the relationship between $\boldsymbol{T}$ and $\boldsymbol{F}$ given above, and the penalty in the original parametrization taking into account the identifiability constraints (3.12). For the coefficient vector $\boldsymbol{\theta}^{(\text{st})}$ (or in array form $\boldsymbol{\Theta}^{(\text{st})}$, $c_t \times c_1 \times c_2$), the penalty becomes:

$$\breve{\boldsymbol{P}}_{(s,t)} = \tau_1 \underbrace{\boldsymbol{I}_{c_2} \otimes \boldsymbol{D}_1' \boldsymbol{D}_1 \otimes \boldsymbol{K}_t}_{\text{(a)}} + \tau_2 \underbrace{\boldsymbol{D}_2' \boldsymbol{D}_2 \otimes \boldsymbol{I}_{c_1} \otimes \boldsymbol{K}_t}_{\text{(b)}} + \tau_t \underbrace{(\boldsymbol{I}_{c_2} \otimes \boldsymbol{I}_{c_1} - \boldsymbol{1}\boldsymbol{1}_2'/c_2 \otimes \boldsymbol{1}\boldsymbol{1}_1'/c_1) \otimes \boldsymbol{D}_t' \boldsymbol{D}_t}_{\text{(c)}}\ , \tag{B.5}$$

where (a) and (b) impose the constraints:

$$\textstyle\sum_i^{c_1} \boldsymbol{\theta}_{t,ij}^{(\text{st})} = 0 \quad \text{and} \quad \sum_j^{c_2} \boldsymbol{\theta}_{t,ij}^{(\text{st})} = 0\ .$$

The last term (c), can be rewritten as $\boldsymbol{K}_s \otimes \boldsymbol{D}_t' \boldsymbol{D}_t$, where $\boldsymbol{K}_s$ is a centering matrix over dimension $c_s = c_1 c_2$, i.e.

$$\boldsymbol{K}_s = (\boldsymbol{I}_{c_s} - \boldsymbol{1}\boldsymbol{1}_s'/c_s)\ ,$$

which leads to the constraint:

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