Working paper

2019-10

Statistics and Econometrics

ISSN 2387-0303

Out-of-sample prediction in multidimensional P-spline models

Alba Carballo, María Durbán, Dae-Jin Lee

uc3m | Universidad Carlos III de Madrid

Out-of-sample prediction in multidimensional P-spline models

Alba Carballo[a], María Durbán[a], Dae-Jin Lee[b]

[a] *Department of Statistics, Universidad Carllos III de Madrid, Spain*
[b] *Basque Center for Applied Mathematics, Bilbao, Spain*

Abstract

Prediction of out-of-sample values is a problem of interest in any regression model. In the context of penalized smooth mixed model regression Carballo *et al.* (2017) have proposed a general framework for prediction in additive models without interaction terms. The aim of this paper is to extend this work, based on the methodology proposed in Currie *et al.* (2004), to models that include interaction terms, i.e. prediction is needed in multidimensional setting.
Our approach fits the data and predicts the new observations simultaneously and uses constraints to ensure a coherent fit or to impose further restrictions on the predictions. We also develop this methodology for the so called smooth-ANOVA models which allow us to include interaction terms that can be decomposed as a sum of several smooth functions. To illustrate the methodology two real data sets are used, one to predict log mortality rates in the Spanish population and another to predict aboveground biomass in Populus trees as a smooth function of height and diameter. We examine the performance of the interaction models in comparison to the Smooth-ANOVA models (both models with and without the restriction the fit has to be maintained) through a simulation study.

Keywords: Prediction, Penalized regression, P-splines, Mixed models

1 Introduction

P-splines with B-splines bases have become popular in applications and in theoretical work. More than twenty years ago that Eilers & Marx (1996) proposed the P-splines methodology and its use has become very popular. The methodology has not been used only to fit data but also to obtain out-of-sample predicted values. Currie *et al.* (2004) showed how the method of P-splines can be extended to the smoothing and forecasting of two-dimensional mortality tables. Etxeberria *et al.* (2015) and Ugarte *et al.* (2012) have also used the P-spline forecasting methodology, they use a three-dimensional space-time P-spline model to forecast mortality cancer risks in future years.

However, prediction with P-splines is still an open area of research. One of the main interests of this paper is to delve into the study of the method proposed in Currie *et al.* (2004) to give relations between the coefficients that determine the fit and the coefficients that determine the forecast, and to extend the methodology to the mixed models framework. For the one dimensional case, this study has been done in Carballo *et al.* (2017).

Studying the properties of the method proposed by Currie *et al.* (2004) we will see that the fit changes when the fit obtained when using the "in sample data" is different from when fitting and predicting simultaneously. The differences between the fits can sometimes be small but as we also see in the examples, the effect on the prediction can be relevant. We will show that for the particular case in which just one of the two covariates is extended the fit can be maintained by modifying the extended penalty matrix. However, when the two covariates are extended the penalty matrix cannot be modified, since the matrices involved in obtaining the estimated parameters are singular. As a general solution to keep the fit we propose to impose restrictions on the coefficients. We achieve this by using the standard Lagrangian formulation of the least squares minimization problem following Greene & Seaks (1991). In this paper, we mostly use restrictions to impose that the fit has to be maintained, but our proposal could be useful for other purposes. For instance, to incorporate known information about the value of the shape of prediction out-of-sample values.

In context such as spatial or spatio-temporal modelling we can be interested in prediction at new locations, this would involve out-of-sample prediction for two covariates (latitude and longitude), and so we extend the proposal of Currie *et al.* (2004) to predict when the two covariates are extended and to predict with Smooth-ANOVA models. These models have been formulated as mixed models by Lee (2010) and Lee & Durbán (2011) and allow us to decompose multidimensional smooth functions as additive terms and interactions.

We organize the remaining of the paper as follows. In Section 2 we give a general approach to out-of-sample prediction in penalized splines additive models with interactions. First we extend Currie *et al.* (2004) to the case in which prediction out-of-sample is needed in both covariates of the interaction. Then, we show the properties satisfied, under certain conditions, by the coefficients that determine the prediction. Furthermore, we propose an easy method, based on Lagrange multipliers, to obtain constrained predictions. Section 3 shows how out-of-sample predictions can be carried out in the context of multidimensional smooth mixed models, we propose different reparametrizations to predict new values and also show how to impose constraints in this context. Our proposal to predict new values in a more flexible context is shown in Section 4, where we give results on out-of-sample prediction for the Smooth-ANOVA model of Lee & Durbán (2011). In Section 5 we carried out an extensive simulation study to compare interaction and S-ANOVA models as well as their restricted versions. Finally we give some conclusions in Section 6.

2 Prediction in additive models based on multidimensional penalized splines

Additive models are a class of non-parametric regression methods which have been found widespread applications in practice. One of the main assumptions of additive models is that the effect of covariates on the dependent variable follows and additive form,

$$\boldsymbol{y} = f_1(\boldsymbol{x}_1) + f_2(\boldsymbol{x}_2) + f_3(\boldsymbol{x}_3) + \ldots + \boldsymbol{\epsilon}, \quad \boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{R}),$$

with $\boldsymbol{R} = \sigma^2_{\boldsymbol{\epsilon}}\mathbf{I}$, i.e., $\boldsymbol{\epsilon}$ are independent and identically distributed errors with variance $\sigma^2_{\boldsymbol{\epsilon}}$, and f_i smooth functions. If there is no interaction between the terms, everything related to prediction under the Gaussian framework has been done in Carballo *et al.* (2017). Therefore, here we focus on the case of interactions, i.e., in additive models that include terms of the form $f(\boldsymbol{x}_1, \boldsymbol{x}_2)$.

In order to study the prediction approach given in Currie *et al.* (2004), we briefly review the P-splines methodology in the two-dimensional case. We consider a general non-parametric two-dimensional regression model:

$$\boldsymbol{y} = f(\boldsymbol{z}, \boldsymbol{x}) + \boldsymbol{\epsilon}, \quad \boldsymbol{\epsilon} \backsim N(\boldsymbol{0}, \boldsymbol{R}), \tag{1}$$

where $\boldsymbol{z}$, $\boldsymbol{x}$ are the regressors, $\boldsymbol{R} = \sigma^2_{\boldsymbol{\epsilon}}\mathbf{I}$, and $f(\cdot)$ is a 2−multidimensional smooth function that depends on the 2 explanatory variables $\boldsymbol{z} = (z_1, ..., z_{n_{\boldsymbol{z}}})'$ and $\boldsymbol{x} = (x_1, ..., x_{n_{\boldsymbol{x}}})'$, and each of them have lengths $n_{\boldsymbol{z}}$ and $n_{\boldsymbol{x}}$, respectively. Although we are assuming i.i.d. errors for simplicity the results can be easily extended to the case of a general variance-covariance matrix $\boldsymbol{R}$. Suppose now that we are interested in fitting model (1), and assume that the function $f(\boldsymbol{z}, \boldsymbol{x})$ can be represented in terms of basis functions:

$$f(\boldsymbol{z}, \boldsymbol{x}) = \boldsymbol{B}\boldsymbol{\theta}, \tag{2}$$

with $\boldsymbol{B}$ a B-spline regression basis, and $\boldsymbol{\theta}$ the vector of coefficients. If we consider array data, the smooth multidimensional surface is constructed from the Kronecker product of the marginal B-spline basis for each covariate, the basis for the model (2) is

$$\boldsymbol{B} = \boldsymbol{B}_{\boldsymbol{x}} \otimes \boldsymbol{B}_{\boldsymbol{z}}, \tag{3}$$

where $\otimes$ is the Kronecker product of two matrices, and $\boldsymbol{B}_{\boldsymbol{x}} = \boldsymbol{B}(\boldsymbol{x})$ and $\boldsymbol{B}_{\boldsymbol{z}} = \boldsymbol{B}(\boldsymbol{z})$, of dimensions $n_{\boldsymbol{x}} \times c_{\boldsymbol{x}}$ and $n_{\boldsymbol{z}} \times c_{\boldsymbol{z}}$, are the marginal B-spline basis for $\boldsymbol{x}$ and $\boldsymbol{z}$, respectively. Then, the dimension of (3) is $n_{\boldsymbol{x}}n_{\boldsymbol{z}} \times c_{\boldsymbol{x}}c_{\boldsymbol{z}}$. On the other hand, if we consider scattered data, the basis is constructed from the Tensor product of marginal B-spline basis defined in Currie *et al.* (2006) as the Box-Product, denoted by symbol $\Box$:

$$\boldsymbol{B} = \boldsymbol{B}_{\boldsymbol{x}} \Box \boldsymbol{B}_{\boldsymbol{z}} = (\boldsymbol{B}_{\boldsymbol{x}} \otimes \mathbf{1}'_{c_{\boldsymbol{z}}}) \odot (\mathbf{1}'_{c_{\boldsymbol{x}}} \otimes \boldsymbol{B}_{\boldsymbol{z}}),$$

where the operator $\odot$ is the element-wise matrix product and $\mathbf{1}_{c_{\boldsymbol{z}}}$ and $\mathbf{1}_{c_{\boldsymbol{x}}}$ are column vectors of ones of lengths $c_{\boldsymbol{z}}$ and $c_{\boldsymbol{x}}$.

In both cases, the vector of coefficients $\boldsymbol{\theta}$ can be arranged into a $c_{\boldsymbol{z}} \times c_{\boldsymbol{x}}$ matrix $\boldsymbol{\Theta}$, that is

$$\boldsymbol{\Theta} = \begin{bmatrix} \boldsymbol{\theta}_{11} & \boldsymbol{\theta}_{12} & \cdots & \boldsymbol{\theta}_{1c_{\boldsymbol{x}}} \\ \boldsymbol{\theta}_{21} & \boldsymbol{\theta}_{22} & \cdots & \boldsymbol{\theta}_{2c_{\boldsymbol{x}}} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{\theta}_{c_{\boldsymbol{z}}1} & \boldsymbol{\theta}_{c_{\boldsymbol{z}}2} & \cdots & \boldsymbol{\theta}_{c_{\boldsymbol{z}}c_{\boldsymbol{x}}} \end{bmatrix}, \tag{4}$$

then, the two-dimensional P-spline model can be written as

$$f(\boldsymbol{z}, \boldsymbol{x}) = (\boldsymbol{B}_{\boldsymbol{x}} \otimes \boldsymbol{B}_{\boldsymbol{z}})\,\boldsymbol{\theta} = \text{vec}(\boldsymbol{B}_{\boldsymbol{z}}\boldsymbol{\Theta}\boldsymbol{B}'_{\boldsymbol{x}}), \tag{5}$$

where vec$(\cdot)$ denotes the vectorization operator.

In the two dimensional case, the penalty on the coefficients vector $\boldsymbol{\theta}$ penalizes the difference between adjacent coefficients of rows and columns of the matrix $\boldsymbol{\Theta}$ (4). The penalty on rows of $\boldsymbol{\Theta}$ is:

$$\sum_{j=1}^{c_{\boldsymbol{z}}} \boldsymbol{\theta}'_j \boldsymbol{D}'_{\boldsymbol{z}} \boldsymbol{D}_{\boldsymbol{z}} \boldsymbol{\theta}_j = \boldsymbol{\theta}'(\boldsymbol{I}_{c_{\boldsymbol{x}}} \otimes \boldsymbol{D}'_{\boldsymbol{z}}\boldsymbol{D}_{\boldsymbol{z}})\boldsymbol{\theta}, \tag{6}$$

and, similarly, on the columns:

$$\sum_{i=1}^{c_{\boldsymbol{x}}} \boldsymbol{\theta}'_i \boldsymbol{D}'_{\boldsymbol{x}} \boldsymbol{D}_{\boldsymbol{x}} \boldsymbol{\theta}_i = \boldsymbol{\theta}'(\boldsymbol{D}'_{\boldsymbol{x}}\boldsymbol{D}_{\boldsymbol{x}} \otimes \boldsymbol{I}_{c_{\boldsymbol{z}}})\boldsymbol{\theta}, \tag{7}$$

where $\boldsymbol{D}_{\boldsymbol{z}}$ and $\boldsymbol{D}_{\boldsymbol{x}}$ are the difference matrices acting on the rows and columns of $\boldsymbol{\Theta}$, respectively. Therefore, the penalty matrix $\boldsymbol{P}$ in two dimensions is:

$$\boldsymbol{P} = \lambda_{\boldsymbol{z}} \underbrace{\boldsymbol{I}_{c_{\boldsymbol{x}}} \otimes \boldsymbol{D}'_{\boldsymbol{z}} \boldsymbol{D}_{\boldsymbol{z}}}_{\boldsymbol{P}^{\boldsymbol{z}}} + \lambda_{\boldsymbol{x}} \underbrace{\boldsymbol{D}'_{\boldsymbol{x}} \boldsymbol{D}_{\boldsymbol{x}} \otimes \boldsymbol{I}_{c_{\boldsymbol{z}}}}_{\boldsymbol{P}^{\boldsymbol{x}}}, \tag{8}$$

where $\lambda_{\boldsymbol{z}}$ and $\lambda_{\boldsymbol{x}}$ are the smoothing parameters for each dimension of the model. Since $\lambda_{\boldsymbol{z}}$ and $\lambda_{\boldsymbol{x}}$ are not necessary equal, the penalty (8) allows for anisotropic smoothing. To estimate the coefficients, Eilers & Marx (1996) minimize the penalized sum of squares:

$$S(\boldsymbol{\theta}) = (\boldsymbol{y} - \boldsymbol{B}\boldsymbol{\theta})'(\boldsymbol{y} - \boldsymbol{B}\boldsymbol{\theta}) + \boldsymbol{\theta}'\boldsymbol{P}\boldsymbol{\theta}. \tag{9}$$

Therefore, for given values of $\lambda_{\boldsymbol{z}}$ and $\lambda_{\boldsymbol{x}}$, the solution of the penalized sum of squares (9), is:

$$\hat{\boldsymbol{\theta}} = (\boldsymbol{B}'\boldsymbol{B} + \boldsymbol{P})^{-1}\boldsymbol{B}'\boldsymbol{y}. \tag{10}$$

The smoothing parameter of each dimension can be estimated using a information criteria (such as Akaike or Bayesian criteria) or a cross-validation criteria method.

Once we have presented a brief introduction of the multidimensional P-splines, in the next section, we detail the prediction methodology. Although we will use B-spline basis and penalties based on differences, the methodology proposed here can be extended to any basis and quadratic penalty.

2.1 Out-of-sample prediction

In this section, we extend the approach given in Currie *et al.* (2004) to obtain the forecast when not only one of the two independent variables but the two extend. In the framework of model (1), given a vector of $n_{\boldsymbol{z}} n_{\boldsymbol{x}}$ observations $\boldsymbol{y}$ of the response variable, suppose that we want to predict $n_p = n_{\boldsymbol{z}} n_{\boldsymbol{x}_p} + n_{\boldsymbol{z}_p} n_{\boldsymbol{x}} + n_{\boldsymbol{z}_p} n_{\boldsymbol{x}_p}$ new values at $(\boldsymbol{z}, \boldsymbol{x}_p)$, $(\boldsymbol{z}_p, \boldsymbol{x})$ and $(\boldsymbol{z}_p, \boldsymbol{x}_p)$, i.e., if we arrange the observations vector into a matrix $\boldsymbol{Y}$ of dimension $n_{\boldsymbol{z}} \times n_{\boldsymbol{x}}$, the observed and predicted values can be arranged into a matrix of dimension $n_{\boldsymbol{z}_+} \times n_{\boldsymbol{x}_+}$ ($n_{\boldsymbol{z}_+} = n_{\boldsymbol{z}} + n_{\boldsymbol{z}_p}$, $n_{\boldsymbol{x}_+} = n_{\boldsymbol{x}} + n_{\boldsymbol{x}_p}$), as:

$$\boldsymbol{Y}_+ = \begin{bmatrix} \boldsymbol{Y} & \boldsymbol{Y}_{\boldsymbol{z}\boldsymbol{x}_p} \\ \boldsymbol{Y}_{\boldsymbol{z}_p\boldsymbol{x}} & \boldsymbol{Y}_{\boldsymbol{z}_p\boldsymbol{x}_p} \end{bmatrix}. \tag{11}$$

Notice that the dimensions of $\boldsymbol{Y}_{\boldsymbol{z}\boldsymbol{x}_p}$, $\boldsymbol{Y}_{\boldsymbol{z}_p\boldsymbol{x}}$ and $\boldsymbol{Y}_{\boldsymbol{z}_p\boldsymbol{x}_p}$ are $n_{\boldsymbol{z}} \times n_{\boldsymbol{x}_p}$, $n_{\boldsymbol{z}_p} \times n_{\boldsymbol{x}}$ and $n_{\boldsymbol{z}_p} \times n_{\boldsymbol{x}_p}$, respectively.

We propose to fit and forecast the model simultaneously considering the following extended model:

$$\boldsymbol{y}_+ = \boldsymbol{B}_+\boldsymbol{\theta}_+ + \boldsymbol{\epsilon}_+, \quad \boldsymbol{\epsilon}_+ \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{R}_+) \tag{12}$$

where $\boldsymbol{y}_+ = \text{vec}(\boldsymbol{Y}_+)$, with $\boldsymbol{Y}_+$ as in (11), where $\boldsymbol{Y}$ are the observed values and $\boldsymbol{Y}_{\boldsymbol{z}\boldsymbol{x}_p}$, $\boldsymbol{Y}_{\boldsymbol{z}_p\boldsymbol{x}}$ and $\boldsymbol{Y}_{\boldsymbol{z}_p\boldsymbol{x}_p}$ are arbitrary values, and $\boldsymbol{R}_+ = \sigma^2_{\boldsymbol{\epsilon}} \tilde{\boldsymbol{R}}_+$ with $\tilde{\boldsymbol{R}}_+ = \tilde{\boldsymbol{R}}_{\boldsymbol{x}_+} \otimes \tilde{\boldsymbol{R}}_{\boldsymbol{z}_+}$, where $\tilde{\boldsymbol{R}}_{\boldsymbol{x}_+}$ and $\tilde{\boldsymbol{R}}_{\boldsymbol{z}_+}$ are diagonal matrices of dimensions $n_{\boldsymbol{x}_+} \times n_{\boldsymbol{x}_+}$ ($n_{\boldsymbol{x}_+} = n_{\boldsymbol{x}} + n_{\boldsymbol{x}_p}$) and $n_{\boldsymbol{z}_+} \times n_{\boldsymbol{z}_+}$ ($n_{\boldsymbol{z}_+} = n_{\boldsymbol{z}} + n_{\boldsymbol{z}_p}$), respectively, with infinity entries if the data is missing and 1 if the data is observed. The quantity infinity expresses that we do not have any information about the data $\boldsymbol{y}_p$. The extended basis is the Kronecker product of the two extended marginal

B-spline basis, $\boldsymbol{B}_{\boldsymbol{x}_+} = \boldsymbol{B}(\boldsymbol{x}_+)$ and $\boldsymbol{B}_{\boldsymbol{z}_+} = \boldsymbol{B}(\boldsymbol{z}_+)$, of dimensions $n_{\boldsymbol{x}_+} \times c_{\boldsymbol{x}_+}$ and $n_{\boldsymbol{z}_+} \times c_{\boldsymbol{z}_+}$, respectively:

$$\boldsymbol{B}_+ = \boldsymbol{B}_{\boldsymbol{x}_+} \otimes \boldsymbol{B}_{\boldsymbol{z}_+} = \begin{bmatrix} \boldsymbol{B}_{\boldsymbol{x}} & \boldsymbol{O} \\ \boldsymbol{B}_{\boldsymbol{x}_{(1)}} & \boldsymbol{B}_{\boldsymbol{x}_{(2)}} \end{bmatrix} \otimes \begin{bmatrix} \boldsymbol{B}_{\boldsymbol{z}} & \boldsymbol{O} \\ \boldsymbol{B}_{\boldsymbol{z}_{(1)}} & \boldsymbol{B}_{\boldsymbol{z}_{(2)}} \end{bmatrix}, \tag{13}$$

where the extended bases $\boldsymbol{B}_{\boldsymbol{x}_+}$ and $\boldsymbol{B}_{\boldsymbol{z}_+}$ are built from a new set of knots that consists of the original knots and extended to cover the full range of $\boldsymbol{x}_+$ and $\boldsymbol{z}_+$, respectively.

To estimate the extended coefficients, we minimize the following function of $\boldsymbol{\theta}_+$:

$$S(\boldsymbol{\theta}_+) = (\boldsymbol{y}_+ - \boldsymbol{B}_+\boldsymbol{\theta}_+)'\tilde{\boldsymbol{R}}_+^{-1}(\boldsymbol{y}_+ - \boldsymbol{B}_+\boldsymbol{\theta}_+) + \boldsymbol{\theta}_+'\boldsymbol{P}_+\boldsymbol{\theta}_+, \tag{14}$$

with extended penalty matrix

$$\boldsymbol{P}_+ = \lambda_{\boldsymbol{z}}\boldsymbol{P}_+^{\boldsymbol{z}+} + \lambda_{\boldsymbol{x}}\boldsymbol{P}_+^{\boldsymbol{x}+}, \tag{15}$$

where $\lambda_{\boldsymbol{z}}$ and $\lambda_{\boldsymbol{x}}$ and $\boldsymbol{P}_+^{\boldsymbol{z}+}$ and $\boldsymbol{P}_+^{\boldsymbol{x}+}$ are the smoothing parameters and the extended penalty matrices for each dimension of the model, respectively. For the particular case of penalties based on differences, we consider:

$$\begin{aligned} \boldsymbol{P}_+^{\boldsymbol{z}+} &= \boldsymbol{I}_{c_{\boldsymbol{x}+}} \otimes \boldsymbol{D}_{\boldsymbol{z}_+}'\boldsymbol{D}_{\boldsymbol{z}_+} = \begin{bmatrix} \boldsymbol{I}_{c_{\boldsymbol{x}}} & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{I}_{c_{\boldsymbol{x}p}} \end{bmatrix} \otimes \boldsymbol{D}_{\boldsymbol{z}_+}'\boldsymbol{D}_{\boldsymbol{z}_+} = \begin{bmatrix} \boldsymbol{I}_{c_{\boldsymbol{x}}} \otimes \boldsymbol{D}_{\boldsymbol{z}_+}'\boldsymbol{D}_{\boldsymbol{z}_+} & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{I}_{c_{\boldsymbol{x}p}} \otimes \boldsymbol{D}_{\boldsymbol{z}_+}'\boldsymbol{D}_{\boldsymbol{z}_+} \end{bmatrix} \\ &= \begin{bmatrix} \boldsymbol{P}_{+11}^{\boldsymbol{z}+} & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{P}_{+22}^{\boldsymbol{z}+} \end{bmatrix}, \end{aligned} \tag{16}$$

and

$$\begin{aligned} \boldsymbol{P}_+^{\boldsymbol{x}+} &= \boldsymbol{D}_{\boldsymbol{x}_+}'\boldsymbol{D}_{\boldsymbol{x}+} \otimes \boldsymbol{I}_{c_{\boldsymbol{z}}} = \begin{bmatrix} (\boldsymbol{D}_{\boldsymbol{x}}'\boldsymbol{D}_{\boldsymbol{x}} + \boldsymbol{D}_{\boldsymbol{x}(1)}'\boldsymbol{D}_{\boldsymbol{x}(1)}) \otimes \boldsymbol{I}_{c_{\boldsymbol{z}}} & \boldsymbol{D}_{\boldsymbol{x}(1)}'\boldsymbol{D}_{\boldsymbol{x}(2)} \otimes \boldsymbol{I}_{c_{\boldsymbol{z}}} \\ \boldsymbol{D}_{\boldsymbol{x}(2)}'\boldsymbol{D}_{\boldsymbol{x}(1)} \otimes \boldsymbol{I}_{c_{\boldsymbol{z}}} & \boldsymbol{D}_{\boldsymbol{x}(2)}'\boldsymbol{D}_{\boldsymbol{x}(2)} \otimes \boldsymbol{I}_{c_{\boldsymbol{z}}} \end{bmatrix} \\ &= \begin{bmatrix} \boldsymbol{P}_{+11}^{\boldsymbol{x}+} & \boldsymbol{P}_{+12}^{\boldsymbol{x}+} \\ \boldsymbol{P}_{+21}^{\boldsymbol{x}+} & \boldsymbol{P}_{+22}^{\boldsymbol{x}+} \end{bmatrix}, \end{aligned} \tag{17}$$

where $\boldsymbol{D}_{\boldsymbol{x}_+}$ and $\boldsymbol{D}_{\boldsymbol{z}_+}$ are the difference matrices acting on the columns and rows of the matrix formed by the extended vector of coefficients $\boldsymbol{\theta}_+$. Notice that $\boldsymbol{D}_{\boldsymbol{z}_+}$ and $\boldsymbol{D}_{\boldsymbol{x}_+}$ are direct extensions of $\boldsymbol{D}_{\boldsymbol{z}}$ and $\boldsymbol{D}_{\boldsymbol{x}}$ but $\boldsymbol{P}_+^{\boldsymbol{z}+}$ and $\boldsymbol{P}_+^{\boldsymbol{x}+}$ are not direct extensions of $\boldsymbol{P}^{\boldsymbol{z}}$ and $\boldsymbol{P}^{\boldsymbol{x}}$. Moreover, if $\boldsymbol{\theta}$ is given in (25) and we are extending the two covariates, $\boldsymbol{\theta}_+ = \text{vec}(\boldsymbol{\Theta}_+)$, with $\boldsymbol{\Theta}_+$:

$$\boldsymbol{\Theta}_+ = \left[\begin{array}{ccccc|ccc} \boldsymbol{\theta}_{11} & \boldsymbol{\theta}_{12} & \cdots & \boldsymbol{\theta}_{1\ c_{\boldsymbol{x}}-1} & \boldsymbol{\theta}_{1c_{\boldsymbol{x}}} & \boldsymbol{\theta}_{1\ c_{\boldsymbol{x}}+1} & \boldsymbol{\theta}_{1\ c_{\boldsymbol{x}}+2} & \cdots \\ \boldsymbol{\theta}_{21} & \boldsymbol{\theta}_{22} & \cdots & \boldsymbol{\theta}_{2\ c_{\boldsymbol{x}}-1} & \boldsymbol{\theta}_{2c_{\boldsymbol{x}}} & \boldsymbol{\theta}_{2\ c_{\boldsymbol{x}}+1} & \boldsymbol{\theta}_{2\ c_{\boldsymbol{x}}+2} & \cdots \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \boldsymbol{\theta}_{c_{\boldsymbol{z}}1} & \boldsymbol{\theta}_{c_{\boldsymbol{z}}2} & \cdots & \boldsymbol{\theta}_{c_{\boldsymbol{z}}\ c_{\boldsymbol{x}}-1} & \boldsymbol{\theta}_{c_{\boldsymbol{z}}c_{\boldsymbol{x}}} & \boldsymbol{\theta}_{c_{\boldsymbol{z}}\ c_{\boldsymbol{x}}+1} & \boldsymbol{\theta}_{c_{\boldsymbol{z}}\ c_{\boldsymbol{x}}+2} & \cdots \\ \hline \boldsymbol{\theta}_{c_{\boldsymbol{z}}+1\ 1} & \boldsymbol{\theta}_{c_{\boldsymbol{z}}+1\ 2} & \cdots & \boldsymbol{\theta}_{c_{\boldsymbol{z}}+1\ c_{\boldsymbol{x}}-1} & \boldsymbol{\theta}_{c_{\boldsymbol{z}}+1\ c_{\boldsymbol{x}}} & \boldsymbol{\theta}_{c_{\boldsymbol{z}}+1\ c_{\boldsymbol{x}}+1} & \boldsymbol{\theta}_{c_{\boldsymbol{z}}+1\ c_{\boldsymbol{x}}+2} & \cdots \\ \boldsymbol{\theta}_{c_{\boldsymbol{z}}+2\ 1} & \boldsymbol{\theta}_{c_{\boldsymbol{z}}+2\ 2} & \cdots & \boldsymbol{\theta}_{c_{\boldsymbol{z}}+2\ c_{\boldsymbol{x}}-1} & \boldsymbol{\theta}_{c_{\boldsymbol{z}}+2\ c_{\boldsymbol{x}}} & \boldsymbol{\theta}_{c_{\boldsymbol{z}}+2\ c_{\boldsymbol{x}}+1} & \boldsymbol{\theta}_{c_{\boldsymbol{z}}+2\ c_{\boldsymbol{x}}+2} & \cdots \\ \vdots & \vdots & \cdots & \vdots & \vdots & \vdots & \vdots & \cdots \end{array}\right].$$

The solution of the extended penalized least squares problem (14) is:

$$\boldsymbol{\theta}_+ = (\boldsymbol{B}_+'\tilde{\boldsymbol{R}}_+^{-1}\boldsymbol{B}_+ + \boldsymbol{P}_+)^{-1}\boldsymbol{B}_+'\tilde{\boldsymbol{R}}_+^{-1}\boldsymbol{y}_+.$$

As mentioned earlier, a information criteria or a cross-validation criteria method might be suitable to choose the optimal values. In practice, following Camarda (2012), the smoothing parameters in (20) are chosen to be the optimal smoothing parameters for the fit.

Notice that as in the case of the fit, the extension to predict new values depends on the structure of the data. If we consider scattered data, we set n_p out-of-sample points (z_{+_i}, x_{+_i}) at which we want to predict new y_{p_i} values for $i = 1, ..., n_p$ and $\boldsymbol{R}_+$ is a diagonal matrix with the first n values equal to 1 and the last n_p values equal to infinity. Everything else is independent of the data structure.

In the next section, we focus on predictions when just one covariate is extended since in this particular case it is possible to obtain expressions that link the coefficients used in the fit with the ones used in the prediction. This is not possible when we extend the two covariates because of the structure introduced by the Kronecker products.

2.1.1 Prediction of a single covariate

As it is shown in Carballo *et al.* (2017), in one dimension the predicted values depend critically on the order of the penalty, since it determines the shape of the prediction function. However, once the observed values were fitted, the number of knots, the degree of the P-spline and the smoothing parameter don't have a huge influence on the predicted values. In this section, we see that this is not the case when we work in two dimensions.

In the framework of model (1), given a vector of $n_{\boldsymbol{z}} \times n_{\boldsymbol{x}}$ observations $\boldsymbol{y}$ of the response variable, suppose that we want to predict $n_p = n_{\boldsymbol{z}} \times n_{\boldsymbol{x}_p}$ new values $\boldsymbol{y}_p$ at $\boldsymbol{z}$ and $\boldsymbol{x}_p$, i.e. we extend just one of the two covariables. Following Currie *et al.* (2004), we fit and predict the model simultaneously, i.e., we consider the following extended model:

$$\boldsymbol{y}_+ = \boldsymbol{B}_+\boldsymbol{\theta}_+ + \boldsymbol{\epsilon}_+, \quad \boldsymbol{\epsilon}_+ \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{R}_+) \tag{18}$$

where $\boldsymbol{y}_+ = (\boldsymbol{y}', \boldsymbol{y}_p')'$, with $\boldsymbol{y}$ the observed values and $\boldsymbol{y}_p$ arbitrary values, and $\boldsymbol{R}_+ = \sigma^2_{\boldsymbol{\epsilon}}\tilde{\boldsymbol{R}}_+$ with $\tilde{\boldsymbol{R}}_+ = \tilde{\boldsymbol{R}}_{\boldsymbol{x}_+} \otimes \tilde{\boldsymbol{R}}_{\boldsymbol{z}}$, with $\tilde{\boldsymbol{R}}_{\boldsymbol{x}_+}$ and $\tilde{\boldsymbol{R}}_{\boldsymbol{z}}$ diagonal matrices of dimensions $n_{\boldsymbol{x}_+} \times n_{\boldsymbol{x}_+}$ and $n_{\boldsymbol{z}} \times n_{\boldsymbol{z}}$, respectively, with infinity entries if the data is to be predicted and 1 if the data is observed, notice that since we are not extending the variable $\boldsymbol{z}$, $\tilde{\boldsymbol{R}}_{\boldsymbol{z}}$ is an identity matrix. In this case, the extended basis is:

$$\boldsymbol{B}_+ = \boldsymbol{B}_{\boldsymbol{x}_+} \otimes \boldsymbol{B}_{\boldsymbol{z}} = \begin{bmatrix} \boldsymbol{B}_{\boldsymbol{x}} & \boldsymbol{O} \\ \boldsymbol{B}_{\boldsymbol{x}_{(1)}} & \boldsymbol{B}_{\boldsymbol{x}_{(2)}} \end{bmatrix} \otimes \boldsymbol{B}_{\boldsymbol{z}} = \begin{bmatrix} \boldsymbol{B}_{\boldsymbol{x}} \otimes \boldsymbol{B}_{\boldsymbol{z}} & \boldsymbol{O} \\ \boldsymbol{B}_{\boldsymbol{x}_{(1)}} \otimes \boldsymbol{B}_{\boldsymbol{z}} & \boldsymbol{B}_{\boldsymbol{x}_{(2)}} \otimes \boldsymbol{B}_{\boldsymbol{z}} \end{bmatrix}, \tag{19}$$

where $\boldsymbol{B}_{\boldsymbol{x}_+} = \boldsymbol{B}(\boldsymbol{x}_+)$ and $\boldsymbol{B}_{\boldsymbol{z}} = \boldsymbol{B}(\boldsymbol{z})$ are the regression bases with $\boldsymbol{x}_+$ and $\boldsymbol{z}$ the two regressors. The new extended B-spline basis, $\boldsymbol{B}_{\boldsymbol{x}_+}$, is built from a new set of knots that consists of the original knots covering x_i, $i = 1, ..., n_{\boldsymbol{x}}$, and extended to the range of the $n_{\boldsymbol{x}_p}$ values of x_{p_j}, $j = 1, ..., n_{\boldsymbol{x}_p}$. I.e., $\boldsymbol{B}_{\boldsymbol{x}_+}$ is a direct extension of $\boldsymbol{B}_{\boldsymbol{x}}$.

Considering the previous extended model we minimize the function of $\boldsymbol{\theta}_+$ given in (14) with $\boldsymbol{B}_+$ defined in (19) and extended penalty matrix:

$$\boldsymbol{P}_+ = \lambda_{\boldsymbol{z}}\boldsymbol{P}^{\boldsymbol{z}}_+ + \lambda_{\boldsymbol{x}}\boldsymbol{P}^{\boldsymbol{x}_+}_+, \tag{20}$$

where, since only the covariate $\boldsymbol{x}$ is extended, $\boldsymbol{P}_+^{\boldsymbol{z}}$ and $\boldsymbol{P}_+^{\boldsymbol{x}+}$ are:

$$\begin{aligned}\boldsymbol{P}_+^{\boldsymbol{z}} &= \boldsymbol{I}_{c_{\boldsymbol{x}+}} \otimes \boldsymbol{D}_{\boldsymbol{z}}'\boldsymbol{D}_{\boldsymbol{z}} = \begin{bmatrix} \boldsymbol{I}_{c_{\boldsymbol{x}}} & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{I}_{c_{\boldsymbol{x}p}} \end{bmatrix} \otimes \boldsymbol{D}_{\boldsymbol{z}}'\boldsymbol{D}_{\boldsymbol{z}} = \begin{bmatrix} \boldsymbol{I}_{c_{\boldsymbol{x}}} \otimes \boldsymbol{D}_{\boldsymbol{z}}'\boldsymbol{D}_{\boldsymbol{z}} & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{I}_{c_{\boldsymbol{x}p}} \otimes \boldsymbol{D}_{\boldsymbol{z}}'\boldsymbol{D}_{\boldsymbol{z}} \end{bmatrix} \\ &= \begin{bmatrix} \boldsymbol{P}_{+11}^{\boldsymbol{z}} & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{P}_{+22}^{\boldsymbol{z}} \end{bmatrix}, \end{aligned} \tag{21}$$

and $\boldsymbol{P}_+^{\boldsymbol{x}+}$ as in (17), where $\boldsymbol{D}_{\boldsymbol{x}_+}$ and $\boldsymbol{D}_{\boldsymbol{z}}$ are the difference matrices acting on the columns and rows of the matrix formed by the extended vector of coefficients $\boldsymbol{\Theta}_+$:

$$\boldsymbol{\Theta}_+ = \left[\begin{array}{ccccc|ccc} \boldsymbol{\theta}_{11} & \boldsymbol{\theta}_{12} & \cdots & \boldsymbol{\theta}_{1\ c_{\boldsymbol{x}}-1} & \boldsymbol{\theta}_{1c_{\boldsymbol{x}}} & \boldsymbol{\theta}_{1\ c_{\boldsymbol{x}}+1} & \boldsymbol{\theta}_{1\ c_{\boldsymbol{x}}+2} & \cdots \\ \boldsymbol{\theta}_{21} & \boldsymbol{\theta}_{22} & \cdots & \boldsymbol{\theta}_{2\ c_{\boldsymbol{x}}-1} & \boldsymbol{\theta}_{2c_{\boldsymbol{x}}} & \boldsymbol{\theta}_{2\ c_{\boldsymbol{x}}+1} & \boldsymbol{\theta}_{2\ c_{\boldsymbol{x}}+2} & \cdots \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \boldsymbol{\theta}_{c_{\boldsymbol{z}}1} & \boldsymbol{\theta}_{c_{\boldsymbol{z}}2} & \cdots & \boldsymbol{\theta}_{c_{\boldsymbol{z}}\ c_{\boldsymbol{x}}-1} & \boldsymbol{\theta}_{c_{\boldsymbol{z}}c_{\boldsymbol{x}}} & \boldsymbol{\theta}_{c_{\boldsymbol{z}}\ c_{\boldsymbol{x}}+1} & \boldsymbol{\theta}_{c_{\boldsymbol{z}}\ c_{\boldsymbol{x}}+2} & \cdots \end{array}\right].$$

As we have said in the previous section, the methodology depends on the structure of the data. If we consider scattered data and extend just the covariate $\boldsymbol{x}$, both bases have to be extended since they have to have the same number of rows, $\boldsymbol{B}_{\boldsymbol{z}}$ is extended by rows to construct $\boldsymbol{B}_{\boldsymbol{z}}^+$ (built from the same knots that $\boldsymbol{B}_{\boldsymbol{z}}$) and $\boldsymbol{B}_{\boldsymbol{x}_+}$ is extended by columns and rows to cover the range of $\boldsymbol{x}_+$. Therefore, $\boldsymbol{B}_{\boldsymbol{z}}^+$ and $\boldsymbol{B}_{\boldsymbol{x}_+}$ have size $n_+ \times c_{\boldsymbol{z}}$ and $n_+ \times c_{\boldsymbol{x}_+}$. The superscript $(+)$ of $\boldsymbol{B}_{\boldsymbol{z}}^+$ indicates that the basis is extended but the prediction is not outside the range of the observed values of the covariable $\boldsymbol{z}$. In this case $\boldsymbol{R}_+$ is a diagonal matrix with the first n values equal to 1 and the last $n_{\boldsymbol{x}_p}$ values equal to infinity.

Since we extend just one of the two covariates and penalties are based on differences between adjacent coefficients, the method satisfies certain important properties. These properties are an immediate consequence of the following theorems.

Theorem 1. *The coefficients obtained from minimization of (14) with extended basis (19), extended error covariance matrix (18) and extended penalty matrix (20) where* $\boldsymbol{P}_+^{\boldsymbol{z}}$ *and* $\boldsymbol{P}_+^{\boldsymbol{x}+}$ *are (21) and (17), respectively, satisfy the following properties:*

I. The first c, $c = c_{\boldsymbol{z}} \times c_{\boldsymbol{x}}$, *coefficients of* $\hat{\boldsymbol{\theta}}_+$, *are:*

$$\hat{\boldsymbol{\theta}}_{+1,\ldots,c} = \left(\boldsymbol{B}'\boldsymbol{B} + \lambda_{\boldsymbol{x}}\boldsymbol{P}_{+11}^{\boldsymbol{x}+} + \lambda_{\boldsymbol{z}}\boldsymbol{P}_{+11}^{\boldsymbol{z}} - \lambda_{\boldsymbol{x}}^2\boldsymbol{P}_{+12}^{\boldsymbol{x}+}(\lambda_{\boldsymbol{x}}\boldsymbol{P}_{+22}^{\boldsymbol{x}+} + \lambda_{\boldsymbol{z}}\boldsymbol{P}_{+22}^{\boldsymbol{z}})^{-1}\boldsymbol{P}_{+21}^{\boldsymbol{x}+}\right)^{-1}\boldsymbol{B}'\boldsymbol{y}, \tag{22}$$

where $\boldsymbol{P}_{+11}^{\boldsymbol{x}+}$, $\boldsymbol{P}_{+12}^{\boldsymbol{x}+}$, $\boldsymbol{P}_{+21}^{\boldsymbol{x}+}$ *and* $\boldsymbol{P}_{+22}^{\boldsymbol{x}+}$ *defined in (17) and* $\boldsymbol{P}_{+11}^{\boldsymbol{z}}$ *and* $\boldsymbol{P}_{+22}^{\boldsymbol{z}}$ *defined in (21).*

II. The coefficients for the $n_p = n_{\boldsymbol{z}} \times n_{\boldsymbol{x}_p}$ *predicted values are*

$$\hat{\boldsymbol{\theta}}_p = -\left(\frac{\lambda_{\boldsymbol{z}}}{\lambda_{\boldsymbol{x}}}\boldsymbol{P}_{+22}^{\boldsymbol{z}} + \boldsymbol{P}_{+22}^{\boldsymbol{x}+}\right)^{-1}\boldsymbol{P}_{+21}^{\boldsymbol{x}+}\hat{\boldsymbol{\theta}}_{+1,\ldots,c}, \tag{23}$$

where $\boldsymbol{P}_{+22}^{\boldsymbol{x}+}$, $\boldsymbol{P}_{+21}^{\boldsymbol{x}+}$ *defined in (17) and* $\boldsymbol{P}_{+22}^{\boldsymbol{z}}$ *defined in (21).*

Proof. Differenciating (14) with respect to $\boldsymbol{\theta}_+$ leads to

$$\frac{\partial S}{\partial \boldsymbol{\theta}_+} = -2\boldsymbol{B}_+'\tilde{\boldsymbol{R}}_+^{-1}(\boldsymbol{y}_+ - \boldsymbol{B}_+\boldsymbol{\theta}_+) + 2(\lambda_{\boldsymbol{z}}\boldsymbol{P}_+^{\boldsymbol{z}} + \lambda_{\boldsymbol{x}}\boldsymbol{P}_+^{\boldsymbol{x}+}) = 0$$

i.e., the penalized least squares solution is given by:

$$\hat{\boldsymbol{\theta}}_+ = (\boldsymbol{B}'_+\tilde{\boldsymbol{R}}_+^{-1}\boldsymbol{B}_+ + \lambda_{\boldsymbol{z}}\boldsymbol{P}_+^{\boldsymbol{z}} + \lambda_{\boldsymbol{x}}\boldsymbol{P}_+^{\boldsymbol{x}+})^{-1}\boldsymbol{B}'_+\tilde{\boldsymbol{R}}_+^{-1}\boldsymbol{y}_+. \qquad (24)$$

Let us define $\boldsymbol{C} = (\boldsymbol{B}'_+\tilde{\boldsymbol{R}}_+^{-1}\boldsymbol{B}_+ + \lambda_{\boldsymbol{z}}\boldsymbol{P}_+^{\boldsymbol{z}} + \lambda_{\boldsymbol{x}}\boldsymbol{P}_+^{\boldsymbol{x}+})$ and $\boldsymbol{C}^{-1} = \begin{bmatrix}\boldsymbol{C}^{11} & \boldsymbol{C}^{12}\\ \boldsymbol{C}^{21} & \boldsymbol{C}^{22}\end{bmatrix}$, with this notation and since $\tilde{\boldsymbol{R}}_+^{-1} = \tilde{\boldsymbol{R}}_{\boldsymbol{x}_+}^{-1} \otimes \tilde{\boldsymbol{R}}_{\boldsymbol{z}}^{-1} = \text{blockdiag}(\boldsymbol{I}, \boldsymbol{O})$, with $\boldsymbol{I}$ an identity matrix of dimension $n_{\boldsymbol{x}}n_{\boldsymbol{z}} \times n_{\boldsymbol{x}}n_{\boldsymbol{z}}$ and $\boldsymbol{O}$ a null matrix of dimension $n_{\boldsymbol{x}_p}n_{\boldsymbol{z}} \times n_{\boldsymbol{x}_p}n_{\boldsymbol{z}}$, equation (24) can be rewritten as

$$\boldsymbol{\theta}_+ = \boldsymbol{C}^{-1}\boldsymbol{B}'_+\tilde{\boldsymbol{R}}_+^{-1}\boldsymbol{y}_+ = \begin{bmatrix}\boldsymbol{C}^{11}\boldsymbol{B}'\boldsymbol{y}\\ \boldsymbol{C}^{21}\boldsymbol{B}'\boldsymbol{y}\end{bmatrix}. \qquad (25)$$

If $\boldsymbol{C} = \begin{bmatrix}\boldsymbol{C}_{11} & \boldsymbol{C}_{12}\\ \boldsymbol{C}_{21} & \boldsymbol{C}_{22}\end{bmatrix}$, by Theorem 8.5.11 given in Harville (2000) we have that:

$$\boldsymbol{C}^{-1} = \begin{bmatrix}\boldsymbol{K}^{-1} & -\boldsymbol{K}^{-1}\boldsymbol{C}_{12}\boldsymbol{C}_{22}^{-1}\\ -\boldsymbol{C}_{22}^{-1}\boldsymbol{C}_{21}\boldsymbol{K}^{-1} & \boldsymbol{C}_{22}^{-1} + \boldsymbol{C}_{22}^{-1}\boldsymbol{C}_{21}\boldsymbol{K}^{-1}\boldsymbol{C}_{12}\boldsymbol{C}_{22}^{-1}\end{bmatrix},$$

with $\boldsymbol{K} = \boldsymbol{C}_{11} - \boldsymbol{C}_{12}\boldsymbol{C}_{22}^{-1}\boldsymbol{C}_{21}$. Therefore:

$$\boldsymbol{C}^{11} = \boldsymbol{K}^{-1} = \left(\boldsymbol{B}'\boldsymbol{B} + \lambda_{\boldsymbol{x}}\boldsymbol{P}_{+11}^{\boldsymbol{x}+} + \lambda_{\boldsymbol{z}}\boldsymbol{P}_{+11}^{\boldsymbol{z}} - \lambda_{\boldsymbol{x}}^2\boldsymbol{P}_{+12}^{\boldsymbol{x}+}(\lambda_{\boldsymbol{x}}\boldsymbol{P}_{+22}^{\boldsymbol{x}+} + \lambda_{\boldsymbol{z}}\boldsymbol{P}_{+22}^{\boldsymbol{z}})^{-1}\boldsymbol{P}_{+21}^{\boldsymbol{x}+}\right)^{-1}$$

and

$$\begin{aligned}\boldsymbol{C}^{21} &= -\boldsymbol{C}_{22}^{-1}\boldsymbol{C}_{21}\boldsymbol{K}^{-1}\\ &= -(\lambda_{\boldsymbol{x}}\boldsymbol{P}_{+22}^{\boldsymbol{x}+} + \lambda_{\boldsymbol{z}}\boldsymbol{P}_{+22}^{\boldsymbol{z}})^{-1}\lambda_{\boldsymbol{x}}\boldsymbol{P}_{+21}^{\boldsymbol{x}+}\boldsymbol{C}^{11}\end{aligned}$$

and by equation (25) the coefficients for the fit and for the prediction are given by equations (22) and (23), respectively, as we wanted to show.

□

Hence, by the previous theorem, when we predict in two dimensions extending one covariate, the predicted values $\boldsymbol{y}_p$, obtained by using the new coefficients, $\boldsymbol{\theta}_p$; depend on the ratio $\frac{\lambda_{\boldsymbol{z}}}{\lambda_{\boldsymbol{x}}}$, unless $\lambda_{\boldsymbol{x}} = \lambda_{\boldsymbol{z}}$, obviously. Therefore, while in one dimension we have that, once the data are fitted, the smoothing parameter does not play any role in the prediction, we have found that in two dimensions the smoothing parameters in both directions, $\lambda_{\boldsymbol{x}}$ and $\lambda_{\boldsymbol{z}}$, determine the prediction.

Notice that we have proved that the coefficients that give the fit when the fit and the prediction are obtained simultaneously (22) are not the same as the solution we obtain only fitting the data (10), property that is verified when we predict in one dimension (Carballo *et al.* (2017)). Although, in the one dimensional case, the extended penalty is not a direct extension of the penalty used to fit the data, the blocks of the extended penalty are simplified and the fit is maintained. This does not occur in the case of two dimensions, unless the block $\boldsymbol{P}_{+22}^{\boldsymbol{z}}$ in (21) is equal to zero, as we will see in the following corollary.

Corollary 2 (Theorem 1)**.** *If $\boldsymbol{P}_{+22}^{\boldsymbol{z}} = \boldsymbol{O}$ in (21), the solution from the minimizing (14) verifies:*

1. *The fit remains invariant when out-of-sample prediction is carried out.*

2. *Considering the matrix of coefficients that give the fit,* $\hat{\boldsymbol{\Theta}}$*, and the matrix of coefficients that give the prediction,* $\hat{\boldsymbol{\Theta}}_p$*, each row* $j = 1, ..., c_{\boldsymbol{z}}$*, of the additional matrix of coefficients is a linear combination of the last* $q_{\boldsymbol{z}}$ *old coefficients of that row (*$q_{\boldsymbol{z}}$ *is the order of the penalty acting on rows and* $c_{\boldsymbol{z}}$ *the number of rows of* $\boldsymbol{B}_{\boldsymbol{z}}$*).*

In particular, $\boldsymbol{P}^{\boldsymbol{z}}_{+22} = \boldsymbol{O}$ *if* $\boldsymbol{I}_{c_{\boldsymbol{x}p}} = \boldsymbol{O}$*.*

For the particular case of penalty orders two and three, i.e. $q_{\boldsymbol{x}} = q_{\boldsymbol{z}} = 2$ and $q_{\boldsymbol{x}} = q_{\boldsymbol{z}} = 3$, the proof is given in Appendix .1.

As we have proved in the previous corollary, setting $\boldsymbol{I}_{c_{\boldsymbol{x}p}}$ equal to zero we preserve the fit and everything is analogous to the one dimensional case. In the literature, there are some works in which $\boldsymbol{I}_{c_{\boldsymbol{x}p}}$ is considered equal to zero, e.g. Ugarte *et al.* (2012). However, in practice, we do not set $\boldsymbol{I}_{c_{\boldsymbol{x}p}}$ as a null matrix since we are not imposing the penalty correctly. Furthermore, we could not extend it to the case in which we want out-of-sample prediction in both dimensions. We can not set $\boldsymbol{I}_{c_{\boldsymbol{z}p}}$ and $\boldsymbol{I}_{c_{\boldsymbol{x}p}}$ equal to zero in (16) and (17), since the matrix $\boldsymbol{B}'_+\tilde{\boldsymbol{R}}_+^{-1}\boldsymbol{B}_+ + \boldsymbol{P}_+$ would be singular.

In the next section, we propose the use of constraints to maintain the fit when the fit and the prediction are obtained simultaneously, the restrictions can be used when out-of-sample prediction is carried out only in one dimension or in more dimensions.

2.2 Constrained out-of-sample prediction

As we have shown in the previous section, natural extensions of penalty matrices provides changes in the fit. To overcome this problem, and as a possible way to incorporate known information about the prediction we propose to use constrained P-splines. In this section we introduce a methodology that allow us to impose constant and fixed restrictions and to impose restrictions that depend on the observed data.

Our proposal to impose constraints in the prediction is to obtain the solution of the extended models (18) and (12) subject to a set of l linear constraints given by the equation

$$\boldsymbol{C}\boldsymbol{\theta}_+ = \boldsymbol{r},$$

where $\boldsymbol{C}$ is a constraint matrix of dimension $l \times c_+$ acting on all coefficients, and $\boldsymbol{r}$ is the restrictions vector of dimension $l \times 1$. I.e., we have the following restricted extended regression model:

$$\boldsymbol{y}_+ = \boldsymbol{B}_+\boldsymbol{\theta}^*_+ + \boldsymbol{\epsilon}_+, \qquad \boldsymbol{\epsilon}_+ \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{R}_+)$$

subject to $\boldsymbol{C}\boldsymbol{\theta}^*_+ = \boldsymbol{r}$. Depending on whether we are predicting out-of-sample in one or two dimensions we extend $\boldsymbol{y}_+$, $\boldsymbol{B}_+$ and $\boldsymbol{R}_+$ defined as in model (18) or as in model (12). As a clarification on the notation used throughout this document, notice that the superscript $(^*)$ refers to the use of constraints.

Following Greene & Seaks (1991), the Lagrangian formulation of the penalized least squares minimization problem is:

$$\mathcal{L}(\boldsymbol{\theta}^*_+, \boldsymbol{\omega}) = (\boldsymbol{y}_+ - \boldsymbol{B}_+\boldsymbol{\theta}^*_+)'\tilde{\boldsymbol{R}}_+^{-1}(\boldsymbol{y}_+ - \boldsymbol{B}_+\boldsymbol{\theta}^*_+) + \boldsymbol{\theta}^{*'}_+\boldsymbol{P}_+\boldsymbol{\theta}^*_+ + 2\boldsymbol{\omega}'(\boldsymbol{C}\boldsymbol{\theta}^*_+ - \boldsymbol{r}), \qquad (26)$$

where $\tilde{\boldsymbol{R}}_+^{-1}$ is defined as in model (18) or as in model (12) depending on if we extend one or two covariates, and $\boldsymbol{P}_+$ is the extended penalty matrix ((20) or (15)), $\boldsymbol{\theta}^*_+$ denotes the restricted least squares (RLS) estimator and $\boldsymbol{\omega}$ is a $l \times 1$ vector of Lagrange multipliers. Differentiating (26) we find

$$\frac{\partial \mathcal{L}}{\partial \boldsymbol{\theta}_{+}^{*}} = -2\boldsymbol{B}_{+}'\tilde{\boldsymbol{R}}_{+}^{-1}(\boldsymbol{y}_{+} - \boldsymbol{B}_{+}\boldsymbol{\theta}_{+}^{*}) + 2\boldsymbol{P}_{+}\boldsymbol{\theta}_{+}^{*} + 2\boldsymbol{C}'\boldsymbol{\omega}, \tag{27}$$

$$\frac{\partial \mathcal{L}}{\partial \boldsymbol{\omega}} = \boldsymbol{C}\boldsymbol{\theta}_{+}^{*} - \boldsymbol{r}. \tag{28}$$

Writing the system as a partitioned matrix the equation yields

$$\begin{bmatrix} \boldsymbol{B}_{+}'\tilde{\boldsymbol{R}}_{+}^{-1}\boldsymbol{B}_{+} + \boldsymbol{P}_{+} & \boldsymbol{C}' \\ \boldsymbol{C} & \boldsymbol{O} \end{bmatrix} \begin{bmatrix} \hat{\boldsymbol{\theta}}_{+}^{*} \\ \hat{\boldsymbol{\omega}} \end{bmatrix} = \begin{bmatrix} \boldsymbol{B}_{+}'\tilde{\boldsymbol{R}}_{+}^{-1}\boldsymbol{y}_{+} \\ \boldsymbol{r} \end{bmatrix}.$$

$\hat{\boldsymbol{\theta}}_{+}^{*}$ and $\hat{\boldsymbol{\omega}}$ can be obtained by solving the previous system or, alternative, by following the steps below.

Setting (27) to 0, $-\boldsymbol{B}_{+}'\tilde{\boldsymbol{R}}_{+}^{-1}\boldsymbol{y}_{+} + \boldsymbol{B}_{+}'\tilde{\boldsymbol{R}}_{+}^{-1}\boldsymbol{B}_{+}\boldsymbol{\theta}_{+}^{*} + \boldsymbol{P}_{+}\boldsymbol{\theta}_{+}^{*} + \boldsymbol{C}'\boldsymbol{\omega} = 0$, therefore:

$$\begin{aligned} \hat{\boldsymbol{\theta}}_{+}^{*} &= (\boldsymbol{B}_{+}'\tilde{\boldsymbol{R}}_{+}^{-1}\boldsymbol{B}_{+} + \boldsymbol{P}_{+})^{-1}(\boldsymbol{B}_{+}'\tilde{\boldsymbol{R}}_{+}^{-1}\boldsymbol{y}_{+} - \boldsymbol{C}'\boldsymbol{\omega}) \\ &= \hat{\boldsymbol{\theta}}_{+} - (\boldsymbol{B}_{+}'\tilde{\boldsymbol{R}}_{+}^{-1}\boldsymbol{B}_{+} + \boldsymbol{P}_{+})^{-1}\boldsymbol{C}'\hat{\boldsymbol{\omega}}, \end{aligned} \tag{29}$$

where $\hat{\boldsymbol{\theta}}_{+} = (\boldsymbol{B}_{+}'\tilde{\boldsymbol{R}}_{+}^{-1}\boldsymbol{B}_{+} + \boldsymbol{P}_{+})^{-1}\boldsymbol{B}_{+}'\tilde{\boldsymbol{R}}_{+}^{-1}\boldsymbol{y}_{+}$ is the unrestricted penalized least squares estimator.

Since $\boldsymbol{C}\boldsymbol{\theta}_{+}^{*} = \boldsymbol{r}$, multiplying equation (29) by $\boldsymbol{C}$, we have that $\boldsymbol{C}\boldsymbol{\theta}_{+} - \boldsymbol{C}(\boldsymbol{B}_{+}'\tilde{\boldsymbol{R}}_{+}^{-1}\boldsymbol{B}_{+} + \boldsymbol{P}_{+})^{-1}\boldsymbol{C}'\boldsymbol{\omega} = \boldsymbol{r}$, i.e.

$$\hat{\boldsymbol{\omega}} = [\boldsymbol{C}(\boldsymbol{B}_{+}'\tilde{\boldsymbol{R}}_{+}^{-1}\boldsymbol{B}_{+} + \boldsymbol{P}_{+})^{-1}\boldsymbol{C}']^{-1}(\boldsymbol{C}\boldsymbol{\theta}_{+} - \boldsymbol{r}). \tag{30}$$

Therefore, the coefficients subject to the restriction, $\hat{\boldsymbol{\theta}}_{+}^{*}$, are obtained by computing the vector of Lagrange multipliers (30) and substituting in (29), i.e. $\hat{\boldsymbol{\theta}}_{+}^{*}$ is the unconstrained solution, $\hat{\boldsymbol{\theta}}_{+}$, plus a multiple of the discrepancy vector.

The constrained fitted and predicted values are

$$\hat{\boldsymbol{y}}_{+}^{*} = \boldsymbol{B}_{+}\hat{\boldsymbol{\theta}}_{+}^{*},$$

defining the matrices $\boldsymbol{A_2} = (\boldsymbol{B}_{+}'\tilde{\boldsymbol{R}}_{+}^{-1}\boldsymbol{B}_{+} + \boldsymbol{P}_{+})^{-1}\boldsymbol{C}'[\boldsymbol{C}(\boldsymbol{B}_{+}'\tilde{\boldsymbol{R}}_{+}^{-1}\boldsymbol{B}_{+} + \boldsymbol{P}_{+})^{-1}\boldsymbol{C}']^{-1}$ and $\boldsymbol{A_1} = (\boldsymbol{B}_{+}'\tilde{\boldsymbol{R}}_{+}^{-1}\boldsymbol{B}_{+} + \boldsymbol{P}_{+})^{-1}\boldsymbol{B}_{+}'$, $\hat{\boldsymbol{y}}_{+}^{*}$ can be written as:

$$\hat{\boldsymbol{y}}_{+}^{*} = \boldsymbol{B}_{+}(\boldsymbol{A_1}\tilde{\boldsymbol{R}}_{+}^{-1}\boldsymbol{y}_{+} - \boldsymbol{A_2}\boldsymbol{C}\boldsymbol{A_1}\tilde{\boldsymbol{R}}_{+}^{-1}\boldsymbol{y}_{+} + \boldsymbol{A_2}\boldsymbol{r}).$$

The variance of $\boldsymbol{y}_{+}^{*}$ depends on the following set of restrictions:

a) If the restrictions are constant and fixed, i.e. $\boldsymbol{r}$ is constant and does not depend on the data, the variance is:

$$\text{Var}[\hat{\boldsymbol{y}}_{+}^{*}] = \sigma_{\epsilon}^{2}\boldsymbol{B}_{+}\boldsymbol{A_3}\tilde{\boldsymbol{R}}_{+}^{-1}\boldsymbol{A_3}'\boldsymbol{B}_{+}',$$

with $\boldsymbol{A_3} = \boldsymbol{A_1} - \boldsymbol{A_2}\boldsymbol{C}\boldsymbol{A_1}$.

b) If the restrictions depend on the data, we have to take into account the variability of $\boldsymbol{r}$. For instance, if the restriction is the fit has to be maintained, $\boldsymbol{r} = \hat{\boldsymbol{\theta}} = (\boldsymbol{B}'\boldsymbol{B}+\boldsymbol{P})^{-1}\boldsymbol{B}'\boldsymbol{y}$, therefore the variance is:

$$\mathrm{Var}[\hat{\boldsymbol{y}}_+^*] = \sigma_\epsilon^2 \boldsymbol{B}_+\boldsymbol{A_4}\tilde{\boldsymbol{R}}_+^{-1}\boldsymbol{A_4}'\boldsymbol{B}_+',$$

with $\boldsymbol{A_4} = \boldsymbol{A_1} - \boldsymbol{A_2}\boldsymbol{C}\boldsymbol{A_1} + \boldsymbol{A_2}\begin{bmatrix}(\boldsymbol{B}'\boldsymbol{B}+\boldsymbol{P})^{-1} \\ \boldsymbol{O}\end{bmatrix}$, with $\boldsymbol{O}$ a null matrix of dimension $c_p \times n_p$, $c_p = c_{\boldsymbol{z}}c_{\boldsymbol{x}_p} + c_{\boldsymbol{z}_p}c_{\boldsymbol{x}} + c_{\boldsymbol{z}_p}c_{\boldsymbol{x}_p}$ the number of new coefficients, and n_p the number of new observations.

Let us explain how the restriction on the fit can be imposed in practice. Suppose that we just carry out out-of-sample prediction in one of the two covariates, that the coefficients matrix from the fit has dimension 4×3, and that the coefficients matrix that gives the fit and the forecast has dimension 4×5 i.e.,

$$\hat{\boldsymbol{\Theta}} = \begin{bmatrix} \hat{\theta}_1 & \hat{\theta}_5 & \hat{\theta}_9 \\ \hat{\theta}_2 & \hat{\theta}_6 & \hat{\theta}_{10} \\ \hat{\theta}_3 & \hat{\theta}_7 & \hat{\theta}_{11} \\ \hat{\theta}_4 & \hat{\theta}_8 & \hat{\theta}_{12} \end{bmatrix}, \qquad \boldsymbol{\Theta}_+^* = \begin{bmatrix} \theta_1 & \theta_5 & \theta_9 & \theta_{13} & \theta_{17} \\ \theta_2 & \theta_6 & \theta_{10} & \theta_{14} & \theta_{18} \\ \theta_3 & \theta_7 & \theta_{11} & \theta_{15} & \theta_{19} \\ \theta_4 & \theta_8 & \theta_{12} & \theta_{16} & \theta_{20} \end{bmatrix},$$

where in $\boldsymbol{\Theta}_+^*$ the coefficients that determine the fit are in red and the coefficients that determine the forecast in blue. If we impose the restriction the fit has to be maintained, we define the restriction equation

$$\boldsymbol{C}\boldsymbol{\theta}_+^* = \boldsymbol{r},$$

where $\boldsymbol{\theta}_+^* = \mathrm{Vec}(\boldsymbol{\Theta}_+^*)$, $\boldsymbol{C} = [\boldsymbol{I}_{12\times 12} \mid \boldsymbol{O}_{12\times 8}]$ ($\boldsymbol{I}_{12\times 12}$ an identity matrix of dimension 12 and $\boldsymbol{O}_{12\times 8}$ a zero matrix of dimension 12×8) and $\boldsymbol{r} = \hat{\boldsymbol{\theta}} = \mathrm{vec}(\hat{\boldsymbol{\Theta}})$.

On the other hand, if we extend the two covariates and the coefficients matrices for the fit and for the fit and the prediction are, respectively:

$$\hat{\boldsymbol{\Theta}} = \begin{bmatrix} \hat{\theta}_1 & \hat{\theta}_5 & \hat{\theta}_9 \\ \hat{\theta}_2 & \hat{\theta}_6 & \hat{\theta}_{10} \\ \hat{\theta}_3 & \hat{\theta}_7 & \hat{\theta}_{11} \\ \hat{\theta}_4 & \hat{\theta}_8 & \hat{\theta}_{12} \end{bmatrix}, \qquad \boldsymbol{\Theta}_+^* = \begin{bmatrix} \theta_1 & \theta_7 & \theta_{13} & \theta_{19} & \theta_{25} \\ \theta_2 & \theta_8 & \theta_{14} & \theta_{20} & \theta_{26} \\ \theta_3 & \theta_9 & \theta_{15} & \theta_{21} & \theta_{27} \\ \theta_4 & \theta_{10} & \theta_{16} & \theta_{22} & \theta_{28} \\ \theta_5 & \theta_{11} & \theta_{17} & \theta_{23} & \theta_{29} \\ \theta_6 & \theta_{12} & \theta_{18} & \theta_{24} & \theta_{30} \end{bmatrix},$$

i.e., $c_{\boldsymbol{z}} = 4$, $c_{\boldsymbol{z}_p} = 2$, $c_{\boldsymbol{x}} = 3$ and $c_{\boldsymbol{x}_p} = 2$. To impose the restriction the fit has to be maintained, we define the restriction equation

$$\boldsymbol{C}\boldsymbol{\theta}_+^* = \boldsymbol{r},$$

where $\boldsymbol{\theta}_+^* = \mathrm{Vec}(\boldsymbol{\Theta}_+^*)$, $\boldsymbol{C} = \mathrm{blockdiag}(\boldsymbol{I}_{4\times 4}, [\boldsymbol{O}_{2\times 2} \mid \boldsymbol{I}_{4\times 4}], [\boldsymbol{O}_{2\times 2} \mid \boldsymbol{I}_{4\times 4}])$ ($\boldsymbol{I}_{4\times 4}$ is an identity matrix of dimension 4 and $\boldsymbol{O}_{2\times 2}$ is a zero matrix of dimension 2) and $\boldsymbol{r} = \hat{\boldsymbol{\theta}} = \mathrm{vec}(\hat{\boldsymbol{\Theta}})$.

In general, regardless of the number of variables that we extend whenever $c_{\boldsymbol{x}_p} \geq c_{\boldsymbol{z}_p}$, if the restriction is the fit has to be maintained, $\boldsymbol{C}$ is a block diagonal matrix with the first block an identity matrix of dimension $c_{\boldsymbol{z}} \times c_{\boldsymbol{z}}$ and $c_{\boldsymbol{x}_p}$ blocks equal to $[\boldsymbol{O}_{c_{\boldsymbol{z}_p} \times c_{\boldsymbol{z}_p}} \mid \boldsymbol{I}_{c_{\boldsymbol{z}} \times c_{\boldsymbol{z}}}]$, i.e.

$$\boldsymbol{C} = \begin{bmatrix} \boldsymbol{I}_{c_{\boldsymbol{z}} \times c_{\boldsymbol{z}}} & & & & \\ & [\boldsymbol{O}_{c_{\boldsymbol{z}_p} \times c_{\boldsymbol{z}_p}} \mid \boldsymbol{I}_{c_{\boldsymbol{z}} \times c_{\boldsymbol{z}}}] & & & \\ & & [\boldsymbol{O}_{c_{\boldsymbol{z}_p} \times c_{\boldsymbol{z}_p}} \mid \boldsymbol{I}_{c_{\boldsymbol{z}} \times c_{\boldsymbol{z}}}] & & \\ & & & [\boldsymbol{O}_{c_{\boldsymbol{z}_p} \times c_{\boldsymbol{z}_p}} \mid \boldsymbol{I}_{c_{\boldsymbol{z}} \times c_{\boldsymbol{z}}}] & \\ & & & & \ddots \end{bmatrix},$$

and $\boldsymbol{r} = \text{vec}(\hat{\boldsymbol{\Theta}})$.

2.3 Example 1: Prediction of mortality data

Although mortality data are often analyzed through a Poisson distribution, for the simple purpose of illustrating the proposed methodology we use a data set on the log mortality rates of US male population considering the log mortality rates as normal data. We use data from the Human Mortality Database (2018), from ages 0 to 110+ over the period 1960-2014, forecasting up to 2050, i.e., we carry out out-of-sample prediction in one of the two covariates, the years.

The Lee Carter method (Lee & Carter (1992)) is one of the most common methods used for estimating and forecasting mortality data, however it has been observed crossover (higher mortality rates for younger ages than for older ages), to avoid this problem Delwarde *et al.* (2007) made a major improvement. In order to compare our method with the solution given by Delwarde *et al.* (2007) we have obtained the fit and the forecast through four different models:

a) Model 1, unrestricted model:

$$\boldsymbol{y}_+ = f(\boldsymbol{age}, \boldsymbol{year}_+) + \boldsymbol{\epsilon}_+, \quad \boldsymbol{\epsilon}_+ \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{R}_+)$$

where $\boldsymbol{y}_+$ is the log mortality rate and $\boldsymbol{R}_+$ is defined as in (18).

b) Model 2: The model defined in a) subject to the restriction the fit is maintained.

c) Model 3: The model defined in a) subject to two restrictions:

- The fit is maintained.
- The structure across ages is preserved. We impose this restriction avoid crossover for ages, i.e. to avoid higher mortality rates for younger ages than for older ages. To do this, we take the coefficients pattern at the last years and we project it. In order to do that we impose that the difference between the coefficients of every two consecutive projections has to be constant and equal to the difference between the corresponding last coefficients from the fit.

  Let us explain with an example how this two restrictions can be imposed at the same time, suppose that the coefficients matrix from the fit has dimension 4×3, and that the coefficients matrix that gives the fit and the forecast has dimension 4×5 i.e.,

$$\hat{\boldsymbol{\Theta}} = \begin{bmatrix} \hat{\theta}_1 & \hat{\theta}_5 & \hat{\theta}_9 \\ \hat{\theta}_2 & \hat{\theta}_6 & \hat{\theta}_{10} \\ \hat{\theta}_3 & \hat{\theta}_7 & \hat{\theta}_{11} \\ \hat{\theta}_4 & \hat{\theta}_8 & \hat{\theta}_{12} \end{bmatrix}, \qquad \boldsymbol{\Theta}_+^* = \begin{bmatrix} \theta_1 & \theta_5 & \theta_9 & \theta_{13} & \theta_{17} \\ \theta_2 & \theta_6 & \theta_{10} & \theta_{14} & \theta_{18} \\ \theta_3 & \theta_7 & \theta_{11} & \theta_{15} & \theta_{19} \\ \theta_4 & \theta_8 & \theta_{12} & \theta_{16} & \theta_{20} \end{bmatrix},$$

in $\boldsymbol{\Theta}_{+}^{*}$ the coefficients that determine the fit are in red and the coefficients that determine the forecast in blue. The restriction equation is

$$\boldsymbol{C}\boldsymbol{\theta}_{+}^{*} = \boldsymbol{r},$$

where $\boldsymbol{\theta}_{+}^{*} = \text{Vec}(\boldsymbol{\Theta}_{+}^{*})$, $\boldsymbol{C} = \text{blockdiag}(\boldsymbol{I}_{12\times 12}, \boldsymbol{U}, \boldsymbol{U})$ with $\boldsymbol{U} = \begin{bmatrix} 1 & -1 & 0 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 1 & -1 \end{bmatrix}$

and $\boldsymbol{r} = \begin{bmatrix} \hat{\boldsymbol{\theta}} \\ \hat{\theta}_9 - \hat{\theta}_{10} \\ \hat{\theta}_{10} - \hat{\theta}_{11} \\ \hat{\theta}_{11} - \hat{\theta}_{12} \\ \hat{\theta}_9 - \hat{\theta}_{10} \\ \hat{\theta}_{10} - \hat{\theta}_{11} \\ \hat{\theta}_{11} - \hat{\theta}_{12} \end{bmatrix}$ with $\hat{\boldsymbol{\theta}} = \text{Vec}(\hat{\boldsymbol{\Theta}})$,

i.e., we are imposing that the coefficients that determine the fit have to be the ones we obtain when only fitting the data, and that the difference between the coefficients that determine the forecast of two consecutive rows has to be equal to the difference between the last coefficients from the fit of those rows.

d) Model 4, is the one given in Delwarde *et al.* (2007), where the model proposed in Lee & Carter (1992) is modified to get regular projected life tables.
The original Lee & Carter (1992) model is:

$$log(m_{x,y}) = \alpha_x + \beta_x k_y + \epsilon_y$$

where $m_{x,y}$ is the central rate of morality at age x in year y and α_x, β_x and k_y are parameters to be estimated, and ϵ_y is the error term with mean zero and variance $\sigma_{\boldsymbol{\epsilon}}^2$. This model is fitted to historical data and the resulting estimated k_t's are then modeled and projected as a stochastic time series using standard Box-Jenkins methods. Delwarde *et al.* (2007) have improved the Lee-Carter model smoothing through penalized splines the estimated β_x's.

Figure 1 shows the fit and the forecast obtained with model 1 (left panel above), model 2 (right panel above), model 3 (left panel below) and model 4 (right panel below).

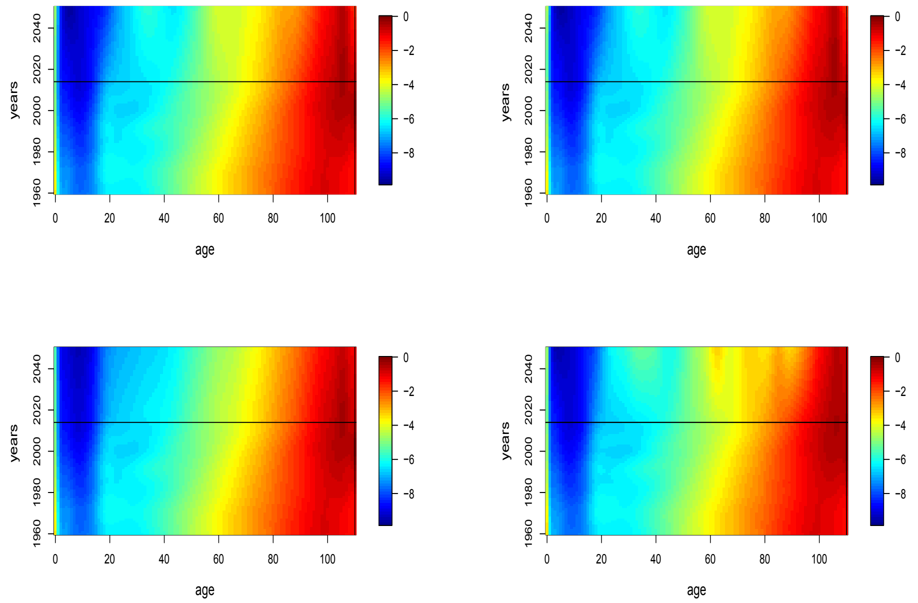


Figure 1: Fit and forecast of a data set on the log mortality rates of US males aged 0-110+ over the period 1960-2014, through model 1 (red line), model 2 (green line), model 3 (blue line) and model 4 (orange line). The vertical line indicates the year from which we predict.

In order to illustrate the differences between the fits, we have selected five ages: 20, 40, 60, 80 and 100, in Figure 2 we show the fit and the forecast for those ages obtained through model 1 (red line), model 2 (green line), model 3 (blue line) and model 4 (orange line). The fit provided by the first three models (models 1, 2 and 3) is almost the same and better than the fit provided by model 4. However, the fit given by model 4 is quite different and worse than the others for ages 40 and 100.
The predictions with model 1 and 2 are almost identical (in Figure 2 we can hardly appreciate the green line because it is below the green line). Despite giving very similar results in the fit, model 3 provides quite different results in the forecast, for age 60 model 1 and model 2 provide an increase in the log mortality rate for the period $2020 - 2050$ since they are forecasting the incrementing trend in mortality between $2010 - 2016$, while the most realistic results are given by models 3 and 4, in which the mortality decreases. The predictions given by model 4 are coherent and similar to those of model 3 for ages 20, 40 and 80. However, model 4 provides an increase in the log mortality rate for age 100, what seems to be inconsistent with the observed log mortality ratio.

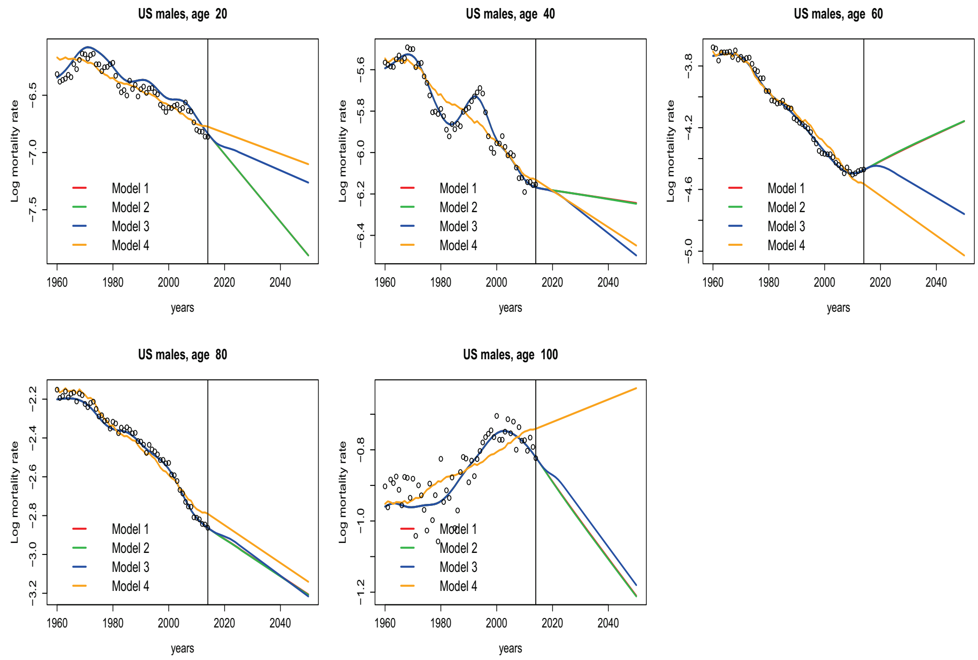


Figure 2: Fit and forecast of selected ages: 20, 40, 60, 80 and 100 obtained through model 1 (red line), model 2 (green line), model 3 (blue line) and model 4 (orange line). The vertical line indicates the year from which we predict (2014).

As we have seen, the most realistic results are provided when we impose two restrictions: the fit is maintained and the structure across ages is preserved, i.e. for model 3. If we do not maintain the coefficients pattern, crossover for ages can happen. We illustrate this fact in Figure 3, where we plot the obtained projections with model 2 (green line) and model 3 (blue line) for ages 46 and 47. We can see that the fit is quite similar for the three models, the log mortality rates for age 46 are lower than for age 47 in the range of known data. However, in the forecast, for model 2 crossover for ages 46 and 47 occurs and in 2050 the log mortality rate is larger for age 46 than for age 47. This does not occur for model 3, in which case the imposed restriction preserves the structure across ages. The crossover problem has been also solved in Delwarde *et al.* (2007) and in Currie (2013), they do not impose restrictions to prevent the crossover but smoothing the β_x's and the β_x's and α_x's of the Lee-Carter model, respectively, they avoid the problem.

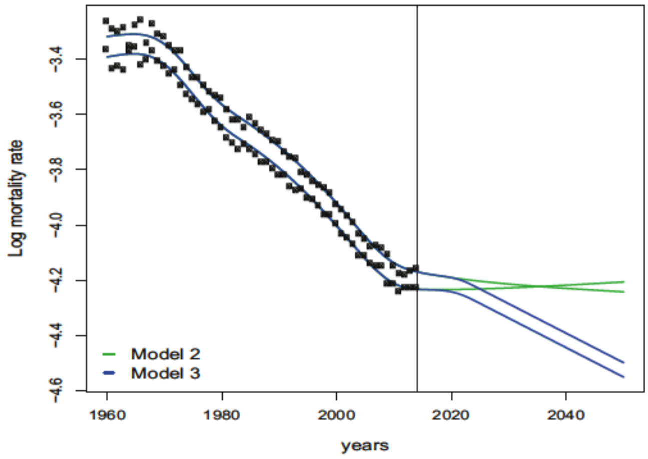


Figure 3: Fit and forecast for ages 66 and 67 obtained through model 2 (green line) and model 3 (blue line). Model 3 prevent crossover for ages. The vertical line indicates the year from which we predict (2014).

3 Out-of-sample prediction in multidimensional smooth mixed models

The connection between penalized smoothing and mixed models was established thirty years ago in Green (1987) (see also Currie & Durbán (2002) and Wand (2003)). The key point of this equivalence is the fact that the smoothing parameter becomes a ratio of variances and variance components can be estimated through restricted maximum likelihood procedure (REML) (see Patterson & Thompson (1971)). The interest on this representation is due to the possibility of including smoothing in a large class of models and the use of the methodology and software already developed for mixed models for estimation and inference. We here exploit the link to mixed models to extend the results of the previous section.

In order to reparameterize a penalized smooth model as a mixed model it is necessary to find a new basis that allows the representation of model (1) as a mixed model. We replace the smooth function by a basis representation which is now written as $\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{Z\alpha} + \boldsymbol{\epsilon}$ where coefficients $\boldsymbol{\alpha}$ are penalized to achieve smoothness. This leads to a mixed model of the form

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{Z\alpha} + \boldsymbol{\epsilon}, \quad \text{with} \quad \boldsymbol{\alpha} \sim \mathcal{N}(0, \boldsymbol{G}) \quad \text{and} \quad \boldsymbol{\epsilon} \sim \mathcal{N}(0, \sigma_\epsilon^2 \boldsymbol{R}), \tag{31}$$

where $\boldsymbol{R} = \boldsymbol{I}$, i.e. the errors are independent and identically distributed with variance σ_ϵ^2, and $\boldsymbol{X}$ and $\boldsymbol{Z}$ are the model matrices and $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$ are the fixed and random effects coefficients respectively. The random effects have the covariance matrix $\boldsymbol{G}$, which depends on the variances of the random effects.

There are different alternatives for the reparameterization of the original smooth model as a mixed model in (31). The idea is to find a transformation $\boldsymbol{\Omega}$ such that:

$$\boldsymbol{B\Omega} = \begin{bmatrix}\boldsymbol{X} \mid \boldsymbol{Z}\end{bmatrix} \text{ and } \boldsymbol{\Omega}'\boldsymbol{\theta} = \begin{bmatrix}\boldsymbol{\beta} \\ \boldsymbol{\alpha}\end{bmatrix} \text{ to have } \boldsymbol{B\theta} = \boldsymbol{X\beta} + \boldsymbol{Z\alpha}.$$

For that, we follow Lee (2010) and consider the SVD of the marginal difference matrices $\mathbf{D}_x'\mathbf{D}_x$ and $\mathbf{D}_z'\mathbf{D}_z$:

$$\mathbf{D}_i'\mathbf{D}_i = \mathbf{U}_i\boldsymbol{\Sigma}_i\mathbf{U}_i',$$

where $\mathbf{\Sigma}_i$ is a diagonal matrix that contains the eigenvalues of the SVD of $\mathbf{D}_i'\mathbf{D}_i$ and $\mathbf{U}_i$ is the corresponding matrix of eigenvectors, for $i = \boldsymbol{z}, \boldsymbol{x}$. The matrix $\mathbf{U}_i$ can be split to have $\mathbf{U}_i = [\mathbf{U}_{if} \mid \mathbf{U}_{ir}]$, where $\mathbf{U}_{if}$ and $\mathbf{U}_{ir}$ are matrices containing the eigenvectors associated to the null and non-null parts, respectively, and $\tilde{\mathbf{\Sigma}}_i$ has $(c_i - q_i)$ positive eigenvalues, for $i = \boldsymbol{z}, \boldsymbol{x}$. We consider the transformation matrix $\mathbf{\Omega}$ as:

$$\mathbf{\Omega} = [\underbrace{\mathbf{U}_{\boldsymbol{x}f} \otimes \mathbf{U}_{\boldsymbol{z}f}}_{\mathbf{\Omega}_f} \mid \underbrace{\mathbf{U}_{\boldsymbol{x}r} \otimes \mathbf{U}_{\boldsymbol{z}r} \mid \mathbf{U}_{\boldsymbol{x}f} \otimes \mathbf{U}_{\boldsymbol{z}r} \mid \mathbf{U}_{\boldsymbol{x}r} \otimes \mathbf{U}_{\boldsymbol{z}r}}_{\mathbf{\Omega}_r}], \tag{32}$$

which is obtained by reordering the block matrices of the matrix $[\mathbf{U}_{\boldsymbol{x}f} \mid \mathbf{U}_{\boldsymbol{x}r}] \otimes [\mathbf{U}_{\boldsymbol{z}f} \mid \mathbf{U}_{\boldsymbol{z}r}]$. Then, given the transformation matrix in (32), the mixed model matrices are:

$$\mathbf{X} = \mathbf{B}\mathbf{\Omega}_f = (\mathbf{B}_{\boldsymbol{x}} \otimes \mathbf{B}_{\boldsymbol{z}})(\mathbf{U}_{\boldsymbol{x}f} \otimes \mathbf{U}_{\boldsymbol{z}f}) = \mathbf{B}_{\boldsymbol{x}}\mathbf{U}_{\boldsymbol{x}f} \otimes \mathbf{B}_{\boldsymbol{z}}\mathbf{U}_{\boldsymbol{z}f},$$

and

$$\mathbf{Z} = \mathbf{B}\mathbf{\Omega}_r = [\mathbf{B}_{\boldsymbol{x}}\mathbf{U}_{\boldsymbol{x}r} \otimes \mathbf{B}_{\boldsymbol{z}}\mathbf{U}_{\boldsymbol{z}f} \mid \mathbf{B}_{\boldsymbol{x}}\mathbf{U}_{\boldsymbol{x}f} \otimes \mathbf{B}_{\boldsymbol{z}}\mathbf{U}_{\boldsymbol{z}r} \mid \mathbf{B}_{\boldsymbol{x}}\mathbf{U}_{\boldsymbol{x}r} \otimes \mathbf{B}_{\boldsymbol{z}}\mathbf{U}_{\boldsymbol{z}r}].$$

Denoting the matrices $\mathbf{X}_i = \mathbf{B}_i\mathbf{U}_{if}$ and $\mathbf{Z}_i = \mathbf{B}_i\mathbf{U}_{ir}$ $(i = \boldsymbol{z}, \boldsymbol{x})$, the previous mixed model matrices can be expressed as:

$$\begin{aligned} \mathbf{X} &= \mathbf{X}_{\boldsymbol{x}} \otimes \mathbf{X}_{\boldsymbol{z}}, \\ \mathbf{Z} &= [\mathbf{Z}_{\boldsymbol{x}} \otimes \mathbf{X}_{\boldsymbol{z}} \mid \mathbf{X}_{\boldsymbol{x}} \otimes \mathbf{Z}_{\boldsymbol{z}} \mid \mathbf{Z}_{\boldsymbol{x}} \otimes \mathbf{Z}_{\boldsymbol{z}}]. \end{aligned} \tag{33}$$

Moreover, given the transformation matrix $\mathbf{\Omega}$ in (32) and the two-dimensional penalty matrix defined in (8), the mixed model penalty matrix is:

$$\mathbf{\Omega}'\boldsymbol{P}\mathbf{\Omega} = \begin{bmatrix} \boldsymbol{O}_q & \\ & \boldsymbol{F} \end{bmatrix}, \text{ with } \mathbf{F} = \begin{bmatrix} \lambda_{\boldsymbol{x}}\tilde{\mathbf{\Sigma}}_{\boldsymbol{x}} \otimes \mathbf{I}_{q_{\boldsymbol{z}}} & & \\ & \lambda_{\boldsymbol{z}}\mathbf{I}_{q_{\boldsymbol{x}}} \otimes \tilde{\mathbf{\Sigma}}_{\boldsymbol{z}} & \\ & & \lambda_{\boldsymbol{x}}\tilde{\mathbf{\Sigma}}_{\boldsymbol{x}} \otimes \mathbf{I}_{c_{\boldsymbol{z}}-q_{\boldsymbol{z}}} + \lambda_{\boldsymbol{x}}\mathbf{I}_{c_{\boldsymbol{x}}-q_{\boldsymbol{x}}} \otimes \tilde{\mathbf{\Sigma}}_{\boldsymbol{z}} \end{bmatrix}, \tag{34}$$

where $q = q_{\boldsymbol{z}}q_{\boldsymbol{x}}$ and the matrices $\tilde{\mathbf{\Sigma}}_i$ $(i = \boldsymbol{z}, \boldsymbol{x})$ diagonal matrices containing the positive eigenvalues of $\boldsymbol{D}_i'\boldsymbol{D}_i$. Then, the covariance matrix $\boldsymbol{G}$ associated to the random effects can be written as:

$$\boldsymbol{G} = \sigma_{\boldsymbol{\epsilon}}^2\boldsymbol{F}^{-1}. \tag{35}$$

The estimation of model in (31) is done using the mixed model equations of Henderson (1975), the solution is:

$$\begin{aligned} \hat{\boldsymbol{\beta}} &= (\boldsymbol{X}'\hat{\boldsymbol{V}}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\hat{\boldsymbol{V}}^{-1}\boldsymbol{y}, \\ \hat{\boldsymbol{\alpha}} &= \hat{\boldsymbol{G}}\boldsymbol{Z}'\hat{\boldsymbol{V}}^{-1}(\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}), \end{aligned}$$

where $\hat{\boldsymbol{V}} = \boldsymbol{Z}\hat{\boldsymbol{G}}\boldsymbol{Z}' + \hat{\sigma}_{\boldsymbol{\epsilon}}^2\boldsymbol{I}$, the estimation of the covariance parameters can be carried out by the Restricted Maximum Likelihood (REML) approach proposed by Patterson & Thompson (1971).

Once we have set the general framework, we extend the results presented in Section 2.1 to the multidimensional mixed model framework. To reformulate the extended model (12) as a mixed model we need to extend the mixed model components to consider the following extended mixed model:

$$\boldsymbol{y}_+ = \boldsymbol{X}_+\boldsymbol{\beta}_{\tilde{+}} + \boldsymbol{Z}_+\boldsymbol{\alpha}_+ + \boldsymbol{\epsilon}_+, \quad \boldsymbol{\epsilon}_+ \sim \mathcal{N}(\mathbf{0}, \boldsymbol{R}_+), \quad \boldsymbol{\alpha}_+ \sim \mathcal{N}(\mathbf{0}, \boldsymbol{G}_+), \tag{36}$$

The subscript of $\boldsymbol{\beta}_{\tilde{+}}$ is $(\tilde{+})$ and is not $(+)$ to indicate that the fixed effects in the extended model (36) are not the same as the fixed effects in the original model (31), however both fixed effects have the same dimension. The variance matrix of the error, $\boldsymbol{R}_+$ is defined as in (12).

Once we have the extended model matrices, $\boldsymbol{X}_+$ and $\boldsymbol{Z}_+$ and the extended covariance matrix $\boldsymbol{G}_+$, the fit and the forecast are obtained solving the extended mixed model equations of Henderson (1975):

$$\begin{bmatrix} \hat{\boldsymbol{\beta}}_{\tilde{+}} \\ \hat{\boldsymbol{\alpha}}_+ \end{bmatrix} = \boldsymbol{L}^{-1} \begin{bmatrix} \boldsymbol{X}'_+ \boldsymbol{R}_+^{-1} \\ \boldsymbol{Z}'_+ \boldsymbol{R}_+^{-1} \end{bmatrix} \boldsymbol{y}_+, \tag{37}$$

where $\boldsymbol{y}_+ = (\boldsymbol{y}', \boldsymbol{y}'_p)'$ is defined as in (18) and matrix $\boldsymbol{L}$ equals

$$\boldsymbol{L} = \begin{bmatrix} \boldsymbol{X}'_+ \boldsymbol{R}_+^{-1} \boldsymbol{X}_+ & \boldsymbol{X}'_+ \boldsymbol{R}_+^{-1} \boldsymbol{Z}_+ \\ \boldsymbol{Z}'_+ \boldsymbol{R}_+^{-1} \boldsymbol{X}_+ & \boldsymbol{Z}'_+ \boldsymbol{R}_+^{-1} \boldsymbol{Z}_+ + \boldsymbol{G}_+^{-1} \end{bmatrix}.$$

Since $\hat{\boldsymbol{y}}_+ = [\boldsymbol{X}_+ \mid \boldsymbol{Z}_+] \begin{bmatrix} \hat{\boldsymbol{\beta}}_{\tilde{+}} \\ \hat{\boldsymbol{\alpha}}_+ \end{bmatrix}$, its variance is:

$$\mathrm{Var}[\hat{\boldsymbol{y}}_+] = [\boldsymbol{X}_+ \mid \boldsymbol{Z}_+] \boldsymbol{L}^{-1} \begin{bmatrix} \boldsymbol{X}'_+ \\ \boldsymbol{Z}'_+ \end{bmatrix}. \tag{38}$$

The variance components can be estimated by maximizing the extended residual maximum log-likelihood (REML):

$$-\frac{1}{2}\log|\boldsymbol{V}_+| - \frac{1}{2}\log|\boldsymbol{X}'_+ \boldsymbol{V}_+^{-1} \boldsymbol{X}_+| - \frac{1}{2}(\boldsymbol{y}_+ - \boldsymbol{X}_+ \boldsymbol{\beta}_{\tilde{+}})' \boldsymbol{V}_+^{-1} (\boldsymbol{y}_+ - \boldsymbol{X}_+ \boldsymbol{\beta}_{\tilde{+}}), \tag{39}$$

where $\boldsymbol{V}_+ = \boldsymbol{R}_+ + \boldsymbol{Z}_+ \boldsymbol{G}_+ \boldsymbol{Z}'_+$.

To obtain the extended mixed model components we need to define an extended transformation matrix. The natural extension of the transformation matrix (32) is to consider the SVD decompositions of the extended difference matrices, i.e. of $\boldsymbol{D}'_{\boldsymbol{z}+} \boldsymbol{D}_{\boldsymbol{z}+}$ and of $\boldsymbol{D}'_{\boldsymbol{x}+} \boldsymbol{D}_{\boldsymbol{x}+}$, but we have to take into account that the extended transformation built from these singular value decompositions does not provide direct extensions of the mixed model matrices from the fit, $\boldsymbol{X}$ and $\boldsymbol{Z}$. This is not a problem, unless we want to impose the restriction that the fit has to be maintained. In this case the fixed effects estimated from the extended model have to be the same as the fixed effects that determine the fit, i.e. $\hat{\boldsymbol{\beta}}_{\tilde{+}} = \hat{\boldsymbol{\beta}}$, and the random effects estimated in the extended model have to be a direct extension of the random effects that determine de fit, i.e. $\hat{\boldsymbol{\alpha}}_+$ has to contain the values of $\hat{\boldsymbol{\alpha}}$, and besides that $\hat{\boldsymbol{\beta}}_{\tilde{+}}$ and $\hat{\boldsymbol{\alpha}}_+$ have to multiplied by model matrices that are direct extensions of the model matrices that determine the fit, $\boldsymbol{X}$ and $\boldsymbol{Z}$.

The natural extended transformation matrix, $\boldsymbol{\Omega}_+$, will not yield the matrices mentioned above, it will return the extended fixed and random effects matrices, $\boldsymbol{X}_+$ and $\boldsymbol{Z}_+$, that are not a direct extension of the model matrices used to obtain the fit, $\boldsymbol{X}$ and $\boldsymbol{Z}$. There are two options that allow us to solve this problem:

- Define the constraint matrix $\boldsymbol{C}$ in the P-spline model framework and reparameterize it to obtain the restrictions matrix for the extended mixed model, $\boldsymbol{C}_{\mathrm{MM}}$. I.e., define $\boldsymbol{C}$ and compute $\boldsymbol{C}_{\mathrm{MM}} = \boldsymbol{C}\boldsymbol{\Omega}_+$.

- Due to identifiability problems the previous proposal can not be used always, as we will see in Section 4. Therefore, we define an extended transformation matrix $\boldsymbol{\Omega}_+^*$ that allow us to obtain extended fixed and random effects matrices that are a direct extension of the mixed model matrices used to obtain the fit.

The first option is straightforward and can be carried out by using any extended transformation matrix $\boldsymbol{\Omega}_+$. However, to implement the second option we define an extended transformation matrix $\boldsymbol{\Omega}_+^*$ that allow us to preserve the model matrices.

We now give the expressions of the extended mixed model components depending on the extended transformation matrix that we use:

- The natural extended transformation $\boldsymbol{\Omega}_+$ based on the SVD of the extended difference matrices.
- An extended transformation matrix $\boldsymbol{\Omega}_+^*$ that preserves the model matrices.

3.0.1 Natural reparameterization of P-splines as mixed models for out-of-sample prediction

The natural extension of the transformation matrix (32) is to consider the SVD decompositions of the extended difference matrices, i.e. $\boldsymbol{D}'_{\boldsymbol{x}_+}\boldsymbol{D}_{\boldsymbol{x}_+} = \boldsymbol{U}_{\boldsymbol{x}_+}\boldsymbol{\Sigma}_{\boldsymbol{x}_+}\boldsymbol{U}'_{\boldsymbol{x}_+}$ and $\boldsymbol{D}'_{\boldsymbol{z}_+}\boldsymbol{D}_{\boldsymbol{z}_+} = \boldsymbol{U}_{\boldsymbol{z}_+}\boldsymbol{\Sigma}_{\boldsymbol{z}_+}\boldsymbol{U}'_{\boldsymbol{z}_+}$, where the matrices $\boldsymbol{U}_i$, for $i = \boldsymbol{z}_+, \boldsymbol{x}_+$, can be splitted in two parts, $\boldsymbol{U}_i = [\boldsymbol{U}_{if} \mid \boldsymbol{U}_{ir}]$, where $\boldsymbol{U}_{if}$ contains the null part (of dimension $c_i \times q_i$) and $\boldsymbol{U}_{ir}$ contains the span or the non-null part of the decomposition (of dimension $c_i \times (c_i - q_i)$), then the extended transformation matrix is:

$$\boldsymbol{\Omega}_+ = [\underbrace{\mathbf{U}_{\boldsymbol{x}+f} \otimes \mathbf{U}_{\boldsymbol{z}f}}_{\boldsymbol{\Omega}_{+f}} \mid \underbrace{\mathbf{U}_{\boldsymbol{x}+r} \otimes \mathbf{U}_{\boldsymbol{z}+r} \mid \mathbf{U}_{\boldsymbol{x}+f} \otimes \mathbf{U}_{\boldsymbol{z}+r} \mid \mathbf{U}_{\boldsymbol{x}+r} \otimes \mathbf{U}_{\boldsymbol{z}+r}}_{\boldsymbol{\Omega}_{+r}}], \tag{40}$$

Then, given the extended transformation matrix in (40), the mixed model matrices for the two-dimensional case are obtained as:

$$\mathbf{X}_+ = \mathbf{B}_+\boldsymbol{\Omega}_{+f} = \mathbf{B}_{\boldsymbol{x}_+}\mathbf{U}_{\boldsymbol{x}+f} \otimes \mathbf{B}_{\boldsymbol{z}_+}\mathbf{U}_{\boldsymbol{z}+f}, \tag{41}$$

and

$$\mathbf{Z}_+ = \mathbf{B}\boldsymbol{\Omega}_{+r} = \left[\mathbf{B}_{\boldsymbol{x}_+}\mathbf{U}_{\boldsymbol{x}+r} \otimes \mathbf{B}_{\boldsymbol{z}_+}\mathbf{U}_{\boldsymbol{z}+f} \mid \mathbf{B}_{\boldsymbol{x}_+}\mathbf{U}_{\boldsymbol{x}+f} \otimes \mathbf{B}_{\boldsymbol{z}_+}\mathbf{U}_{\boldsymbol{z}+r} \mid \mathbf{B}_{\boldsymbol{x}_+}\mathbf{U}_{\boldsymbol{x}+r} \otimes \mathbf{B}_{\boldsymbol{z}_+}\mathbf{U}_{\boldsymbol{z}+r}\right]. \tag{42}$$

Moreover, given the transformation matrix $\boldsymbol{\Omega}_+$ in (40) and the two-dimensional penalty matrix defined in (15), the mixed model penalty matrix is obtained as:

$$\boldsymbol{\Omega}'_+\boldsymbol{P}_+\boldsymbol{\Omega}_+ = \begin{bmatrix} \boldsymbol{O}_q & \\ & \boldsymbol{F}_+ \end{bmatrix}, \text{ with } \mathbf{F}_+ = \begin{bmatrix} \lambda_{\boldsymbol{x}}\tilde{\boldsymbol{\Sigma}}_{\boldsymbol{x}_+} \otimes \mathbf{I}_{q_{\boldsymbol{z}}} & & \\ & \lambda_{\boldsymbol{z}}\mathbf{I}_{q_{\boldsymbol{x}}} \otimes \tilde{\boldsymbol{\Sigma}}_{\boldsymbol{z}_+} & \\ & & \lambda_{\boldsymbol{x}}\tilde{\boldsymbol{\Sigma}}_{\boldsymbol{x}_+} \otimes \mathbf{I}_{c_{\boldsymbol{z}_+}-q_{\boldsymbol{z}}} + \lambda_{\boldsymbol{x}}\mathbf{I}_{c_{\boldsymbol{x}_+}-q_{\boldsymbol{x}}} \otimes \tilde{\boldsymbol{\Sigma}}_{\boldsymbol{z}_+} \end{bmatrix}, \tag{43}$$

where $q = q_{\boldsymbol{z}}q_{\boldsymbol{x}}$ and the matrices $\tilde{\boldsymbol{\Sigma}}_i$ contains the positive eigenvalues of the SVD of $\boldsymbol{D}'_i\boldsymbol{D}_i$, for $i = \boldsymbol{z}_+, \boldsymbol{x}_+$. Then,

$$\boldsymbol{G}_+ = \sigma^2_{\boldsymbol{\epsilon}}\boldsymbol{F}_+^{-1}. \tag{44}$$

As a particular case, suppose that we extend just one independent covariable, in this case the natural extension of the trasformation matrix (32) is

$$\boldsymbol{\Omega}_+ = [\underbrace{\mathbf{U}_{\boldsymbol{x}+f} \otimes \mathbf{U}_{\boldsymbol{z}f}}_{\boldsymbol{\Omega}_{+f}} \mid \underbrace{\mathbf{U}_{\boldsymbol{x}+r} \otimes \mathbf{U}_{\boldsymbol{z}f} \mid \mathbf{U}_{\boldsymbol{x}+f} \otimes \mathbf{U}_{\boldsymbol{z}r} \mid \mathbf{U}_{\boldsymbol{x}+r} \otimes \mathbf{U}_{\boldsymbol{z}r}}_{\boldsymbol{\Omega}_{+r}}], \tag{45}$$

which is based on the SVD decompositions $\boldsymbol{D}'_{\boldsymbol{x}_+}\boldsymbol{D}_{\boldsymbol{x}_+} = \boldsymbol{U}_{\boldsymbol{x}_+}\boldsymbol{\Sigma}_{\boldsymbol{x}_+}\boldsymbol{U}'_{\boldsymbol{x}_+}$ and $\boldsymbol{D}'_{\boldsymbol{z}}\boldsymbol{D}_{\boldsymbol{z}} = \boldsymbol{U}_{\boldsymbol{z}}\boldsymbol{\Sigma}_{\boldsymbol{z}}\boldsymbol{U}'_{\boldsymbol{z}}$. Given the extended transformation matrix in (45), the mixed model matrices are:

$$\mathbf{X}_+ = \mathbf{B}_+\boldsymbol{\Omega}_{+f} = (\mathbf{B}_{\boldsymbol{x}_+} \otimes \mathbf{B}_{\boldsymbol{z}})(\mathbf{U}_{\boldsymbol{x}+f} \otimes \mathbf{U}_{\boldsymbol{z}f}) = \mathbf{B}_{\boldsymbol{x}_+}\mathbf{U}_{\boldsymbol{x}+f} \otimes \mathbf{B}_{\boldsymbol{z}}\mathbf{U}_{\boldsymbol{z}f}, \tag{46}$$

and

$$\mathbf{Z}_+ = \mathbf{B}\boldsymbol{\Omega}_{+r} = \left[\mathbf{B}_{\boldsymbol{x}_+}\mathbf{U}_{\boldsymbol{x}+r} \otimes \mathbf{B}_{\boldsymbol{z}}\mathbf{U}_{\boldsymbol{z}f} \mid \mathbf{B}_{\boldsymbol{x}_+}\mathbf{U}_{\boldsymbol{x}+f} \otimes \mathbf{B}_{\boldsymbol{z}}\mathbf{U}_{\boldsymbol{z}r} \mid \mathbf{B}_{\boldsymbol{x}_+}\mathbf{U}_{\boldsymbol{x}+r} \otimes \mathbf{B}_{\boldsymbol{z}}\mathbf{U}_{\boldsymbol{z}r}\right]. \tag{47}$$

For the transformation matrix $\boldsymbol{\Omega}_+$ in (45) and the penalty matrix defined in (20), the mixed model penalty matrix is:

$$\boldsymbol{\Omega}'_+\boldsymbol{P}_+\boldsymbol{\Omega}_+ = \begin{bmatrix}\boldsymbol{O}_q & \\ & \boldsymbol{F}_+\end{bmatrix}, \text{ with } \mathbf{F}_+ = \begin{bmatrix}\lambda_{\boldsymbol{x}}\tilde{\boldsymbol{\Sigma}}_{\boldsymbol{x}_+} \otimes \mathbf{I}_{q_{\boldsymbol{z}}} & & \\ & \lambda_{\boldsymbol{z}}\mathbf{I}_{q_{\boldsymbol{x}}} \otimes \tilde{\boldsymbol{\Sigma}}_{\boldsymbol{z}} & \\ & & \lambda_{\boldsymbol{x}}\tilde{\boldsymbol{\Sigma}}_{\boldsymbol{x}_+} \otimes \mathbf{I}_{c_{\boldsymbol{z}}-q_{\boldsymbol{z}}} + \lambda_{\boldsymbol{x}}\mathbf{I}_{c_{\boldsymbol{x}_+}-q_{\boldsymbol{x}}} \otimes \tilde{\boldsymbol{\Sigma}}_{\boldsymbol{z}}\end{bmatrix}, \tag{48}$$

where $q = q_{\boldsymbol{z}}q_{\boldsymbol{x}}$ and the matrices $\tilde{\boldsymbol{\Sigma}}_i$ $(i = \boldsymbol{z}, \boldsymbol{x}_+)$ were defined above. And,

$$\boldsymbol{G}_+ = \sigma^2_{\epsilon}\boldsymbol{F}^{-1}_+. \tag{49}$$

Ugarte *et al.* (2012) also carry out multidimensional out-of-sample prediction when only one covariate is extended, in this work the authors set $\boldsymbol{I}_{c_{\boldsymbol{x}_p}}$ equal to zero in (20) and propose to use an extended transformation matrix that preserves the transformation used to obtain the fit. They consider the extended transformation matrix $\boldsymbol{\Omega}_+$ defined as $\begin{bmatrix}\boldsymbol{\Omega} & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{D}^{-1}_{\boldsymbol{x}+(2)} \otimes \boldsymbol{I}_{c_{\boldsymbol{z}}}\end{bmatrix}$, the problem is that with the previous extended transformation matrix we would not differentiate between fixed and random effects, the fixed part would also be penalized since the first q $(q = q_{\boldsymbol{x}}q_{\boldsymbol{z}})$ rows and columns of $\boldsymbol{\Omega}'_+\boldsymbol{P}_+\boldsymbol{\Omega}_+$ are not zero.

3.0.2 Reparameterization of P-splines as mixed models for coherent prediction

To preserve the model matrices used to obtain the fit, our proposal is to define the following extended transformation matrix:

$$\boldsymbol{\Omega}^*_+ = [\underbrace{\mathbf{U}^*_{\boldsymbol{x}+f} \otimes \mathbf{U}^*_{\boldsymbol{z}+f}}_{\boldsymbol{\Omega}^*_{+f}} \mid \underbrace{\mathbf{U}^*_{\boldsymbol{x}+r} \otimes \mathbf{U}^*_{\boldsymbol{z}+f} \mid \mathbf{U}^*_{\boldsymbol{x}+f} \otimes \mathbf{U}^*_{\boldsymbol{z}+r} \mid \mathbf{U}^*_{\boldsymbol{x}+r} \otimes \mathbf{U}^*_{\boldsymbol{z}+r}}_{\boldsymbol{\Omega}^*_{+r}}], \tag{50}$$

which is obtained by reordering the block matrices of the matrix $[\mathbf{U}^*_{\boldsymbol{x}+f} \mid \mathbf{U}^*_{\boldsymbol{x}+r}] \otimes [\mathbf{U}^*_{\boldsymbol{z}+f} \mid \mathbf{U}^*_{\boldsymbol{z}+r}]$, where

$$[\mathbf{U}^*_{\boldsymbol{z}+f} \mid \mathbf{U}^*_{\boldsymbol{z}+r}] = \left[\begin{array}{c|cc}\boldsymbol{U}_{\boldsymbol{z}f} & \boldsymbol{U}_{\boldsymbol{z}r} & \boldsymbol{O} \\ -\boldsymbol{D}^{-1}_{\boldsymbol{z}(2)}\boldsymbol{D}_{\boldsymbol{z}(1)}\boldsymbol{U}_{\boldsymbol{z}f} & \boldsymbol{O} & \boldsymbol{D}^{-1}_{\boldsymbol{z}(2)}\end{array}\right],$$

$$[\mathbf{U}^*_{\boldsymbol{x}+f} \mid \mathbf{U}^*_{\boldsymbol{x}+r}] = \left[\begin{array}{c|cc}\boldsymbol{U}_{\boldsymbol{x}f} & \boldsymbol{U}_{\boldsymbol{x}r} & \boldsymbol{O} \\ -\boldsymbol{D}^{-1}_{\boldsymbol{x}(2)}\boldsymbol{D}_{\boldsymbol{x}(1)}\boldsymbol{U}_{\boldsymbol{x}f} & \boldsymbol{O} & \boldsymbol{D}^{-1}_{\boldsymbol{x}(2)}\end{array}\right],$$

where $\boldsymbol{U}_{\boldsymbol{z}f}$, $\boldsymbol{U}_{\boldsymbol{z}r}$, $\boldsymbol{U}_{\boldsymbol{x}f}$ and $\boldsymbol{U}_{\boldsymbol{x}r}$ are defined as in (32), $\boldsymbol{D}_{\boldsymbol{z}(2)}$ $\boldsymbol{D}_{\boldsymbol{z}(1)}$ are blocks of the extended difference matrix $\boldsymbol{D}_{\boldsymbol{z}_+}$ (see (21)) and $\boldsymbol{D}_{\boldsymbol{x}(2)}$ $\boldsymbol{D}_{\boldsymbol{x}(1)}$ are blocks of the extended difference matrix $\boldsymbol{D}_{\boldsymbol{x}_+}$ (see (17)). Notice that this definition of $\boldsymbol{U}^*_{if}$ for $i = \boldsymbol{x}_+, \boldsymbol{z}_+$, verifies $\boldsymbol{D}_i\boldsymbol{U}^*_{if} = \boldsymbol{O}$, i.e., the fixed part is not penalized.

Then, given the transformation matrix in (50) and denoting the matrices $\mathbf{X}^*_{\boldsymbol{z}_+} = \mathbf{B}_{\boldsymbol{z}_+}\mathbf{U}^*_{\boldsymbol{z}+f}$, $\mathbf{Z}^*_{\boldsymbol{z}_+} = \mathbf{B}_{\boldsymbol{z}_+}\mathbf{U}^*_{\boldsymbol{z}+r}$, $\mathbf{X}^*_{\boldsymbol{x}_+} = \mathbf{B}_{\boldsymbol{x}_+}\mathbf{U}^*_{\boldsymbol{x}+f}$ and $\mathbf{Z}^*_{\boldsymbol{x}_+} = \mathbf{B}_{\boldsymbol{x}_+}\mathbf{U}^*_{\boldsymbol{x}+r}$, the mixed model matrices for the two dimensional case are obtained as:

$$\begin{aligned}\mathbf{X}^*_+ &= \mathbf{X}^*_{\boldsymbol{x}_+} \otimes \mathbf{X}^*_{\boldsymbol{z}_+}, \\ \mathbf{Z}^*_+ &= [\mathbf{Z}_{\boldsymbol{x}_+} \otimes \mathbf{X}^*_{\boldsymbol{z}_+} \mid \mathbf{X}^*_{\boldsymbol{x}_+} \otimes \mathbf{Z}^*_{\boldsymbol{z}_+} \mid \mathbf{Z}^*_{\boldsymbol{x}_+} \otimes \mathbf{Z}^*_{\boldsymbol{z}_+}].\end{aligned} \tag{51}$$

Notice that $\boldsymbol{X}^*_{i_+}$ and $\boldsymbol{Z}^*_{i_+}$, for $i = \boldsymbol{z}, \boldsymbol{x}$, are direct extensions of $\boldsymbol{X}_i$ and $\boldsymbol{Z}_i$, i.e., they have the following form:

$$\boldsymbol{X}^*_{i_+} = \begin{bmatrix} \boldsymbol{X}_i \\ \boldsymbol{X}_{i_p} \end{bmatrix}, \; \boldsymbol{Z}^*_{i_+} = \begin{bmatrix} \boldsymbol{Z}_i & \boldsymbol{O} \\ \boldsymbol{Z}_{i(1)} & \boldsymbol{Z}_{i(2)} \end{bmatrix}. \tag{52}$$

Therefore, $\boldsymbol{X}^*_+$ and $\boldsymbol{Z}^*_+$ are also direct extensions of $\boldsymbol{X}$ and $\boldsymbol{Z}$, respectively, they are:

$$\begin{aligned} \boldsymbol{X}^*_+ &= \begin{bmatrix} \boldsymbol{X}_{\boldsymbol{x}} \\ \boldsymbol{X}_{\boldsymbol{x}_p} \end{bmatrix} \otimes \begin{bmatrix} \boldsymbol{X}_{\boldsymbol{z}} \\ \boldsymbol{X}_{\boldsymbol{z}_p} \end{bmatrix}, \\ \boldsymbol{Z}^*_{\boldsymbol{x}_+} &= \begin{bmatrix} \boldsymbol{Z}_{\boldsymbol{x}} & \boldsymbol{O} \\ \boldsymbol{Z}_{\boldsymbol{x}(1)} & \boldsymbol{Z}_{\boldsymbol{x}(2)} \end{bmatrix} \otimes \begin{bmatrix} \boldsymbol{Z}_{\boldsymbol{z}} & \boldsymbol{O} \\ \boldsymbol{Z}_{\boldsymbol{z}(1)} & \boldsymbol{Z}_{\boldsymbol{z}(2)} \end{bmatrix}. \end{aligned}$$

The following theorem gives the covariance matrix of the random effects for the transformation matrix given in (50) and the extended penalty matrix given in (15). aaa

Theorem 3. *Given the extended transformation $\boldsymbol{\Omega}^*_+$ in two dimensions defined in (50) and the extended penalty matrix in (15). The mixed model block-diagonal precision matrix $\boldsymbol{F}^*_+$ is*

$$\boldsymbol{F}^*_+ = \begin{bmatrix} \boldsymbol{F}^*_{+11} & \boldsymbol{O} & \boldsymbol{F}^*_{+13} \\ \boldsymbol{O} & \boldsymbol{F}^*_{+22} & \boldsymbol{F}^*_{+23} \\ \boldsymbol{F}^*_{+31} & \boldsymbol{F}^*_{+32} & \boldsymbol{F}^*_{+33} \end{bmatrix} \tag{53}$$

with

$$\begin{aligned} \boldsymbol{F}^*_{+11} &= \lambda_{\boldsymbol{x}} \boldsymbol{U}^{*'}_{\boldsymbol{x}+r} \boldsymbol{D}'_{\boldsymbol{x}_+} \boldsymbol{D}_{\boldsymbol{x}_+} \boldsymbol{U}^*_{\boldsymbol{x}+r} \otimes \boldsymbol{U}^{*'}_{\boldsymbol{z}+f} \boldsymbol{U}^*_{\boldsymbol{z}+r}, \\ \boldsymbol{F}^*_{+13} &= \lambda_{\boldsymbol{x}} \boldsymbol{U}^{*'}_{\boldsymbol{x}+r} \boldsymbol{D}'_{\boldsymbol{x}_+} \boldsymbol{D}_{\boldsymbol{x}_+} \boldsymbol{U}^*_{\boldsymbol{x}+r} \otimes \boldsymbol{U}^{*'}_{\boldsymbol{z}+f} \boldsymbol{U}^*_{\boldsymbol{z}+r}, \\ \boldsymbol{F}^*_{+22} &= \lambda_{\boldsymbol{z}} \boldsymbol{U}^{*'}_{\boldsymbol{x}+f} \boldsymbol{U}^*_{\boldsymbol{x}+f} \otimes \boldsymbol{U}^{*'}_{\boldsymbol{z}+r} \boldsymbol{D}'_{\boldsymbol{z}_+} \boldsymbol{D}_{\boldsymbol{z}_+} \boldsymbol{U}^*_{\boldsymbol{z}+r}, \\ \boldsymbol{F}^*_{+23} &= \lambda_{\boldsymbol{z}} \boldsymbol{U}^{*'}_{\boldsymbol{x}+f} \boldsymbol{U}^*_{\boldsymbol{x}+r} \otimes \boldsymbol{U}^{*'}_{\boldsymbol{z}+r} \boldsymbol{D}'_{\boldsymbol{z}_+} \boldsymbol{D}_{\boldsymbol{z}_+} \boldsymbol{U}^*_{\boldsymbol{z}+r}, \\ \boldsymbol{F}^*_{+31} &= \boldsymbol{F}^*_{+13}, \\ \boldsymbol{F}^*_{+32} &= \boldsymbol{F}^*_{+23}, \\ \boldsymbol{F}^*_{+33} &= \lambda_{\boldsymbol{z}} \boldsymbol{U}'_{\boldsymbol{x}+r} \boldsymbol{U}_{\boldsymbol{x}+r} \otimes \boldsymbol{U}'_{\boldsymbol{z}+r} \boldsymbol{D}'_{\boldsymbol{z}_+} \boldsymbol{D}_{\boldsymbol{z}_+} \boldsymbol{U}_{\boldsymbol{z}+r} + \lambda_{\boldsymbol{x}} \boldsymbol{U}'_{\boldsymbol{x}+r} \boldsymbol{D}'_{\boldsymbol{x}_+} \boldsymbol{D}_{\boldsymbol{x}_+} \boldsymbol{U}_{\boldsymbol{x}+r} \otimes \boldsymbol{U}'_{\boldsymbol{z}+r} \boldsymbol{U}_{\boldsymbol{z}+r}, \end{aligned}$$

and the covariance matrix of the random effects is $\boldsymbol{G}^*_+ = \sigma^2_{\boldsymbol{\epsilon}} \boldsymbol{F}^{*-1}$.

The proof of the previous Theorem is shown in Appendix .2.

If we extend just one of the two covariates, to preserve the model matrices we define the following extended transformation matrix $\boldsymbol{\Omega}^*_+$:

$$\boldsymbol{\Omega}^*_+ = [\underbrace{\mathbf{U}^*_{\boldsymbol{x}+f} \otimes \mathbf{U}_{\boldsymbol{z}f}}_{\boldsymbol{\Omega}^*_{+f}} \mid \underbrace{\mathbf{U}^*_{\boldsymbol{x}+r} \otimes \mathbf{U}_{\boldsymbol{z}r} \mid \mathbf{U}^*_{\boldsymbol{x}+f} \otimes \mathbf{U}_{\boldsymbol{z}r} \mid \mathbf{U}^*_{\boldsymbol{x}+r} \otimes \mathbf{U}_{\boldsymbol{z}r}}_{\boldsymbol{\Omega}^*_{+r}}], \tag{54}$$

which is obtained by reordering the block matrices of the matrix $[\mathbf{U}^*_{\boldsymbol{x}+f} \mid \mathbf{U}^*_{\boldsymbol{x}+r}] \otimes [\mathbf{U}_{\boldsymbol{z}f} \mid \mathbf{U}_{\boldsymbol{z}r}]$, where $[\mathbf{U}_{\boldsymbol{z}f} \mid \mathbf{U}_{\boldsymbol{z}r}]$ is defined as in (32) and $[\mathbf{U}^*_{\boldsymbol{x}+f} \mid \mathbf{U}^*_{\boldsymbol{x}+r}]$ defined as in (50).

In this case, the model components are:

- Extended mixed model matrices:

$$\mathbf{X}^*_+ = \mathbf{B}_{\boldsymbol{x}_+} \mathbf{U}^*_{\boldsymbol{x}+f} \otimes \mathbf{B}_{\boldsymbol{z}} \mathbf{U}_{\boldsymbol{z}f},$$

and

$$\mathbf{Z}^*_+ = \mathbf{B}\boldsymbol{\Omega}^*_{+r} = \left[\mathbf{B}_{\boldsymbol{x}_+} \mathbf{U}^*_{\boldsymbol{x}+r} \otimes \mathbf{B}_{\boldsymbol{z}} \mathbf{U}_{\boldsymbol{z}f} \mid \mathbf{B}_{\boldsymbol{x}_+} \mathbf{U}^*_{\boldsymbol{x}+f} \otimes \mathbf{B}_{\boldsymbol{z}} \mathbf{U}_{\boldsymbol{z}r} \mid \mathbf{B}_{\boldsymbol{x}_+} \mathbf{U}^*_{\boldsymbol{x}+r} \otimes \mathbf{B}_{\boldsymbol{z}} \mathbf{U}_{\boldsymbol{z}r}\right].$$

- Extended random effects covariance matrix $\boldsymbol{G}^*_+ = \sigma^2_{\boldsymbol{\epsilon}} \boldsymbol{F}^{*-1}$, with

$$\boldsymbol{F}^*_+ = \begin{bmatrix} \lambda_{\boldsymbol{x}} \boldsymbol{U}^{*'}_{\boldsymbol{x}_+ r} \boldsymbol{D}'_{\boldsymbol{x}_+} \boldsymbol{D}_{\boldsymbol{x}_+} \boldsymbol{U}^*_{\boldsymbol{x}_+ r} \otimes \boldsymbol{I}_{q_{\boldsymbol{z}}} & & \\ & \lambda_{\boldsymbol{z}} \boldsymbol{U}^{*'}_{\boldsymbol{x}_+ f} \boldsymbol{U}^*_{\boldsymbol{x}_+ f} \otimes \widetilde{\boldsymbol{\Sigma}}_{\boldsymbol{z}} & \lambda_{\boldsymbol{z}} \boldsymbol{U}^{*'}_{\boldsymbol{x}_+ f} \boldsymbol{U}^*_{\boldsymbol{x}_+ r} \otimes \widetilde{\boldsymbol{\Sigma}}_{\boldsymbol{z}} \\ & \lambda_{\boldsymbol{z}} \boldsymbol{U}^{*'}_{\boldsymbol{x}_+ r} \boldsymbol{U}^*_{\boldsymbol{x}_+ f} \otimes \widetilde{\boldsymbol{\Sigma}}_{\boldsymbol{z}} & \lambda_{\boldsymbol{z}} \boldsymbol{U}'_{\boldsymbol{x}_+ r} \boldsymbol{U}_{\boldsymbol{x}_+ r} \otimes \widetilde{\boldsymbol{\Sigma}}_{\boldsymbol{z}} + \lambda_{\boldsymbol{x}} \boldsymbol{U}'_{\boldsymbol{x}_+ r} \boldsymbol{D}'_{\boldsymbol{x}_+} \boldsymbol{D}_{\boldsymbol{x}_+} \boldsymbol{U}_{\boldsymbol{x}_+ r} \otimes \boldsymbol{I}_{c_{\boldsymbol{z}} - q_{\boldsymbol{z}}} \end{bmatrix} \tag{55}$$

where $\tilde{\boldsymbol{\Sigma}}_{\boldsymbol{z}}$, of dimensions $(c_{\boldsymbol{z}} - q_{\boldsymbol{z}}) \times (c_{\boldsymbol{z}} - q_{\boldsymbol{z}})$ is the diagonal matrix of positive eigenvalues of $\boldsymbol{D}'_{\boldsymbol{z}} \boldsymbol{D}_{\boldsymbol{z}}$.

3.1 Constrained smooth mixed models for coherent out-of-sample prediction

As in the case of 2D P-spline models, constraints need to be imposed in order to ensure coherent fit and prediction. In this section we explain how predictions (subject to the restriction that the fit is kept) are carried out in the context of mixed models. Suppose that we consider the following restricted extended mixed model:

$$\boldsymbol{y}_+ = \boldsymbol{X}_+ \boldsymbol{\beta}^*_{\tilde{+}} + \boldsymbol{Z}_+ \boldsymbol{\alpha}^*_+ + \boldsymbol{\epsilon}_+, \quad \boldsymbol{\epsilon}_+ \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{R}_+), \quad \boldsymbol{\alpha}^*_+ \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{G}_+),$$

subject to $\boldsymbol{C}_{\text{MM}} \begin{bmatrix} \boldsymbol{\beta}^*_{\tilde{+}} \\ \boldsymbol{\alpha}^*_+ \end{bmatrix} = \boldsymbol{r}_{\text{MM}}$, where $\boldsymbol{C}_{\text{MM}}$ is a constraint matrix of dimension $l \times c_+$ acting on all coefficients and built analogously to how $\boldsymbol{C}$ is built in Section 2.2, and $\boldsymbol{r}_{\text{MM}} = \begin{bmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\boldsymbol{\alpha}} \end{bmatrix}$ is the restrictions vector of dimension $l \times 1$. Notice that we use the superscript $(*)$ to indicate that we are imposing restrictions.

To estimate the restricted parameters we minimize the following constrained penalized likelihood:

$$\begin{aligned} \mathcal{L}(\boldsymbol{\beta}^*_{\tilde{+}}, \boldsymbol{\alpha}^*_+, \boldsymbol{w}) &= (\boldsymbol{y}_+ - \boldsymbol{X}_+ \boldsymbol{\beta}^*_{\tilde{+}} - \boldsymbol{Z}_+ \boldsymbol{\alpha}^*_+)' \boldsymbol{R}^{-1}_+ (\boldsymbol{y}_+ - \boldsymbol{X}_+ \boldsymbol{\beta}^*_{\tilde{+}} - \boldsymbol{Z}_+ \boldsymbol{\alpha}^*_+) + \boldsymbol{\alpha}^{*'}_+ \boldsymbol{G}^{-1}_+ \boldsymbol{\alpha}^*_+ \\ &\quad + 2\boldsymbol{w}' \left(\boldsymbol{C}_{MM} \begin{bmatrix} \boldsymbol{\beta}^*_{\tilde{+}} \\ \boldsymbol{\alpha}^*_+ \end{bmatrix} - \boldsymbol{r}_{MM} \right) \end{aligned} \tag{56}$$

Since we have to take derivatives with respect to $\boldsymbol{\beta}^*_{\tilde{+}}$ and with respect to $\boldsymbol{\alpha}^*_+$, we divide the matrix of constraints into two parts, one associated with the fixed effects and the other one associated with the random effects, $\boldsymbol{C}_{MM} = \left[\boldsymbol{C}_{MM_f} \mid \boldsymbol{C}_{MM_r} \right]$. Formula (56) is rewritten as:

$$\begin{aligned} \mathcal{L}(\boldsymbol{\beta}^*_{\tilde{+}}, \boldsymbol{\alpha}^*_+, \boldsymbol{w}) &= (\boldsymbol{y}_+ - \boldsymbol{X}_+ \boldsymbol{\beta}^*_{\tilde{+}} - \boldsymbol{Z}_+ \boldsymbol{\alpha}^*_+)' \boldsymbol{R}^{-1}_+ (\boldsymbol{y}_+ - \boldsymbol{X}_+ \boldsymbol{\beta}^*_{\tilde{+}} - \boldsymbol{Z}_+ \boldsymbol{\alpha}^*_+) + \boldsymbol{\alpha}^{*'}_+ \boldsymbol{G}^{-1}_+ \boldsymbol{\alpha}^*_+ \\ &\quad + 2\boldsymbol{\omega}' (\boldsymbol{C}_{MM_f} \boldsymbol{\beta}^*_{\tilde{+}} + \boldsymbol{C}_{MM_r} \boldsymbol{\alpha}^*_+ - \boldsymbol{r}_{MM}). \end{aligned}$$

The first order conditions yield

$$\frac{\partial \mathcal{L}}{\partial \boldsymbol{\beta}^*_{\tilde{+}}} = -2\boldsymbol{X}'_+ \boldsymbol{R}^{-1}_+ (\boldsymbol{y}_+ - \boldsymbol{X}_+ \boldsymbol{\beta}^*_{\tilde{+}} - \boldsymbol{Z}_+ \boldsymbol{\alpha}^*_+) + 2\boldsymbol{C}'_{MM_f} \boldsymbol{\omega} \tag{57}$$

$$\frac{\partial \mathcal{L}}{\partial \boldsymbol{\alpha}^*_+} = -2\boldsymbol{Z}'_+ \boldsymbol{R}^{-1}_+ (\boldsymbol{y}_+ - \boldsymbol{X}_+ \boldsymbol{\beta}^*_{\tilde{+}} - \boldsymbol{Z}_+ \boldsymbol{\alpha}^*_+) + 2\boldsymbol{G}^{-1}_+ \boldsymbol{\alpha}_+ + 2\boldsymbol{C}'_{MM_r} \boldsymbol{\omega} \tag{58}$$

$$\frac{\partial \mathcal{L}}{\partial \boldsymbol{w}} = \boldsymbol{C}_{MM_f} \boldsymbol{\beta}^*_{\tilde{+}} + \boldsymbol{C}_{MM_r} \boldsymbol{\alpha}^*_+ - \boldsymbol{r}_{MM} \tag{59}$$

Therefore, we have the system:

$$\begin{bmatrix} \boldsymbol{X}'_+ \boldsymbol{R}^{-1}_+ \boldsymbol{X}_+ & \boldsymbol{X}'_+ \boldsymbol{R}^{-1}_+ \boldsymbol{Z}_+ & \boldsymbol{C}'_{MM_f} \\ \boldsymbol{Z}'_+ \boldsymbol{R}^{-1}_+ \boldsymbol{X}_+ & \boldsymbol{Z}'_+ \boldsymbol{R}^{-1}_+ \boldsymbol{Z}_+ + \boldsymbol{G}^{-1}_+ & \boldsymbol{C}'_{MM_r} \\ \boldsymbol{C}_{MM_f} & \boldsymbol{C}_{MM_r} & \boldsymbol{O} \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta}^*_{\tilde{+}} \\ \boldsymbol{\alpha}^*_+ \\ \boldsymbol{\omega} \end{bmatrix} = \begin{bmatrix} \boldsymbol{X}'_+ \boldsymbol{R}^{-1}_+ \boldsymbol{y}_+ \\ \boldsymbol{Z}'_+ \boldsymbol{R}^{-1}_+ \boldsymbol{y}_+ \\ \boldsymbol{r}_{MM} \end{bmatrix}$$

We can also compute the fixed and random effects and the Lagrangean multipliers following the next steps. By (57) and (58) we know that

$$\begin{bmatrix}\hat{\boldsymbol{\beta}}^*_{\tilde{+}}\\ \hat{\boldsymbol{\alpha}}^*_{+}\end{bmatrix} = \begin{bmatrix}\hat{\boldsymbol{\beta}}_{\tilde{+}}\\ \hat{\boldsymbol{\alpha}}_{+}\end{bmatrix} - \boldsymbol{L}_+^{-1}\boldsymbol{C'}_{MM}\hat{\boldsymbol{w}}, \tag{60}$$

where $\boldsymbol{L}_+ = \begin{bmatrix}\boldsymbol{X'}_+\boldsymbol{R}_+^{-1}\boldsymbol{X}_+^* & \boldsymbol{X'}_+\boldsymbol{R}_+^{-1}\boldsymbol{Z}_+\\ \boldsymbol{Z'}_+\boldsymbol{R}_+^{-1}\boldsymbol{X}_+ & \boldsymbol{Z'}_+\boldsymbol{R}_+^{-1}\boldsymbol{Z}_+ + \boldsymbol{G}_+^{-1}\end{bmatrix}$ and $\begin{bmatrix}\hat{\boldsymbol{\beta}}_{\tilde{+}}\\ \hat{\boldsymbol{\alpha}}_{+}\end{bmatrix}$ are the unrestricted penalized least squares estimators, and

$$\hat{\boldsymbol{\omega}} = \left[\boldsymbol{C}_{MM}\boldsymbol{L}_+^{-1}\boldsymbol{C'}_{MM}\right]^{-1}\left[\boldsymbol{C}_{MM}\begin{bmatrix}\boldsymbol{\beta}_{\tilde{+}}\\ \boldsymbol{\alpha}_{+}\end{bmatrix} - \boldsymbol{r}_{MM}\right] \tag{61}$$

Therefore, the coefficients subject to the restriction, $\begin{bmatrix}\hat{\boldsymbol{\beta}}^*_{\tilde{+}}\\ \hat{\boldsymbol{\alpha}}^*_{+}\end{bmatrix}$, are obtained by computing the vector of Lagrange multipliers (61) and substituting in (60), i.e. $\begin{bmatrix}\hat{\boldsymbol{\beta}}^*_{\tilde{+}}\\ \hat{\boldsymbol{\alpha}}^*_{+}\end{bmatrix}$ is the unconstrained solution, $\begin{bmatrix}\hat{\boldsymbol{\beta}}_{\tilde{+}}\\ \hat{\boldsymbol{\alpha}}_{+}\end{bmatrix}$, plus a multiple of the discrepancy vector.

The restricted fitted and predicted values are

$$\hat{\boldsymbol{y}}^*_+ = [\boldsymbol{X}_+ \mid \boldsymbol{Z}_+]\begin{bmatrix}\hat{\boldsymbol{\beta}}^*_{*}\\ \hat{\boldsymbol{\alpha}}^*_{+}\end{bmatrix},$$

defining the matrices $\boldsymbol{L}_+ = \begin{bmatrix}\boldsymbol{X'}_+\boldsymbol{R}_+^{-1}\boldsymbol{X}_+ & \boldsymbol{X'}_+\boldsymbol{R}_+^{-1}\boldsymbol{Z}_+\\ \boldsymbol{Z'}_+\boldsymbol{R}_+^{-1}\boldsymbol{X}_+ & \boldsymbol{Z'}_+\boldsymbol{R}_+^{-1}\boldsymbol{Z}_+ + \boldsymbol{G}_+^{-1}\end{bmatrix}$, $\boldsymbol{A_1}_{MM} = \boldsymbol{L}_+^{-1}\begin{bmatrix}\boldsymbol{X}_+\\ \boldsymbol{Z}_+\end{bmatrix}$ and $\boldsymbol{A_2}_{MM} = \boldsymbol{L}_+^{-1}\boldsymbol{C'}_{MM}[\boldsymbol{C}_{MM}\boldsymbol{L}_+^{-1}\boldsymbol{C'}_{MM}]^{-1}$, $\hat{\boldsymbol{y}}_+$ can be written as:

$$\hat{\boldsymbol{y}}^*_+ = [\boldsymbol{X}_+ \mid \boldsymbol{Z}_+](\boldsymbol{A_1}_{MM}\boldsymbol{R}_+^{-1}\boldsymbol{y}_+ - \boldsymbol{A_2}_{MM}\boldsymbol{C}_{MM}\boldsymbol{A_1}_{MM}\boldsymbol{R}_+^{-1}\boldsymbol{y}_+ + \boldsymbol{A_2}_{MM}\boldsymbol{r}_{MM}).$$

Since $\boldsymbol{r}_{MM} = \begin{bmatrix}\hat{\boldsymbol{\beta}}\\ \hat{\boldsymbol{\alpha}}\end{bmatrix} = \boldsymbol{L}\begin{bmatrix}\boldsymbol{X'}\boldsymbol{R}^{-1}\\ \boldsymbol{Z'}\boldsymbol{R}^{-1}\end{bmatrix}\boldsymbol{y}$, with $\boldsymbol{L} = \begin{bmatrix}\boldsymbol{X'}\boldsymbol{R}^{-1}\boldsymbol{X} & \boldsymbol{X}\boldsymbol{R}^{-1}\boldsymbol{Z}\\ \boldsymbol{Z'}\boldsymbol{R}^{-1}\boldsymbol{X} & \boldsymbol{Z}\boldsymbol{R}^{-1}\boldsymbol{Z} + \boldsymbol{G}^{-1}\end{bmatrix}$, taking into account the variability of $\boldsymbol{r}_{MM}$ the variance is:

$$\text{Var}[\hat{\boldsymbol{y}}^*_+] = [\boldsymbol{X}_+ \mid \boldsymbol{Z}_+]\boldsymbol{A_4}_{MM}\boldsymbol{R}_+^{-1}\boldsymbol{A'_4}_{MM}\begin{bmatrix}\boldsymbol{X'}_+\\ \boldsymbol{Z'}_+\end{bmatrix},$$

with $\boldsymbol{A_4}_{MM} = \left(\boldsymbol{A_1}_{MM} - \boldsymbol{A_2}_{MM}\boldsymbol{C}_{MM}\boldsymbol{A_1}_{MM} + \boldsymbol{A_2}_{MM}\begin{bmatrix}\boldsymbol{L}\begin{bmatrix}\boldsymbol{X'}\\ \boldsymbol{Z'}\end{bmatrix} & \\ & \boldsymbol{O}\end{bmatrix}\right)$, with $\boldsymbol{O}$ a null matrix of dimension $c_p \times n_p$, c_p the number of new coefficients and n_p the number of new values we predict.

4 Component-wise prediction with P-spline Smooth-ANOVA models

Sometimes, model (1) will not be flexible enough and it might force unnecessary complexity (i.e. use too many degrees of freedom to fit the data). In order to add more flexibility and drop unnecessary terms if they are not relevant, Lee & Durbán (2011) propose the use of P-spline smooth-ANOVA models. This approach decomposes the sum of smooth functions similarly as on analysis of variance decomposition. In this section, we show how out-of-sample prediction can be carried out in this context. First of all we briefly review how we can fit a smooth function as a decomposition of smooth functions and bases which are identifiable. Suppose we have a data vector $\boldsymbol{y}$ of length $n \times 1$, where $n = n_{\boldsymbol{z}} n_{\boldsymbol{x}}$, and the regressors $\boldsymbol{z} = (z_1, z_2, ...z_{n_{\boldsymbol{z}}})'$ and $\boldsymbol{x} = (x_1, x_2, ...x_{n_{\boldsymbol{x}}})'$. Let us consider the following Smooth-ANOVA model:

$$\boldsymbol{y} = \gamma + f_1(\boldsymbol{z}) + f_2(\boldsymbol{x}) + f_{1,2}(\boldsymbol{z}, \boldsymbol{x}) + \boldsymbol{\epsilon} = \boldsymbol{B\theta} + \boldsymbol{\epsilon}, \quad \boldsymbol{\epsilon} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{R}) \tag{62}$$

where $\boldsymbol{R} = \sigma^2_{\boldsymbol{\epsilon}} \boldsymbol{I}$, i.e. the errors are independent and identically distributed and the B-spline basis $\boldsymbol{B}$ is defined as:

$$\boldsymbol{B} = [\boldsymbol{1}_n \mid \boldsymbol{1}_{n_{\boldsymbol{x}}} \otimes \boldsymbol{B}_{\boldsymbol{z}} \mid \boldsymbol{B}_{\boldsymbol{x}} \otimes \boldsymbol{1}_{n_{\boldsymbol{z}}} \mid \boldsymbol{B}_{\boldsymbol{x}} \otimes \boldsymbol{B}_{\boldsymbol{z}}], \tag{63}$$

of dimension $n \times (1 + c_{\boldsymbol{z}} + c_{\boldsymbol{x}} + c_{\boldsymbol{z}} c_{\boldsymbol{x}})$, and where $\boldsymbol{1}_{n_{\boldsymbol{z}}}$ and $\boldsymbol{1}_{n_{\boldsymbol{x}}}$ are column vectors of ones of length $n_{\boldsymbol{z}}$ and $n_{\boldsymbol{x}}$ respectively, and the vector of regression coefficients is $\boldsymbol{\theta} = (\gamma, \boldsymbol{\theta}'_{\boldsymbol{z}}, \boldsymbol{\theta}'_{\boldsymbol{x}}, \boldsymbol{\theta}'_{\boldsymbol{s}})'$, where $\boldsymbol{\theta}_{\boldsymbol{z}}$ and $\boldsymbol{\theta}_{\boldsymbol{x}}$ are the vectors of coefficients for the main effects, of dimension $c_{\boldsymbol{z}} \times 1$ and $c_{\boldsymbol{x}} \times 1$, respectively, and $\boldsymbol{\theta}_s$ is the vector of coefficients for the interaction, of dimension $c_{\boldsymbol{z}} c_{\boldsymbol{x}} \times 1$. Therefore, model (62) is written as:

$$\boldsymbol{y} = \gamma \boldsymbol{1}_n + (\boldsymbol{1}_{n_{\boldsymbol{x}}} \otimes \boldsymbol{B}_{\boldsymbol{z}})\boldsymbol{\theta}_{\boldsymbol{z}} + (\boldsymbol{B}_{\boldsymbol{x}} \otimes \boldsymbol{1}_{n_{\boldsymbol{z}}})\boldsymbol{\theta}_{\boldsymbol{x}} + (\boldsymbol{B}_{\boldsymbol{x}} \otimes \boldsymbol{B}_{\boldsymbol{z}})\boldsymbol{\theta}_{\boldsymbol{s}}, \tag{64}$$

with penalty the following block-diagonal matrix:

$$\boldsymbol{P} = \begin{bmatrix} 0 & & & \\ & \lambda_{\boldsymbol{z}} \boldsymbol{D}'_{\boldsymbol{z}} \boldsymbol{D}_{\boldsymbol{z}} & & \\ & & \lambda_{\boldsymbol{x}} \boldsymbol{D}'_{\boldsymbol{x}} \boldsymbol{D}_{\boldsymbol{x}} & \\ & & & \tau_{\boldsymbol{z}} \boldsymbol{I}_{c_{\boldsymbol{x}}} \otimes \boldsymbol{D}'_{\boldsymbol{z}} \boldsymbol{D}_{\boldsymbol{z}} + \tau_{\boldsymbol{x}} \boldsymbol{D}'_{\boldsymbol{x}} \boldsymbol{D}_{\boldsymbol{x}} \otimes \boldsymbol{I}_{c_{\boldsymbol{z}}} \end{bmatrix}, \tag{65}$$

of dimension $(1 + c_{\boldsymbol{z}} + c_{\boldsymbol{x}} + c_{\boldsymbol{z}} c_{\boldsymbol{x}}) \times (1 + c_{\boldsymbol{z}} + c_{\boldsymbol{x}} + c_{\boldsymbol{z}} c_{\boldsymbol{x}})$, where each block corresponds to the penalty over each of the coefficients of the model. Lee (2010) pointed out that the model (64) and the penalty (65) should be modified in order to preserve identifiability, their proposal is to construct identifiability model bases and penalties reparameterizing the model as a mixed model instead of imposing numerical constraints as other authors have proposed (Wood (2006)). They define the following transformation matrix $\boldsymbol{\Omega} = [\boldsymbol{\Omega}_f \mid \boldsymbol{\Omega}_r]$ with dimension $(1 + c_{\boldsymbol{z}} + c_{\boldsymbol{x}} + c_{\boldsymbol{z}} c_{\boldsymbol{x}})$, where:

$$\boldsymbol{\Omega}_f = \begin{bmatrix} 1 & & & \\ & 1 \otimes \boldsymbol{u}^{(2)}_{\boldsymbol{z}f} & & \\ & & \boldsymbol{u}^{(2)}_{\boldsymbol{x}f} \otimes 1 & \\ & & & \boldsymbol{u}^{(2)}_{\boldsymbol{x}f} \otimes \boldsymbol{u}^{(2)}_{\boldsymbol{z}f} \end{bmatrix}, \tag{66}$$

$$\boldsymbol{\Omega}_r = \begin{bmatrix} 1 & & & \\ & 1 \otimes \boldsymbol{U}_{\boldsymbol{z}r} & & \\ & & \boldsymbol{U}_{\boldsymbol{x}r} \otimes 1 & \\ & & & \boldsymbol{u}^{(2)}_{\boldsymbol{x}f} \otimes \boldsymbol{U}_{\boldsymbol{z}r} \mid \boldsymbol{U}_{\boldsymbol{x}r} \otimes \boldsymbol{u}^{(2)}_{\boldsymbol{z}f} \mid \boldsymbol{U}_{\boldsymbol{x}r} \otimes \boldsymbol{U}_{\boldsymbol{z}r} \end{bmatrix}, \tag{67}$$

and $\boldsymbol{u}^{(2)}_{\boldsymbol{z}f}$ and $\boldsymbol{u}^{(2)}_{\boldsymbol{x}f}$ are the second columns of $\boldsymbol{U}_{\boldsymbol{z}r}$ and $\boldsymbol{U}_{\boldsymbol{x}r}$, respectively, and $\boldsymbol{U}_{\boldsymbol{z}f}$ and $\boldsymbol{U}_{\boldsymbol{x}f}$ are the eigenvectors corresponding to the positive values of the SVD of $\boldsymbol{D}'_{\boldsymbol{z}} \boldsymbol{D}_{\boldsymbol{z}}$ and $\boldsymbol{D}'_{\boldsymbol{x}} \boldsymbol{D}_{\boldsymbol{x}}$, respectively.

Given the previous transformation matrix, Lee (2010) show that the fixed and random effects matrices $\boldsymbol{X}$ and $\boldsymbol{Z}$ are:

$$\begin{aligned}
\boldsymbol{X} &= [\mathbf{1}_n \mid \mathbf{1}_{n_{\boldsymbol{x}}} \otimes \tilde{\boldsymbol{z}} \mid \tilde{\boldsymbol{x}} \otimes \mathbf{1}_{n_{\boldsymbol{z}}} \mid \tilde{\boldsymbol{x}} \otimes \tilde{\boldsymbol{z}}] \qquad (68)\\
\boldsymbol{Z} &= [\mathbf{1}_n \mid \mathbf{1}_{n_{\boldsymbol{x}}} \otimes \boldsymbol{Z}_{\boldsymbol{z}} \mid \boldsymbol{Z}_{\boldsymbol{x}} \otimes \mathbf{1}_{n_{\boldsymbol{z}}} \mid \tilde{\boldsymbol{x}} \otimes \boldsymbol{Z}_{\boldsymbol{z}} \mid \boldsymbol{Z}_{\boldsymbol{x}} \otimes \tilde{\boldsymbol{z}} \mid \boldsymbol{Z}_{\boldsymbol{x}} \otimes \boldsymbol{Z}_{\boldsymbol{z}}] \qquad (69)
\end{aligned}$$

where $\tilde{\boldsymbol{z}} = \boldsymbol{B}_{\boldsymbol{z}}\boldsymbol{u}^{(2)}_{\boldsymbol{z}f}$, $\tilde{\boldsymbol{x}} = \boldsymbol{B}_{\boldsymbol{x}}\boldsymbol{u}^{(2)}_{\boldsymbol{x}f}$, $\boldsymbol{Z}_{\boldsymbol{z}} = \boldsymbol{B}_{\boldsymbol{z}}\boldsymbol{U}_{\boldsymbol{z}r}$ and $\boldsymbol{Z}_{\boldsymbol{x}} = \boldsymbol{B}_{\boldsymbol{x}}\boldsymbol{U}_{\boldsymbol{x}r}$. Moreover, the mixed model penalty is:

$$\boldsymbol{F} = \text{blockdiag}(\boldsymbol{F}_{(1)}, \boldsymbol{F}_{(2)}, \boldsymbol{F}_{(1,2)}), \qquad (70)$$

where for a second order penalty, it has size $(c_{\boldsymbol{z}}c_{\boldsymbol{x}} - 4) \times (c_{\boldsymbol{z}}c_{\boldsymbol{x}} - 4)$, and with:

$$\begin{aligned}
\boldsymbol{F}_{(1)} &= \lambda_{\boldsymbol{z}}\tilde{\boldsymbol{\Sigma}}_{\boldsymbol{z}},\\
\boldsymbol{F}_{(2)} &= \lambda_{\boldsymbol{x}}\tilde{\boldsymbol{\Sigma}}_{\boldsymbol{x}},\\
\boldsymbol{F}_{(1,2)} &= \text{blockdiag}(\tau_{\boldsymbol{z}}\tilde{\boldsymbol{\Sigma}}_{\boldsymbol{z}}, \tau_{\boldsymbol{x}}\tilde{\boldsymbol{\Sigma}}_{\boldsymbol{x}}, \tau_{\boldsymbol{z}}\boldsymbol{I}_{c_{\boldsymbol{x}}-2} \otimes \tilde{\boldsymbol{\Sigma}}_{\boldsymbol{z}} + \tau_{\boldsymbol{x}}\tilde{\boldsymbol{\Sigma}}_{\boldsymbol{x}} \otimes \boldsymbol{I}_{c_{\boldsymbol{z}}-2})
\end{aligned}$$

where $\tilde{\boldsymbol{\Sigma}}_{\boldsymbol{z}}$ and $\tilde{\boldsymbol{\Sigma}}_{\boldsymbol{x}}$ are the nonzero eigenvalues of the SVD of $\boldsymbol{D}'_{\boldsymbol{z}}\boldsymbol{D}_{\boldsymbol{z}}$ and $\boldsymbol{D}'_{\boldsymbol{x}}\boldsymbol{D}_{\boldsymbol{x}}$, respectively. With the previous representation of model (62), Lee (2010) avoid the identifiability problem removing the column vector of $\mathbf{1}$'s in the random effects matrix (69). For more details, see Lee (2010). Then we will detail our proposal to obtain predictions with S-ANOVA models.

Although the out-of-sample prediction will be carried out in the context of mixed models, we present the approach in the original P-splines formulation, since the reparameterization needed for the out-of-sample prediction will be based on this formulation. In particular, we need to know what the extended penalty matrix is in order to calculate the precision matrix of the random effects.

In the framework of model (62), given a vector of $n_{\boldsymbol{z}}n_{\boldsymbol{x}}$ observations $\boldsymbol{y}$ of the response variable, suppose that we want to predict $n_p = n_{\boldsymbol{z}}n_{\boldsymbol{x}_p} + n_{\boldsymbol{z}_p}n_{\boldsymbol{x}} + n_{\boldsymbol{z}_p}n_{\boldsymbol{x}_p}$ new values at $(\boldsymbol{z}, \boldsymbol{x}_p)$, $(\boldsymbol{z}_p, \boldsymbol{x})$ and $(\boldsymbol{z}_p, \boldsymbol{x}_p)$. I.e., the matrix $\boldsymbol{Y}_+$ of observed and predicted values can be arrenged as in (11). For this case we consider the following extended Smooth-ANOVA model:

$$\boldsymbol{y}_+ = \gamma + f_1(\boldsymbol{z}_+) + f_2(\boldsymbol{x}_+) + f_{1,2}(\boldsymbol{z}_+, \boldsymbol{x}_+) + \boldsymbol{\epsilon}_+, \quad \boldsymbol{\epsilon}_+ \sim \mathcal{N}(\mathbf{0}, \sigma^2_{\boldsymbol{\epsilon}}\boldsymbol{R}_+) \qquad (71)$$

where $\boldsymbol{R}_+$ is defined as in (12) and we assume:

$$\boldsymbol{y}_+ = \boldsymbol{B}_+\boldsymbol{\theta}_+, \quad \boldsymbol{\epsilon}_+ \sim \mathcal{N}(\mathbf{0}, \boldsymbol{R}_+)$$

where the extended B-spline basis $\boldsymbol{B}_+$ is defined as:

$$\boldsymbol{B}_+ = [\mathbf{1}_{n_+} \mid \mathbf{1}_{n_{\boldsymbol{x}_+}} \otimes \boldsymbol{B}_{\boldsymbol{z}_+} \mid \boldsymbol{B}_{\boldsymbol{x}_+} \otimes \mathbf{1}_{n_{\boldsymbol{z}_+}} \mid \boldsymbol{B}_{\boldsymbol{x}_+} \otimes \boldsymbol{B}_{\boldsymbol{z}_+}], \qquad (72)$$

of dimension $n \times (1 + c_{\boldsymbol{z}_+} + c_{\boldsymbol{x}_+} + c_{\boldsymbol{z}_+}c_{\boldsymbol{x}_+})$, and where $\mathbf{1}_{n_{\boldsymbol{z}_+}}$ and $\mathbf{1}_{n_{\boldsymbol{x}_+}}$ are column vectors of ones of length $n_{\boldsymbol{z}_+}$ and $n_{\boldsymbol{x}_+}$ respectively. Therefore, model (71) is written as:

$$\boldsymbol{y}_+ = \gamma\mathbf{1}_{n_+} + (\mathbf{1}_{n_{\boldsymbol{x}_+}} \otimes \boldsymbol{B}_{\boldsymbol{z}_+})\boldsymbol{\theta}_{\boldsymbol{z}_+} + (\boldsymbol{B}_{\boldsymbol{x}_+} \otimes \mathbf{1}_{n_{\boldsymbol{z}_+}})\boldsymbol{\theta}_{\boldsymbol{x}_+} + (\boldsymbol{B}_{\boldsymbol{x}_+} \otimes \boldsymbol{B}_{\boldsymbol{z}_+})\boldsymbol{\theta}_{\boldsymbol{s}_+}, \qquad (73)$$

with extended penalty the following block-diagonal matrix:

$$\boldsymbol{P}_+ = \begin{bmatrix} 0 & & & \\ & \lambda_{\boldsymbol{z}}\boldsymbol{D}'_{\boldsymbol{z}_+}\boldsymbol{D}_{\boldsymbol{z}_+} & & \\ & & \lambda_{\boldsymbol{x}}\boldsymbol{D}'_{\boldsymbol{x}_+}\boldsymbol{D}_{\boldsymbol{x}_+} & \\ & & & \tau_{\boldsymbol{z}}\boldsymbol{I}_{c_{\boldsymbol{x}_+}} \otimes \boldsymbol{D}'_{\boldsymbol{z}_+}\boldsymbol{D}_{\boldsymbol{z}_+} + \tau_{\boldsymbol{x}}\boldsymbol{D}'_{\boldsymbol{x}_+}\boldsymbol{D}_{\boldsymbol{x}_+} \otimes \boldsymbol{I}_{c_{\boldsymbol{z}_+}} \end{bmatrix}, \qquad (74)$$

of dimension $(1 + c_{\boldsymbol{z}_+} + c_{\boldsymbol{x}_+} + c_{\boldsymbol{z}_+} c_{\boldsymbol{x}_+}) \times (1 + c_{\boldsymbol{z}_+} + c_{\boldsymbol{x}_+} + c_{\boldsymbol{z}_+} c_{\boldsymbol{x}_+})$, where each block corresponds to the penalty over each of the coefficients of the model.

To reformulate the extended P-spline S-ANOVA model (71) as a mixed model we need to define an extended transformation matrix. As in Section **??**, we can use the natural extended transformation matrix based on the SVD of the extended difference matrices, $\boldsymbol{\Omega}_+$, or an extended transformation matrix, $\boldsymbol{\Omega}_+^*$, that allow us to obtain extended mixed model matrices that are direct extensions of the model matrices that give the fit.

S-ANOVA model given in 62 is not identifiable but its mixed model representation allow ut to impose easily the necessary constraints. To obtain predictions with S-ANOVA models subject to the constraint that the fit is maintained we have to use the extended transformation matrix $\boldsymbol{\Omega}_+^*$. In the following two Sections we define the transformation matrices $\boldsymbol{\Omega}_+$ and $\boldsymbol{\Omega}_+^*$ and the model components associated to each case.

4.1 Natural reparametization of the S-ANOVA model into a mixed model for prediction

To reparameterize (71) as a mixed model the natural extended transformation matrix is $\boldsymbol{\Omega}_+ = [\boldsymbol{\Omega}_{+f} \mid \boldsymbol{\Omega}_{+r}]$ with dimension $(1 + c_{\boldsymbol{z}_+} + c_{\boldsymbol{x}_+} + c_{\boldsymbol{z}_+} c_{\boldsymbol{x}_+})$, where:

$$\boldsymbol{\Omega}_{+f} = \begin{bmatrix} 1 & & & \\ & 1 \otimes \boldsymbol{u}^{(2)}_{\boldsymbol{z}+f} & & \\ & & \boldsymbol{u}^{(2)}_{\boldsymbol{x}+f} \otimes 1 & \\ & & & \boldsymbol{u}^{(2)}_{\boldsymbol{x}+f} \otimes \boldsymbol{u}^{(2)}_{\boldsymbol{z}+f} \end{bmatrix},$$

$$\boldsymbol{\Omega}_{+r} = \begin{bmatrix} 1 & & & \\ & 1 \otimes \boldsymbol{U}_{\boldsymbol{z}+r} & & \\ & & \boldsymbol{U}_{\boldsymbol{x}+r} \otimes 1 & \\ & & & \boldsymbol{u}^{(2)}_{\boldsymbol{x}+f} \otimes \boldsymbol{U}_{\boldsymbol{z}+r} \mid \boldsymbol{U}_{\boldsymbol{x}+r} \otimes \boldsymbol{u}^{(2)}_{\boldsymbol{z}+f} \mid \boldsymbol{U}_{\boldsymbol{x}+r} \otimes \boldsymbol{U}_{\boldsymbol{z}+r} \end{bmatrix}, \quad (75)$$

where $\boldsymbol{u}^{(2)}_{\boldsymbol{z}+f}$ and $\boldsymbol{u}^{(2)}_{\boldsymbol{x}+f}$ are the second columns of $\boldsymbol{U}_{\boldsymbol{z}+f}$ and $\boldsymbol{U}_{\boldsymbol{x}+f}$, respectively, and $\boldsymbol{U}_{if}$ and $\boldsymbol{U}_{ir}$ are the eigenvectors corresponding to the zero values and positive values of the SVD of $\boldsymbol{D}_i'\boldsymbol{D}_i$, respectively, for $i = \boldsymbol{z}_+, \boldsymbol{x}_+$.

We obtain the fixed effects matrix as:

$$\boldsymbol{X}_+ = \boldsymbol{B}_+ \boldsymbol{\Omega}_{+f} = [\boldsymbol{1}_{n_+} \mid \boldsymbol{1}_{n_{\boldsymbol{x}_+}} \otimes \tilde{\boldsymbol{z}}_+ \mid \tilde{\boldsymbol{x}}_+ \otimes \boldsymbol{1}_{n_{\boldsymbol{z}_+}} \mid \tilde{\boldsymbol{x}}_+ \otimes \tilde{\boldsymbol{z}}_+], \quad (76)$$

where $\tilde{\boldsymbol{z}}_+$ and $\tilde{\boldsymbol{x}}_+$ are $\boldsymbol{B}_{\boldsymbol{z}_+} \boldsymbol{u}^{(2)}_{\boldsymbol{z}+f}$ and $\boldsymbol{B}_{\boldsymbol{x}_+} \boldsymbol{u}^{(2)}_{\boldsymbol{x}+f}$, respectively. The random effects matrix is obtained is:

$$\boldsymbol{Z}_+ = \boldsymbol{B}_+ \boldsymbol{\Omega}_{+r} = [\boldsymbol{1}_{n_+} \mid \boldsymbol{1}_{n_{\boldsymbol{x}_+}} \otimes \boldsymbol{Z}_{\boldsymbol{z}_+} \mid \boldsymbol{Z}_{\boldsymbol{x}_+} \otimes \boldsymbol{1}_{n_{\boldsymbol{z}_+}} \mid \tilde{\boldsymbol{x}}_+ \otimes \boldsymbol{Z}_{\boldsymbol{z}_+} \mid \boldsymbol{Z}_{\boldsymbol{x}_+} \otimes \tilde{\boldsymbol{z}}_+ \mid \boldsymbol{Z}_{\boldsymbol{x}_+} \otimes \boldsymbol{Z}_{\boldsymbol{z}_+}], \quad (77)$$

where $\boldsymbol{Z}_{\boldsymbol{z}_+}$ and $\boldsymbol{Z}_{\boldsymbol{x}_+}$ are $\boldsymbol{B}_{\boldsymbol{z}_+} \boldsymbol{U}_{\boldsymbol{z}+r}$ and $\boldsymbol{B}_{\boldsymbol{x}_+} \boldsymbol{U}_{\boldsymbol{x}+r}$, respectively.

For the extended transformation matrix given in (75) and extended penalty given in (74) the mixed model precision matrix of the random effects is given by the following theorem.

Theorem 4. *The extended precision matrix of random effects for S-ANOVA model in (71) with extended transformation matrix given in (75) and extended penalty given in (74) is the block-diagonal defined by:*

$$\boldsymbol{F}_+ = \text{blockdiag}(\boldsymbol{F}_+^{(1)}, \boldsymbol{F}_+^{(2)}, \boldsymbol{F}_+^{(1,2)}), \tag{78}$$

where for a second order penalty, it has size $(c_{\boldsymbol{z}_+}c_{\boldsymbol{x}_+} - 4) \times (c_{\boldsymbol{z}_+}c_{\boldsymbol{x}_+} - 4)$*, and where:*

$$\begin{aligned}
\boldsymbol{F}_+^{(1)} &= \lambda_{\boldsymbol{z}}\tilde{\boldsymbol{\Sigma}}_{\boldsymbol{z}_+},\\
\boldsymbol{F}_+^{(2)} &= \lambda_{\boldsymbol{x}}\tilde{\boldsymbol{\Sigma}}_{\boldsymbol{x}_+},\\
\boldsymbol{F}_+^{(1,2)} &= \text{blockdiag}(\tau_{\boldsymbol{z}}\tilde{\boldsymbol{\Sigma}}_{\boldsymbol{z}_+}, \tau_{\boldsymbol{x}}\tilde{\boldsymbol{\Sigma}}_{\boldsymbol{x}_+}, \tau_{\boldsymbol{z}}\boldsymbol{I}_{c_{\boldsymbol{z}_+}-2} \otimes \tilde{\boldsymbol{\Sigma}}_{\boldsymbol{z}_+} + \tau_{\boldsymbol{x}}\tilde{\boldsymbol{\Sigma}}_{\boldsymbol{x}_+} \otimes \boldsymbol{I}_{c_{\boldsymbol{x}_+}-2}).
\end{aligned}$$

where $\tilde{\boldsymbol{\Sigma}}_{\boldsymbol{z}_+}$ *and* $\tilde{\boldsymbol{\Sigma}}_{\boldsymbol{x}_+}$ *are the nonzero eigenvalues of the SVD of* $\boldsymbol{D}'_{\boldsymbol{z}_+}\boldsymbol{D}_{\boldsymbol{z}_+}$ *and* $\boldsymbol{D}'_{\boldsymbol{x}_+}\boldsymbol{D}_{\boldsymbol{x}_+}$*, respectively. The random effects covariance matrix is therefore* $\boldsymbol{G}_+ = \sigma_{\boldsymbol{\epsilon}}^2\boldsymbol{F}_+^{-1}$*.*

The proof of the previous Theorem is given in Appendix .3.

Once the model components are defined, the fit and the prediction are obtained simultaneously, the estimation of the fixed and random effects and the variance components would also be carried out by solving the extended Henderson system of equations (37) and maximizing the extended REML (39).

In the case in which just one covariate is extended, the models components are:

- Extended mixed model matrices:

$$\boldsymbol{X}_+ = \boldsymbol{B}_+\boldsymbol{\Omega}_{+f} = [\boldsymbol{1}_{n_+} \mid \boldsymbol{1}_{n_{\boldsymbol{x}_+}} \otimes \tilde{\boldsymbol{z}} \mid \tilde{\boldsymbol{x}}_+ \otimes \boldsymbol{1}_{n_{\boldsymbol{z}}} \mid \tilde{\boldsymbol{x}}_+ \otimes \tilde{\boldsymbol{z}}], \tag{79}$$

  where $\tilde{\boldsymbol{z}}$ and $\tilde{\boldsymbol{x}}_+$ are $\boldsymbol{B}_{\boldsymbol{z}}\boldsymbol{u}_{\boldsymbol{z}f}^{(2)}$ and $\boldsymbol{B}_{\boldsymbol{x}_+}\boldsymbol{u}_{\boldsymbol{x}_+f}^{(2)}$, respectively. And

$$\boldsymbol{Z}_+ = \boldsymbol{Z}_+ = \boldsymbol{B}_+\boldsymbol{\Omega}_{+r} = [\boldsymbol{1}_{n_+} \mid \boldsymbol{1}_{n_{\boldsymbol{x}_+}} \otimes \boldsymbol{Z}_{\boldsymbol{z}} \mid \boldsymbol{Z}_{\boldsymbol{x}_+}\boldsymbol{U}_{\boldsymbol{x}_+r} \mid \boldsymbol{Z}_{\boldsymbol{x}_+} \otimes \boldsymbol{Z}_{\boldsymbol{z}}], \tag{80}$$

  where $\boldsymbol{Z}_{\boldsymbol{z}}$ and $\boldsymbol{Z}_{\boldsymbol{x}_+}$ are $\boldsymbol{B}_{\boldsymbol{z}}\boldsymbol{U}_{\boldsymbol{z}r}$ and $\boldsymbol{B}_{\boldsymbol{x}_+}\boldsymbol{U}_{\boldsymbol{x}_+r}$, respectively.

- Extended random effects covariance matrix $\boldsymbol{G}_+ = \sigma_{\boldsymbol{\epsilon}}^2\boldsymbol{F}_+^{-1}$, with:

$$\boldsymbol{F}_+ = \text{blockdiag}(\boldsymbol{F}^{(1)}, \boldsymbol{F}_+^{(2)}, \boldsymbol{F}_+^{(1,2)}), \tag{81}$$

  where for a second order penalty, $q_{\boldsymbol{x}} = q_{\boldsymbol{z}} = 2$, it has size $(c_{\boldsymbol{z}}c_{\boldsymbol{x}_+} - 4) \times (c_{\boldsymbol{z}}c_{\boldsymbol{x}_+} - 4)$, and where:

$$\begin{aligned}
\boldsymbol{F}^{(1)} &= \lambda_{\boldsymbol{z}}\tilde{\boldsymbol{\Sigma}}_{\boldsymbol{z}},\\
\boldsymbol{F}_+^{(2)} &= \lambda_{\boldsymbol{x}}\tilde{\boldsymbol{\Sigma}}_{\boldsymbol{x}_+},\\
\boldsymbol{F}_+^{(1,2)} &= \text{blockdiag}(\tau_{\boldsymbol{z}}\tilde{\boldsymbol{\Sigma}}_{\boldsymbol{z}}, \tau_{\boldsymbol{x}}\tilde{\boldsymbol{\Sigma}}_{\boldsymbol{x}_+}, \tau_{\boldsymbol{z}}\boldsymbol{I}_{c_{\boldsymbol{z}}-q_{\boldsymbol{z}}} \otimes \tilde{\boldsymbol{\Sigma}}_{\boldsymbol{z}} + \tau_{\boldsymbol{x}}\tilde{\boldsymbol{\Sigma}}_{\boldsymbol{x}_+} \otimes \boldsymbol{I}_{c_{\boldsymbol{x}_+}-q_{\boldsymbol{x}}}),
\end{aligned}$$

  with $\tilde{\boldsymbol{\Sigma}}_{\boldsymbol{z}}$ and $\tilde{\boldsymbol{\Sigma}}_{\boldsymbol{x}_+}$ the nonzero eigenvalues of the SVD of $\boldsymbol{D}'_{\boldsymbol{z}}\boldsymbol{D}_{\boldsymbol{z}}$ and $\boldsymbol{D}'_{\boldsymbol{x}_+}\boldsymbol{D}_{\boldsymbol{x}_+}$, respectively.

However, the method presented above does not ensure the invariance of the fit, i.e., if out-of-sample prediction is performed.

4.2 Coherent prediction with S-ANOVA model

To predict with S-ANOVA models subject to the constraint that the fit has to be maintained an extended transformation matrix that preserves the model matrices has to be used. For the case in which the two covariates are extended, we define the following extended transformation matrix $\boldsymbol{\Omega}^*_+ = [\boldsymbol{\Omega}^*_{+f} \mid \boldsymbol{\Omega}^*_{+r}]$ with dimension $(1+c_{\boldsymbol{z}+}+c_{\boldsymbol{x}+}+c_{\boldsymbol{z}+}c_{\boldsymbol{x}+})$, and:

$$\boldsymbol{\Omega}^*_{+f} = \begin{bmatrix} 1 & & & \\ & 1 \otimes \boldsymbol{u}^{*(2)}_{\boldsymbol{z}+f} & & \\ & & \boldsymbol{u}^{*(2)}_{\boldsymbol{x}+f} \otimes 1 & \\ & & & \boldsymbol{u}^{*(2)}_{\boldsymbol{x}+f} \otimes \boldsymbol{u}^{*(2)}_{\boldsymbol{z}+f} \end{bmatrix},$$

$$\boldsymbol{\Omega}^*_{+r} = \begin{bmatrix} 1 & & & \\ & 1 \otimes \boldsymbol{U}^*_{\boldsymbol{z}+r} & & \\ & & \boldsymbol{U}^*_{\boldsymbol{x}+r} \otimes 1 & \\ & & & \boldsymbol{u}^{*(2)}_{\boldsymbol{x}+f} \otimes \boldsymbol{U}^*_{\boldsymbol{z}+r} \mid \boldsymbol{U}^*_{\boldsymbol{x}+r} \otimes \boldsymbol{u}^{*(2)}_{\boldsymbol{z}+f} \mid \boldsymbol{U}^*_{\boldsymbol{x}+r} \otimes \boldsymbol{U}^*_{\boldsymbol{z}+r} \end{bmatrix}, \quad (82)$$

where $\boldsymbol{u}^{*(2)}_{\boldsymbol{z}+f}$ and $\boldsymbol{u}^{*(2)}_{\boldsymbol{x}+f}$ are the second columns of $\boldsymbol{U}^*_{\boldsymbol{z}+f}$ and $\boldsymbol{U}^*_{\boldsymbol{x}+f}$, respectively, $\boldsymbol{U}^*_{\boldsymbol{z}+f} = \begin{bmatrix} \boldsymbol{U}_{\boldsymbol{z}f} \\ -\boldsymbol{D}^{-1}_{\boldsymbol{z}(2)}\boldsymbol{D}_{\boldsymbol{z}(1)}\boldsymbol{U}_{\boldsymbol{z}f} \end{bmatrix}$, with $\boldsymbol{U}_{\boldsymbol{z}f}$ the eigenvectors corresponding to the positive values of the SVD of $\boldsymbol{D}'_{\boldsymbol{z}}\boldsymbol{D}_{\boldsymbol{z}}$ and $\boldsymbol{D}_{\boldsymbol{z}(2)}$ and $\boldsymbol{D}_{\boldsymbol{z}(1)}$ blocks of and (16) $\boldsymbol{U}^*_{\boldsymbol{z}+f}$ are the eigenvectors corresponding to the positive values of the SVD of $\boldsymbol{D}'_{\boldsymbol{z}+}\boldsymbol{D}_{\boldsymbol{z}+}$ and $\boldsymbol{U}^*_{\boldsymbol{x}+f} = \begin{bmatrix} \boldsymbol{U}_{\boldsymbol{x}f} \\ -\boldsymbol{D}^{-1}_{\boldsymbol{x}(2)}\boldsymbol{D}_{\boldsymbol{x}(1)}\boldsymbol{U}_{\boldsymbol{x}f} \end{bmatrix}$, with $\boldsymbol{U}_{\boldsymbol{x}f}$ the eigenvectors corresponding to the positive values of the SVD of $\boldsymbol{D}'_{\boldsymbol{x}}\boldsymbol{D}_{\boldsymbol{x}}$ and $\boldsymbol{D}_{\boldsymbol{x}(2)}$ and $\boldsymbol{D}_{\boldsymbol{x}(1)}$ blocks of (17).

The fixed effects matrix is:

$$\boldsymbol{X}^*_+ = \boldsymbol{B}_+\boldsymbol{\Omega}^*_{+f} = [\boldsymbol{1}_{n_+} \mid \boldsymbol{1}_{n_{\boldsymbol{x}_+}} \otimes \tilde{\boldsymbol{z}}_+ \mid \tilde{\boldsymbol{x}}_+ \otimes \boldsymbol{1}_{n_{\boldsymbol{z}_+}} \mid \tilde{\boldsymbol{x}}_+ \otimes \tilde{\boldsymbol{z}}_+], \quad (83)$$

where $\tilde{\boldsymbol{z}}_+$ and $\tilde{\boldsymbol{x}}_+$ are $\boldsymbol{B}_{\boldsymbol{z}+}\boldsymbol{u}^{*(2)}_{\boldsymbol{z}f}$ and $\boldsymbol{B}_{\boldsymbol{x}+}\boldsymbol{u}^{*(2)}_{\boldsymbol{x}+f}$, respectively, and the random effects matrix is:

$$\boldsymbol{Z}^*_+ = \boldsymbol{B}_+\boldsymbol{\Omega}^*_{+r} = [\boldsymbol{1}_{n_+} \mid \boldsymbol{1}_{n_{\boldsymbol{x}_+}} \otimes \boldsymbol{Z}^*_{\boldsymbol{z}+} \mid \boldsymbol{Z}^*_{\boldsymbol{x}+} \otimes \boldsymbol{1}_{n_{\boldsymbol{z}_+}} \mid \tilde{\boldsymbol{x}}_+ \otimes \boldsymbol{Z}^*_{\boldsymbol{z}+} \mid \boldsymbol{Z}^*_{\boldsymbol{x}+} \otimes \tilde{\boldsymbol{z}}_+ \mid \boldsymbol{Z}^*_{\boldsymbol{x}+} \otimes \boldsymbol{Z}^*_{\boldsymbol{z}+}], \quad (84)$$

where $\boldsymbol{Z}^*_{\boldsymbol{z}+}$ and $\boldsymbol{Z}^*_{\boldsymbol{x}+}$ are $\boldsymbol{B}_{\boldsymbol{z}+}\boldsymbol{U}^*_{\boldsymbol{z}+r}$ and $\boldsymbol{B}_{\boldsymbol{x}+}\boldsymbol{U}^*_{\boldsymbol{x}+r}$, respectively. The covariance matrix of random effects is $\boldsymbol{G}^*_+ = \sigma^2_{\boldsymbol{\epsilon}}\boldsymbol{F}^{*-1}_+$, with $\boldsymbol{F}^*_+$ given by the following Theorem.

Theorem 5. *The extended precision matrix of random effects for S-ANOVA model in (71) with extended transformation matrix given in (82) and extended penalty given in (74) is the block-diagonal defined by:*

$$\boldsymbol{F}^*_+ = \text{blockdiag}(\boldsymbol{F}^{(1)}, \boldsymbol{F}^{(2)}_+, \boldsymbol{F}^{(1,2)}_+), \quad (85)$$

where for a second order penalty, it has size $(c_{\boldsymbol{z}} c_{\boldsymbol{x}_+} - 4) \times (c_{\boldsymbol{z}} c_{\boldsymbol{x}_+} - 4)$*, and where:*

$$\begin{aligned}
\boldsymbol{F}^{(1)} &= \lambda_{\boldsymbol{z}} \boldsymbol{U}^{*\prime}_{\boldsymbol{z}+r} \boldsymbol{D}'_{\boldsymbol{z}_+} \boldsymbol{D}_{\boldsymbol{z}_+} \boldsymbol{U}^{*}_{\boldsymbol{z}+r}, \\
\boldsymbol{F}^{(2)}_{+} &= \lambda_{\boldsymbol{x}} \boldsymbol{U}^{*\prime}_{\boldsymbol{x}+r} \boldsymbol{D}'_{\boldsymbol{x}_+} \boldsymbol{D}_{\boldsymbol{x}_+} \boldsymbol{U}^{*}_{\boldsymbol{x}+r}, \\
\boldsymbol{F}^{(1,2)}_{+} &= \begin{bmatrix} \boldsymbol{F}^{(1,2)}_{+11} & \boldsymbol{O} & \boldsymbol{F}^{(1,2)}_{+13} \\ \boldsymbol{O} & \boldsymbol{F}^{(1,2)}_{+22} & \boldsymbol{F}^{(1,2)}_{+23} \\ \boldsymbol{F}^{(1,2)\prime}_{+13} & \boldsymbol{F}^{(1,2)\prime}_{+23} & \boldsymbol{F}^{(1,2)}_{+33} \end{bmatrix},
\end{aligned}$$

with

$$\begin{aligned}
\boldsymbol{F}^{(1,2)}_{+11} &= \tau_{\boldsymbol{z}} \boldsymbol{u}^{*(2)\prime}_{\boldsymbol{x}+f} \boldsymbol{u}^{*(2)}_{\boldsymbol{x}+f} \otimes \boldsymbol{U}^{*\prime}_{\boldsymbol{z}+r} \boldsymbol{D}'_{\boldsymbol{z}_+} \boldsymbol{D}_{\boldsymbol{z}_+} \boldsymbol{U}^{*}_{\boldsymbol{z}+r} \\
\boldsymbol{F}^{(1,2)}_{+13} &= \tau_{\boldsymbol{z}} \boldsymbol{u}^{*(2)\prime}_{\boldsymbol{x}+f} \boldsymbol{U}^{*}_{\boldsymbol{x}+r} \otimes \boldsymbol{U}^{*\prime}_{\boldsymbol{z}+r} \boldsymbol{D}'_{\boldsymbol{z}_+} \boldsymbol{D}_{\boldsymbol{z}_+} \boldsymbol{U}^{*}_{\boldsymbol{z}+r}, \\
\boldsymbol{F}^{(1,2)}_{+22} &= \tau_{\boldsymbol{x}} \boldsymbol{U}^{*\prime}_{\boldsymbol{x}+r} \boldsymbol{D}'_{\boldsymbol{x}_+} \boldsymbol{D}_{\boldsymbol{x}_+} \boldsymbol{U}^{*}_{\boldsymbol{x}+r} \otimes \boldsymbol{u}^{*(2)\prime}_{\boldsymbol{z}+r} \boldsymbol{u}^{*(2)}_{\boldsymbol{z}+r}, \\
\boldsymbol{F}^{(1,2)}_{+23} &= \tau_{\boldsymbol{x}} \boldsymbol{U}^{*\prime}_{\boldsymbol{x}+r} \boldsymbol{D}'_{\boldsymbol{x}_+} \boldsymbol{D}_{\boldsymbol{x}_+} \boldsymbol{U}^{*}_{\boldsymbol{x}+r} \otimes \boldsymbol{u}^{*(2)\prime}_{\boldsymbol{z}+f} \boldsymbol{U}^{*}_{\boldsymbol{z}+r}. \\
\boldsymbol{F}^{(1,2)}_{+33} &= \tau_{\boldsymbol{z}} \boldsymbol{U}^{*\prime}_{\boldsymbol{x}+r} \boldsymbol{U}^{*}_{\boldsymbol{x}+r} \otimes \boldsymbol{U}^{*\prime}_{\boldsymbol{z}+r} \boldsymbol{D}'_{\boldsymbol{z}_+} \boldsymbol{D}_{\boldsymbol{z}_+} \boldsymbol{U}^{*}_{\boldsymbol{z}+r} + \tau_{\boldsymbol{x}} \boldsymbol{U}^{*\prime}_{\boldsymbol{x}+r} \boldsymbol{D}'_{\boldsymbol{x}_+} \boldsymbol{D}_{\boldsymbol{x}_+} \boldsymbol{U}^{*}_{\boldsymbol{x}+r} \otimes \boldsymbol{U}^{*\prime}_{\boldsymbol{z}+r} \boldsymbol{U}^{*}_{\boldsymbol{z}+r}.
\end{aligned}$$

The proof is given in Appendix .4.

If just one covariate is extended the extended mixed model components are:

- Extended mixed model matrices:

$$\boldsymbol{X}^{*}_{+} = [\boldsymbol{1}_{n_+} \mid \boldsymbol{1}_{n_{\boldsymbol{x}_+}} \otimes \tilde{\boldsymbol{z}} \mid \tilde{\boldsymbol{x}}_+ \otimes \boldsymbol{1}_{n_{\boldsymbol{z}}} \mid \tilde{\boldsymbol{x}}_+ \otimes \tilde{\boldsymbol{z}}], \tag{86}$$

  where $\tilde{\boldsymbol{z}}$ and $\tilde{\boldsymbol{x}}_+$ are $\boldsymbol{B}_{\boldsymbol{z}} \boldsymbol{u}^{(2)}_{\boldsymbol{z}f}$ and $\boldsymbol{B}_{\boldsymbol{x}_+} \boldsymbol{u}^{*(2)}_{\boldsymbol{x}+f}$, respectively. The random effects matrix is $\boldsymbol{Z}^{*}_{+} = \boldsymbol{B}_{+} \boldsymbol{\Omega}^{*}_{+r}$, i.e.:

$$\boldsymbol{Z}^{*}_{+} = [\boldsymbol{1}_{n_+} \mid \boldsymbol{1}_{n_{\boldsymbol{x}_+}} \otimes \boldsymbol{Z}_{\boldsymbol{z}} \mid \boldsymbol{Z}^{*}_{\boldsymbol{x}_+} \otimes \boldsymbol{1}_{n_{\boldsymbol{z}}} \mid \boldsymbol{Z}^{*}_{\boldsymbol{x}_+} \otimes \boldsymbol{Z}_{\boldsymbol{z}}], \tag{87}$$

  where $\boldsymbol{Z}_{\boldsymbol{z}}$ and $\boldsymbol{Z}_{\boldsymbol{x}_+}$ are $\boldsymbol{B}_{\boldsymbol{z}} \boldsymbol{U}_{\boldsymbol{z}r}$ and $\boldsymbol{B}_{\boldsymbol{x}_+} \boldsymbol{U}^{*}_{\boldsymbol{x}+r}$, respectively.

- Extended random effects covariance matrix $\boldsymbol{G}^{*}_{+} = \sigma^2_{\boldsymbol{\epsilon}} \boldsymbol{F}^{*-1}_{+}$ with

$$\boldsymbol{F}^{*}_{+} = \text{blockdiag}(\boldsymbol{F}^{(1)}, \boldsymbol{F}^{(2)}_{+}, \boldsymbol{F}^{(1,2)}_{+}), \tag{88}$$

  where for a second order penalty, $q_{\boldsymbol{x}} = q_{\boldsymbol{z}} = 2$, it has size $(c_{\boldsymbol{z}} c_{\boldsymbol{x}_+} - 4) \times (c_{\boldsymbol{z}} c_{\boldsymbol{x}_+} - 4)$, and where:

$$\begin{aligned}
\boldsymbol{F}^{(1)} &= \lambda_{\boldsymbol{z}} \tilde{\boldsymbol{\Sigma}}_{\boldsymbol{z}}, \\
\boldsymbol{F}^{(2)}_{+} &= \lambda_{\boldsymbol{x}} \boldsymbol{U}^{*\prime}_{\boldsymbol{x}+r} \boldsymbol{D}'_{\boldsymbol{x}_+} \boldsymbol{D}_{\boldsymbol{x}_+} \boldsymbol{U}^{*}_{\boldsymbol{x}+r}, \\
\boldsymbol{F}^{(1,2)}_{+} &= \begin{bmatrix} \boldsymbol{F}^{(1,2)}_{+11} & \boldsymbol{O} & \boldsymbol{F}^{(1,2)}_{+13} \\ \boldsymbol{O} & \boldsymbol{F}^{(1,2)}_{+22} & \boldsymbol{O} \\ \boldsymbol{F}^{(1,2)\prime}_{+13} & \boldsymbol{O} & \boldsymbol{F}^{(1,2)}_{+33} \end{bmatrix},
\end{aligned}$$

  with

$$\begin{aligned}
\boldsymbol{F}^{(1,2)}_{+11} &= \tau_{\boldsymbol{z}} \boldsymbol{u}^{*(2)\prime}_{\boldsymbol{x}+f} \boldsymbol{u}^{*(2)}_{\boldsymbol{x}+f} \otimes \tilde{\boldsymbol{\Sigma}}_{\boldsymbol{z}} \\
\boldsymbol{F}^{(1,2)}_{+13} &= \tau_{\boldsymbol{z}} \boldsymbol{u}^{*(2)\prime}_{\boldsymbol{x}+f} \boldsymbol{U}^{*}_{\boldsymbol{x}+r} \otimes \tilde{\boldsymbol{\Sigma}}_{\boldsymbol{z}}, \\
\boldsymbol{F}^{(1,2)}_{+22} &= \tau_{\boldsymbol{x}} \boldsymbol{U}^{*\prime}_{\boldsymbol{x}+r} \boldsymbol{D}'_{\boldsymbol{x}_+} \boldsymbol{D}_{\boldsymbol{x}_+} \boldsymbol{U}^{*}_{\boldsymbol{x}+r} \otimes \boldsymbol{I}_{q_{\boldsymbol{z}}}, \\
\boldsymbol{F}^{(1,2)}_{+33} &= \tau_{\boldsymbol{z}} \boldsymbol{U}^{*\prime}_{\boldsymbol{x}+r} \boldsymbol{U}^{*}_{\boldsymbol{x}+r} \otimes \tilde{\boldsymbol{\Sigma}}_{\boldsymbol{z}} + \tau_{\boldsymbol{x}} \boldsymbol{U}^{*\prime}_{\boldsymbol{x}+r} \boldsymbol{D}'_{\boldsymbol{x}_+} \boldsymbol{D}_{\boldsymbol{x}_+} \boldsymbol{U}^{*}_{\boldsymbol{x}+r} \otimes \boldsymbol{I}_{c_{\boldsymbol{z}} - q_{\boldsymbol{z}}}.
\end{aligned}$$

  where $\tilde{\boldsymbol{\Sigma}}_{\boldsymbol{z}}$ is the nonzero eigenvalues of the SVD of $\boldsymbol{D}'_{\boldsymbol{z}} \boldsymbol{D}_{\boldsymbol{z}}$.

4.3 Example 2: Prediction of aboveground biomass

In this section, we apply the proposed 2D interaction P-spline and S-ANOVA models to a real data set. The data came from three trials carried out in Spain (Rivas-Martínez *et al.* (2002)), where nine clones belonging to the European Catalog of Basic Materials for the Populus genes were included. The data set was also used in Sánchez-González *et al.* (2016). They point out that aboveground biomass estimation in short-rotation forestry plantations is essential to determine the economic viability of the crop prior to harvesting. Hence, it is important to obtain accurate predictions using only a minimum set of easily obtainable information i.e., height and diameter (denoted by $\boldsymbol{z}$ and $\boldsymbol{x}$). We propose to predict out-of-sample aboveground biomass as a smooth function of height and diameter using the following two models:

- 2D interaction P-spline model:

$$\boldsymbol{y}_+ = f(\boldsymbol{z}_+, \boldsymbol{x}_+) + \boldsymbol{\epsilon}_+, \quad \boldsymbol{\epsilon}_+ \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{R}_+)$$

- Smooth-ANOVA model:

$$\boldsymbol{y}_+ = f(\boldsymbol{z}_+) + f(\boldsymbol{x}_+) + f(\boldsymbol{z}_+, \boldsymbol{x}_+) + \boldsymbol{\epsilon}_+, \quad \boldsymbol{\epsilon}_+ \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{R}_+)$$

From the original data we have selected the clone *verde* and predicted the weight for diameter and height out-of-sample values. The observed data consists of 315 observations, for diameter values measured at 1.30 m breast height within the interval $[0.2, 7.3]$ and height values measured within the interval $[1.35, 9.32]$.

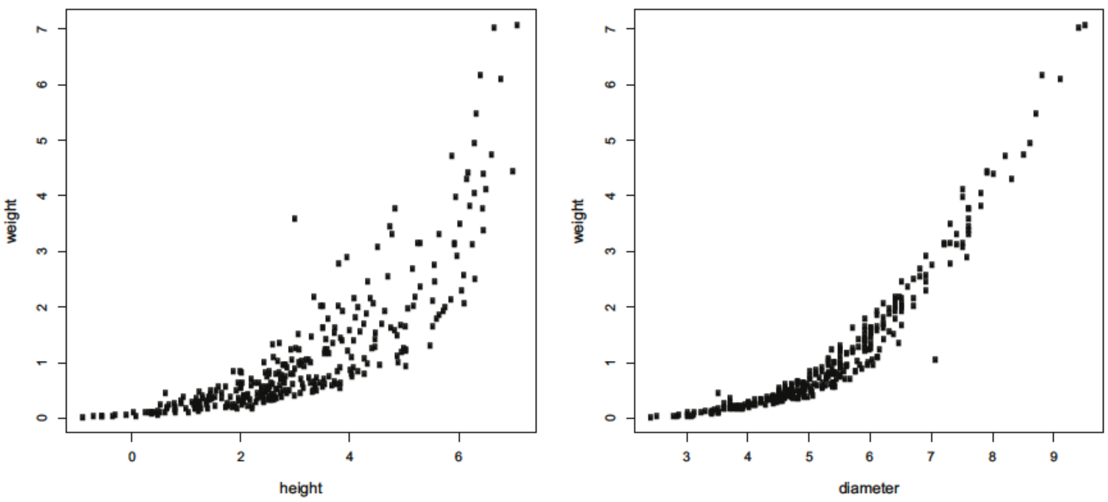


Figure 4: Plot of weight versus height (left panel) and plot of total weight versus diameter (right panel).

We have predicted weight for 10 new out-of-sample values for diameter and height. In Figure 5 we plot the smooth trend for height (left panel) and for diameter (right panel) obtained after fitting and predicting with the S-ANOVA model imposing that the fit is maintained. As it shown, the effect of diameter is stronger than height, but both smooth terms are significantly different from zero. Figure 6 shows the fitted and predicted interaction function ($f(\boldsymbol{z}_+, \boldsymbol{x}_+)$) for the restricted S-ANOVA model.

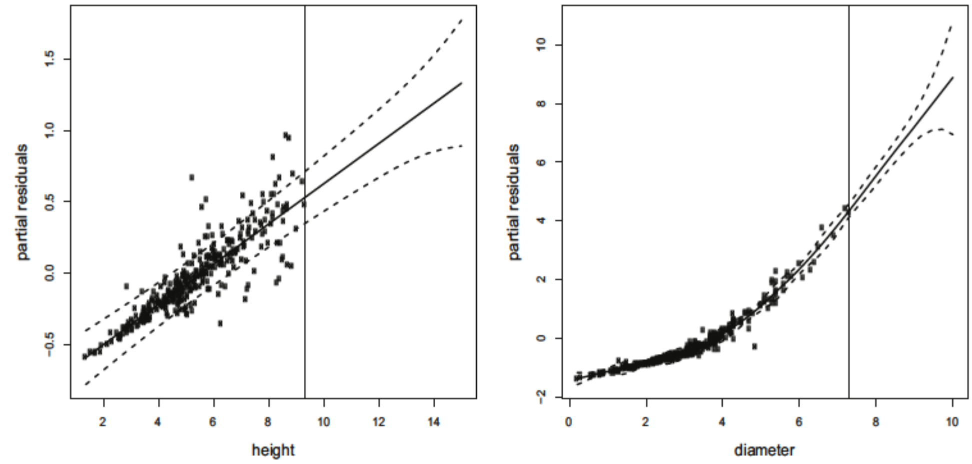


Figure 5: Fitted and predicted smooth curves for height (left panel) and for diameter (right panel) using the restricted S-ANOVA model. The vertical line indicates the height and diameter values from which we predict (9.32 and 7.3).

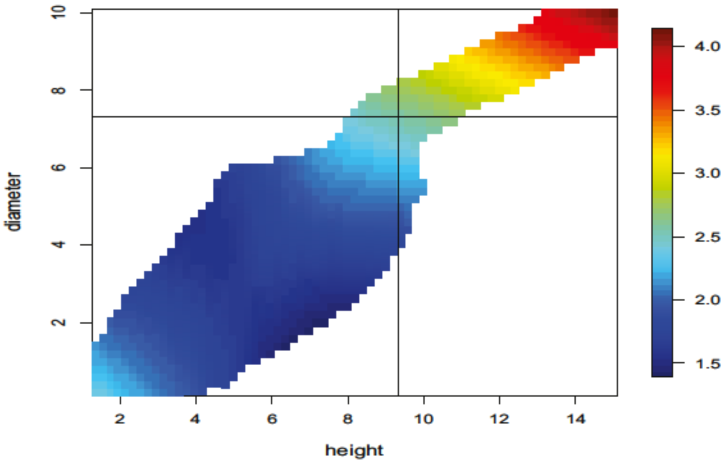


Figure 6: Fitted and predicted interaction function for the the restricted S-ANOVA model. The vertical line indicates the height value from which we predict (9.32) and the horizontal line indicates the diameter value from which we predict (7.3).

Figure 7 illustrates the solution when the fit and the forecast are obtained with the S-ANOVA model (top left panel), with the S-ANOVA model imposing that the fit is maintained (top right panel), with the 2D interaction P-spline model (bottom left panel) and with the 2D interaction P-spline models imposing that the fit is maintained (bottom right panel). As the figure shows, the results obtained from the S-ANOVA model and from the restricted S-ANOVA model are almost equal. However, the solutions obtained from the 2D interaction P-spline models are different depending on if the restriction the fit is maintained is imposed or not. As we can appreciate, the fit changes significantly if the restriction is not imposed. In the prediction, the most significant difference is that the 2D interaction P-spline model gives lower weight values for the largest diameter and height values than the 2D interaction P-spline model.

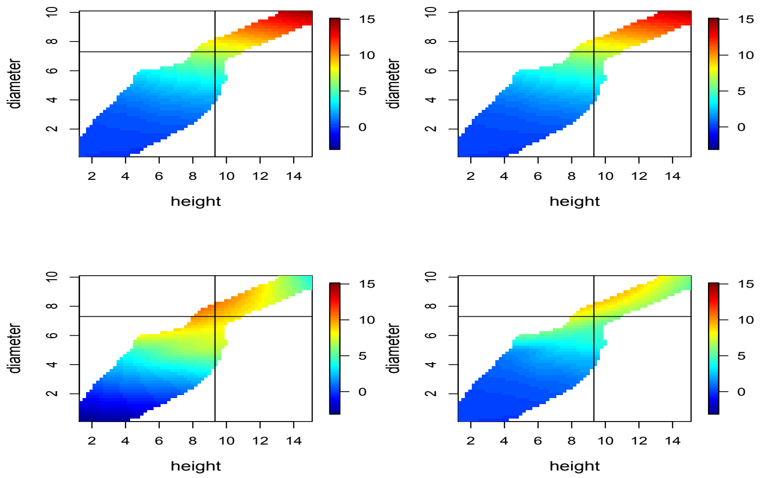


Figure 7: Fit and prediction with the S-ANOVA model (top left panel), with the S-ANOVA model imposing that the fit is maintained (top right panel), with the 2D interaction P-spline model (bottom left panel) and with the 2D interaction P-spline models imposing that the fit is maintained (bottom right panel) at out-of-sample values of diameter ($[7.3, 10]$) and height ($[9.32, 15]$).

Comparing the S-ANOVA models and the 2D interaction P-spline models, we conclude that the most coherent solution is given by the S-ANOVA models, since it gives the largest weight values for the highest values of height and diameter.

5 Simulation study

In previous sections, we have shown how to predict with interaction models (from P-spline and mixed models points of view) and with Smooth-ANOVA models, and how to impose restrictions. In this Section, we examine the performance of the interaction and Smooth-ANOVA models in comparison to interaction and Smooth-ANOVA models in which we impose the constraint that the fit has to be the same as the fit we obtain when only fitting the data. For this propose we have simulated the data in two different scenarios:

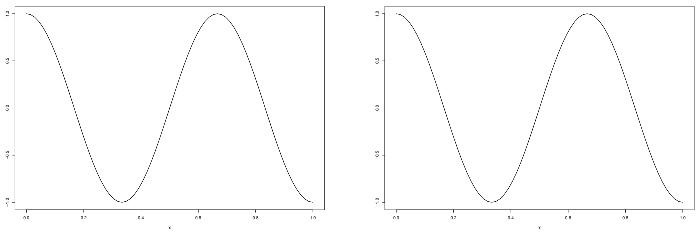
a) Scenario 1. From an interaction model:

$$\boldsymbol{S}^1 = f_{1,2}(\boldsymbol{z}, \boldsymbol{x}).$$

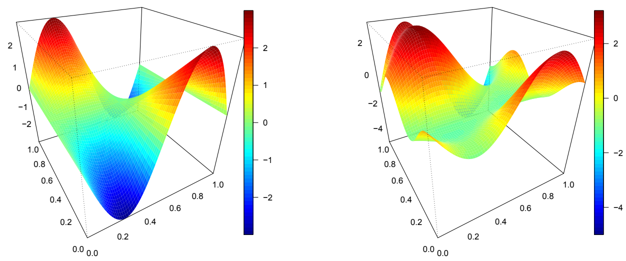
b) Scenario 2. From a two main effects with interaction model:

$$\boldsymbol{S}^2 = f_1(\boldsymbol{z}) + f_2(\boldsymbol{x}) + f_{1,2}(\boldsymbol{z}, \boldsymbol{x}).$$

In both cases $f_1(\boldsymbol{z}) = \sin(2\pi \boldsymbol{z})$, $f_2(\boldsymbol{x}) = \cos(3\pi \boldsymbol{x})$ and $f_{1,2}(\boldsymbol{z}, \boldsymbol{x}) = 3\,\sin(2\pi \boldsymbol{z})(2\boldsymbol{z} - 1)$. To simulate the data we have generated a grid of 4900 values. Both covariates, $\boldsymbol{z}$ and $\boldsymbol{x}$, take 70 equidistant values in the interval $[0, 1]$, and the errors are independent and identically distributed, with mean 0 and variance $\sigma^2_\epsilon = 0.25$. Figure 8 shows $f_1(\boldsymbol{z})$, $f_2(\boldsymbol{x})$ and the surfaces proposed in scenarios 1 and 2.



(a) $f_1(\boldsymbol{z})$ (b) $f_2(\boldsymbol{x})$



(c) $f_{1,2}(\boldsymbol{z}, \boldsymbol{x})$ (d) $f_1(\boldsymbol{z}) + f_2(\boldsymbol{x}) + f_{1,2}(\boldsymbol{z}, \boldsymbol{x})$

Figure 8: Functions (a) and (b) are the nonlinear main effects of $\boldsymbol{z}$ and $\boldsymbol{x}$, (c) is the interaction surface and (d) is the sum of the main effects and the interaction surfaces

For each scenario, we fit and predict with four models, the models and their components are listed below:

- 2D interaction P-spline model, i.e.

$$\begin{aligned} \boldsymbol{y}_+ &= f(\boldsymbol{z}_+, \boldsymbol{x}_+) + \boldsymbol{\epsilon}_+ \\ &= \boldsymbol{X}_+\boldsymbol{\beta}_{\tilde{+}} + \boldsymbol{Z}_+\boldsymbol{\alpha}_+ + \boldsymbol{\epsilon}_+, \quad \boldsymbol{\epsilon}_+ \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{R}_+), \quad \boldsymbol{\alpha}_+ \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{G}_+), \end{aligned}$$

  where the model components depend on how many covariates are extended:

  - Extending one covariate, the model components are $\boldsymbol{X}_+$ and $\boldsymbol{Z}_+$ defined in (46) and (47), $\boldsymbol{R}_+$ defined in (18) and $\boldsymbol{G}_+$ defined in (49).
  - Extending two covariates, the model components are $\boldsymbol{X}_+$ and $\boldsymbol{Z}_+$ defined in (41) and (42), $\boldsymbol{R}_+$ defined in (12) and $\boldsymbol{G}_+$ defined in (44).

  In both cases, after defining the model components, to estimate the model we maximize the extended REML (39) and solve the extended mixed model system of equations of Henderson (37) to estimate the model parameters.

- 2D interaction P-spline model with restriction, i.e.

$$\begin{aligned} \boldsymbol{y}_+ &= f(\boldsymbol{z}_+, \boldsymbol{x}_+) + \boldsymbol{\epsilon}_+ \\ &= \boldsymbol{X}_+\boldsymbol{\beta}^*_{\tilde{+}} + \boldsymbol{Z}_+\boldsymbol{\alpha}^*_+ + \boldsymbol{\epsilon}_+, \quad \boldsymbol{\epsilon}_+ \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{R}_+), \quad \boldsymbol{\alpha}_+ \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{G}_+), \end{aligned}$$

  subject to the fit has to be maintained, restriction imposed through a equation $\boldsymbol{C}_{MM}\begin{bmatrix}\boldsymbol{\beta}^*_{\tilde{+}} \\ \boldsymbol{\alpha}^*_+\end{bmatrix} = \boldsymbol{r}_{MM}$, where the model components depend on how many covariates are extended:

– Extending one covariate, the model components are $\boldsymbol{X}_+$ and $\boldsymbol{Z}_+$ defined in (46) and (47), $\boldsymbol{R}_+$ defined in (18) and $\boldsymbol{G}_+$ defined in (49).
– Extending two covariates, the model components are $\boldsymbol{X}_+$ and $\boldsymbol{Z}_+$ defined in (41) and (42), $\boldsymbol{R}_+$ defined in (12) and $\boldsymbol{G}_+$ defined in (44).

In both cases, we define the constraints matrix $\boldsymbol{C}_{MM}$ and the constraints vector $\boldsymbol{r}_{MM}$ in the P-splines context and use the extended transformations ($\boldsymbol{\Omega}_+$ given in (45) if we extend one covariate and $\boldsymbol{\Omega}_+$ given in (40) if we extend the two covariates). To obtain the constraints matrix in the context of mixed models, $\boldsymbol{C}_{MM} = \boldsymbol{C}\boldsymbol{\Omega}_+$. Since we are imposing the restriction that the fit has to be maintained, the covariance parameters used to obtain the fit and the forecast simultaneously are the ones estimated to compute the fit. Once the model components are defined the fixed and random effects are computed solving the system (60).

- Smooth-ANOVA model, i.e.

$$\begin{aligned} \boldsymbol{y}_+ &= f_1(\boldsymbol{z}_+) + f_2(\boldsymbol{x}_+) + f_{1,2}(\boldsymbol{z}_+, \boldsymbol{x}_+) + \boldsymbol{\epsilon}_+ \\ &= \boldsymbol{X}_+\boldsymbol{\beta}_+^{\tilde{}} + \boldsymbol{Z}_+\boldsymbol{\alpha}_+ + \boldsymbol{\epsilon}_+, \quad \boldsymbol{\epsilon}_+ \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{R}_+), \quad \boldsymbol{\alpha}_+ \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{G}_+), \end{aligned}$$

where the model components are defined depending on how many covariates we extend.

– Extending one covariate, the components are $\boldsymbol{X}_+$ and $\boldsymbol{Z}_+$ defined in (79) and (80), $\boldsymbol{R}_+$ defined in (18) and $\boldsymbol{G}_+$ defined through the equation of $\boldsymbol{F}_+$ in (81).
– Extending two covariates, the components are $\boldsymbol{X}_+$ and $\boldsymbol{Z}_+$ defined in (76) and (77), $\boldsymbol{R}_+$ defined in (12) and $\boldsymbol{G}_+$ defined through the equation of $\boldsymbol{F}_+$ in (78).

In both cases, after defining the model components, to estimate the model we maximize the extended REML (39) and solve the extended mixed model system of equations of Henderson (37).

- Smooth-ANOVA model with restriction, i.e.

$$\begin{aligned} \boldsymbol{y}_+ &= f_1(\boldsymbol{z}_+) + f_2(\boldsymbol{x}_+) + f_{1,2}(\boldsymbol{z}_+, \boldsymbol{x}_+) + \boldsymbol{\epsilon}_+ \\ &= \boldsymbol{X}_+\boldsymbol{\beta}_+^{*} + \boldsymbol{Z}_+\boldsymbol{\alpha}_+^* + \boldsymbol{\epsilon}_+, \quad \boldsymbol{\epsilon}_+ \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{R}_+), \quad \boldsymbol{\alpha}_+^* \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{G}_+), \end{aligned}$$

subject to the fit has to be maintained, restriction imposed through a equation $\boldsymbol{C}_{MM}\begin{bmatrix}\boldsymbol{\beta}_+^*\\ \boldsymbol{\alpha}_+^*\end{bmatrix} = \boldsymbol{r}_{MM}$, where the model components depend on how many covariates are extended:

– Extending one covariate, the components are $\boldsymbol{X}_+$ and $\boldsymbol{Z}_+$ defined in (86) and (87), $\boldsymbol{R}_+$ defined in (18) and $\boldsymbol{G}_+$ defined through $\boldsymbol{F}_+$ in (88).
– Extending two covariates, the components are $\boldsymbol{X}_+$ and $\boldsymbol{Z}_+$ defined in (83) and (84), $\boldsymbol{R}_+$ defined in (12) and $\boldsymbol{G}_+$ defined through $\boldsymbol{F}_+$ in (85).

Again the variance parameters are the ones estimated to compute the fit. Once the model components are defined the fixed and random effects are computed solving the system (60).

For each scenario and each model, we have repeated the following 100 times:

- Start with a data set of observations arranged into a matrix $\boldsymbol{Y}_+$ of dimension $n_{\boldsymbol{z}_+} \times n_{\boldsymbol{x}_+}$:
$$\boldsymbol{Y}_+ = \begin{bmatrix} \boldsymbol{Y} & \boldsymbol{Y}_{\boldsymbol{z}\boldsymbol{x}_p} \\ \boldsymbol{Y}_{\boldsymbol{z}_p\boldsymbol{x}} & \boldsymbol{Y}_{\boldsymbol{z}_p\boldsymbol{x}_p} \end{bmatrix},$$
where $\boldsymbol{Y}$, $\boldsymbol{Y}_{\boldsymbol{z}\boldsymbol{x}_p}$, $\boldsymbol{Y}_{\boldsymbol{z}_p\boldsymbol{x}}$ and $\boldsymbol{Y}_{\boldsymbol{z}_p\boldsymbol{x}_p}$ have dimension $n_{\boldsymbol{z}} \times n_{\boldsymbol{x}}$, $n_{\boldsymbol{z}} \times n_{\boldsymbol{x}_p}$, $n_{\boldsymbol{z}_p} \times n_{\boldsymbol{x}}$ and $n_{\boldsymbol{z}_p} \times n_{\boldsymbol{x}_p}$, respectively.
- Split our data into two groups: the training data $\boldsymbol{Y}$ and the test data $\boldsymbol{Y}_{\boldsymbol{z}\boldsymbol{x}_p}$, $\boldsymbol{Y}_{\boldsymbol{z}_p\boldsymbol{x}}$ and $\boldsymbol{Y}_{\boldsymbol{z}_p\boldsymbol{x}_p}$.
- Use the training data to predict $4900 - n$ new observations ($n = n_{\boldsymbol{z}} \times n_{\boldsymbol{x}}$).
- Check the accuracy of each model.

The marginal B-spline bases $\boldsymbol{B}_{\boldsymbol{z}}$ and $\boldsymbol{B}_{\boldsymbol{x}}$ are constructed with 15 knots and cubic splines and then extended to cover the whole range of $\boldsymbol{z}_+$ and $\boldsymbol{x}_+$, the penalty orders are two.

As it is known, a model which fits the data well does not necessarily forecast well, therefore we have compared the fit performance, the forecast performance and the overall performance of the methods. To check the accuracy of the methods we follow Hyndman (2006) and take the errors as the difference between the function values from which we simulate the data and the fit and forecast produced using only the data in the training set:
$$\boldsymbol{E}_+ = \boldsymbol{S}^k - \hat{\boldsymbol{Y}}_+, \quad k = 1, 2$$
i.e, we have the errors matrix $\boldsymbol{E}_+ = \begin{bmatrix} \boldsymbol{E} & \boldsymbol{E}_{\boldsymbol{z}\boldsymbol{x}_p} \\ \boldsymbol{E}_{\boldsymbol{z}_p\boldsymbol{x}} & \boldsymbol{E}_{\boldsymbol{z}_p\boldsymbol{x}_p} \end{bmatrix}$, and therefore the vectors containing the errors in the fit, in the forecast and in the overall performance:

- Vector with the fit errors: $\boldsymbol{e}^{(f)} = \mathrm{vec}(\boldsymbol{E})$
- Vector with the prediction errors: $\boldsymbol{e}^{(p)} = (\mathrm{vec}(\boldsymbol{E}_{\boldsymbol{z}\boldsymbol{x}_p}), \mathrm{vec}(\boldsymbol{E}_{\boldsymbol{z}_p\boldsymbol{x}}), \mathrm{vec}(\boldsymbol{E}_{\boldsymbol{z}_p\boldsymbol{x}_p}))$
- Vector with the total errors: $\boldsymbol{e}^{(t)} = \mathrm{vec}(\boldsymbol{E}_+)$

The errors measure that we use is the mean absolute error because as it is said in Hyndman (2006) it is less sensitive to outliers than the root mean square error:

$$\text{Mean absolute error: MAE} = \frac{\sum_{i=1}^{N} |e_i^{(l)}|}{N}, \quad N = \mathrm{length}(\boldsymbol{e}^{(l)})$$

for $l = f, p, t$, i.e. for the errors in the fit, in the forecast or in total.

We have divided the results of the simulation study in two Sections, one to show the obtained results in Scenario 1 and other to show the obtained results in Scenario 2.

5.1 Simulations results for Scenario 1

Below we show the obtained results for Scenario 1. We have made boxplots for the MAE values in the fit, in the forecast and in the overall performance. For different values of $n_{\boldsymbol{z}_p}$ and $n_{\boldsymbol{x}_p}$, we fit and predict with the four smooth mixed models: interaction, interaction with restriction, S-ANOVA and S-ANOVA with restriction.

The x-axis labels of the boxpots refer to the models: *unrestricted* (2D interaction P-spline model), *restricted* (2D interaction P-spline model with restriction), *ANOVA* (S-ANOVA model) and *restricted ANOVA* (S-ANOVA model with restriction).

Notice that for the different values of n_{z_p} and n_{x_p} we consider as the training data an observations matrix of dimension $(70 - n_{z_p}) \times (70 - n_{x_p})$ and we predict $(70 - n_{z_p}) \times n_{x_p} + n_{z_p} \times (70 - n_{x_p}) + n_{z_p} \times n_{x_p}$ new values. The results are listed in Appendix .5 for the different values of n_{z_p} and n_{x_p}, to illustrate them below we show the figure for the particular case of $n_{z_p} = 0$ and $n_{x_p} = 5$:

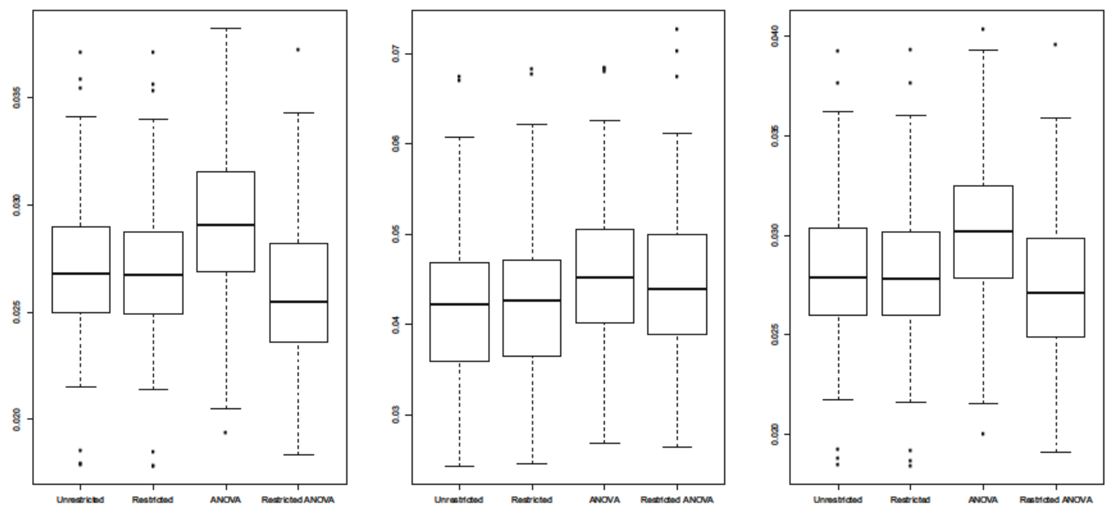


Figure 9: MAE in the fit (left panel), in the forecast (middle panel) and in total (right panel) of smooth models in scenario 1 and $n_{z_p} = 0$ and $n_{x_p} = 5$.

The fit remains relatively invariant in the case of the 2D P-spline models, with and without restriction, and in the restricted S-ANOVA model. While the unrestricted S-ANOVA model is modified in most scenarios.

In the prediction, the best performance is also made by the 2D interaction models (restricted and unrestricted) and by the restricted S-ANOVA model, except for the particular case in which we predict in one of the two covariates. In Figures .12 and .13 we can see that for the particular values of (n_{z_p}, n_{x_p}), $(0, 15)$ and $(0, 20)$, the most accurate model (lower MAE values) in the prediction is the unrestricted interaction model followed by the S-ANOVA model and by the restricted interaction model, respectively.

We conclude that simultaneous fit and prediction with S-ANOVA models produces significant changes in the fit and that in the overall performance (including fit and prediction), the 2D P-spline models (with and without restriction) and the restricted S-ANOVA models are the most accurate.

5.2 Simulations results for Scenario 2

In Appendix .5 we show the results obtained for Scenario 2, i.e., for the case in which the true surface is constructed from a model with two main effects and with an interaction. Below you can see the boxplot for the particular case of $n_{z_p} = 0$ and $n_{x_p} = 5$:

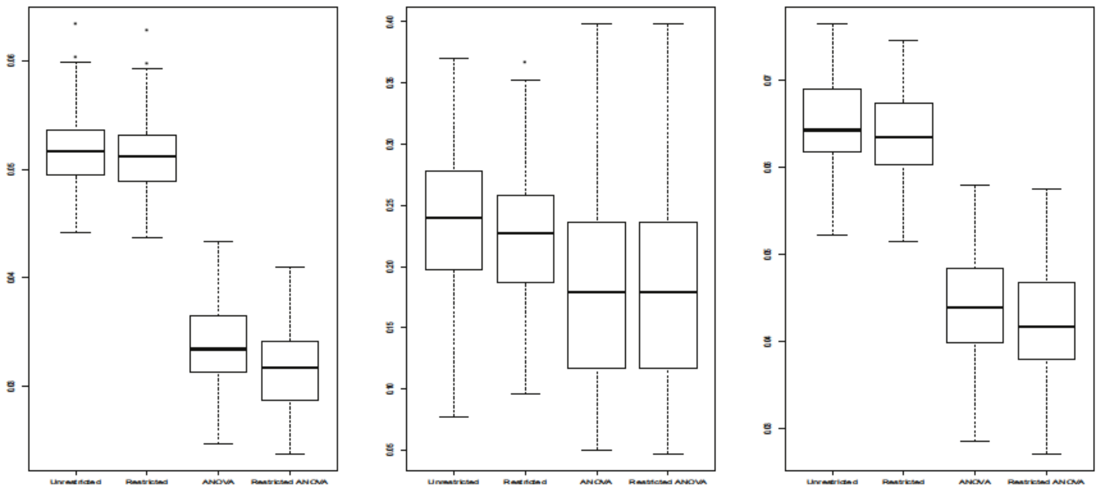


Figure 10: MAE in the fit (left panel), in the forecast (middle panel) and in total (right panel) of smooth models in scenario 2 and $n_{z_p} = 0$ and $n_{x_p} = 5$.

As a summary we can conclude that there is a significant difference between the results obtained with the interaction models and with the S-ANOVA models: 2D interaction P-splines models (restricted and unrestricted) are always less accurate than S-ANOVA models. This fact is due to interaction models are constraint to fit the true model without taking into account the main effects.

In the fit, the performances of the 2D P-spline interaction models, with and without restrictions, are quite similar. However in the prediction, the restricted interaction model is better than the unrestricted interaction model, the major difference between the two models can be seen for most scenarios.

In the case of the S-ANOVA models, the restricted model is always better than the unrestricted one, in the fit, in the prediction and therefore in the overall performance. Although, in the particular case in which we only predict in one dimension, the prediction errors for both models are almost equal (see Figures .22, .23 and .24).

The conclusion for this scenario is that the restricted S-ANOVA model is clearly the most accurate. Therefore, our suggestion in models with interaction is to use the restricted S-ANOVA model always.

6 Conclusions

In this paper we have presented a general framework for out-of-sample prediction in smooth additive models with interaction terms. We build our proposal from the method proposed in Currie *et al.* (2004), we have extended their approach to the case in which out-of-sample prediction is necessary in both directions of the interaction terms.

The method proposed in Currie *et al.* (2004) deal with out-of-sample predictions as missing values with 0 weights and carry out fit and prediction simultaneously. We have shown that this approach yields different fitted values depending on whether only fitting or fitting and prediction is carried out.

To solve this incoherence we propose to maximize the penalized likelihood subject to linear constraints which ensure that the coefficients obtained in the fit of the data remain the same when fit and out-of-sample prediction is carried out simultaneously. To do so we use Lagrange multipliers. This general approach can be used to impose other relevant constraints such as the case of mortality forecasting when structure across ages needs to be preserved.

The methodology proposed is extended to the case of smooth mixed models since it will allow us to predict out-of-sample in a wide class of models. We have shown that attention needs to be imposed since the matrices of fixed and random effects for out-of-sample prediction need to be direct extensions of the matrices used in the fit.

We have developed a method for out-of-sample prediction for the Smooth ANOVA model proposed in Lee & Durbán (2011). A simulation study has been carried out to compare constrained and unconstrained out-of-sample prediction when using 2D interaction models and S-ANOVA models. From the results of the simulation study, we have concluded that in most situations the constrained S-ANOVA model behaves better in the fit and out-of-sample predictions when the prediction is carried out in both dimensions of the interaction term, if prediction is needed in only one of the covariates results depend on the simulated scenario, although in most cases the S-ANOVA model outperformed the full interaction (restricted or not) model.

The constrained prediction method proposed has also been used in two real data examples, one in which mortality rates are forecasted over the years and the importance of imposing constraints on the coefficients to ensure coherent forecast is shown. The other example predicts tree biomass as a function of the tree height and diameter.

Currently we are working on out-of-sample prediction in the context of generalized additive and mixed models.

References

Camarda, C.G. 2012. MortalitySmooth: An R Package for Smoothing Poisson Counts with P-splines. *Journal of Statistical Software*, **50**(1).

Carballo, A., Durbán, M., & Lee, D.-J. 2017. A general framework for prediction in penalized regression. *UC3M Working Papers. Statistics and Econometrics.*

Currie, I. D., & Durbán, M. 2002. Flexible Smoothing with P-Splines: A Unified Approach. *Statistical Modelling*, **2**, 333–349.

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

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

Currie, I.D. 2013. Smoothing constrained generalized linear models with an applicaiton to the Lee-Carter model. *Statistical Modelling*, **13**(1), 69–93.

Delwarde, A., Denuit, M., & Eilers, P. H. C. 2007. Smoothing the Lee-Carter and Poisson log-bilinear models for mortality forecasting: a penalized log-likelihood approach. *Statistics and Modelling*, **7**, 29–48.

Eilers, P. H. C., & Marx, B. D. 1996. Flexible smoothing with B-splines and penalties. *Statistical Science*, **11**(2), 89–121.

Etxeberria, J., Ugarte, M.D., Goicoa, T., & A.F., Militino. 2015. On predicting cancer mortality using anova-type P-splines models. *Revstat - Statistical Journal*, **13**(1), 21–40.

Green, P. J. 1987. Penalized Likelihod for General Semi-Parametric Regression Models. *International Statistical Review*, **55**(3), 245–259.

Greene, W. H., & Seaks, G. T. 1991. The Restricted Least Squares Estimator: A Pedagogical Note. *The Review of Economics and Statistics*, **73**(3), 563–567.

Harville, D. 2000. *Matrix Algebra from a Statistician's Perspective*. Springer.

Henderson, C. R. 1975. Best linear unbiased estimation and prediction under a selection model. *Biometrics*, **31**, 423–447.

Human Mortality Database. 2018. *University of California, Berkeley (USA), and Max Planck Institute for Demographic Research (Germany)*. https://mortality.org.

Hyndman, R. J. and. Koehler A. B. 2006. Another look at measures of forecast accuracy. *International Journal of Forecasting*, 679 – 688.

Lee, D. J. 2010. *Smoothing mixed models for spatial and spatio-temporal data*. Ph.D. thesis, Department of Statistics, Universidad Carlos III de Madrid, Spain.

Lee, D-J., & Durbán, M. 2011. P-spline ANOVA-type interaction models for spatio-temporal smoothing. *Statistical Modelling*, **11**(1), 49–69.

Lee, R.D., & Carter, L. 1992. Modelling and forecasting the time series of US mortality. *Journal of the American Statistical Association*, **87**, 659–71.

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

Rivas-Martínez, Díaz, Díaz, T.E., Fernández-González, F., Izco, J., Loidi, J., Lousã, M., & Penas, A. 2002. Vascular plant communities of Spain and Portugal. Addenda to the syntaxonomical checklist of 2001. *Itinera Geobotanica*, 15, 1–2, 5–22.

Sánchez-González, M., Durbán, M., Lee, D.J., Cañellas, I., & Sixto, H. 2016. Smooth additive mixed models for predicting aboveground biomass. *Journal of Agricultural, Biological and Environmental Statistics*, **22**, 23–41.

Ugarte, M. D., Goicoa, T., Etxeberria, J., & Militino, A. F. 2012. Projections of cancer mortality risks using spatio-temporal P-spline models. *Statistical Methods in Medical Research*, **21**(5), 545–560.

Wand, M.P. 2003. Smoothing and mixed models. *Computational statistics*, **18**, 223–249.

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

Appendix

Appendix .1 Proof of Corollary 2

Proof. Notice that if $\boldsymbol{P}^{\boldsymbol{z}}_{+22} = \boldsymbol{O}$, by (22), the coefficients that give the fit are:

$$\hat{\boldsymbol{\theta}}_{+1,\dots,c} = \left(\boldsymbol{B}'\boldsymbol{B} + \lambda_{\boldsymbol{x}}(\boldsymbol{D}'_{\boldsymbol{x}}\boldsymbol{D}_{\boldsymbol{x}} \otimes \boldsymbol{I}_{c_{\boldsymbol{z}}}) + \lambda_{\boldsymbol{z}}(\boldsymbol{I}_{c_{\boldsymbol{x}}} \otimes \boldsymbol{D}'_{\boldsymbol{z}}\boldsymbol{D}_{\boldsymbol{z}})\right)^{-1}\boldsymbol{B}'\boldsymbol{y},$$

i.e., the same as the coefficients we obtain only fitting the data without a prediction, (10). Let us see which are the coefficients that determine the forecast when the penalty orders are two or three.

- Differences of order 2.

  Suppose a difference matrix with second order penalty $\boldsymbol{D}_{\boldsymbol{x}_+}$ of dimensions $(c_{\boldsymbol{x}_+} - 2) \times c_{\boldsymbol{x}_+}$:

$$\boldsymbol{D}_{\boldsymbol{x}_+} = \begin{bmatrix} \boldsymbol{D} & \boldsymbol{O} \\ \boldsymbol{D}_{\boldsymbol{x}(1)} & \boldsymbol{D}_{\boldsymbol{x}(2)} \end{bmatrix} = \begin{bmatrix} 1 & -2 & 1 & 0 & 0 & 0 & 0 & \cdots \\ 0 & 1 & -2 & 1 & 0 & 0 & 0 & \cdots \\ 0 & 0 & 1 & -2 & 1 & 0 & 0 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & 0 & 1 & -2 & 1 \end{bmatrix},$$

where $\boldsymbol{D}_{\boldsymbol{x}(2)}$:

$$\boldsymbol{D}_{\boldsymbol{x}(2)} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & \cdots & 0 \\ -2 & 1 & 0 & 0 & 0 & \cdots & 0 \\ 1 & 2 & 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & -2 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & -2 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & 1 & -2 & 1 \end{bmatrix},$$

of dimension $c_{\boldsymbol{x}_p} \times c_{\boldsymbol{x}_p}$, where $c_{\boldsymbol{x}_p}$ is the number of columns of $\boldsymbol{B}_{\boldsymbol{x}(2)}$. Therefore, $(\boldsymbol{D}_{\boldsymbol{x}(2)} \otimes \boldsymbol{I}_{c_{\boldsymbol{z}}})^{-1}$ has the form

$$\begin{bmatrix}
1 & & & & & & & & & & & & \\
 & \ddots & & & & & & & & & & & \\
 & & 1 & & & & & & & & & & \\
2 & & & 1 & & & & & & & & & \\
 & \ddots & & & \ddots & & & & & & & & \\
 & & 2 & & & 1 & & & & & & & \\
3 & & & 2 & & & 1 & & & & & & \\
 & \ddots & & & \ddots & & & \ddots & & & & & \\
 & & 3 & & & 2 & & & 1 & & & & \\
4 & & & 3 & & & 2 & & & 1 & & & \\
 & \ddots & & & \ddots & & & \ddots & & & \ddots & & \\
 & & 4 & & & 3 & & & 2 & & & 1 & \\
 & \vdots & & & \vdots & & & \vdots & & & & \vdots & \ddots
\end{bmatrix}$$

of dimension $(c_{\boldsymbol{x}_p} \cdot c_{\boldsymbol{z}}) \times (c_{\boldsymbol{x}_p} \cdot c_{\boldsymbol{z}})$, each block has dimension $c_{\boldsymbol{z}} \times c_{\boldsymbol{z}}$. Moreover,

$$\boldsymbol{D}_{\boldsymbol{x}(1)} \otimes \boldsymbol{I}_{c_{\boldsymbol{z}}} = \begin{bmatrix}
0 & & & 0 & & & & 1 & & & -2 & & \\
 & \ddots & & & \ddots & & \cdots & & \ddots & & & \ddots & \\
 & & 0 & & & 0 & & & & 1 & & & -2 \\
0 & & & 0 & & & & 0 & & & 1 & & \\
 & \ddots & & & \ddots & & \cdots & & \ddots & & & \ddots & \\
 & & 0 & & & 0 & & & & 0 & & & 1 \\
0 & & & 0 & & & & 0 & & & 0 & & \\
 & \ddots & & & \ddots & & \cdots & & \ddots & & & \ddots & \\
 & & 0 & & & 0 & & & & 0 & & & 0 \\
 & \vdots & & & \vdots & & \cdots & & \vdots & & & \vdots &
\end{bmatrix},$$

i.e., is a matrix of dimension $(c_{\boldsymbol{x}_p} \cdot c_{\boldsymbol{z}}) \times (c_{\boldsymbol{x}} \cdot c_{\boldsymbol{z}})$ with just three blocks of dimension

$c_{\boldsymbol{z}} \times c_{\boldsymbol{z}}$ that are not blocks of zeros. Therefore,

$$(\boldsymbol{D}_{\boldsymbol{x}(2)} \otimes \boldsymbol{I}_{c_{\boldsymbol{z}}})^{-1}(\boldsymbol{D}_{\boldsymbol{x}(1)} \otimes \boldsymbol{I}_{c_{\boldsymbol{z}}}) = \left[\begin{array}{ccccccccccccc}
0 & & & 0 & & & & 1 & & & -2 & & \\
 & \ddots & & & \ddots & & \cdots & & \ddots & & & \ddots & \\
 & & 0 & & & 0 & & & & 1 & & & -2 \\
0 & & & 0 & & & & 2 & & & -3 & & \\
 & \ddots & & & \ddots & & \cdots & & \ddots & & & \ddots & \\
 & & 0 & & & 0 & & & & 2 & & & -3 \\
0 & & & 0 & & & & 3 & & & -4 & & \\
 & \ddots & & & \ddots & & \cdots & & \ddots & & & \ddots & \\
 & & 0 & & & 0 & & & & 3 & & & -4 \\
0 & & & 0 & & & & 4 & & & -5 & & \\
 & \ddots & & & \ddots & & \cdots & & \ddots & & & \ddots & \\
 & & 0 & & & 0 & & & & 4 & & & -5 \\
 & \vdots & & & \vdots & & \cdots & & \vdots & & & \vdots &
\end{array}\right],$$

with dimension $(c_{\boldsymbol{x}_p} \cdot c_{\boldsymbol{z}}) \times (c_{\boldsymbol{x}} \cdot c_{\boldsymbol{z}})$. Hence, considering the matrix of coefficients that give the fit, $\hat{\boldsymbol{\Theta}}$, and the matrix of coefficients that give the forecast, $\hat{\boldsymbol{\Theta}}_p$, each row $j = 1, ..., c_{\boldsymbol{z}}$, of the additional matrix of coefficients is a linear combination of two old coefficients of that row:

$$\hat{\boldsymbol{\Theta}}_{j\ \cdot} = \hat{\theta}_{j\ c_{\boldsymbol{x}}} \begin{bmatrix} 1 \\ 1 \\ 1 \\ \vdots \end{bmatrix} + (\hat{\theta}_{j\ c_{\boldsymbol{x}}} - \hat{\theta}_{j\ c_{\boldsymbol{x}}-1}) \begin{bmatrix} 1 \\ 2 \\ 3 \\ \vdots \end{bmatrix}.$$

- Differences of order 3.
  Suppose a difference matrix with third order penalty, $\boldsymbol{D}_{\boldsymbol{x}_+}$ of dimensions $(c_{\boldsymbol{x}_+} - 3) \times c_{\boldsymbol{x}_+}$:

$$\boldsymbol{D}_+ = \begin{bmatrix} -1 & 3 & -3 & 1 & 0 & 0 & \cdots & 0 & 0 & 0 & 0 \\ 0 & -1 & 3 & -3 & 1 & 0 & \cdots & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 3 & -3 & 1 & \cdots & 0 & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \cdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & 0 & 0 & \cdots & -1 & 3 & -3 & 1 \end{bmatrix}.$$

In this case, $\boldsymbol{D}_{\boldsymbol{x}(2)}$ is:

$$\boldsymbol{D}_{\boldsymbol{x}(2)} = \begin{bmatrix} 1 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & 0 \\ -3 & 1 & 0 & 0 & \cdots & 0 & 0 & 0 & 0 \\ 3 & -3 & 1 & 0 & \cdots & 0 & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \cdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & -1 & 3 & -3 & 1 \end{bmatrix}.$$

$(\boldsymbol{D}_{\boldsymbol{x}(2)} \otimes \boldsymbol{I}_{c_{\boldsymbol{z}}})^{-1}$ and $\boldsymbol{D}_{\boldsymbol{x}(1)} \otimes \boldsymbol{I}_{c_{\boldsymbol{z}}}$ are:

$$
(\boldsymbol{D}_{\boldsymbol{x}(2)} \otimes \boldsymbol{I}_{c_{\boldsymbol{z}}})^{-1} = \left[\begin{array}{ccccccccccccc}
1 & & & & & & & & & & & & \\
 & \ddots & & & & & & & & & & & \\
 & & 1 & & & & & & & & & & \\
3 & & & 1 & & & & & & & & & \\
 & \ddots & & & \ddots & & & & & & & & \\
 & & 3 & & & 1 & & & & & & & \\
6 & & & 3 & & & 1 & & & & & & \\
 & \ddots & & & \ddots & & & \ddots & & & & & \\
 & & 6 & & & 3 & & & 1 & & & & \\
10 & & & 6 & & & 3 & & & 1 & & & \\
 & \ddots & & & \ddots & & & \ddots & & & \ddots & & \\
 & & 10 & & & 6 & & & 3 & & & 1 & \\
 & \vdots & & & \vdots & & & \vdots & & & & \vdots & \ddots
\end{array}\right],
$$

$$
\boldsymbol{D}_{\boldsymbol{x}(1)} \otimes \boldsymbol{I}_{c_{\boldsymbol{z}}} = \left[\begin{array}{ccccccccccccc}
0 & & & & -1 & & & 3 & & & -3 & & \\
 & \ddots & & \dots & & \ddots & & & \ddots & & & \ddots & \\
 & & 0 & & & & -1 & & & 3 & & & -3 \\
0 & & & & 0 & & & -1 & & & 3 & & \\
 & \ddots & & \dots & & \ddots & & & \ddots & & & \ddots & \\
 & & 0 & & & & 0 & & & -1 & & & 3 \\
0 & & & & 0 & & & 0 & & & -1 & & \\
 & \ddots & & \dots & & \ddots & & & \ddots & & & \ddots & \\
 & & 0 & & & & 0 & & & 0 & & & -1 \\
 & \vdots & & \dots & & \vdots & & & \vdots & & & \vdots &
\end{array}\right].
$$

Then,

$$
(\boldsymbol{D}_{\boldsymbol{x}(2)} \otimes \boldsymbol{I}_{c_{\boldsymbol{z}}})^{-1}(\boldsymbol{D}_{\boldsymbol{x}(1)} \otimes \boldsymbol{I}_{c_{\boldsymbol{z}}}) = \left[\begin{array}{ccccccccccccc}
0 & & & & -1 & & & 3 & & & -3 & & \\
 & \ddots & & \dots & & \ddots & & & \ddots & & & \ddots & \\
 & & 0 & & & & -1 & & & 3 & & & -3 \\
0 & & & & -3 & & & 8 & & & -6 & & \\
 & \ddots & & \dots & & \ddots & & & \ddots & & & \ddots & \\
 & & 0 & & & & -3 & & & 8 & & & -6 \\
0 & & & & -6 & & & 15 & & & -10 & & \\
 & \ddots & & \dots & & \ddots & & & \ddots & & & \ddots & \\
 & & 0 & & & & -6 & & & 15 & & & -10 \\
0 & & & & -10 & & & 24 & & & -15 & & \\
 & \ddots & & \dots & & \ddots & & & \ddots & & & \ddots & \\
 & & 0 & & & & -10 & & & 24 & & & -15 \\
 & \vdots & & \dots & & \vdots & & & \vdots & & & \vdots &
\end{array}\right].
$$

Therefore, each row, $j = 1, ..., c_{\boldsymbol{z}}$, of the additional matrix of coefficients is a linear combination of three old coefficients of that row:

$$
\hat{\boldsymbol{\Theta}}_{p_j\ \cdot} = \hat{\theta}_{j\ c_{\boldsymbol{x}}} \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ \vdots \end{bmatrix} + \frac{3\hat{\theta}_{j\ c_{\boldsymbol{x}}} - 4\hat{\theta}_{j\ c_{\boldsymbol{x}}-1} + \hat{\theta}_{j\ c_{\boldsymbol{x}}-2}}{2} \begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \\ \vdots \end{bmatrix} + \frac{\hat{\theta}_{j\ c_{\boldsymbol{x}}} - 2\hat{\theta}_{j\ c_{\boldsymbol{x}}-1} + \hat{\theta}_{j\ c_{\boldsymbol{x}}-2}}{2} \begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \\ \vdots \end{bmatrix}^2 .
$$

□

Appendix .2 Proof of Theorem 3

Proof. Given the extended transformation $\boldsymbol{\Omega}^*_+$ in two dimensions defined in (50) and the extended penalty matrix in (15) the extended mixed model penalty is:

$$
\begin{aligned}
\boldsymbol{\Phi}^*_+ &= \boldsymbol{\Omega}^{*'}_+\boldsymbol{P}_+\boldsymbol{\Omega}^*_+ = \boldsymbol{\Omega}^{*'}_+(\lambda_{\boldsymbol{z}}\boldsymbol{P}^{\boldsymbol{z}+}_+ + \lambda_{\boldsymbol{x}}\boldsymbol{P}^{\boldsymbol{x}+}_+)\boldsymbol{\Omega}^*_+ = \lambda_{\boldsymbol{z}}\boldsymbol{\Omega}^{*'}_+\boldsymbol{P}^{\boldsymbol{z}+}_+\boldsymbol{\Omega}^*_+ + \lambda_{\boldsymbol{x}}\boldsymbol{\Omega}^{*'}_+\boldsymbol{P}^{\boldsymbol{x}+}_+\boldsymbol{\Omega}^*_+ \\
&= \lambda_{\boldsymbol{z}}\begin{bmatrix} \boldsymbol{U}^{*'}_{\boldsymbol{x}+f}\otimes\boldsymbol{U}^{*'}_{\boldsymbol{z}+f} \\ \boldsymbol{U}^{*'}_{\boldsymbol{x}+r}\otimes\boldsymbol{U}^{*'}_{\boldsymbol{z}+f} \\ \boldsymbol{U}^{*'}_{\boldsymbol{x}+f}\otimes\boldsymbol{U}^{*'}_{\boldsymbol{z}+r} \\ \boldsymbol{U}^{*'}_{\boldsymbol{x}+r}\otimes\boldsymbol{U}^{*'}_{\boldsymbol{z}+r} \end{bmatrix}(\boldsymbol{I}_{c_{\boldsymbol{x}+}}\otimes\boldsymbol{D}'_{\boldsymbol{z}+}\boldsymbol{D}_{\boldsymbol{z}+})\left[\boldsymbol{U}^*_{\boldsymbol{x}+f}\otimes\boldsymbol{U}^*_{\boldsymbol{z}+f} \mid \boldsymbol{U}^*_{\boldsymbol{x}+r}\otimes\boldsymbol{U}^*_{\boldsymbol{z}+f} \mid \boldsymbol{U}^*_{\boldsymbol{x}+f}\otimes\boldsymbol{U}^*_{\boldsymbol{z}+r} \mid \boldsymbol{U}^*_{\boldsymbol{x}+r}\otimes\boldsymbol{U}^*_{\boldsymbol{z}+r}\right] \\
&+ \lambda_{\boldsymbol{x}}\begin{bmatrix} \boldsymbol{U}^{*'}_{\boldsymbol{x}+f}\otimes\boldsymbol{U}^{*'}_{\boldsymbol{z}+f} \\ \boldsymbol{U}^{*'}_{\boldsymbol{x}+r}\otimes\boldsymbol{U}^{*'}_{\boldsymbol{z}+f} \\ \boldsymbol{U}^{*'}_{\boldsymbol{x}+f}\otimes\boldsymbol{U}^{*'}_{\boldsymbol{z}+r} \\ \boldsymbol{U}^{*'}_{\boldsymbol{x}+r}\otimes\boldsymbol{U}^{*'}_{\boldsymbol{z}+r} \end{bmatrix}(\boldsymbol{D}'_{\boldsymbol{x}+}\boldsymbol{D}_{\boldsymbol{x}+}\otimes\boldsymbol{I}_{c_{\boldsymbol{z}+}})\left[\boldsymbol{U}^*_{\boldsymbol{x}+f}\otimes\boldsymbol{U}^*_{\boldsymbol{z}+f} \mid \boldsymbol{U}^*_{\boldsymbol{x}+r}\otimes\boldsymbol{U}^*_{\boldsymbol{z}+f} \mid \boldsymbol{U}^*_{\boldsymbol{x}+f}\otimes\boldsymbol{U}^*_{\boldsymbol{z}+r} \mid \boldsymbol{U}^*_{\boldsymbol{x}+r}\otimes\boldsymbol{U}^*_{\boldsymbol{z}+r}\right] \\
&= \lambda_{\boldsymbol{z}}\begin{bmatrix}
\boldsymbol{U}^{*'}_{\boldsymbol{x}+f}\boldsymbol{I}_{c_{\boldsymbol{x}+}}\boldsymbol{U}^*_{\boldsymbol{x}+f}\otimes\boldsymbol{U}^{*'}_{\boldsymbol{z}+f}\boldsymbol{D}'_{\boldsymbol{z}}\boldsymbol{D}_{\boldsymbol{z}+}\boldsymbol{U}^*_{\boldsymbol{z}+f} & \boldsymbol{O} & \boldsymbol{O} & \boldsymbol{O} \\
\boldsymbol{O} & \boldsymbol{U}^{*'}_{\boldsymbol{x}+r}\boldsymbol{I}_{c_{\boldsymbol{x}+}}\boldsymbol{U}^*_{\boldsymbol{x}+r}\otimes\boldsymbol{U}^{*'}_{\boldsymbol{z}+f}\boldsymbol{D}'_{\boldsymbol{z}+}\boldsymbol{D}_{\boldsymbol{z}+}\boldsymbol{U}^*_{\boldsymbol{z}+f} & \boldsymbol{O} & \boldsymbol{O} \\
\boldsymbol{O} & \boldsymbol{O} & \boldsymbol{U}^{*'}_{\boldsymbol{x}+f}\boldsymbol{I}_{c_{\boldsymbol{x}+}}\boldsymbol{U}^*_{\boldsymbol{x}+f}\otimes\boldsymbol{U}^{*'}_{\boldsymbol{z}+r}\boldsymbol{D}'_{\boldsymbol{z}+}\boldsymbol{D}_{\boldsymbol{z}+}\boldsymbol{U}^*_{\boldsymbol{z}+r} & \boldsymbol{U}^{*'}_{\boldsymbol{x}+f}\boldsymbol{I}_{c_{\boldsymbol{x}+}}\boldsymbol{U}^*_{\boldsymbol{x}+f}\otimes\boldsymbol{U}^{*'}_{\boldsymbol{z}+r}\boldsymbol{D}'_{\boldsymbol{z}+}\boldsymbol{D}_{\boldsymbol{z}+}\boldsymbol{U}^*_{\boldsymbol{z}+r} \\
 & \boldsymbol{O} & \boldsymbol{U}^{*'}_{\boldsymbol{x}+r}\boldsymbol{I}_{c_{\boldsymbol{x}+}}\boldsymbol{U}^*_{\boldsymbol{x}+f}\otimes\boldsymbol{U}^{*'}_{\boldsymbol{z}r}\boldsymbol{D}'_{\boldsymbol{z}+}\boldsymbol{D}_{\boldsymbol{z}+}\boldsymbol{U}^*_{\boldsymbol{z}r} & \boldsymbol{U}^{*'}_{\boldsymbol{x}+r}\boldsymbol{I}_{c_{\boldsymbol{x}+}}\boldsymbol{U}^*_{\boldsymbol{x}+r}\otimes\boldsymbol{U}^{*'}_{\boldsymbol{z}+r}\boldsymbol{D}'_{\boldsymbol{z}+}\boldsymbol{D}_{\boldsymbol{z}+}\boldsymbol{U}_{\boldsymbol{z}+r}
\end{bmatrix} \\
&+ \lambda_{\boldsymbol{x}}\begin{bmatrix}
\boldsymbol{U}^{*'}_{\boldsymbol{x}+f}\boldsymbol{D}'_{\boldsymbol{x}+}\boldsymbol{D}_{\boldsymbol{x}+}\boldsymbol{U}^*_{\boldsymbol{x}+f}\otimes\boldsymbol{U}^{*'}_{\boldsymbol{z}+f}\boldsymbol{U}^*_{\boldsymbol{z}+f} & \boldsymbol{O} & \boldsymbol{O} & \boldsymbol{U}^{*'}_{\boldsymbol{x}+r}\boldsymbol{D}'_{\boldsymbol{x}+}\boldsymbol{D}_{\boldsymbol{x}+}\boldsymbol{U}^*_{\boldsymbol{x}+r}\otimes\boldsymbol{U}^{*'}_{\boldsymbol{z}+f}\boldsymbol{U}^*_{\boldsymbol{z}+r} \\
\boldsymbol{O} & \boldsymbol{U}^{*'}_{\boldsymbol{x}+r}\boldsymbol{D}'_{\boldsymbol{x}+}\boldsymbol{D}_{\boldsymbol{x}+}\boldsymbol{U}^*_{\boldsymbol{x}+r}\otimes\boldsymbol{U}^{*'}_{\boldsymbol{z}+f}\boldsymbol{U}^*_{\boldsymbol{z}+f} & \boldsymbol{O} & \boldsymbol{O} \\
\boldsymbol{O} & \boldsymbol{O} & \boldsymbol{U}^{*'}_{\boldsymbol{x}+f}\boldsymbol{D}'_{\boldsymbol{x}+}\boldsymbol{D}_{\boldsymbol{x}+}\boldsymbol{U}^*_{\boldsymbol{x}+f}\otimes\boldsymbol{U}^{*'}_{\boldsymbol{z}+r}\boldsymbol{U}^*_{\boldsymbol{z}+r} & \boldsymbol{O} \\
\boldsymbol{U}^{*'}_{\boldsymbol{x}+r}\boldsymbol{D}'_{\boldsymbol{x}+}\boldsymbol{D}_{\boldsymbol{x}+}\boldsymbol{U}^*_{\boldsymbol{x}+r}\otimes\boldsymbol{U}^{*'}_{\boldsymbol{z}+r}\boldsymbol{U}^*_{\boldsymbol{z}+f} & \boldsymbol{O} & \boldsymbol{O} & \boldsymbol{U}^{*'}_{\boldsymbol{x}+r}\boldsymbol{D}'_{\boldsymbol{x}+}\boldsymbol{D}_{\boldsymbol{x}+}\boldsymbol{U}^*_{\boldsymbol{x}+r}\otimes\boldsymbol{U}^{*'}_{\boldsymbol{z}+r}\boldsymbol{U}^*_{\boldsymbol{z}+r}
\end{bmatrix} \\
&= \lambda_{\boldsymbol{z}}\begin{bmatrix}
\boldsymbol{U}^{*'}_{\boldsymbol{x}+f}\boldsymbol{I}_{c_{\boldsymbol{x}+}}\boldsymbol{U}^*_{\boldsymbol{x}+f}\otimes\boldsymbol{O}_{q_{\boldsymbol{z}}} & \boldsymbol{O} & \boldsymbol{O} & \boldsymbol{O} \\
\boldsymbol{O} & \boldsymbol{U}^{*'}_{\boldsymbol{x}+r}\boldsymbol{I}_{c_{\boldsymbol{x}+}}\boldsymbol{U}^*_{\boldsymbol{x}+r}\otimes\boldsymbol{O}_{q_{\boldsymbol{z}}} & \boldsymbol{O} & \boldsymbol{O} \\
\boldsymbol{O} & \boldsymbol{O} & \boldsymbol{U}^{*'}_{\boldsymbol{x}+f}\boldsymbol{U}^*_{\boldsymbol{x}+f}\otimes\boldsymbol{U}^{*'}_{\boldsymbol{z}+r}\boldsymbol{D}'_{\boldsymbol{z}+}\boldsymbol{D}_{\boldsymbol{z}+}\boldsymbol{U}^*_{\boldsymbol{z}+r} & \boldsymbol{U}^{*'}_{\boldsymbol{x}+f}\boldsymbol{U}^*_{\boldsymbol{x}+r}\otimes\boldsymbol{U}^{*'}_{\boldsymbol{z}+r}\boldsymbol{D}'_{\boldsymbol{z}+}\boldsymbol{D}_{\boldsymbol{z}+}\boldsymbol{U}^*_{\boldsymbol{z}+r} \\
\boldsymbol{O} & \boldsymbol{O} & \boldsymbol{U}^{*'}_{\boldsymbol{x}+r}\boldsymbol{U}^*_{\boldsymbol{x}+f}\otimes\boldsymbol{U}^{*'}_{\boldsymbol{z}+r}\boldsymbol{D}'_{\boldsymbol{z}+}\boldsymbol{D}_{\boldsymbol{z}+}\boldsymbol{U}^*_{\boldsymbol{z}+r} & \boldsymbol{U}^{*'}_{\boldsymbol{x}+r}\boldsymbol{U}^*_{\boldsymbol{x}+r}\otimes\boldsymbol{U}^{*'}_{\boldsymbol{z}+r}\boldsymbol{D}'_{\boldsymbol{z}+}\boldsymbol{D}_{\boldsymbol{z}+}\boldsymbol{U}^*_{\boldsymbol{z}+r}
\end{bmatrix} \\
&+ \lambda_{\boldsymbol{x}}\begin{bmatrix}
\boldsymbol{O}_{q_{\boldsymbol{x}}}\otimes\boldsymbol{U}^{*'}_{\boldsymbol{z}+f}\boldsymbol{I}_{c_{\boldsymbol{z}+}}\boldsymbol{U}^{*'}_{\boldsymbol{z}+f} & \boldsymbol{O} & \boldsymbol{O} & \boldsymbol{O} \\
\boldsymbol{O} & \boldsymbol{U}^{*'}_{\boldsymbol{x}+r}\boldsymbol{D}'_{\boldsymbol{x}+}\boldsymbol{D}_{\boldsymbol{x}+}\boldsymbol{U}^*_{\boldsymbol{x}+r}\otimes\boldsymbol{U}^{*'}_{\boldsymbol{z}+f}\boldsymbol{U}^*_{\boldsymbol{z}+f} & \boldsymbol{O} & \boldsymbol{U}^{*'}_{\boldsymbol{x}+r}\boldsymbol{D}'_{\boldsymbol{x}+}\boldsymbol{D}_{\boldsymbol{x}+}\boldsymbol{U}^*_{\boldsymbol{x}+r}\otimes\boldsymbol{U}^{*'}_{\boldsymbol{z}+f}\boldsymbol{U}^*_{\boldsymbol{z}+r} \\
\boldsymbol{O} & \boldsymbol{O} & \boldsymbol{O}_{q_{\boldsymbol{x}}}\otimes\boldsymbol{U}^{*'}_{\boldsymbol{z}+r}\boldsymbol{U}^*_{\boldsymbol{z}+r} & \boldsymbol{O} \\
\boldsymbol{O} & \boldsymbol{U}^{*'}_{\boldsymbol{x}+r}\boldsymbol{D}'_{\boldsymbol{x}+}\boldsymbol{D}_{\boldsymbol{x}+}\boldsymbol{U}^*_{\boldsymbol{x}+r}\otimes\boldsymbol{U}^{*'}_{\boldsymbol{z}+r}\boldsymbol{U}^*_{\boldsymbol{z}+f} & \boldsymbol{O} & \boldsymbol{U}^{*'}_{\boldsymbol{x}+r}\boldsymbol{D}'_{\boldsymbol{x}+}\boldsymbol{D}_{\boldsymbol{x}+}\boldsymbol{U}^*_{\boldsymbol{x}+r}\otimes\boldsymbol{U}^{*'}_{\boldsymbol{z}+r}\boldsymbol{U}^*_{\boldsymbol{z}+r}
\end{bmatrix}
\end{aligned}
$$

Therefore, $\boldsymbol{\Phi}^*_+ = blockdiag(\boldsymbol{O}_{q_{\boldsymbol{z}}q_{\boldsymbol{x}}}, \boldsymbol{F}^*_+)$, with $\boldsymbol{F}^*_+$ given in (53). Then, the extended covariance matrix of the random effects is $\boldsymbol{G}^*_+ = \sigma^2_{\boldsymbol{\epsilon}}\boldsymbol{F}^{*-1}_+$. □

Appendix .3 Proof of Theorem 4

Proof. Given the extended transformation matrix for the random part $\boldsymbol{\Omega}_{+r}$ and the extended penalty matrix $\boldsymbol{P}_+$ defined in (75) and (74), respectively, $\boldsymbol{F}_+$ is:

$$
\begin{aligned}
\boldsymbol{F}_+ &= \boldsymbol{\Omega}'_{r+}\boldsymbol{P}_+\boldsymbol{\Omega}_{r+} \\
&= \begin{bmatrix} 0 & \boldsymbol{U}'_{\boldsymbol{z}r} & \cdots & & & \\ \vdots & & \boldsymbol{U}'_{\boldsymbol{x}_+r} & & & \\ & & & \boldsymbol{u}^{(2)\prime}_{\boldsymbol{x}_+f} \otimes \boldsymbol{U}'_{\boldsymbol{z}_+r} & & \\ & & & \boldsymbol{U}'_{\boldsymbol{x}_+r} \otimes \boldsymbol{u}^{(2)\prime}_{\boldsymbol{z}_+f} & & \\ & & & \boldsymbol{U}'_{\boldsymbol{x}_+r} \otimes \boldsymbol{U}'_{\boldsymbol{z}_+r} & & \end{bmatrix} \begin{bmatrix} 0 & & \cdots & \\ \vdots & \lambda_{\boldsymbol{z}}\boldsymbol{D}'_{\boldsymbol{z}_+}\boldsymbol{D}_{\boldsymbol{z}_+} & & \\ & & \lambda_{\boldsymbol{x}}\boldsymbol{D}'_{\boldsymbol{x}_+}\boldsymbol{D}_{\boldsymbol{x}_+} & \\ & & & \tau_{\boldsymbol{x}}\boldsymbol{D}'_{\boldsymbol{x}_+}\boldsymbol{D}_{\boldsymbol{x}_+} \otimes \boldsymbol{I}_{c_{\boldsymbol{z}_+}} + \tau_{\boldsymbol{z}}\boldsymbol{I}_{c_{\boldsymbol{x}_+}} \otimes \boldsymbol{D}'_{\boldsymbol{z}_+}\boldsymbol{D}_{\boldsymbol{z}_+} \end{bmatrix} \\
&\quad \begin{bmatrix} 0 & \cdots & & \\ \boldsymbol{U}_{\boldsymbol{z}r} & & & \\ \vdots & \boldsymbol{U}_{\boldsymbol{x}_+r} & & \\ & & \boldsymbol{u}^{(2)}_{\boldsymbol{x}_+f} \otimes \boldsymbol{U}_{\boldsymbol{z}_+r} \mid \boldsymbol{U}_{\boldsymbol{x}_+r} \otimes \boldsymbol{u}^{(2)}_{\boldsymbol{z}_+f} \mid \boldsymbol{U}_{\boldsymbol{x}_+r} \otimes \boldsymbol{U}_{\boldsymbol{z}_+r} & \end{bmatrix} \\
&= \begin{bmatrix} \lambda_{\boldsymbol{z}}\boldsymbol{U}'_{\boldsymbol{z}_+f}\boldsymbol{D}'_{\boldsymbol{z}_+}\boldsymbol{D}_{\boldsymbol{z}_+}\boldsymbol{U}_{\boldsymbol{z}_+f} & \boldsymbol{O} & \boldsymbol{O} \\ \boldsymbol{O} & \lambda_{\boldsymbol{x}}\boldsymbol{U}'_{\boldsymbol{x}_+f}\boldsymbol{D}'_{\boldsymbol{x}_+}\boldsymbol{D}_{\boldsymbol{x}_+}\boldsymbol{U}_{\boldsymbol{x}_+f} & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{O} & \tau_{\boldsymbol{x}}\boldsymbol{u}^{(2)\prime}_{\boldsymbol{x}_+f}\boldsymbol{D}'_{\boldsymbol{x}_+}\boldsymbol{D}_{\boldsymbol{x}_+}\boldsymbol{u}^{(2)}_{\boldsymbol{x}_+} \otimes \boldsymbol{U}'_{\boldsymbol{z}_+r}\boldsymbol{U}_{\boldsymbol{z}_+r} + \tau_{\boldsymbol{z}}\boldsymbol{u}^{(2)\prime}_{\boldsymbol{x}_+f}\boldsymbol{u}^{(2)}_{\boldsymbol{x}_+f} \otimes \boldsymbol{U}'_{\boldsymbol{z}_+r}\boldsymbol{D}'_{\boldsymbol{z}_+}\boldsymbol{D}_{\boldsymbol{z}_+}\boldsymbol{U}_{\boldsymbol{z}_+r} \end{bmatrix} \\
&\quad \begin{bmatrix} \boldsymbol{O} & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{O} \\ \tau_{\boldsymbol{x}}\boldsymbol{U}'_{\boldsymbol{x}_+r}\boldsymbol{D}'_{\boldsymbol{x}_+}\boldsymbol{D}_{\boldsymbol{x}_+}\boldsymbol{U}_{\boldsymbol{x}_+r} \otimes \boldsymbol{u}^{(2)\prime}_{\boldsymbol{z}_+f}\boldsymbol{u}^{(2)}_{\boldsymbol{z}_+f} + \tau_{\boldsymbol{z}}\boldsymbol{U}'_{\boldsymbol{x}_+r}\boldsymbol{U}_{\boldsymbol{x}_+r} \otimes \boldsymbol{u}^{(2)\prime}_{\boldsymbol{z}_+f}\boldsymbol{D}'_{\boldsymbol{z}_+}\boldsymbol{D}_{\boldsymbol{z}_+}\boldsymbol{u}^{(2)}_{\boldsymbol{z}_+f} & \boldsymbol{O} \\ \boldsymbol{O} & \tau_{\boldsymbol{x}}\boldsymbol{U}'_{\boldsymbol{x}_+r}\boldsymbol{D}'_{\boldsymbol{x}_+}\boldsymbol{D}_{\boldsymbol{x}_+}\boldsymbol{U}_{\boldsymbol{x}_+r} \otimes \boldsymbol{U}'_{\boldsymbol{z}_+r}\boldsymbol{U}_{\boldsymbol{z}_+r} + \tau_{\boldsymbol{z}}\boldsymbol{U}'_{\boldsymbol{x}_+r}\boldsymbol{U}_{\boldsymbol{x}_+r} \otimes \boldsymbol{U}'_{\boldsymbol{z}_+r}\boldsymbol{D}'_{\boldsymbol{z}_+}\boldsymbol{D}_{\boldsymbol{z}_+}\boldsymbol{U}_{\boldsymbol{z}_+r} \end{bmatrix},
\end{aligned}
$$

using $\boldsymbol{U}'_{ir}\boldsymbol{D}'_i\boldsymbol{D}_i\boldsymbol{U}_{ir} = \tilde{\boldsymbol{\Sigma}}_i$, $\boldsymbol{u}^{(2)\prime}_{if}\boldsymbol{D}'_i = \boldsymbol{O}$, $\boldsymbol{u}^{(2)\prime}_{if}\boldsymbol{u}^{(2)}_{if} = 1$ and $\boldsymbol{U}'_{ir}\boldsymbol{U}_{ir} = \boldsymbol{I}_{c_i-q_i}$, for $i = \boldsymbol{z}_+, \boldsymbol{x}_+$, we obtain the extended mixed model penalty $\boldsymbol{F}_+$ in (78). □

Appendix .4 Proof of Theorem 5

Proof. Given the extended transformation matrix for the random part $\boldsymbol{\Omega}_{+r}$ and the extended penalty matrix $\boldsymbol{P}_+$ defined in (82) and (74), respectively, $\boldsymbol{F}^*_+$ is:

$$
\begin{aligned}
\boldsymbol{F}^*_+ = \boldsymbol{\Omega}^{*\prime}_{r+}\boldsymbol{P}_+\boldsymbol{\Omega}^*_{r+} &= \begin{bmatrix} 0 & \boldsymbol{U}^{*\prime}_{\boldsymbol{z}_+r} & \cdots & \\ \vdots & & \boldsymbol{U}^{*\prime}_{\boldsymbol{x}_+r} & \\ & & & \boldsymbol{u}^{*(2)\prime}_{\boldsymbol{x}_+f} \otimes \boldsymbol{U}^{*\prime}_{\boldsymbol{z}_+r} \\ & & & \boldsymbol{U}^{*\prime}_{\boldsymbol{x}_+r} \otimes \boldsymbol{u}^{*(2)\prime}_{\boldsymbol{z}_+f} \\ & & & \boldsymbol{U}^{*\prime}_{\boldsymbol{x}_+r} \otimes \boldsymbol{U}^{*\prime}_{\boldsymbol{z}_+r} \end{bmatrix} \begin{bmatrix} 0 & & \cdots & \\ \vdots & \lambda_{\boldsymbol{z}}\boldsymbol{D}'_{\boldsymbol{z}_+}\boldsymbol{D}_{\boldsymbol{z}_+} & & \\ & & \lambda_{\boldsymbol{x}}\boldsymbol{D}'_{\boldsymbol{x}_+}\boldsymbol{D}_{\boldsymbol{x}_+} & \\ & & & \lambda_3\boldsymbol{D}'_{\boldsymbol{x}_+}\boldsymbol{D}_{\boldsymbol{x}_+} \otimes \boldsymbol{I}_{c_{\boldsymbol{z}_+}} + \lambda_4\boldsymbol{I}_{c_{\boldsymbol{x}_+}} \otimes \boldsymbol{D}'_{\boldsymbol{z}_+}\boldsymbol{D}_{\boldsymbol{z}_+} \end{bmatrix} \\
&\quad \begin{bmatrix} 0 & \cdots & \\ \boldsymbol{U}_{\boldsymbol{z}r} & & \\ \vdots & \boldsymbol{U}^*_{\boldsymbol{x}_+r} & \\ & & \boldsymbol{u}^{*(2)}_{\boldsymbol{x}_+f} \otimes \boldsymbol{U}^*_{\boldsymbol{z}_+r} \mid \boldsymbol{U}^*_{\boldsymbol{x}_+r} \otimes \boldsymbol{u}^{*(2)}_{\boldsymbol{z}_+f} \mid \boldsymbol{U}^*_{\boldsymbol{x}_+r} \otimes \boldsymbol{U}^*_{\boldsymbol{z}_+r} \end{bmatrix} \\
&= \begin{bmatrix} \lambda_{\boldsymbol{z}}\boldsymbol{U}^{*\prime}_{\boldsymbol{z}_+r}\boldsymbol{D}'_{\boldsymbol{z}_+}\boldsymbol{D}_{\boldsymbol{z}_+}\boldsymbol{U}^*_{\boldsymbol{z}_+r} & & \\ & \lambda_{\boldsymbol{x}}\boldsymbol{U}^{*\prime}_{\boldsymbol{x}_+r}\boldsymbol{D}'_{\boldsymbol{x}_+}\boldsymbol{D}_{\boldsymbol{x}_+}\boldsymbol{U}^*_{\boldsymbol{x}_+r} & \\ & & \boldsymbol{F}^{(1,2)}_+ \end{bmatrix},
\end{aligned}
$$

where $\boldsymbol{F}^{(1,2)}_+ = \begin{bmatrix} \boldsymbol{F}^{(1,2)}_{+11} & \boldsymbol{O} & \boldsymbol{F}^{(1,2)}_{+13} \\ \boldsymbol{O} & \boldsymbol{F}^{(1,2)}_{+22} & \boldsymbol{O} \\ \boldsymbol{F}^{(1,2)\prime}_{+13} & \boldsymbol{O} & \boldsymbol{F}^{(1,2)}_{+33} \end{bmatrix}$, with

$$
\begin{aligned}
\boldsymbol{F}^{(1,2)}_{+11} &= \tau_{\boldsymbol{x}}\boldsymbol{u}^{*(2)\prime}_{\boldsymbol{x}_+f}\boldsymbol{D}'_{\boldsymbol{x}_+}\boldsymbol{D}_{\boldsymbol{x}_+}\boldsymbol{u}^{*(2)}_{\boldsymbol{x}_+f} \otimes \boldsymbol{U}^{*\prime}_{\boldsymbol{z}r}\boldsymbol{U}^*_{\boldsymbol{z}_+r} + \tau_{\boldsymbol{z}}\boldsymbol{u}^{*(2)\prime}_{\boldsymbol{x}_+f}\boldsymbol{u}^{*(2)}_{\boldsymbol{x}_+f} \otimes \boldsymbol{U}^{*\prime}_{\boldsymbol{z}_+r}\boldsymbol{D}'_{\boldsymbol{z}_+}\boldsymbol{D}_{\boldsymbol{z}_+}\boldsymbol{U}^*_{\boldsymbol{z}_+r} \\
\boldsymbol{F}^{(1,2)}_{+13} &= \tau_{\boldsymbol{z}}\boldsymbol{u}^{*(2)\prime}_{\boldsymbol{x}_+f}\boldsymbol{U}^*_{\boldsymbol{x}_+r} \otimes \boldsymbol{U}^*_{\boldsymbol{z}_+r}\boldsymbol{D}'_{\boldsymbol{z}_+}\boldsymbol{D}_{\boldsymbol{z}_+}\boldsymbol{U}^*_{\boldsymbol{z}_+r} + \tau_{\boldsymbol{x}}\boldsymbol{u}^{*(2)\prime}_{\boldsymbol{x}_+f}\boldsymbol{D}'_{\boldsymbol{x}_+}\boldsymbol{D}_{\boldsymbol{x}_+}\boldsymbol{U}^*_{\boldsymbol{x}_+r} \otimes \boldsymbol{U}^{*\prime}_{\boldsymbol{z}_+r}\boldsymbol{U}^*_{\boldsymbol{z}_+r} \\
\boldsymbol{F}^{(1,2)}_{+22} &= \tau_{\boldsymbol{x}}\boldsymbol{U}^{*\prime}_{\boldsymbol{x}_+r}\boldsymbol{D}'_{\boldsymbol{x}_+}\boldsymbol{D}_{\boldsymbol{x}_+}\boldsymbol{U}^*_{\boldsymbol{x}_+r} \otimes \boldsymbol{u}^{(*2)\prime}_{\boldsymbol{z}_+f}\boldsymbol{u}^{*(2)}_{\boldsymbol{z}_+f} + \tau_{\boldsymbol{z}}\boldsymbol{U}^{*\prime}_{\boldsymbol{x}_+r}\boldsymbol{U}^*_{\boldsymbol{x}_+r} \otimes \boldsymbol{u}^{*(2)\prime}_{\boldsymbol{z}_+f}\boldsymbol{D}'_{\boldsymbol{z}_+}\boldsymbol{D}_{\boldsymbol{z}_+}\boldsymbol{u}^{*(2)}_{\boldsymbol{z}_+f} \\
\boldsymbol{F}^{(1,2)}_{+23} &= \tau_{\boldsymbol{x}}\boldsymbol{U}^{*\prime}_{\boldsymbol{x}_+r}\boldsymbol{D}'_{\boldsymbol{x}_+}\boldsymbol{D}_{\boldsymbol{x}_+}\boldsymbol{U}^*_{\boldsymbol{x}_+r} \otimes \boldsymbol{u}^{*(2)\prime}_{\boldsymbol{z}_+f}\boldsymbol{U}^*_{\boldsymbol{z}_+r} + \tau_{\boldsymbol{z}}\boldsymbol{U}^{*\prime}_{\boldsymbol{x}_+r}\boldsymbol{U}^*_{\boldsymbol{x}_+r} \otimes \boldsymbol{u}^{*(2)\prime}_{\boldsymbol{z}_+f}\boldsymbol{D}'_{\boldsymbol{z}_+}\boldsymbol{D}_{\boldsymbol{z}_+}\boldsymbol{U}^*_{\boldsymbol{z}_+r} \\
\boldsymbol{F}^{(1,2)}_{+33} &= \tau_{\boldsymbol{z}}\boldsymbol{U}^{*\prime}_{\boldsymbol{x}_+r}\boldsymbol{U}^*_{\boldsymbol{x}_+r} \otimes \boldsymbol{U}^{*\prime}_{\boldsymbol{z}_+r}\boldsymbol{D}'_{\boldsymbol{z}_+}\boldsymbol{D}_{\boldsymbol{z}_+}\boldsymbol{U}^*_{\boldsymbol{z}_+r} + \tau_{\boldsymbol{x}}\boldsymbol{U}^{*\prime}_{\boldsymbol{x}_+r}\boldsymbol{D}'_{\boldsymbol{x}_+}\boldsymbol{D}_{\boldsymbol{x}_+}\boldsymbol{U}^*_{\boldsymbol{x}_+r} \otimes \boldsymbol{U}^{*\prime}_{\boldsymbol{z}_+r}\boldsymbol{U}^*_{\boldsymbol{z}_+r}
\end{aligned}
$$

using $\boldsymbol{u}^{*(2)\prime}_{if}\boldsymbol{D}'_i = \boldsymbol{O}$ for $i = \boldsymbol{z}_+, \boldsymbol{x}_+$, we obtain the extended mixed model penalty $\boldsymbol{F}^*_+$ in (85). □

Appendix .5 Simulation study results

Simulations results for Scenario 1:

- $n_{z_p} = 0$, $n_{x_p} = 10$

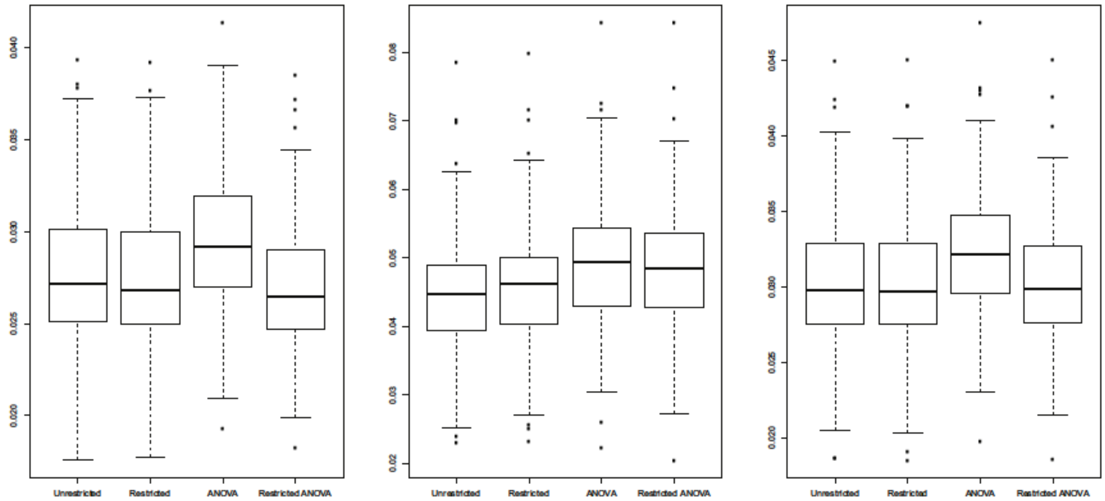


Figure .11: MAE in the fit (left panel), in the forecast (middle panel) and in total (right panel) of smooth models in scenario 1 and $n_{z_p} = 0$ and $n_{x_p} = 10$.

- $n_{z_p} = 0$, $n_{x_p} = 15$

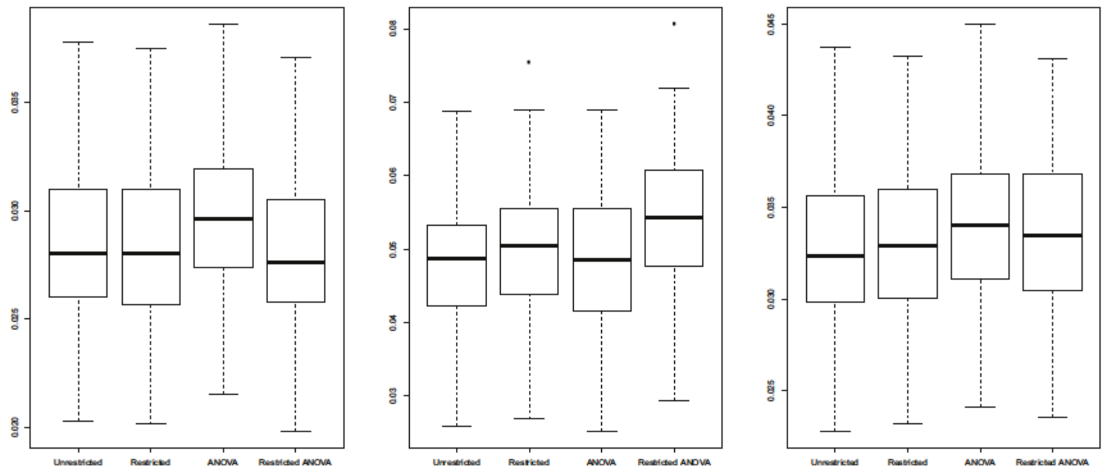


Figure .12: MAE in the fit (left panel), in the forecast (middle panel) and in total (right panel) of smooth models in scenario 1 and $n_{z_p} = 0$ and $n_{x_p} = 15$.

- $n_{z_p} = 0$, $n_{x_p} = 20$

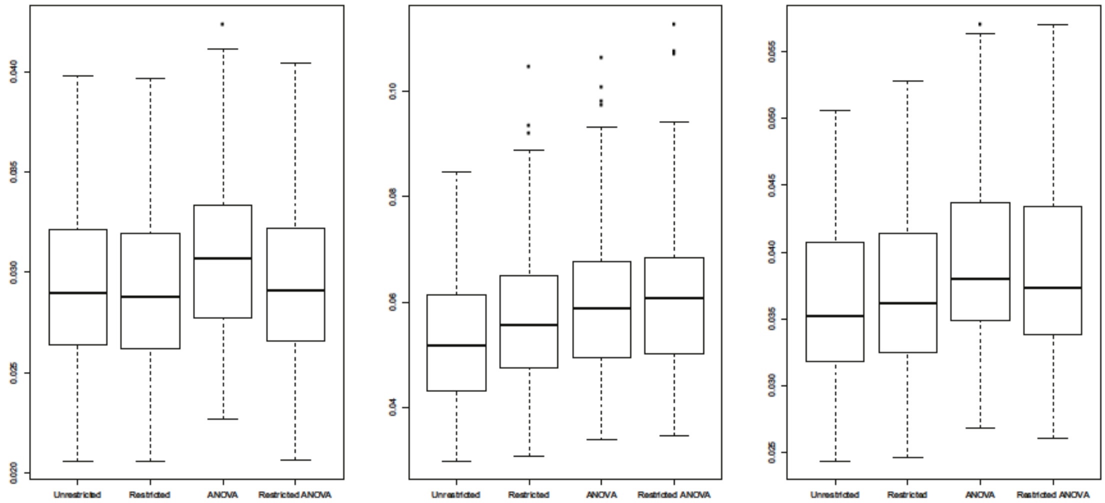


Figure .13: MAE in the fit (left panel), in the forecast (middle panel) and in total (right panel) of smooth models in scenario 1 and $n_{z_p} = 0$ and $n_{x_p} = 20$.

- $n_{z_p} = 10$, $n_{x_p} = 5$

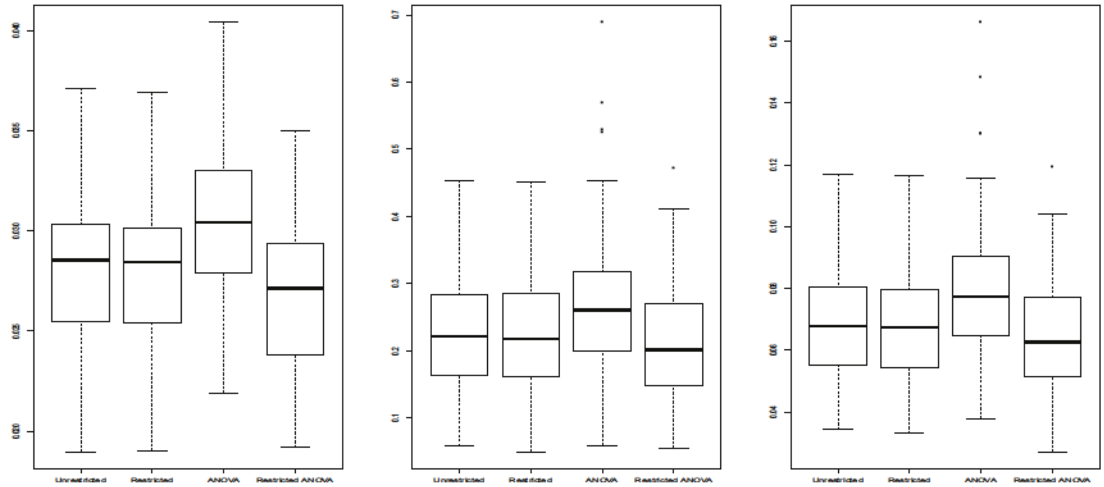


Figure .14: MAE in the fit (left panel), in the forecast (middle panel) and in total (right panel) of smooth models in scenario 2 and $n_{z_p} = 10$, $n_{x_p} = 5$.

- $n_{z_p} = n_{x_p} = 10$

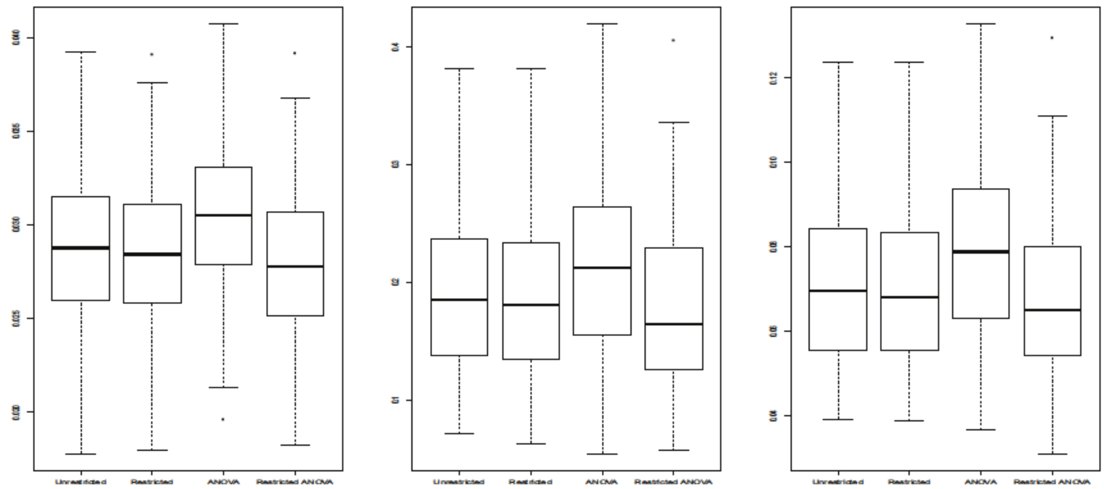


Figure .15: MAE in the fit (left panel), in the forecast (middle panel) and in total (right panel) of smooth models in scenario 2 and $n_{z_p} = 10$, $n_{x_p} = 10$.

- $n_{z_p} = 10$, $n_{x_p} = 15$

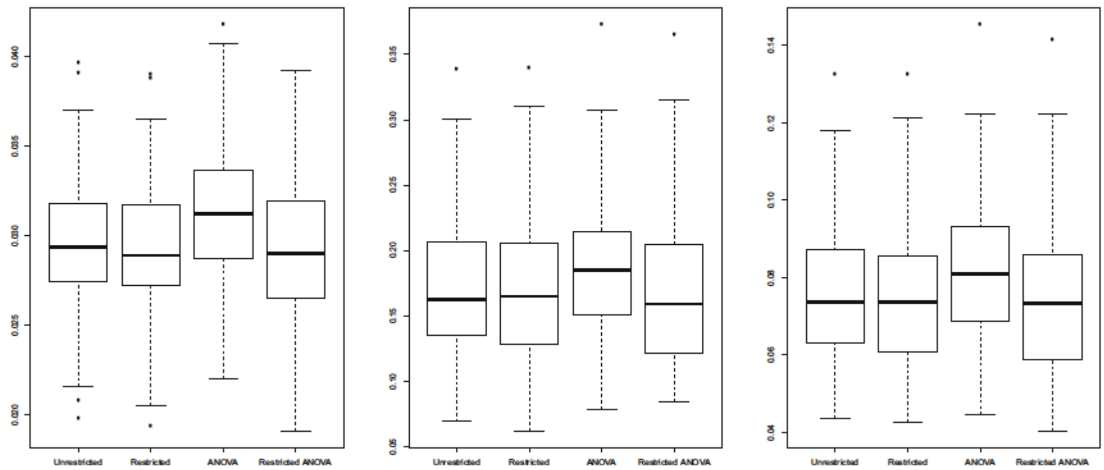


Figure .16: MAE in the fit (left panel), in the forecast (middle panel) and in total (right panel) of smooth models in scenario 2 and $n_{z_p} = 10$, $n_{x_p} = 15$.

- $n_{z_p} = 10$, $n_{x_p} = 20$

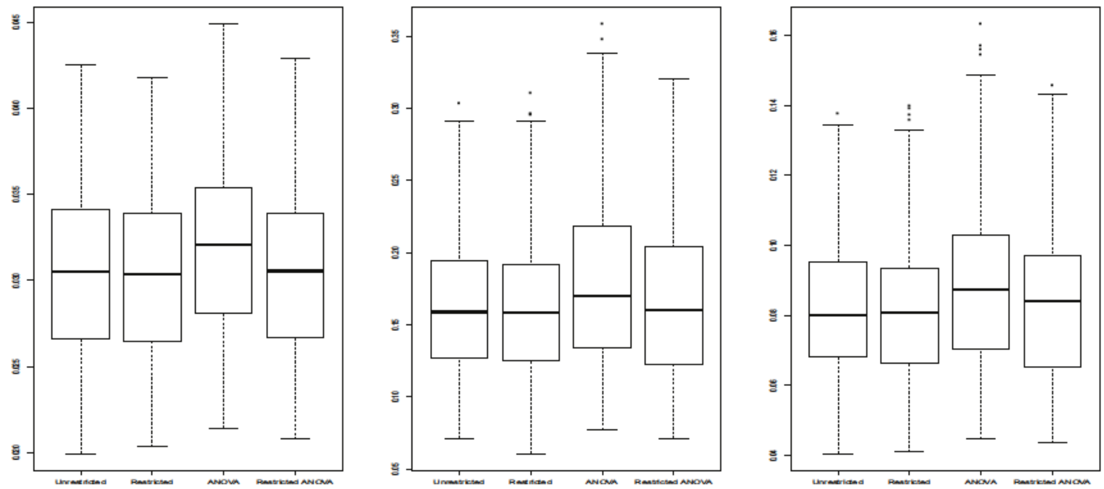


Figure .17: MAE in the fit (left panel), in the forecast (middle panel) and in total (right panel) of smooth models in scenario 2 and $n_{z_p} = 10$, $n_{x_p} = 30$.

- $n_{z_p} = 20$, $n_{x_p} = 5$

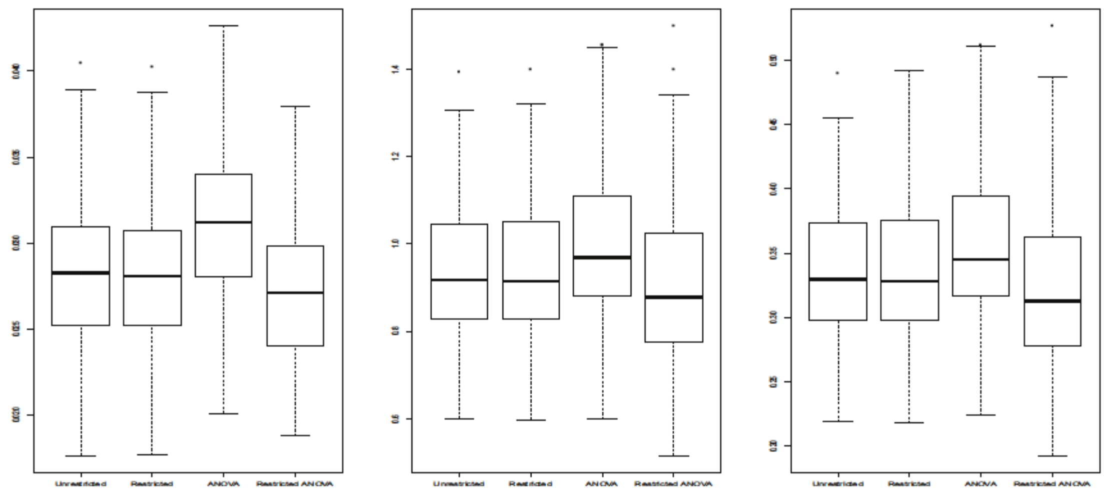


Figure .18: MAE in the fit (left panel), in the forecast (middle panel) and in total (right panel) of smooth models in scenario 2 and $n_{z_p} = 20$, $n_{x_p} = 5$.

- $n_{z_p} = 20$, $n_{x_p} = 10$

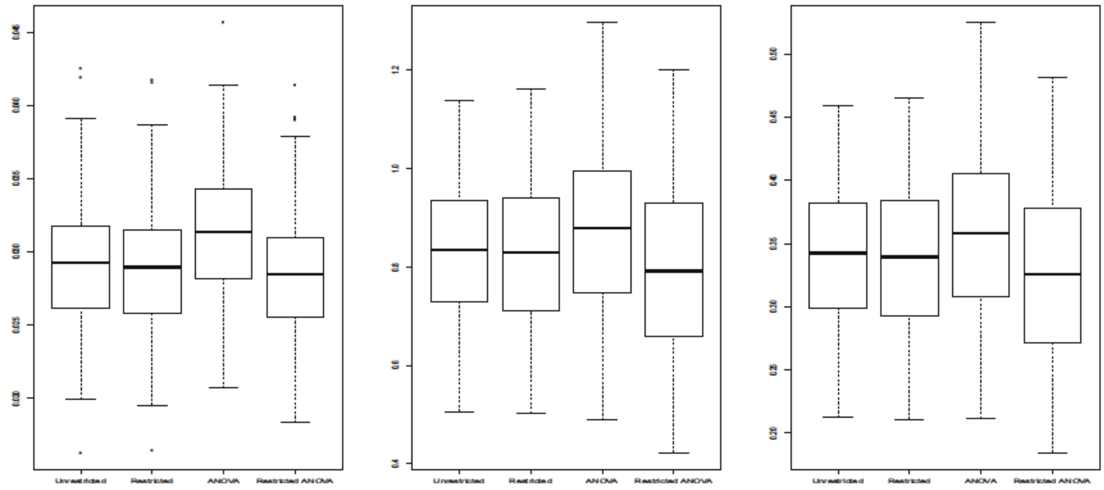


Figure .19: MAE in the fit (left panel), in the forecast (middle panel) and in total (right panel) of smooth models in scenario 2 and $n_{z_p} = 20$, $n_{x_p} = 10$.

- $n_{z_p} = 20$, $n_{x_p} = 15$

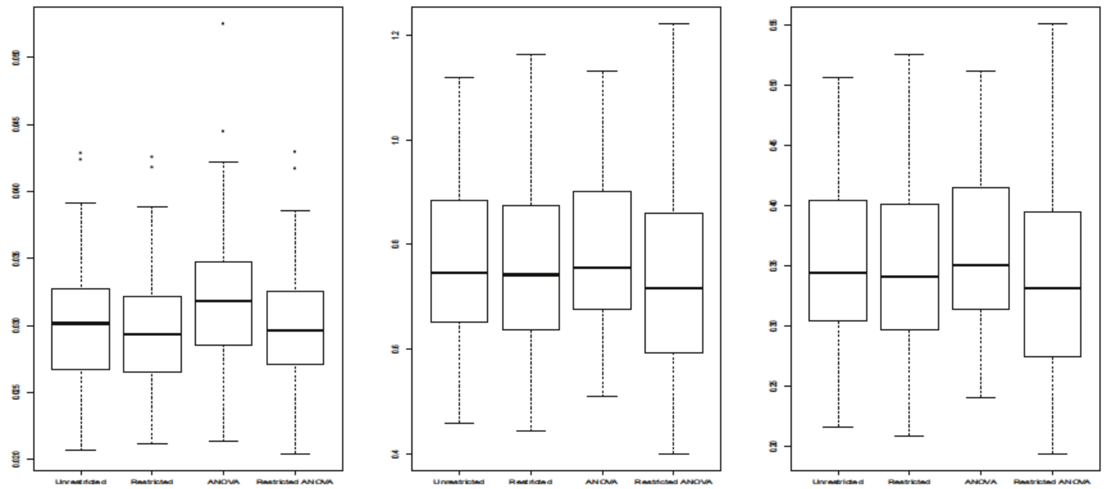


Figure .20: MAE in the fit (left panel), in the forecast (middle panel) and in total (right panel) of smooth models in scenario 2 and $n_{z_p} = 20$, $n_{x_p} = 15$.

- $n_{z_p} = 20$, $n_{x_p} = 20$

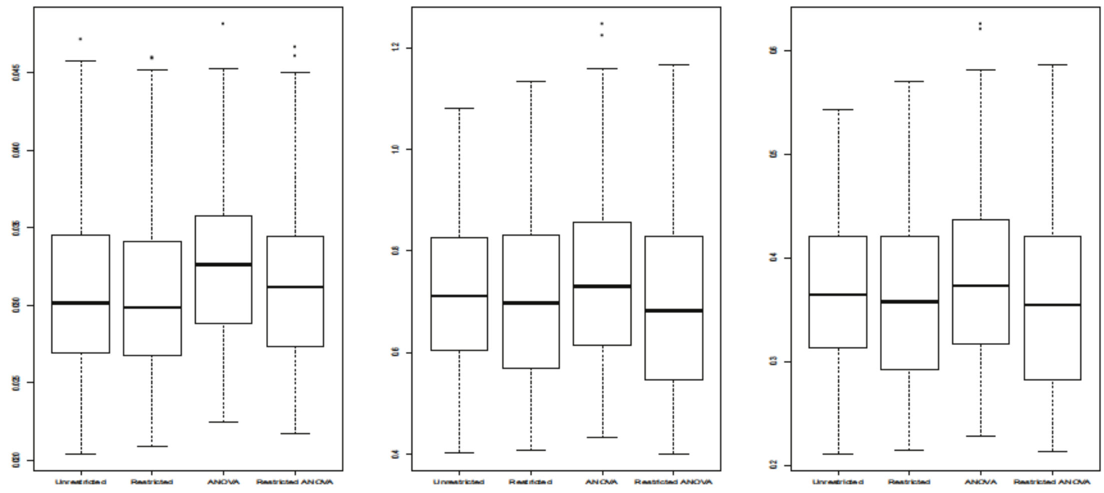


Figure .21: MAE in the fit (left panel), in the forecast (middle panel) and in total (right panel) of smooth models in scenario 2 and $n_{z_p} = 20$, $n_{x_p} = 20$.

Simulations results for Scenario 2:

- $n_{z_p} = 0$, $n_{x_p} = 10$

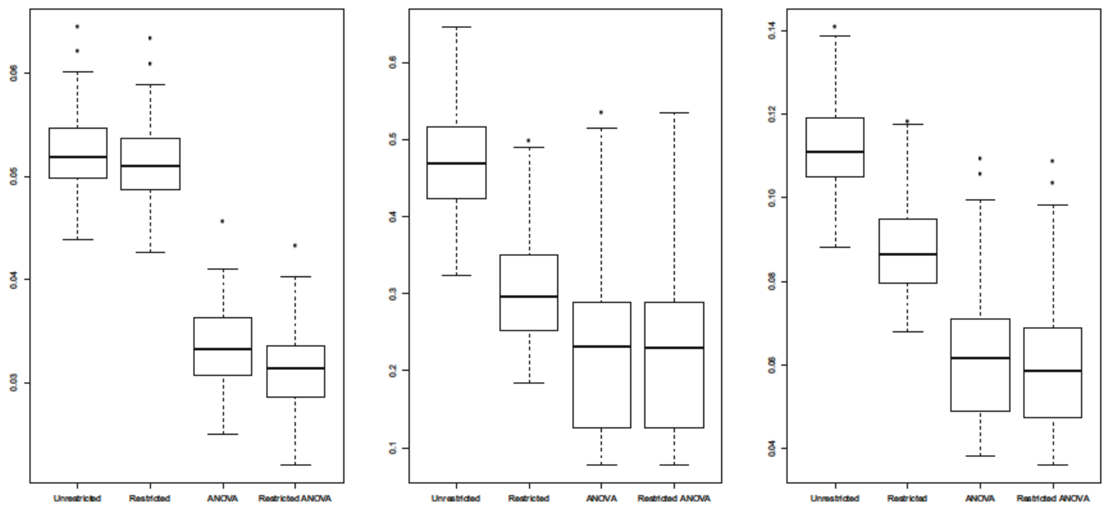


Figure .22: MAE in the fit (left panel), in the forecast (middle panel) and in total (right panel) of smooth models in scenario 2 and $n_{z_p} = 0$ and $n_{x_p} = 10$.

- $n_{z_p} = 0$, $n_{x_p} = 15$

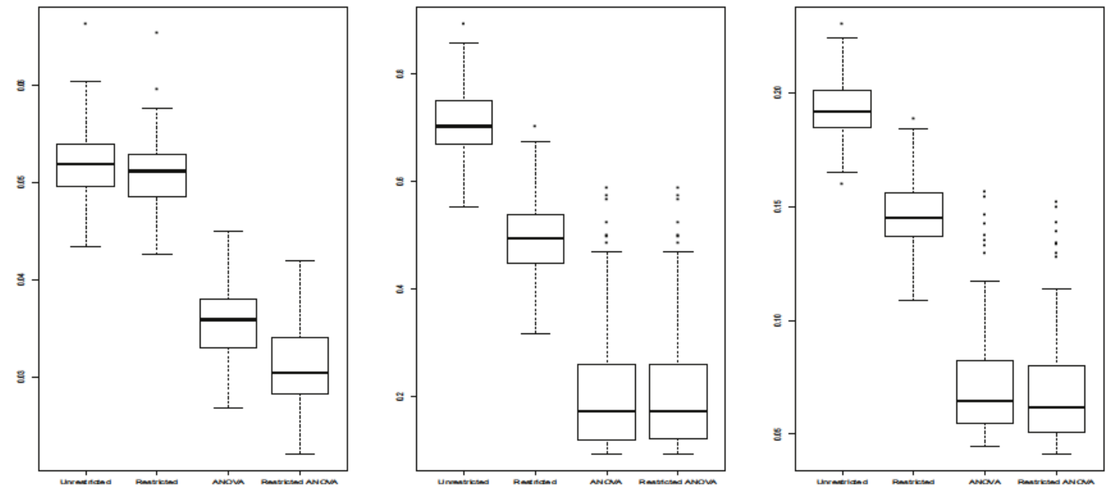


Figure .23: MAE in the fit (left panel), in the forecast (middle panel) and in total (right panel) of smooth models in scenario 2 and $n_{z_p} = 0$ and $n_{x_p} = 15$.

- $n_{z_p} = 0$, $n_{x_p} = 20$

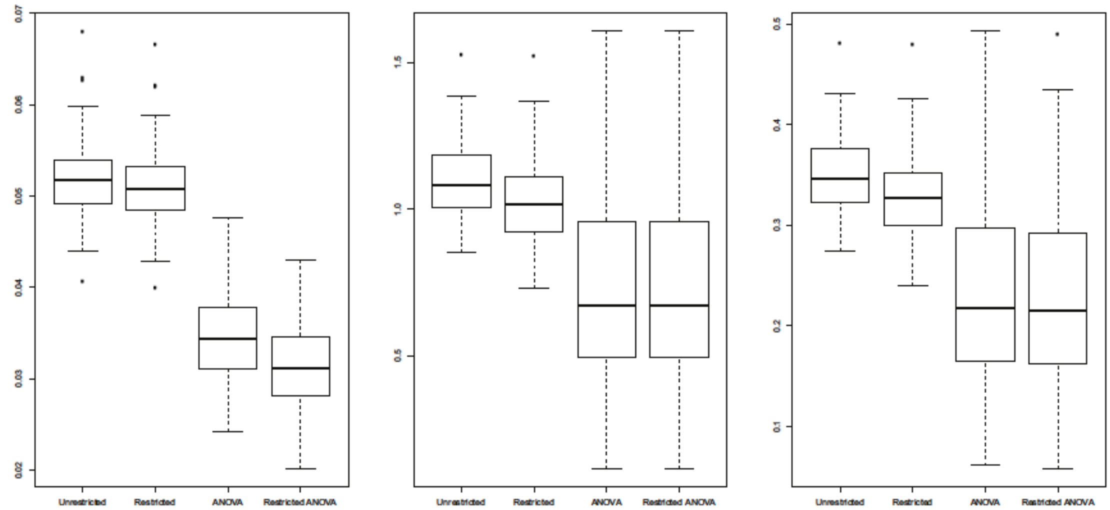


Figure .24: MAE in the fit (left panel), in the forecast (middle panel) and in total (right panel) of smooth models in scenario 2 and $n_{z_p} = 0$ and $n_{x_p} = 20$.

- $n_{z_p} = 10$, $n_{x_p} = 5$

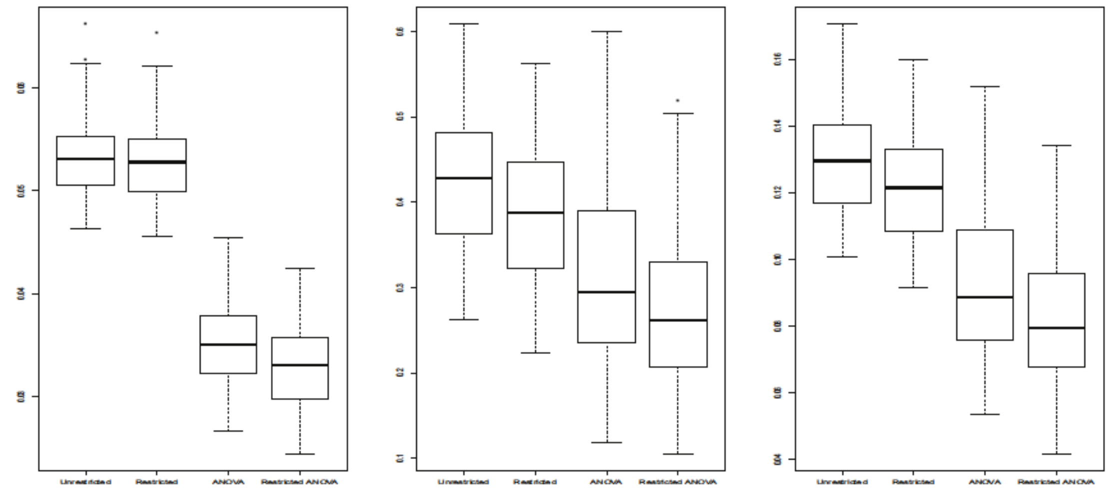


Figure .25: MAE in the fit (left panel), in the forecast (middle panel) and in total (right panel) of smooth models in scenario 2 and $n_{z_p} = 10$ and $n_{x_p} = 5$.

- $n_{z_p} = n_{x_p} = 10$

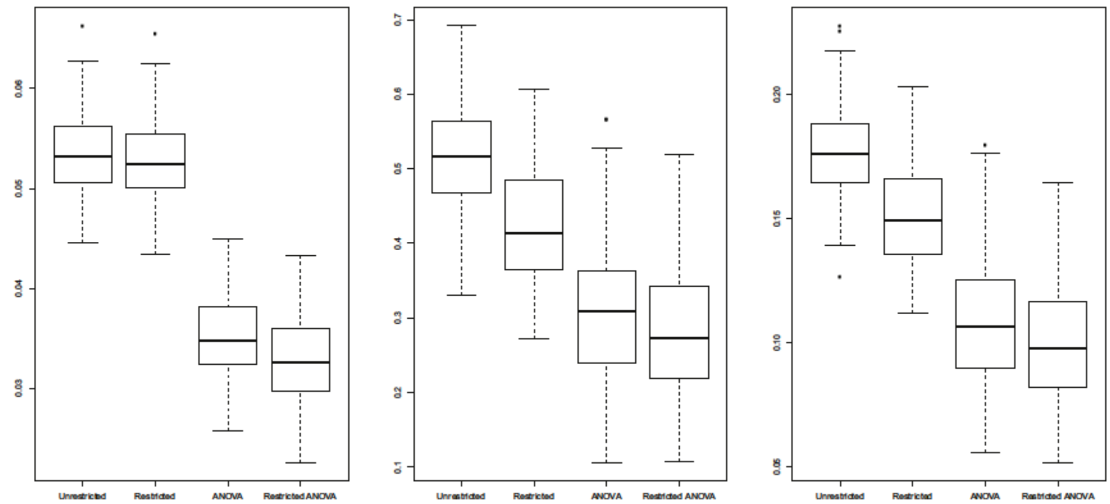


Figure .26: MAE in the fit (left panel), in the forecast (middle panel) and in total (right panel) of smooth models in scenario 2 and $n_{z_p} = n_{x_p} = 10$.

- $n_{z_p} = 10$, $n_{x_p} = 15$

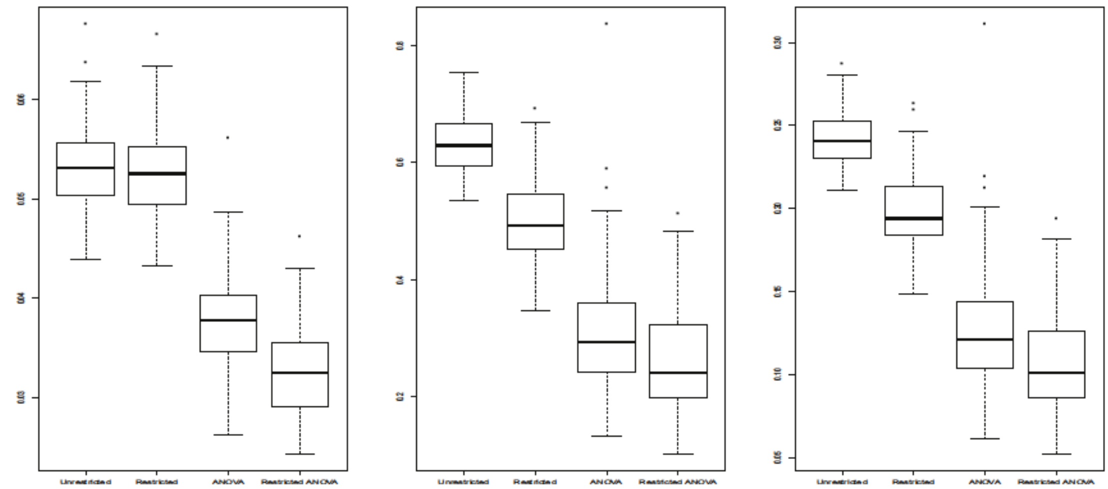


Figure .27: MAE in the fit (left panel), in the forecast (middle panel) and in total (right panel) of smooth models in scenario 2 and $n_{z_p} = 10$ and $n_{x_p} = 15$.

- $n_{z_p} = 10$, $n_{x_p} = 20$

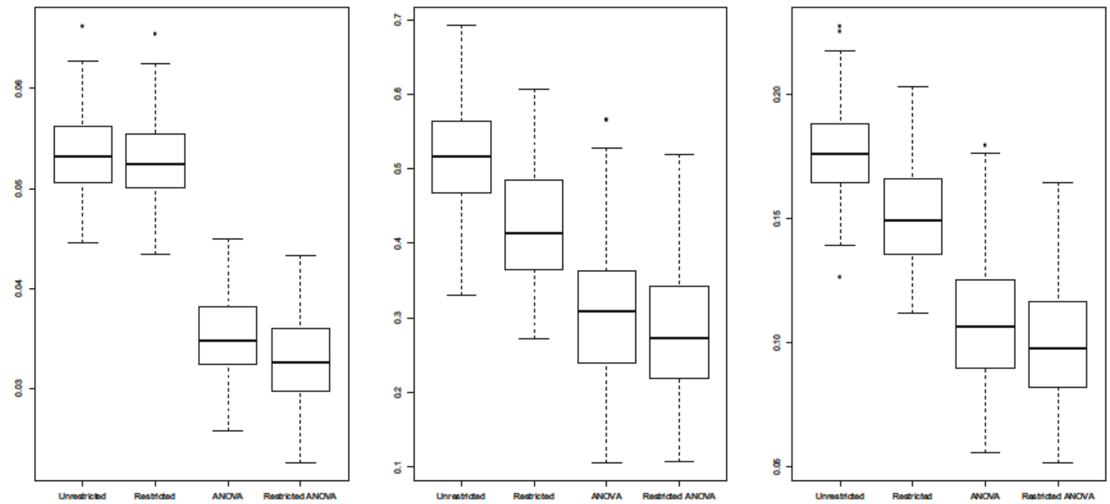


Figure .28: MAE in the fit (left panel), in the forecast (middle panel) and in total (right panel) of smooth models in scenario 2 and $n_{z_p} = 10$ and $n_{x_p} = 20$.

- $n_{z_p} = 20$, $n_{x_p} = 5$

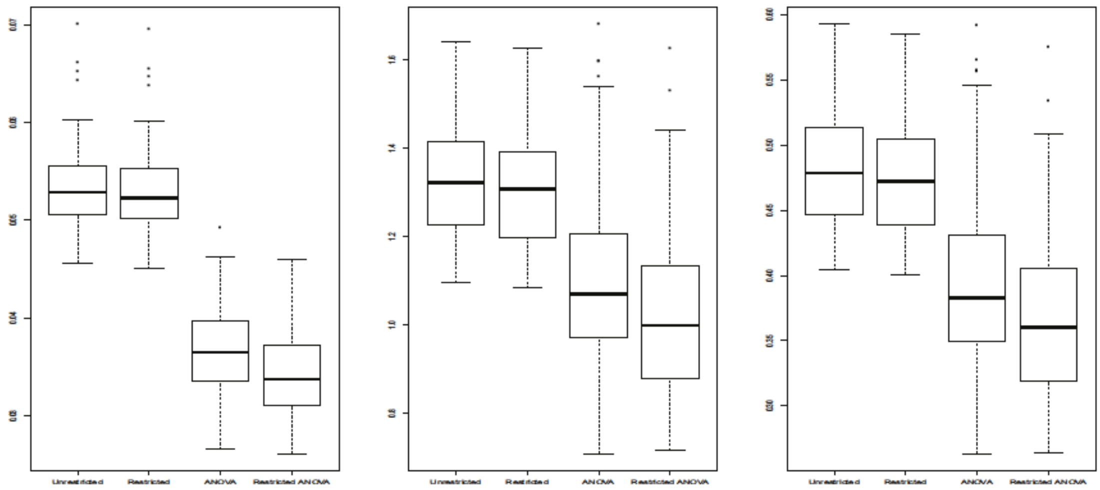


Figure .29: MAE in the fit (left panel), in the forecast (middle panel) and in total (right panel) of smooth models in scenario 2 and $n_{z_p} = 20$ and $n_{x_p} = 5$.

- $n_{z_p} = 20$, $n_{x_p} = 10$

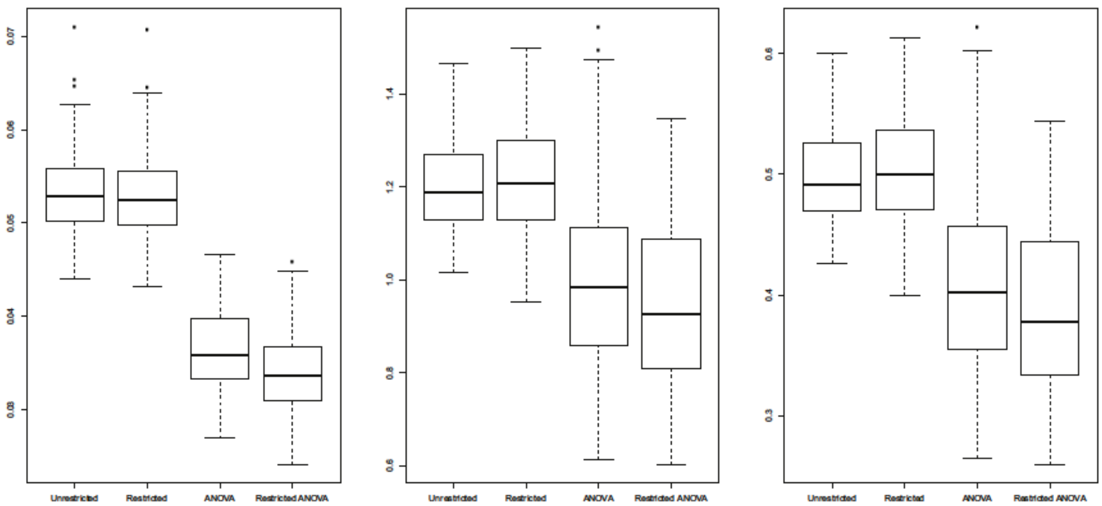


Figure .30: MAE in the fit (left panel), in the forecast (middle panel) and in total (right panel) of smooth models in scenario 2 and $n_{z_p} = 20$ and $n_{x_p} = 10$.

- $n_{z_p} = 20$, $n_{x_p} = 15$

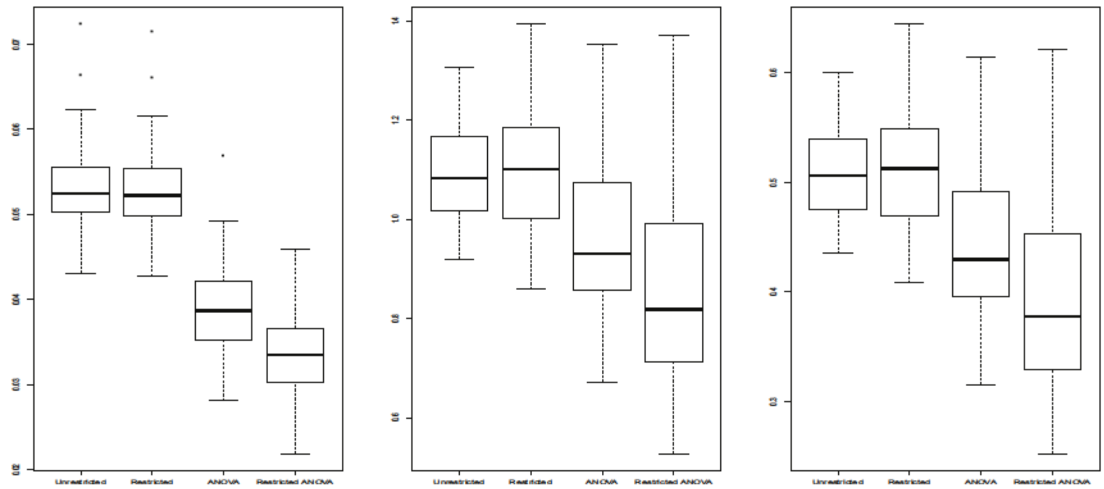


Figure .31: MAE in the fit (left panel), in the forecast (middle panel) and in total (right panel) of smooth models in scenario 2 and $n_{z_p} = 20$ and $n_{x_p} = 15$.

- $n_{z_p} = n_{x_p} = 20$

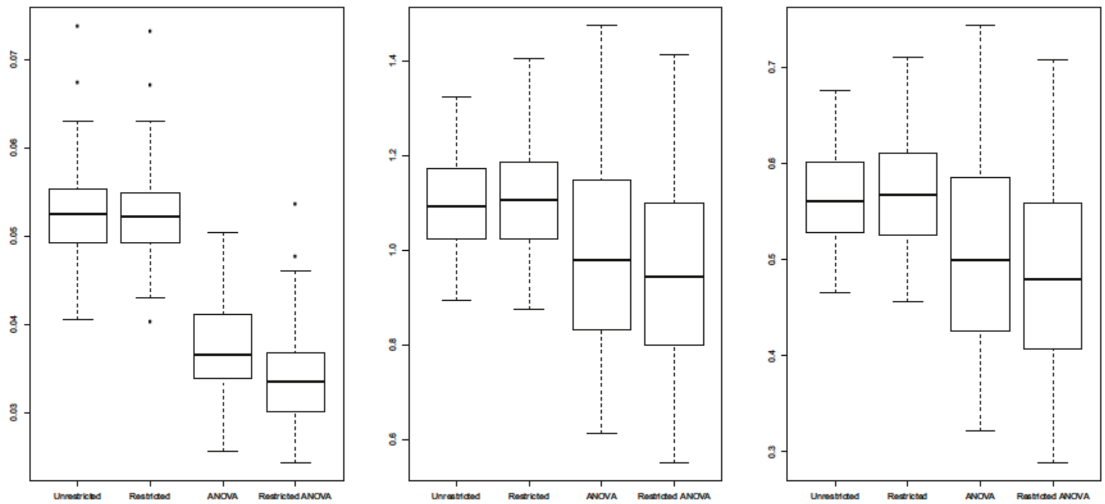


Figure .32: MAE in the fit (left panel), in the forecast (middle panel) and in total (right panel) of smooth models in scenario 2 and $n_{z_p} = n_{x_p} = 20$.