



UNIVERSIDAD CARLOS III DE MADRID

working  
papers

Working Paper 12-25-(19)  
Statistics and Econometrics Series  
November 2012

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

## SEASONAL MODULATION MIXED MODELS FOR TIME SERIES FORECASTING

Dae-Jin Lee<sup>1</sup> and Maria Durbán<sup>2</sup>

### Abstract

---

We propose an extension of a seasonal modulation smooth model with P-splines for times series data using a mixed model formulation. A smooth trend with seasonality decomposition can be estimated simultaneously. We extend the model to consider the forecasting of new future observations in the mixed model framework. Two different approaches are used for forecasting in the context of mixed models, and the equivalence of both methods is shown. The methodology is illustrated with monthly sulphur dioxide (SO<sub>2</sub>) levels in a selection of monitoring sites in Europe from January 1990 to December 2001.

---

**Keywords:** P-splines; Mixed Models; times series forecasting; varying-coefficient models; harmonic regression

**Acknowledgement:** This research was funded by the Spanish Ministry of Science and Innovation (projects MTM2008-02901, and MTM2011-28285-C02-02). The research of Dae-Jin Lee was also funded by an NIH grant for the Superfund Metal Mixtures, Biomarkers and Neurodevelopment project 1PA2ES016454-01A2.

<sup>1</sup> CSIRO, Mathematics, Informatics and Statistics, Clayton 3169 VIC, Australia, email: Dae-Jin.Lee@csiro.au

<sup>2</sup> Departamento de Estadística, Universidad Carlos III de Madrid, Avda. Universidad 30, 28911 Leganés, Madrid, email: marialuz.durban@uc3m.es

# Seasonal modulation smoothing mixed models for time series forecasting

Dae-Jin Lee<sup>a</sup> and María Durbán<sup>b</sup>

<sup>a</sup> CSIRO MATHEMATICS, INFORMATICS AND STATISTICS, CLAYTON 3169 VIC, AUSTRALIA

<sup>b</sup> UNIVERSIDAD CARLOS III DE MADRID, ESCUELA POLITÉCNICA SUPERIOR, LEGANÉS, SPAIN

October 4, 2012

## Abstract

We propose an extension of a seasonal modulation smooth model with  $P$ -splines for times series data using a mixed model formulation. A smooth trend with seasonality decomposition can be estimated simultaneously. We extend the model to consider the forecasting of new future observations in the mixed model framework. Two different approaches are used for forecasting in the context of mixed models, and the equivalence of both methods is shown. The methodology is illustrated with monthly sulphur dioxide ( $SO_2$ ) levels in a selection of monitoring sites in Europe from January 1990 to December 2001.

**Keywords:**  $P$ -splines; Mixed Models; times series forecasting; varying-coefficient models; harmonic regression

## 1 Introduction

The decomposition of trend and seasonality is a classic problem in time series analysis. For instance, in environmental monitoring, measurements of variables such as temperature, humidity, wind direction or air pollution are collected by devices positioned over a region of interest and data are collected in a regular temporal pattern. In this paper, we considered measurements on sulphur dioxide ( $SO_2$ ) concentration levels (in  $\mu g/m^3$ ) over Europe from January 1990 to December 2001. The data were collected through the 'European monitoring and evaluation programme' (EMEP) under the Co-operative Programme for Monitoring and Evaluation of the Long-range Transmission of Air Pollutants in Europe (see further information available at <http://www.emep.int>). The main sources of sulphur dioxide are combustion of sulphur containing fuels and industrial processes, which is emitted to the atmosphere as a gas. Residential combustion for heating (principally fossil fuels as coal and heavy oils) is also a source of sulphure dioxide in particular in winter.

Since 1980 EMEP's main task has been to estimate air pollution emission across members of the European Union. The Council Directive of July 15, 1980 on Air Quality Limit Values and Guide Values for Sulphur Dioxide and Suspended Particulates<sup>1</sup> were

---

<sup>1</sup>80/779/EEC, O.J. L229, 30.08.1980, pp. 30-48

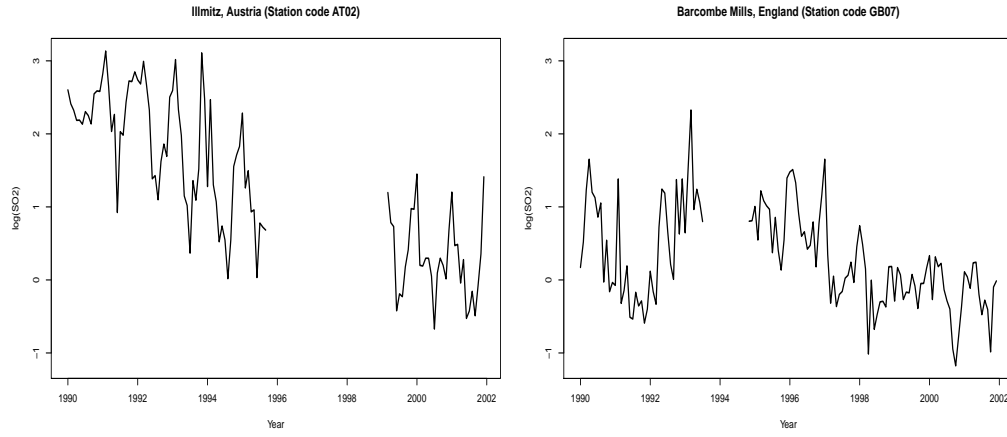


Figure 1.1: Time series plot of  $\log(SO_2)$  data.

adopted to protect human health and the environment against adverse effects from  $SO_2$  and Suspended Particulates. The study of the trends confirmed the decrease of  $SO_2$  emissions which have been reduced by an average of 20% between 1980 and 1990 in the Member States. There are several impacts of high levels of  $SO_2$  for human health (asthma), environment (vegetation) and deterioration of materials and objects of cultural heritage.

Figure 1.1 shows the time series plots of monthly averages of sulphur dioxide ( $SO_2$ ) concentration levels in  $\mu g/m^3$  measured in two monitoring sites in Illmitz (Austria, with station code AT02) and Barcombe Mills (England, with station code GB07). The data are in log scale to remove skewness. Notice that there are some missing observations (mostly due to equipment failure, replacement or calibration). In particular, for station AT02 there is a big gap between October 1995 and March 1999, for station GB07 the gap is between August 1993 and October 1994. There is clear evidence of temporal trends and seasonal effects. These effects are shown in Figure 1.2. Both series present some for of dynamics, possibly interpretable as non-stationarity. For AT02 data, the mean levels of  $\log SO_2$  shows a decreasing pattern across the years and a seasonal pattern with lower levels in the summer months of (June to August). The plots for GB07 data present a different behaviour, in particular during the period 1991 to the end of 1993, the seasonal effect is less evident.

Consider a sequence of  $x_1, \dots, x_n$  time points, the response variable  $y$  can be modelled as:

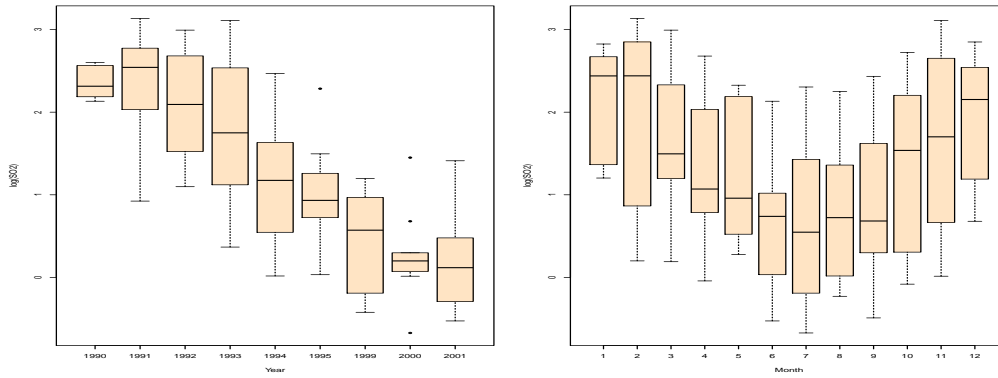
$$y_i = t(x_i) + s(x_i) + \epsilon_i, \quad i = 1, \dots, n \quad (1.1)$$

where  $t(\cdot)$  represents the temporal trend,  $s(\cdot)$  the seasonal pattern, and  $\epsilon_i$  is an error term. A simple approach for seasonal time series consist of a parametric regression model, where  $t(\cdot)$  is a polynomial regression model of order  $m^{th}$ , i.e.

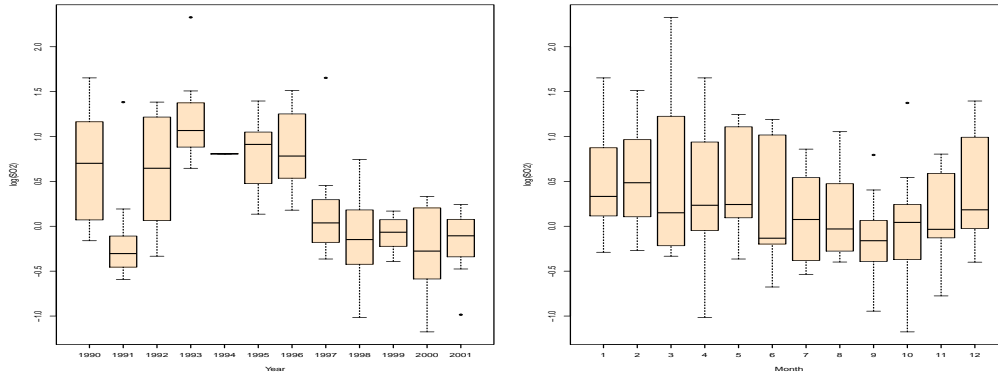
$$t(x_i) = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \dots + \beta_m x_i^m, \quad (1.2)$$

and  $s(\cdot)$  can be expressed as a sum of  $J$  Fourier series, i.e.

$$s(x_i) = \sum_{j=1}^J a_j \cos(2\pi j x_i / p) + b_j \sin(2\pi j x_i / p), \quad (1.3)$$



(a) AT02 data plotted against year (left) and month (right)



(b) GB07 data plotted against year (left) and month (right)

Figure 1.2: Boxplots of  $\log SO_2$  levels by year and month.

where  $p$  is the period (e.g.  $p = 12$  for monthly data),  $a_j$  and  $b_j$  are the regression coefficients for each harmonic. The estimation of this harmonic regression model is done by ordinary least squares. However, the parametric assumption is too restrictive, and hence, more flexibility might be needed to specify the form of the functions in (1.1). Flexibility can be provided using a non-parametric approach, letting  $t(\cdot)$  and  $s(\cdot)$  be smooth functions. In the non-parametric framework, Cleveland et al. (1990) proposed a trend and seasonality decomposition based on loess scatterplot smoother that might be useful for descriptive purposes. However, when there are big gaps in the data (as for AT02), the performance of loess or other local linear smoothers as kernels has to be carefully considered as they can run into difficulties with very small weights/bandwidths.

In this paper, we consider a flexible modelling approach based on the popular method of penalized splines (or  $P$ -splines) proposed by Eilers and Marx (1996). In particular, we consider the extension proposed in Eilers and Marx, 2002 and Eilers et al. (2008) where trend and seasonality in (1.1) are modelled as smooth terms using basis functions. Seasonality is accounted for by trigonometric terms based on Fourier series, combined with a varying-coefficients model (Hastie and Tibshirani, 1993). Both components can be estimated in the generalized additive model framework. The aim of this paper is to investigate smoothing methods to decompose seasonal time series into time trend and seasonality that can be estimated simultaneously in an unified framework, and additionally, forecast future values.

The paper is organized as follows. In Section 2, we introduce the modulation model in Eilers et al. (2008) and extend its formulation into a mixed model. The trend and seasonality of the log  $SO_2$  data for the selected stations is analysed in Section 3. In Section 4, we propose two alternative methods for forecasting new future observations using smooth modulation mixed models. We show the relationship between both methods, and showed that they yield the same results. Section 5 we evaluate the forecast procedures for the log  $SO_2$  data. Some final discussion is given in Section 6.

## 2 Penalized spline modulation model

For data with seasonal patterns Eilers et al. (2008) proposed a *smooth modulation model* given by:

$$\mathbb{E}[y_i] = f(x_i) + \sum_{j=1}^J \{g_j(x_i) \cos(j\omega x_i) + h_j(x_i) \sin(j\omega x_i)\}, \quad (2.1)$$

where  $f(\cdot)$  accounts for the smooth trend, and  $g(\cdot)$  and  $h(\cdot)$  are smooth series that describe the local amplitudes of cosine and sine waves, and  $\omega = 2\pi/p$ . The number of harmonics  $J$  required for the seasonal component is usually taken as 1 or 2 to reduce the number of parameters to be estimated. For  $J = 1$ , we can rewrite model (2.1) in a compact matrix form:

$$\mathbf{y} = \mathbf{B}\boldsymbol{\theta} + \boldsymbol{\epsilon}, \quad \boldsymbol{\epsilon} \sim \mathcal{N}(0, \boldsymbol{\Lambda}), \quad (2.2)$$

where  $\boldsymbol{\Lambda}$  is a covariance matrix of the error term. In this paper we assume uncorrelated i.i.d. errors, i.e.  $\boldsymbol{\Lambda} = \sigma^2 \mathbf{I}$ . However correlated errors can be included using the method proposed by Durbán and Currie (2003). The full regression matrix  $\mathbf{B}$  is formed by blocks (for trend and (co)-sine components), i.e.

$$\mathbf{B} = [\mathbf{B}|\mathbf{CB}|\mathbf{SB}], \quad (2.3)$$

where  $\mathbf{B}$  is a  $B$ -spline basis of size  $n \times c$  (here, for simplicity, we have used basis of the same size for every term in the model). The number of  $B$ -spline basis functions  $c$  depends on the number of knots and the order of the spline. In general we consider  $ndx$  equidistant inner knots and cubic splines, and hence  $c = ndx + 3$ . For the modulation component we have the varying-coefficient terms  $\mathbf{CB}$  and  $\mathbf{SB}$  where  $\mathbf{C} = \text{diag}\{\cos(\omega x_i)\}$ , and  $\mathbf{S} = \text{diag}\{\sin(\omega x_i)\}$ . Figure 2.1 show the components of the regression basis  $\mathbf{B}$  in (2.3) on the  $x$ -axis with  $ndx = 10$ , and hence, a total of 13 basis functions. Notice that,  $\mathbf{B}$  is the  $B$ -spline basis for the trend component, and modulation component bases  $\mathbf{CB}$  and  $\mathbf{SB}$  basically scale the rows of  $\mathbf{B}$  in the (co)-sine frequency domain (middle and bottom panels of Figure 2.1).

The vector of regression coefficients  $\boldsymbol{\theta} = (\theta, \theta_c, \theta_s)'$ , are penalized by a block-diagonal matrix, such that  $\boldsymbol{\theta}'\mathbf{P}\boldsymbol{\theta}$ , with the block-diagonal matrix  $\mathbf{P} = \boldsymbol{\lambda} \otimes D_q' D_q$ , with  $\boldsymbol{\lambda} = \text{diag}(\lambda, \check{\lambda}, \check{\lambda})$ . In practice, a single smoothing parameter  $\check{\lambda}$  is set for the (co)-sine modulation. The matrix  $D_q$  is a difference matrix of penalty order  $q$ . Examples of difference matrices for  $q = 1$  and  $q = 2$  are

$$D_1 = \begin{bmatrix} -1 & 1 & 0 & 0 & \cdots \\ 0 & -1 & 1 & 0 & \cdots \\ 0 & 0 & -1 & 1 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix} \quad \text{or} \quad D_2 = \begin{bmatrix} 1 & -2 & 1 & 0 & \cdots \\ 0 & 1 & -2 & 1 & \cdots \\ 0 & 0 & 1 & -2 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix}.$$

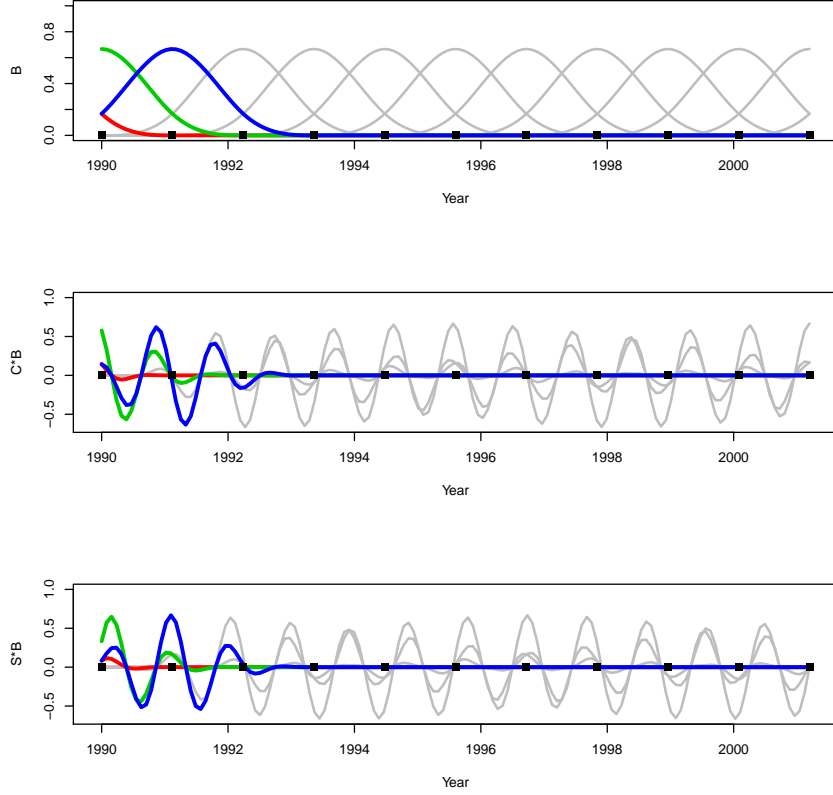


Figure 2.1: Component of regression matrix  $B$  in (2.2), with  $ndx = 10$ . The first three columns of the bases are shown to illustrate the scaling of the basis  $B$  by  $C$  and  $S$ .

Marx et al. (2010) extend the modulation model to a bilinear model where the shape of each period is the same, but the size varies smoothly. In this paper we consider model in (2.1), which also allows the amplitudes to vary smoothly. Assuming normality for the log  $SO_2$  concentration levels, we minimize the penalized log-likelihood

$$S_p = \|\mathbf{y} - \mathbf{B}\boldsymbol{\theta}\|^2 + \mathbf{P}, \quad (2.4)$$

with explicit solution for given  $\lambda$ :

$$\hat{\boldsymbol{\theta}} = (\mathbf{B}'\mathbf{B} + \mathbf{P})^{-1}\mathbf{B}'\mathbf{y}. \quad (2.5)$$

The optimization of  $\lambda$  and  $\check{\lambda}$  are usually done by cross-validation or Akaike or Bayesian information criteria. An important expression to calculate is the so-called hat-matrix  $\mathbf{H}$ , such that

$$\hat{\mathbf{y}} = \mathbf{B}\hat{\boldsymbol{\theta}} = \mathbf{H}\mathbf{y},$$

with

$$\mathbf{H} = \mathbf{B}(\mathbf{B}'\mathbf{B} + \mathbf{P})^{-1}\mathbf{B}'\mathbf{y}. \quad (2.6)$$

The effective dimension (ED) of the fitted model (Hastie and Tibshirani, 1990) is approximately the  $trace(\mathbf{H})$ , and can be efficiently computed as

$$ED = trace\{(\mathbf{B}'\mathbf{B} + \mathbf{P})^{-1}\mathbf{B}'\mathbf{B}\}. \quad (2.7)$$

In model (2.2), the ED of each term in the model is the trace of the portion of diagonal terms of  $\mathbf{H}$  corresponding to each term. Standard error bands for given  $\lambda$  can also be constructed easily:

$$\text{Var}(\hat{\theta}) = \sigma^2(\mathbf{B}'\mathbf{B} + \mathbf{P})^{-1}\mathbf{B}'\mathbf{B}(\mathbf{B}'\mathbf{B} + \mathbf{P})^{-1}.$$

Thus for each component of the model, we can obtain the associated covariance matrix as:

$$\mathbf{C}_k = \sigma^2 \mathbf{B}_k \{ (\mathbf{B}'\mathbf{B} + \mathbf{P})^{-1} \mathbf{B}'\mathbf{B} (\mathbf{B}'\mathbf{B} + \mathbf{P})^{-1} \}_k \mathbf{B}_k,$$

where  $\{\cdot\}_k$  denotes the diagonal block associated with the  $k$ th component. Hence, for  $J = 1$ , we have  $k = 1, 2, 3$  and  $\mathbf{B}_1 = \mathbf{B}$ ,  $\mathbf{B}_2 = \mathbf{C}\mathbf{B}$  and  $\mathbf{B}_3 = \mathbf{S}\mathbf{B}$ . The square root of the diagonal elements of  $\mathbf{C}_k$  are used for the error bands for each component.

## 2.1 Smooth modulation mixed model formulation

The representation of a penalized spline model as a mixed model has become very popular in recent years (see Durbán and Currie (2003) or Wood (2006) for a detailed review). The aim is to find a new basis that allows the representation of model (2.2) with its associated penalty as a mixed model:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\alpha} + \boldsymbol{\epsilon}, \quad \boldsymbol{\epsilon} \sim \mathcal{N}(0, \sigma^2 \mathbf{I}) \quad \boldsymbol{\alpha} \sim \mathcal{N}(0, \mathbf{G}), \quad (2.8)$$

where  $\mathbf{G}$  is the covariance matrix of the random effects that depends on the variance components. Now, we have that  $\mathbf{X}\boldsymbol{\beta}$  are the *fixed effects* term, and  $\mathbf{Z}\boldsymbol{\alpha}$  is the (smooth) *random component*. The smoothing parameters  $\lambda$  and  $\tilde{\lambda}$  becomes the ratio between the error term and the random effect variances,  $\tau^2$  and  $\tilde{\tau}^2$ , for the trend and modulation terms respectively, and they can be estimated by residual maximum likelihood (REML).

There are different alternatives for the reparameterization of the original  $P$ -spline model into a mixed model depending on the bases and the penalty used. The idea is to find a transformation  $\boldsymbol{\Omega}$  such that:

$$\mathbf{B}\boldsymbol{\Omega} = [\mathbf{X} : \mathbf{Z}],$$

where  $\boldsymbol{\Omega}$  is an orthogonal matrix. We split the matrix  $\boldsymbol{\Omega}$  into two submatrices (for the fixed and the random components respectively), i.e.  $\boldsymbol{\Omega} = [\boldsymbol{\Omega}_f : \boldsymbol{\Omega}_r]$ , and such that  $\mathbf{X} = \mathbf{B}\boldsymbol{\Omega}_f$  and  $\mathbf{Z} = \mathbf{B}\boldsymbol{\Omega}_r$ . Since the fixed effects are unpenalized, the matrix  $\mathbf{X}$ , may be replaced by any sub-matrix such that: (i) the composed matrix  $[\mathbf{X} : \mathbf{Z}]$  has full rank (this also implies that both  $\mathbf{X}$  and  $\mathbf{Z}$  have full column rank) and (ii)  $\mathbf{X}$  and  $\mathbf{Z}$  are orthogonal, i.e.  $\mathbf{X}'\mathbf{Z} = 0$ . For the sub-matrix  $\boldsymbol{\Omega}_r$ , there are different alternatives, we follow the approach by Currie et al. (2006) and Lee (2010), and use the singular value decomposition (SVD) on the penalty matrix  $D_q' D_q$ , i.e.

$$D_q' D_q = [U_f : U_r] \begin{bmatrix} \mathbf{O}_q & \\ & \Sigma \end{bmatrix} \begin{bmatrix} U_f' \\ U_r' \end{bmatrix},$$

where  $\mathbf{O}_q$  is square matrix of zeroes of order  $q$  and  $\Sigma$  are the  $(c - q)$  positive eigenvalues.  $U_f$  contains the null part (of dimension  $c \times q$ ), and  $U_r$  contains the span or the non-null part of the decomposition (of dimension  $c \times (c - q)$ ).

For simplicity, let us first consider the trend component, i.e.  $B$ , and coefficient  $\theta$ . The fixed part for the trend can be taken as the design matrix of a polynomial of order  $q - 1$ , i.e.

$$X = [\mathbf{1}_n : x_i : x_i^2 : \dots : x_i^{q-1}], \quad (2.9)$$

where  $\mathbf{1}_n$  is a column vector of ones. For the random part, the random effects matrix  $Z$  is defined as:

$$Z = B\Omega_r \quad \text{with } \Omega_r = U_r\Sigma^{-1/2}, \text{ of dimension } c \times (c - q). \quad (2.10)$$

With this reparameterization, the penalty  $\theta D'_q D_q \theta$  becomes  $\alpha' \alpha$ , and the covariance matrix for the random effects becomes a multiple of an identity matrix, i.e.  $G = \tau^2 \mathbf{I}_{c-q}$ . Given that  $\Omega$  is orthogonal, it is straightforward to obtain the relationship between the inverse of the covariance matrix  $G$  of the random effects and the penalty  $D'D$ , i.e.:

$$G^{-1} = \Omega'_r D' D \Omega_r \iff D' D = \Omega_r G^{-1} \Omega'_r. \quad (2.11)$$

The order of the penalty  $q$  denotes the  $q - 1$  grade polynomial of the fixed part when the smoothing parameters are very large (the *null* model), and hence, the random part can be considered as smooth deviates from the null model. In model (2.2), we consider a first order penalty for the modulation terms, then, the *null* terms for the modulation are  $\cos(\omega x_i)$  and  $\sin(\omega x_i)$  respectively. Then for  $J = 1$ , we have a fixed effect matrix for the smooth modulation model in (2.2) given by:

$$\mathbf{X} = [\mathbf{1}_n | x_i | \dots | x_i^{q-1} | \cos(\omega x_i) | \sin(\omega x_i)], \quad (2.12)$$

which is a design matrix of a harmonic regression model. Now, we define the random part for the modulation components. As in the case the trend component, we use the SVD of the penalty with first order differences, i.e.  $\check{D}'\check{D}$ . The complete random effects matrix for the model in (2.2) is

$$\mathbf{Z} = [Z | C\check{Z} | S\check{Z}], \quad (2.13)$$

where  $\check{Z} = B\check{\Omega}$ . Now  $\check{\Omega} = \check{U}_r\check{\Sigma}^{-1/2}$ , of size  $c \times (c - 1)$  ( $\check{U}$  and  $\check{\Sigma}$  are obtained from the svd of  $\check{D}'\check{D}$ ). Finally, the covariance matrix for the modulation components is  $\check{G} = \check{\tau}^2 \mathbf{I}_{2(c-1)}$ , therefore,  $\mathbf{G} = \text{blockdiag}(G, \check{G})$ . As shown in Section 2, standard errors and confidence bands can be easily obtained by approximating the variance of each smooth term  $\hat{f}_k$ . In the mixed model framework Ruppert et al. (2003) suggest an approximation calculated with respect to the conditional distribution to account the randomness in the random effects  $\alpha$ .

### 3 Trend and seasonality analysis of the $\log SO_2$ data

In this Section, we analyze the  $\log SO_2$  levels for the stations in Austria (AT02) and England (GB07) introduced in Section 1. To construct the model we used  $ndx = 20$  inner knots for  $B$ , and a third order penalty  $q = 3$  for the trend. We considered a first Fourier frequency for the modulation component (i.e.  $J = 1$ ). Estimated smoothing parameters and  $\hat{\sigma}^2$  are shown in Table 3.1, as well as the effective dimension of each component (trend and seasonality). Note that both quantities include the effective degrees of freedom for the fixed components, i.e. 3 for the polynomial trend, and 2 the (co)-sine. The number of Fourier series can be chosen using an information criteria, here we use AIC



Table 3.1: Estimated  $\hat{\sigma}^2$ ,  $\lambda$ ,  $\check{\lambda}$ , and ED for fitted models for AT02 and GB07 stations.

$\log SO_2$	$\hat{\sigma}^2$	$\lambda$	$\check{\lambda}$	ED (trend)	ED (modulation)	AIC	BIC
AT02	0.08	23.63	4.22	6.06	10.43	43.80	92.78
GB07	0.11	0.29	4.21	11.54	11.95	60.67	149.07

(Akaike information criteria) and BIC (Bayesian information criteria), they are defined as:

$$IC = RSS + \delta ED,$$

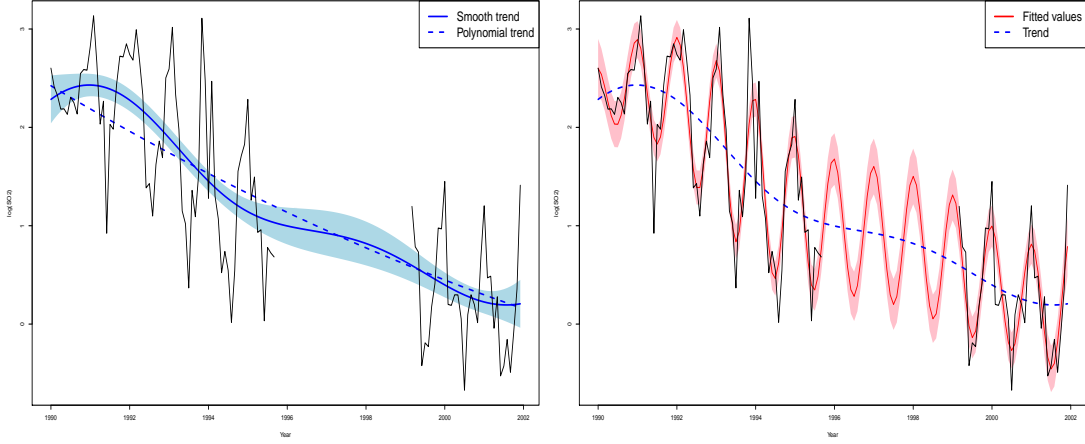
where  $RSS$  is the residual sum of squares, and  $\delta = 2$  for AIC and  $\delta = \log(n)$  for BIC. AIC and BIC values are also useful for model selection for the number of Fourier series. We also fitted the models with  $J = 2$  and obtained greater AIC/BIC values of 47.02 and 100.68 for AT02 and 67.57, and 149.07 for GB07. This results show that for these data a first Fourier frequency might be enough for the seasonal effects.

Figure 3.1 show the trend and seasonality decomposition of the selected data. The smooth fitted trends for AT02 and GB07 stations respectively are represented in Figure 3.1a and Figure 3.1c. The dashed lines represents the polynomial trends of grade  $q - 1$  (i.e.  $X\hat{\beta}$ ) and the solid line is the smooth trend obtained as the sum of the polynomial trend and the smooth deviations from the polynomial (i.e.  $X\hat{\beta} + Z\hat{\alpha}$ ). The figures show that the trend for AT02 is smoother than the GB07 station trend. For AT02 the large gap between October 1995 and March 1999 is interpolated by the model with wider error bands around that period. A similar result can be noticed for GB07 in the smaller gap between August 1993 and October 1994. Figure 3.1b and Figure 3.1d show the fitted values of the smooth modulation models with the corresponding confidence errors bands. The modulation component is shown in top panels of Figure 3.2, the dashed lines correspond to the *fixed* harmonic terms, and the full line is the sum of the fixed part the *random* modulation effect. Bottom panels of Figure 3.2 show only the random effects modulation component. These figures illustrate how random modulation component account for the deviates from the harmonic fixed effect to accommodate the smooth modulation component. Notice that, for the periods with gaps, the magnitude of this random component is smaller than in the other periods of the time series, and these periods are close to the fixed harmonic fit.

## 4 Forecasting with $P$ -splines

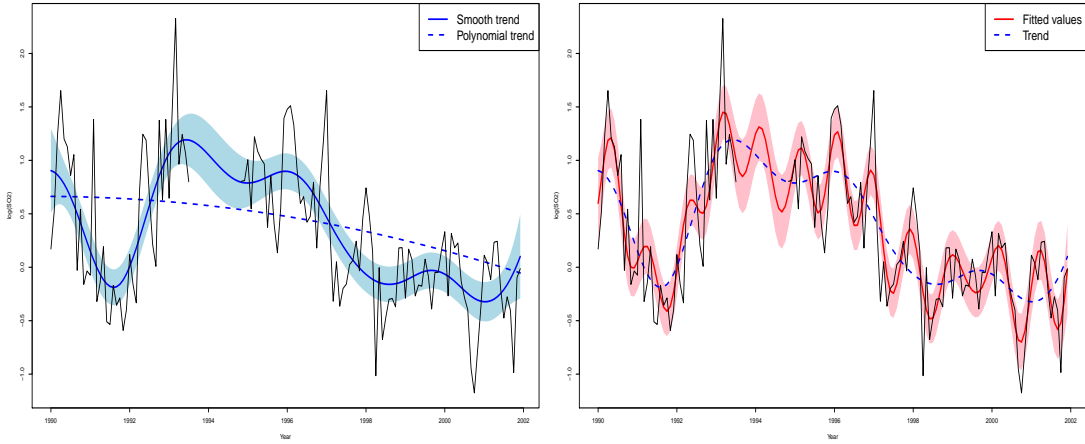
In times series data, it is important to extrapolate or forecast future observations. Currie et al. (2004) proposed a method for fitting and forecasting simultaneously with  $P$ -splines when the coefficients are estimated using penalized least squares. They treat the forecasting of future values as a missing value problem and estimate the fitted and forecasted values simultaneously.

For simplicity, we first illustrate the forecast of the trend component, and then, we show that it is straightforward to include the forecast of the modulation component. Suppose, given  $n$  observations of the response variable  $y$ , that we want to predict new  $n_1$  values  $y_1$  at  $x_1$ . The new vector of observations is  $y^* = (y, y_1)'$ . We need a new extended  $B$ -spline basis  $B^*$  constructed from a new set of knots that extends the original



(a) Smooth and polynomial trends for AT02.

(b) Fitted values and fitted trend for AT02.



(c) Smooth and polynomial trends for GB07

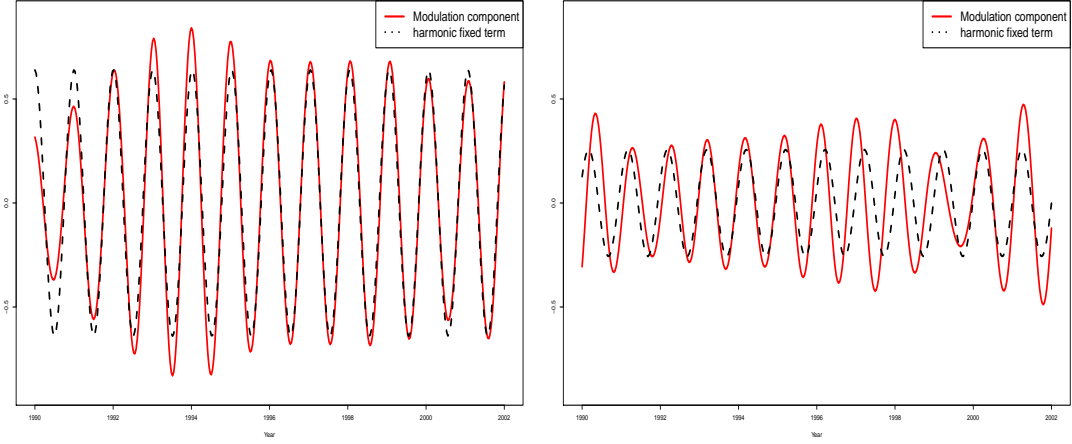
(d) Fitted values and fitted trend for GB07.

Figure 3.1: Trend and Fitted values for AT02 and GB07 stations.

knots used to fit the observed data  $y$ , and also includes a basis for the  $n_1$  observations to forecast. Let us consider this new augmented basis  $B^*$  as

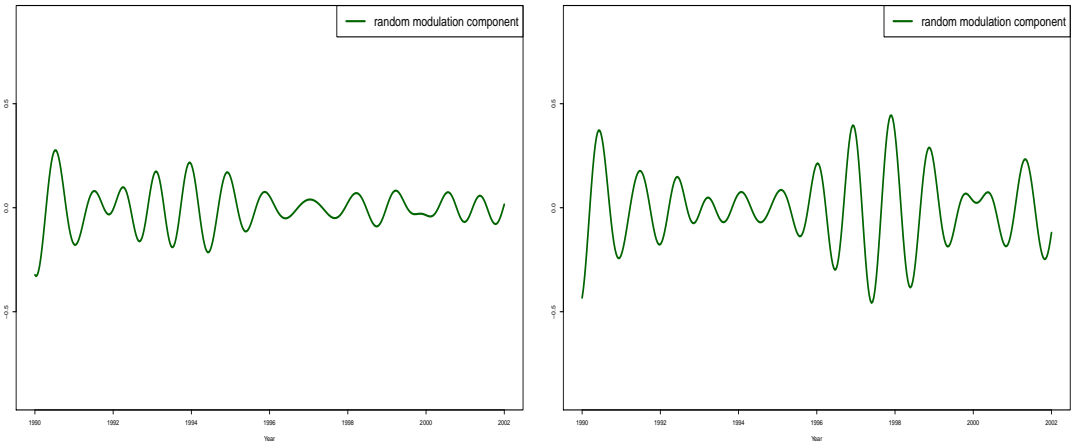
$$B^* = \begin{bmatrix} B & \mathbf{O} \\ B_{(1)} & B_{(2)} \end{bmatrix}, \text{ of size } n^* \times c^*, \quad (4.1)$$

where  $B$  is the  $n \times c$  basis used for fitting the trend component,  $B_{(1)}$  and  $B_{(2)}$  are auxiliary  $B$ -spline basis for prediction up to  $n^* = n + n_1$  values, of sizes  $n_1 \times c$  and  $n_1 \times c_1$  respectively, and  $c^* = c + c_1$ . Figure 4.1 illustrates the component of the new basis (4.1) for forecasting data from January 2002 to December 2004 (the grey shaded area), where  $B_{(1)}$  extends the original basis  $B$  for the new  $n_1$  observations, and  $B_{(2)}$  for the new  $c_1$  coefficients. Associated to the new basis  $B^*$ , we define a new vector of coefficients,  $\theta^*$ , of length  $c^* \times 1$ . Since, we need to increase the number of knots to cover the new range of the covariate values  $x_1$ , we also have to define a new penalty matrix, let's say  $D_q^* D_q^*$  built from a difference matrix  $D^*$  of size  $(c^* - q) \times c^*$ . Following the construction of the



(a) Fixed and smooth modulation for AT02.

(b) Fixed and smooth modulation for GB07.



(c) random modulation for AT02.

(d) random modulation for GB07.

Figure 3.2: Smooth modulation components for AT02 and GB07 stations.

basis in (4.1), we can split the new difference matrix  $D^*$  as:

$$D^* = \begin{bmatrix} D & \mathbf{O} \\ D_{(1)} & D_{(2)} \end{bmatrix}, \quad (4.2)$$

where  $D$  is the difference matrix of size  $(c-q) \times c$ , the sub-matrix  $D_{(1)}$  is of size  $(c_1-q) \times c$ , and  $D_{(2)}$  is a square matrix of size  $(c_1-q) \times (c_1-q)$ .

Currie et al. (2004) showed that defining a diagonal weight matrix  $W$  of size  $n^*$  with diagonal elements equal to 0 if the data is missing/forecasted and 1 if the data is observed. Then, a convenient form to estimate and forecast simultaneously is to use the penalized least square solution given by:

$$\hat{\theta}^* = (B^{*T}WB^* + \lambda D_q^{*T}D_q^*)^{-1} B^*W y^*, \quad (4.3)$$

where arbitrary future values  $y_1$  are chosen for  $y^*$ . This approach emphasizes the role of the penalty as it determines the form of the forecast. Indeed, the choice of  $q$  has no

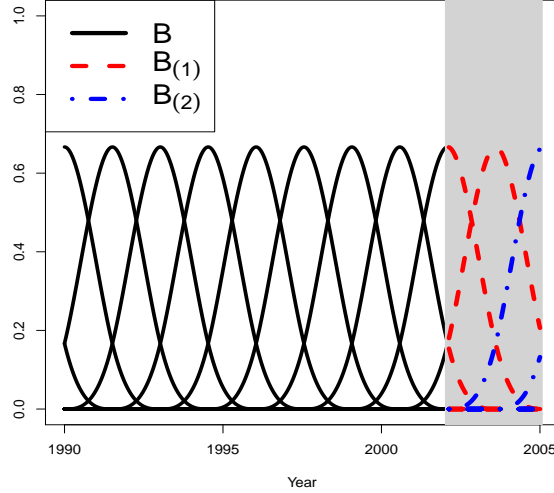


Figure 4.1: Construction of the new extended  $B$ -spline basis for forecasting new values.

discernible effect on the observed values but a dramatic effect on the forecasted values. For  $q = 1, 2, 3$  the extrapolation is constant, linear and quadratic respectively. We discuss the performance of the forecast with different penalty orders in Section 5.

Currie et al. (2004) addressed some interesting invariance properties of the forecasting method. Using the partition of the new basis in (4.1) and the partition of the new difference matrix  $D^*$  in (4.2) it can be proven that: (i) the first  $1, \dots, c_0$  coefficients of  $\hat{\theta}^*$ , are exactly those obtained from fit of  $y_0$ , i.e.:  $\hat{\theta}_{1, \dots, c}^* = \hat{\theta}_0$ ; (ii) the new predicted values are  $\hat{y}_1 = B_{(2)}\hat{\theta}_1$ , where  $\hat{\theta}_1$  are the last  $c_1$  coefficients of  $\hat{\theta}^*$ , and (iii) the coefficients for the predicted  $n^*$  values are:  $\hat{\theta}^* = (\hat{\theta}, -D_{(2)}^{-1}D_{(1)}\hat{\theta})'$ . Hence, the estimation procedure with missing values in (4.3) can be divided in a two-stage procedure: first fit the actual data, and then forecast new values.

## 4.1 Forecasting with smooth mixed models

Now, we extend the forecasting method with  $P$ -splines to the smooth modulation model in (2.2) using the mixed model formulation in Section 2.1. In order to obtain the mixed model reparameterization, we need to reparameterize the extended  $B$ -spline basis in (4.1) with a transformation matrix  $\Omega^*$  such that:

$$B^*\Omega^* = [X^* : Z^*]. \quad (4.4)$$

We define this new matrix  $\Omega^* = (\Omega_f^* : \Omega_r^*)$ . We propose two alternative methods to perform the forecasts. As we are in the mixed model framework, we start by using the standard method for predicting future observations in mixed models, as it is shown in Gilmour et al. (2004). However, the extension of this approach to a more general setting (e.g. multidimensional case) can be complicated and computationally demanding. Therefore, we propose a second approach based on the reparameterization of the penalized spline model proposed in Currie et al. (2004) with missing values. Both methods differ on which transformation matrix  $\Omega^*$  is chosen, and implicitly, in the covariance

structure between the random effects of the observed and predicted values. We show the relationship between both methods.

#### 4.1.1 Method 1: smoothing mixed models forecasting as BLUP's

Prediction in linear mixed models has been a topic of discussion by some authors (see Gilmour et al. (2004) and Welham et al. (2004)). These authors define the prediction to be a linear function of the best linear unbiased predictor (BLUP) of random effects with the best linear unbiased estimate (BLUE) of the fixed effects in the model. To illustrate this approach, we follow Gilmour et al. (2004) and consider the augmented model:

$$\begin{pmatrix} y \\ y_1 \end{pmatrix} = \begin{pmatrix} X \\ X_1 \end{pmatrix} \beta + \begin{pmatrix} Z & \mathbf{O} \\ Z_{(1)} & Z_{(2)} \end{pmatrix} \begin{pmatrix} \alpha \\ \alpha_1 \end{pmatrix} + \begin{pmatrix} \epsilon \\ \epsilon_1 \end{pmatrix}. \quad (4.5)$$

The augmented random effects  $\alpha^* = (\alpha, \alpha_1)'$  have covariance matrix

$$\text{Var}[\alpha^*] = G^* = \begin{pmatrix} G & G_{01} \\ G_{10} & G_{11} \end{pmatrix}. \quad (4.6)$$

Gilmour et al. (2004) showed that the new predicted values are

$$\hat{y}_1 = X_1 \hat{\beta} + Z_1 \hat{\alpha}, \quad (4.7)$$

with

$$Z_1 = (Z_{(1)} + Z_{(2)} G_{10} G^{-1}). \quad (4.8)$$

The estimates of  $\hat{\alpha}_1$  can be obtained by simply augmenting the original set of random effects as:

$$\hat{\alpha}_1 = G_{10} G^{-1} \hat{\alpha}. \quad (4.9)$$

Following this approach, for the reparametrization of the  $P$ -spline model into (4.5), we define the sub-matrix corresponding to the random part as the block-diagonal matrix:

$$\Omega_r^* = \begin{pmatrix} \Omega_r & 0 \\ 0 & \Omega_1 \end{pmatrix}, \quad (4.10)$$

where  $\Omega_r$  is the transformation matrix used for the observed data, and  $\Omega_1$  the one for the predicted values. Hence, we have

$$Z_{(1)} = B_{(1)} \Omega_r, \quad \text{and} \quad Z_{(2)} = B_{(2)} \Omega_1. \quad (4.11)$$

Using the result in (2.11), the covariance matrix  $G^*$  in (4.6), is obtained as

$$G^* = (\Omega_r^{*'} D^{*'} D^* \Omega_r^*)^{-1}, \quad (4.12)$$

where  $D^*$  is the extended difference matrix in (4.2). There are many ways in which  $\Omega_1$  may be chosen, for simplicity (see Appendix A.1) we chose  $\Omega_1 = D_{(2)}^{-1}$ , then, it can be shown that (4.12) becomes:

$$G^* = \begin{pmatrix} G & G_{01} \\ G_{10} & G_{11} \end{pmatrix} = \hat{\tau}^2 \begin{pmatrix} I & -\Omega_r' D_{(1)}' \\ -D_{(1)} \Omega_r & I + D_{(1)} \Omega_r \Omega_r' D_{(1)}' \end{pmatrix}.$$

Then, the random effects matrix  $Z_1$  in (4.8) is

$$Z_1 = Z_{(1)} - Z_{(2)}D_{(1)}\Omega_r, \quad (4.13)$$

and the random effects for the predicted values in (4.9) becomes:

$$\hat{\alpha}_1 = -D_{(1)}\Omega_r\hat{\alpha}. \quad (4.14)$$

Observe that this first method can be viewed as a two-stages approach where first the actual data are fitted, and then with the fitted parameters and variance components, Equation (4.5) is used to obtain the predictions.

#### 4.1.2 Method 2: smoothing mixed models forecasting as a missing values problem

The second alternative method for forecasting is based on the idea in Currie et al. (2004) of considering the forecasting as a missing values problem. We use the SVD of the penalty  $D^*D^*$ , and define  $\Omega_r^*$  directly as

$$\Omega_r^* = U_r^*\Sigma_r^{*-1/2}.$$

Then, the model becomes

$$y^* = X^*\beta + Z^*\alpha^* + \epsilon^*,$$

where

$$X^* = \begin{pmatrix} X \\ X_1 \end{pmatrix}, \text{ and } Z^* = B^*\Omega_r^*. \quad (4.15)$$

The estimation is done using the mixed model system of equations:

$$\begin{pmatrix} \hat{\beta} \\ \hat{\alpha}^* \end{pmatrix} = \begin{pmatrix} X^{*'}WX^* & X^{*'}WZ^* \\ Z^{*'}WX^* & Z^{*'}WZ^* + G^{*-1} \end{pmatrix}^{-1} \begin{pmatrix} X^{*'}W \\ Z^{*'}W \end{pmatrix} y^*,$$

where  $W$  is the diagonal matrix of length  $n^*$  with 0 entries if the data is missing/forecasted and 1 if the data is observed, and variance components are estimated by REML. Notice that, the main difference with the other method in Section 4.1.1 relies on the definition of  $\Omega_r^*$ , which also defines a different random effects matrix,  $Z^*$  in (4.15) is different from  $Z_1$  in (4.13). With this second method, the covariance matrix of the random effects  $\alpha^*$  is a multiple of an identity matrix, i.e.

$$G^* = \hat{\tau}^2 I_{n^*}, \quad (4.16)$$

and then, the observed and predicted random effects are uncorrelated. An important result of both methods is that the fitted and forecasted values (i.e.  $\hat{y}^*$ ), and the estimated variance components ( $\hat{\tau}^2$ ) are exactly the same, and hence, the invariance property of the forecasts addressed in Currie et al. (2004) is maintained.

## 4.2 Forecasting with smooth modulation mixed models

Finally, we use the methods introduced in Section 4.2 to forecast the modulation component. We have to extend the  $B$ -spline basis for the modulation components. The full smooth modulation regression model matrix has the form:

$$B^* = [B^*|C^*B^*|S^*B^*], \quad (4.17)$$

where the (co)-sine modulation components are added block-wise. Basically, we only have to define the varying-coefficient matrices for the range of the  $x^*$  time points, i.e.  $C^* = \text{diag}(\cos(\omega x^*))$  and  $S^* = \text{diag}(\sin(\omega x^*))$ , for the additive modulation blocks  $C^* B^*$  and  $S^* B^*$ , and then use the methods proposed in previous sections to forecast the modulation components. The vector of coefficient is  $\theta^* = (\theta^*, \theta_c^*, \theta_r^*)'$ , and the penalty matrix for the modulation component is the first order difference matrix  $\check{D}^*$ .

$$\check{D}^* = \begin{pmatrix} \check{D}_1 & 0 \\ \check{D}_{(1)} & \check{D}_{(2)} \end{pmatrix}. \quad (4.18)$$

We can use any of the two methods proposed. For method 1 in Section 4.1.1, we define

$$\check{\Omega}_r^* = \begin{pmatrix} \check{\Omega}_r & 0 \\ 0 & \check{\Omega}_1 \end{pmatrix}, \quad (4.19)$$

with  $\check{\Omega}_1 = \check{D}_{(2)}^{-1}$ . The fixed term has the columns corresponding to the fourier series, i.e.  $\cos(\omega x^*)$ , and  $\sin(\omega x^*)$ . We obtain the modulation mixed models prediction matrix  $\check{Z}_1$  with method 1 by using Equations (4.11), (4.13) and (4.14), with  $\check{D}^*$  and  $\check{\Omega}_r$ . Hence, the random effect matrix for the (co)-sine prediction values are

$$C^* \check{Z}_1 \quad \text{and} \quad S^* \check{Z}_1.$$

For method 2 in Section 4.1.2, we compute the SVD on  $\check{D}^{*'} \check{D}^*$  and define:

$$\check{\Omega}_r^* = \check{U}_r^* \check{\Sigma}_r^{*-1/2}. \quad (4.20)$$

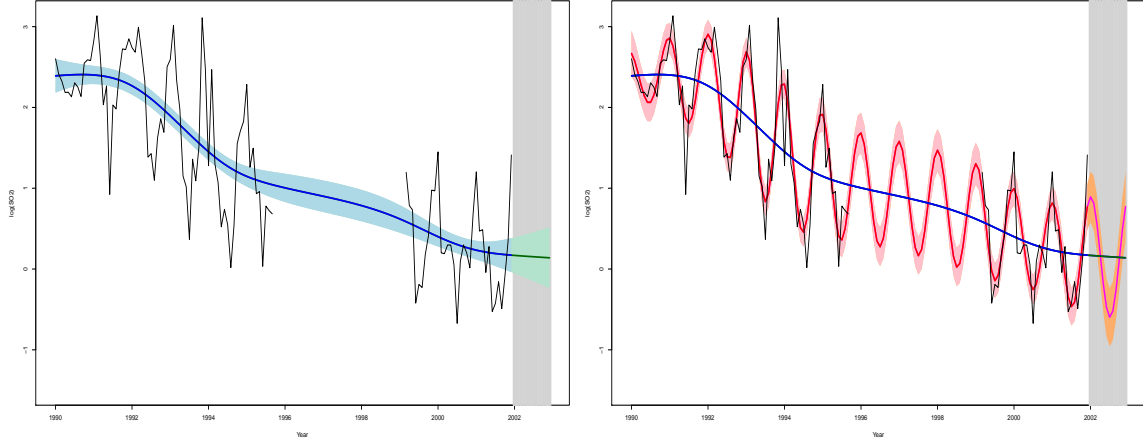
Then modulation random effects prediction matrix for the (co)-sine terms are

$$C^* \check{Z}^* \quad \text{and} \quad S^* \check{Z}^*.$$

To illustrate the forecast, we predict the log  $SO_2$  levels for year 2002. Figure 4.2 and Figure 4.3 show the forecasted trend and final predictions (including the seasonal projections) for AT02 and GB07 stations with second and third penalty orders for the trend component. Notice that, for second order penalty ( $q = 2$ ) the trend forecast is linear, and for third order ( $q = 3$ ) the trend forecast is quadratic. For AT02 the choice of the penalty order for the new observations has a small effect in the forecasts, but for GB07 this choice has a dramatic impact.

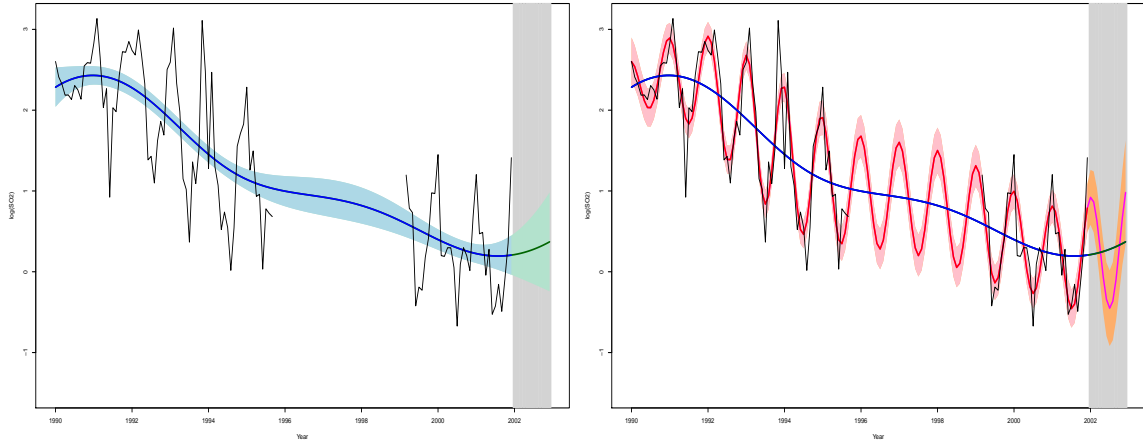
## 5 Forecast analysis of the log $SO_2$ data

To evaluate the performance of the forecasting method with different penalty orders, we divided the available data set into two subsamples. The first subsample of the data (from January 1990 to December 2000) is the *estimation subsample* and is used to fit the model, and the second subsample (from January to December 2001) is the *forecasting subsample* and is used to evaluate the point forecasts. There are different ways we can analyse the performance of a forecasting method. The standard procedure consists of consider a common origin  $T$  (for example January 2001) and compute forecast for a sequence of  $h$  horizons, based on data from the estimation subsample.



(a) Trend forecast for station AT02 with  $q = 2$ .

(b) Forecast for station AT02 with  $q = 2$ .



(c) Trend forecast for station AT02 with  $q = 3$ .

(d) Forecast for station AT02 with  $q = 3$ .

Figure 4.2: Trend and Fitted values for AT02 and GB07 stations aaa.

We consider three accuracy measures: the mean absolute deviation (MAD), the square root mean square error (RMSE) and the Mean absolute percentage error (MAPE). For  $h$ -step ahead forecasts, these measures are defined as:

$$\text{MAD}(h) = \frac{1}{m} \sum_{j=0}^{m-1} |\mathbf{y}_{T+h+j} - \hat{\mathbf{y}}_{T+j}(h)|, \quad (5.1)$$

$$\text{RMSE}(h) = \sqrt{\frac{1}{m} \sum_{j=0}^{m-1} (\mathbf{y}_{T+h+j} - \hat{\mathbf{y}}_{T+j}(h))^2}, \quad (5.2)$$

$$\text{MAPE}(h) = \frac{1}{m} \sum_{j=0}^{m-1} \left| \frac{\mathbf{y}_{T+h+j} - \hat{\mathbf{y}}_{T+j}(h)}{\mathbf{y}_{T+h+j}} \right|, \quad (5.3)$$

where  $m$  is the number of  $h$ -step ahead forecasts available in the forecasting subsample. In practice, one often chooses one measure, and the model with the smallest magnitude



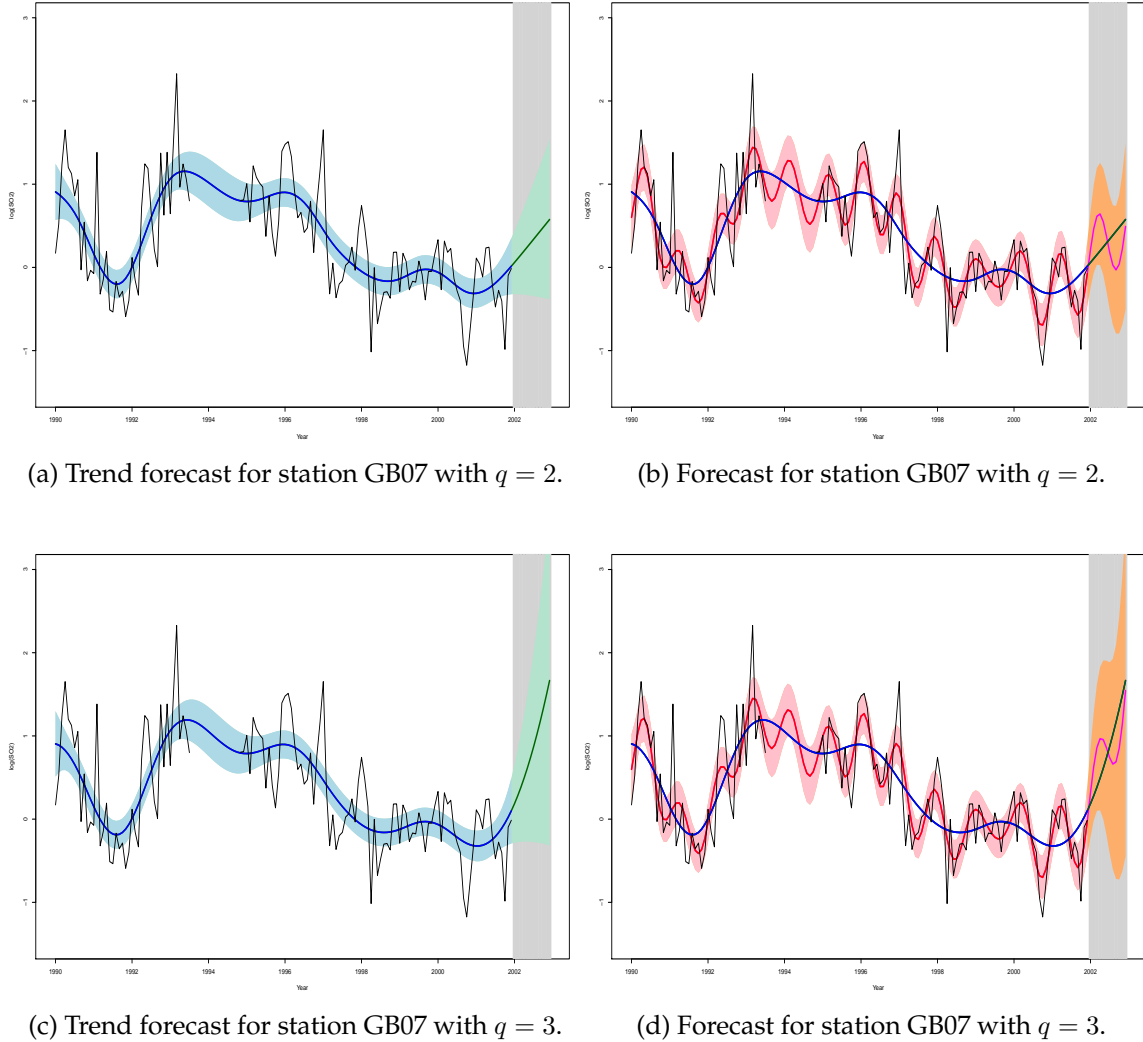
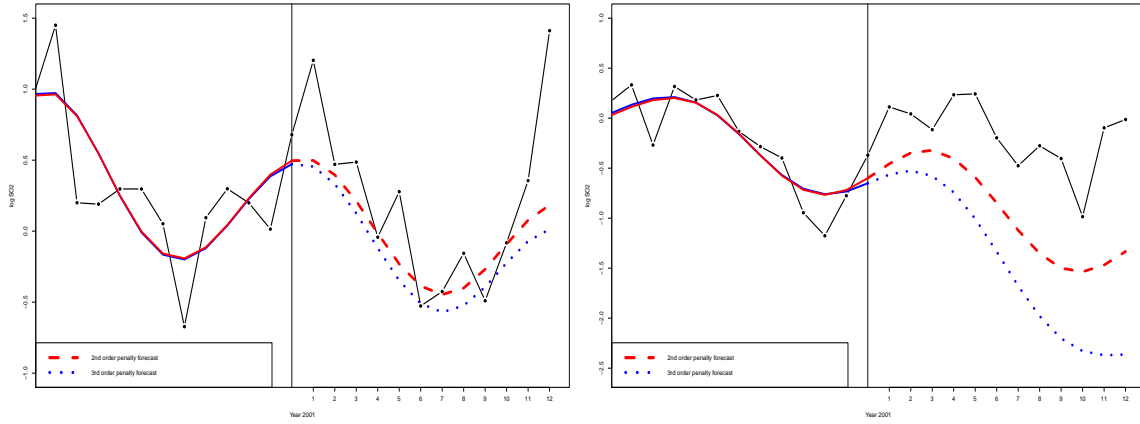


Figure 4.3: Trend and Fitted values for AT02 and GB07 stations.

is proposed as the best  $h$ -step ahead forecasting model (it is possible that different  $h$  may result in selecting different models (Hyndman and Koehler, 2006)). Our aim is to compare the second and third order penalties for different prediction horizons ( $h = 1, 2, \dots, 6$  and  $h = 12$ ). Table 5.1 shows the accuracy measures for AT02 and GB07 stations. We found that for both cases the forecasts with second order penalties has a better performance. Figure 5.2 shows the comparison of second and third order penalty forecasts for the selected stations and  $h = 12$ . Figure 5.1b shows the forecasts for station GB07, the results for this particular station are very poor due to the strong non-linear behaviour of the trend (as shown in Figure 4.3). Notice that, the forecasts also depends on the origin  $T$  chosen for the forecast. To check the forecasting in different situations, we also evaluate the forecast performance varying the forecast origin  $T$  and maintaining a consistent forecast horizon to  $h = 6$ . This procedure draws attention to the forecast errors at a particular horizon  $h$  but also show how the forecast error changes as the horizon lengthens (see Makridakis et al., 1998, for details). The accuracy measures for forecast origin at June 2001 are shown in Table 5.2. Observe that the results are more accurate for  $h = 6$ .



(a)  $h = 12$  forecast for AT02.

(b)  $h = 12$  forecasts for GB07

Figure 5.1: Comparison of second and third order penalty forecast for  $h = 6$  and  $h = 12$  horizons.

Table 5.1: Accuracy measures for forecasting  $h$  periods ahead with  $q$  penalty orders.

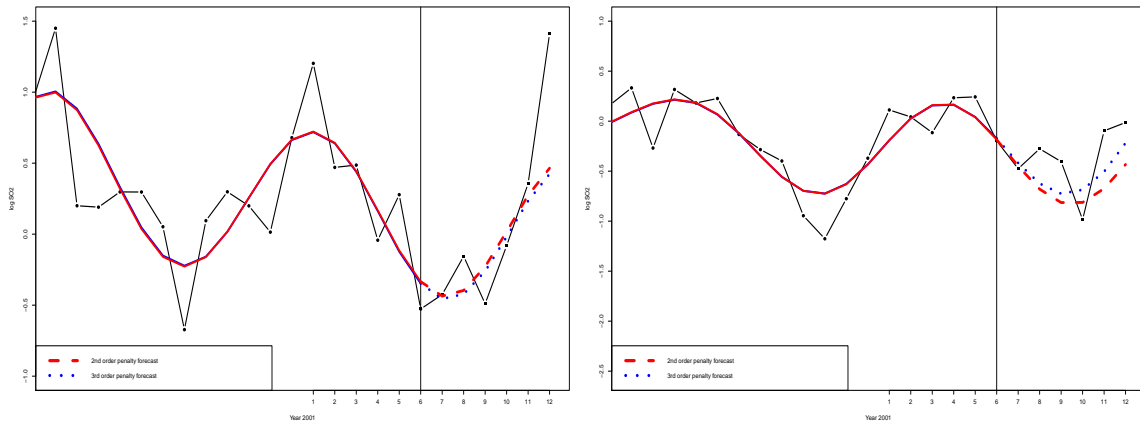
AT02		Forecast horizon ( $h$ )						
	Penalty order ( $q$ )	1	2	3	4	5	6	12
MAD	2	0.705	0.389	0.351	0.270	0.318	0.288	0.311
	3	0.748	0.443	0.416	0.331	0.390	0.328	0.379
RMSE	2	0.705	0.501	0.439	0.380	0.409	0.378	0.461
	3	0.748	0.538	0.486	0.423	0.470	0.429	0.535
MAPE	2	0.586	0.371	0.436	0.489	0.758	0.676	0.658
	3	0.622	0.458	0.553	0.868	1.145	0.959	1.048

GB07		Forecast horizon ( $h$ )						
	Penalty order ( $q$ )	1	2	3	4	5	6	12
MAD	2	0.569	0.479	0.388	0.451	0.527	0.548	0.779
	3	0.676	0.622	0.569	0.671	0.786	0.845	1.312
RMSE	2	0.569	0.488	0.416	0.481	0.569	0.584	0.854
	3	0.676	0.624	0.576	0.697	0.837	0.894	1.440
MAPE	2	5.079	7.121	5.346	4.692	4.440	4.252	12.106
	3	6.038	9.687	7.802	6.894	6.545	6.419	20.778

Station	Penalty order ( $q$ )	MAD	RMSE	MAPE
AT02	2	0.236	0.249	0.941
	3	0.313	0.437	0.739
GB07	2	0.297	0.337	1.544
	3	0.279	0.298	3.798

Table 5.2: Accuracy measures for AT06 and GB07 considering the forecast origin at 06/2001 for  $h = 6$ .



(a)  $h = 6$  forecasts for AT02.

(b)  $h = 6$  forecasts for GB07.

Figure 5.2: Comparison of second and third order penalty forecast for  $h = 6$  and forecast origin at 06/2001.

## 6 Discussion

Smoothing techniques have become a very popular tool for estimation of trends. However, for time series data with seasonality, simultaneous smoothing and forecasting is still an open subject of research. In this paper we have proposed a mixed model formulation of the seasonal modulation model using varying-coefficient terms. This approach allows us to decompose the fitted curve into trend and seasonality.

Welham et al. (2006) proposed a similar mixed model approach based on the so-called  $L$ -splines (Heckman and Ramsay, 2000; Gu, 2002) with smoothing splines and truncated polynomials as basis functions.  $L$ -splines are a large family of smoothing splines defined in terms of a linear differential operator to construct the appropriate kernel function. This differential operator gives the set of core functions that we obtain in the fixed effects matrix  $X$  in (2.12), but we use the  $C$  and  $S$  diagonal matrices for the varying coefficients terms that rescales the original  $B$ -spline basis  $B$ . We found our approach easier to implement, and also easy to extend for forecasting future observations.

We showed some examples from environmental air monitoring data, where the interest relies on both interpolation and extrapolation of monthly time series, and we extend this formulation to forecast future observations. We showed how using the mixed model reparameterization we can use two methods for forecasting seasonal time series, both methods are equivalent giving the same results for the forecasts. Further, the estimates of the coefficients and the fitted values within the range of the training set, the trace of the fitted model and the optimal value of the smoothing parameter do not change whether we include or exclude the missing values to be forecast. However, the method presented in Section 4.1.2 is computationally more efficient; also, when we treat the forecasting of future values as a missing value problem we can estimate the fitted and forecast values simultaneously, and forecasting becomes a natural consequence of the smoothing process.

We have seen that the order of penalty function, which is less relevant in the smoothing of data, is now critical, because it is the penalty function that determines the form of

the forecast. In both examples a penalty of order 2 was preferred (for any lag forecasted) by all the accuracy measures, however, this might not be the case in some situations, and should always be checked. The issue of forecast origin was addressed, and more rigorous a more rigorous evaluation of the forecasts can also be applied as for example parametric bootstrap (see Tsay, 2001, for details). The model fit well for both stations, but the forecasting of future observations performed better for the station in Austria, due to the strong non-stationarity of the data in England.

Other approaches can be combined or more fully explored, including: periodic smoothing with circular B-spline bases or special periodic penalties as shown in Eilers and Marx (2010), forecasting with the bilinear model proposed in Marx et al. (2010), and comparing with other forecasting methods (mostly from the econometrics field). The mixed model formulation also allows for a straightforward hierarchical Bayesian approach using MCMC. In this paper, we assumed uncorrelated errors, we checked there were no autocorrelation structure in the residuals. However, the approach proposed also provides a unified framework for smoothing, forecasting and also the incorporation autocorrelated errors as in Durbán and Currie (2003).

## Acknowledgements

This research was funded by the Spanish Ministry of Science and Innovation (projects MTM 2008-02901, and MTM2011-28285-C02-02). The research of Dae-Jin Lee was funded by an NIH grant for the Superfund Metal Mixtures, Biomarkers and Neurodevelopment project 1PA2ES016454-01A2. The authors are also grateful to Iain Currie for discussion on the projections with  $P$ -splines, and Prof. Adrian W. Bowman for providing the  $SO_2$  data.

## References

- Cleveland, R. B., Cleveland, W. S., McRae, J. E., and Terpenning, I. (1990). Stl: A seasonal-trend decomposition based on loess. *Journal of Official Statistics*, 6(1):3–73.
- Currie, I. D., Durbán, M., and Eilers, P. H. C. (2004). Smoothing and forecasting mortality rates. *Statistical Modelling*, 4(4):279–298.
- Currie, I. D., Durbán, M., and Eilers, P. H. C. (2006). Generalized linear array models with applications to multidimensional smoothing. *J. R. Statist. Soc. B*, 68:1–22.
- Durbán, M. and Currie, I. D. (2003). A note on  $P$ -spline additive models with correlated errors. *Computational Statistics*, 18:251–262.
- Eilers, P. H. C., Gampe, J., Marx, B. D., and Rau, R. (2008). Modulation models for seasonal time series and incidence tables. *Statist. Med.*, 27:3430–3441.
- Eilers, P. H. C. and Marx, B. D. (1996). Flexible smoothing with  $B$ -splines and penalties. *Stat. Sci.*, 11:89–121.
- Eilers, P. H. C. and Marx, B. D. (2002). Generalized linear additive smooth structures. *Journal of Computational and Graphical Statistics*, 11(4):758–783.

- Eilers, P. H. C. and Marx, B. D. (2010). Splines, knots and penalties. *Wiley Interdisciplinary Reviews: Computational Statistics*, 2(6):637–653.
- Gilmour, A., Cullis, B., Welham, S., Gogel, B., and Thompson, R. (2004). An efficient computing strategy for prediction in mixed linear models. *Computational Statistics & Data Analysis*, 44:571–586.
- Gu, C. (2002). *Smoothing Spline ANOVA Models*. Springer Series in Statistics. Springer.
- Harville, D. A. (1997). *Matrix Algebra from a Statistician's Perspective*. Springer-Verlag.
- Hastie, T. and Tibshirani, R. (1993). Varying-coefficient models. *J. R. Statist. Soc. B*, 55(4):757–796.
- Heckman, N. E. and Ramsay, J. O. (2000). Penalized regression with model-based penalties. *The Canadian Journal of Statistics*, 28(2):241–258.
- Hyndman, R. J. and Koehler, A. B. (2006). Another look at measures of forecast accuracy. *International Journal of Forecasting*, 22(4):679–688.
- Lee, D.-J. (2010). *Smoothing mixed model for spatial and spatio-temporal data*. PhD thesis, Department of Statistics, Universidad Carlos III de Madrid, Spain.
- Makridakis, S., Wheelwright, S., and Hyndman, R. J. (1998). *Forecasting: methods and applications*. John Wiley & Sons: New York, 3rd edition.
- Marx, B. D., Eilers, P. H. C., Gampe, J., and Rau, R. (2010). Bilinear modulation models for seasonal tables of counts. *Statistics and Computing*, 20:191–202.
- Ruppert, D., Wand, M. P., and Carroll, R. J. (2003). *Semiparametric Regression*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, UK. ISBN: 0521785162.
- Tsay, R. S. (2001). *Analysis of Financial Time Series*. Wiley Series in Probability and Statistics.
- Welham, S., Cullis, B., Gogel, B., Gilmour, A., and Thompson, R. (2004). Prediction in mixed linear models. *Australian and New Zealand Journal of Statistics*, 46(3):325–347.
- Welham, S., Cullis, B., Kenward, M. G., and Thompson, R. (2006). The analysis of longitudinal data using mixed models L-splines. *Biometrics*, 62(2):392–401.
- Wood, S. N. (2006). *Generalized Additive Models - An introduction with R*. Texts in Statistical Science. Chapman & Hall.

# A Appendix

## A.1 Estimation of new observations by method 1 in Section 4.1.1

As we showed in Section 2.8, in order to obtain the mixed model reparameterization of the  $P$ -spline model for prediction of new observation in (4.5), we need to find a transformation matrix such that

$$B^*\Omega^* = [X^* : Z^*], \quad (\text{A.1})$$

where  $\Omega^* = (\Omega_f^* : \Omega_r^*)$ . For method 1, we define  $\Omega_r^*$  as the block-diagonal matrix:

$$\Omega_r^* = \begin{pmatrix} \Omega_r & 0 \\ 0 & \Omega_1 \end{pmatrix}, \quad (\text{A.2})$$

where  $\Omega$  is the transformation matrix used for the observed data, and  $\Omega_1$  the one for the predicted values. Hence, we have that (A.1) becomes:

$$B^*\Omega^* = \begin{pmatrix} B & \mathbf{O} \\ B_{(1)} & B_{(2)} \end{pmatrix} \begin{pmatrix} \Omega_r & 0 \\ 0 & \Omega_1 \end{pmatrix} = \begin{pmatrix} X & Z & \mathbf{O} \\ X_1 & Z_{(1)} & Z_{(2)} \end{pmatrix}$$

Now, we need to obtain the covariance matrix  $G^*$  of the augmented random effects  $\alpha^*$ . Given the result shown in (2.11), we have

$$G^* = (\Omega_r^{*'} D^{*'} D^* \Omega_r^*)^{-1} = \begin{pmatrix} G & G_{01} \\ G_{10} & G_{11} \end{pmatrix}, \quad (\text{A.3})$$

with  $D^*$  defined in (4.2).

Let be  $A = D^*\Omega^*$ ,  $A$  is a square matrix, so the inverse of  $A'A$  is

$$(A'A)^{-1} = A^{-1}A'^{-1} = A^{-1}(A^{-1})'.$$

Given the result in (Harville, 1997, page 99) for the inverse of block-partitioned matrices:

$$A^{-1} = \begin{pmatrix} U & 0 \\ V & W \end{pmatrix}^{-1} = \begin{pmatrix} U^{-1} & 0 \\ -W^{-1}VU^{-1} & W^{-1} \end{pmatrix}^{-1}$$

We have that:

$$(D^*\Omega^*)^{-1} = \begin{pmatrix} D\Omega_r & 0 \\ D_{(1)}\Omega_r & D_{(2)}\Omega_1 \end{pmatrix}^{-1} = \begin{pmatrix} (D\Omega_r)^{-1} & 0 \\ -(D_{(2)}\Omega_1)^{-1}D_{(1)}\Omega(D\Omega_r)^{-1} & (D_{(2)}\Omega_1)^{-1} \end{pmatrix}.$$

Then

$$\begin{aligned} G^* &= \begin{pmatrix} (D\Omega_r)^{-1} & 0 \\ -(D_{(2)}\Omega_1)^{-1}D_{(1)}\Omega(D\Omega_r)^{-1} & (D_{(2)}\Omega_1)^{-1} \end{pmatrix} \begin{pmatrix} (D\Omega_r)^{-1} & 0 \\ -(D_{(2)}\Omega_1)^{-1}D_{(1)}\Omega(D\Omega_r)^{-1} & (D_{(2)}\Omega_1)^{-1} \end{pmatrix}' = \\ &= \begin{pmatrix} (D\Omega_r)^{-1} & 0 \\ -(D_{(2)}\Omega_1)^{-1}D_{(1)}\Omega(D\Omega_r)^{-1} & (D_{(2)}\Omega_1)^{-1} \end{pmatrix} \begin{pmatrix} ((D\Omega_r)^{-1})' & (-D_{(2)}\Omega_1)^{-1}D_{(1)}\Omega(D\Omega_r)^{-1})' \\ 0 & ((D_{(2)}\Omega_1)^{-1})' \end{pmatrix}' = \\ &= \begin{pmatrix} G_{00} & G_{01} \\ G_{10} & G_{22} \end{pmatrix} \end{aligned}$$

where

$$\begin{aligned}
G &= (D\Omega_r)^{-1}((D\Omega_r)^{-1})' = (D\Omega_r)^{-1}((D\Omega_r)')^{-1} = I \\
G_{01} &= (D\Omega_r)^{-1} \left( -(D_{(2)}\Omega_1)^{-1} D_{(1)}\Omega_r (D\Omega_r)^{-1} \right)' = \\
&= -(D\Omega_r)^{-1} ((D\Omega_r)^{-1})' (D_{(1)}\Omega_r)' ((D_{(2)}\Omega_1)^{-1})' = \\
&= -(D_{(1)}\Omega_r)' ((D_{(2)}\Omega_1)^{-1})' = -(D_{(1)}\Omega_r)' ((D_{(2)}\Omega_1)')^{-1} \\
G_{10} &= G_{01}' = \\
&= -(D_{(2)}\Omega_1)^{-1} D_{(1)}\Omega_r (D\Omega_r)^{-1} ((D\Omega_r)^{-1})' = \\
&= -(D_{(2)}\Omega_1)^{-1} D_{(1)}\Omega_r (D\Omega_r)^{-1} ((D\Omega_r)')^{-1} = \\
&= -(D_{(2)}\Omega_1)^{-1} D_{(1)}\Omega_r \\
G_{22} &= -(D_{(2)}\Omega_1)^{-1} D_{(1)}\Omega_r (D\Omega_r)^{-1} \left( -(D_{(2)}\Omega_1)^{-1} D_{(1)}\Omega_r (D\Omega_r)^{-1} \right)' + (D_{(2)}\Omega_1)^{-1} ((D_{(2)}\Omega_1)^{-1})' = \\
&= (D_{(2)}\Omega_1)^{-1} D_{(1)}\Omega_r (D\Omega_r)^{-1} ((D\Omega_r)')^{-1} ((D\Omega_r)')^{-1} ((D_{(2)}\Omega_1)^{-1} D_{(1)}\Omega_r)' + (D_{(2)}\Omega_1)^{-1} ((D_{(2)}\Omega_1)^{-1})' = \\
&= \dots = \\
&= (D_{(2)}\Omega_1)^{-1} D_{(1)}\Omega_r (D_{(1)}\Omega_r)' ((D_{(2)}\Omega_1)')^{-1} + (D_{(2)}\Omega_1)^{-1} ((D_{(2)}\Omega_1)^{-1})'.
\end{aligned}$$

If we choose  $\Omega_1 = D_{(2)}^{-1}$ , we obtain a simpler expression for the covariance matrix  $G^*$  as:

$$G^* = \begin{pmatrix} G & G_{01} \\ G_{10} & G_{11} \end{pmatrix} = \hat{\tau}^2 \begin{pmatrix} I & -\Omega_r' D_{(1)}' \\ -D_{(1)}\Omega_r & I + D_{(1)}\Omega_r \Omega_r' D_{(1)}' \end{pmatrix}.$$

Similarly we can obtain its inverse as:

$$G^{*-1} = \begin{pmatrix} G^{00} & G^{01} \\ G^{10} & G^{11} \end{pmatrix} = \frac{1}{\hat{\tau}^2} \begin{pmatrix} I + \Omega_r' D_{(1)}' D_{(1)}\Omega_r & \Omega_r' D_{(1)}' \\ D_{(1)}\Omega_r & I \end{pmatrix}.$$

Gilmour et al. (2004) demonstrate that the random effects vector for the predicted values are:

$$\hat{\alpha}_1 = -(G^{11})^{-1} G^{01} \hat{\alpha} = G_{10} G^{-1} \hat{\alpha}.$$

In our case, with the definitions of  $G^{11} = I$ ,  $G^{01} = D_{(1)}\Omega_r$ , and  $G_{01} = -D_{(1)}\Omega_r$ , we obtain that

$$\hat{\alpha}_1 = D_{(1)}\Omega_r \hat{\alpha}.$$