Flexible Spatio-temporal smoothing with array methods

Dae-Jin Lee
CSIRO, Mathematics, Informatics and Statistics
Private Bag 33, Gate 5
Clayton VIC 3168, AUSTRALIA
E-mail: dae-jin.lee@csiro.au

María Durbán
Universidad Carlos III de Madrid, Department of Statistics
Avd de la Universidad 30, Edif. Juan Benet
Leganés, Madrid 28911, SPAIN
E-mail: mdurban@est-econ.uc3m.es

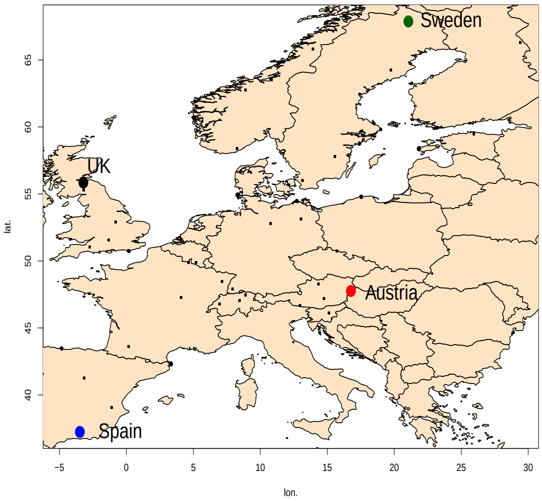
ABSTRACT

In recent years, spatio-temporal modelling has become a challenging area of research in many fields (e.g. epidemiology, environmental studies, and disease mapping). However, most of the models developed are constrained by the large amount of data available. Smoothing methods present very attractive and flexible modelling tools for this type of data set. In the context of environmental studies, where data often present a strong seasonal trend, and the interaction of spatial and temporal processes may be strong, the size of the regression basis needed to capture the temporal trend is large and, as a consequence, the estimation of the spatio-temporal interaction is computationally intensive. We propose the use of Penalized Splines as mixed models for smoothing spatio-temporal data. The array properties of the regression bases allow us to fit Smooth-ANOVA-type models, imposing identifiability constraints over the coefficients. These models are fitted taking advantage of the array structure of the space-time interaction and the use of the GLAM (generalized linear array methods) algorithms. We illustrate the methodology with the analysis of real environmental problems.

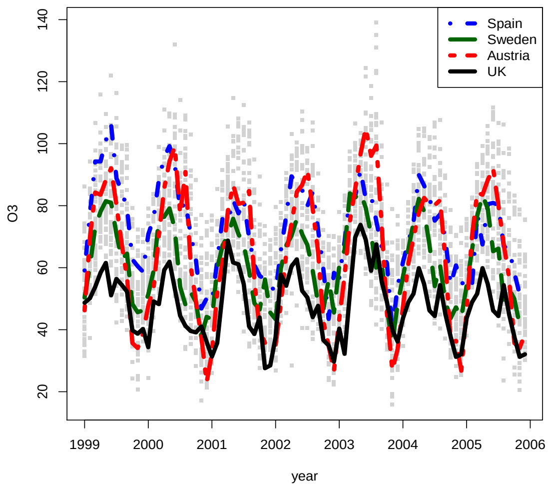
1. Introduction

Spatio-temporal data structure arise in many contexts such as, meteorology, environmental sciences, epidemiology or demography, among others. This wide variety of settings has generated a considerable interest in the development of spatio-temporal models. Recently, Lee and Durbán (2011) proposed the use of multidimensional penalized splines (Eilers and Marx, 1996) for smoothing spatio-temporal data using tensor products of B-splines bases. In this paper, we show *generalized linear array methods*, or GLAM (Currie et al. (2006) and Eilers et al. (2006)) can be used in the spatio-temporal smoothing context. The model is treated in a compact array notation in which the space-time interaction is modelled using a tensor product of a marginal B-spline basis for space ($2d$) and time. The GLAM algorithms take advantages of the structure of the data, avoiding computational issues in storage and allow managing huge amount of data also with high speed and efficient computations in model estimation.

Our methodological development is motivated by the analysis of air pollution levels study in Europe between January 1999 and December 2005. Figure 1 presents the locations of the monitoring



(a) Monitoring stations.



(b) Time series plot in four countries.

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

stations, and the seasonal pattern of ozone levels in four different countries (Spain, Sweden, Austria and UK). The plots show that the stations cover a large area where spatial trends are likely to appear (mostly due to climate conditions), and a clear seasonal pattern is present along the years. Therefore, a smoothing spatio-temporal models seems suitable to estimate simultaneously the spatial and temporal trends.

The paper is organized as follows: in Section 2, we present the interaction model for spatio-temporal smoothing, and show how it can be viewed as GLAM. We present a new model, the Smooth-ANOVA model, where additive effects of space and time are estimated simultaneously with the space-time interaction. In Section 3, we present the estimated models and finally we conclude with some discussion.

2. Spatio-temporal data smoothing with P-splines

Consider data are located in n geographical locations, $s = (\boldsymbol{x}_1, \boldsymbol{x}_2)$, and measured over t time periods $\boldsymbol{x}_t$. Most of the common approaches in spatio-temporal smoothing assume an additive model with two components: a two-dimensional term for the spatial surface and a one-dimensional term for the temporal dimension, of the form:

$$(1) \quad \boldsymbol{y} = f_s(\boldsymbol{x}_1, \boldsymbol{x}_2) + f_t(\boldsymbol{x}_t) + \boldsymbol{\epsilon}, \quad \text{where } \boldsymbol{\epsilon} \sim \mathcal{N}(0, \sigma^2 \boldsymbol{I})$$

Therefore, they impose a separable structure for the spatio-temporal process that, in many cases, will not represent the real structure of the data, as the interaction between space and time is completely ignored.

As an alternative, Lee and Durbán (2011) proposed a class of non-separable models of the form

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

where the smooth function of space and time, $f_{st}(\cdot)$, is modelled using the P-spline methodology as:

(3) $f_{st}(\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{x}_t) = \boldsymbol{B}\boldsymbol{\theta}$,

where $\boldsymbol{B}$ is a regression B-spline basis, constructed from the model covariates $(\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{x}_t)$, and $\boldsymbol{\theta}$ a vector of regression coefficients penalized by a discrete penalty matrix $\boldsymbol{P}$ and controlled by smoothness parameters. The coefficients are obtained by penalized likelihood as:

(4) $\widehat{\boldsymbol{\theta}} = (\boldsymbol{B}'\boldsymbol{W}_\delta\boldsymbol{B} + \boldsymbol{P})^{-1}\boldsymbol{B}'\boldsymbol{W}_\delta\boldsymbol{y}$,

where $\boldsymbol{W}_\delta$ is a diagonal matrix of weights (for Gaussian data $\boldsymbol{W}_\delta$ is an identity matrix). In next Section, we focus on how to construct the appropiate basis and penalty for model (3).

2.1 GLAM for spatio-temporal data

In general, spatial data are scattered rather than in a regular grid. Therefore, for the spatial dimension, the model basis is constructed by the *row-wise* kronecker product, or *box*-product (Eilers, et al. 2006) of the marginal B-splines bases of geographical coordinates, i.e.:

(5) $\boldsymbol{B}_s = \boldsymbol{B}_1 \square \boldsymbol{B}_2$,

where $\boldsymbol{B}_1(\boldsymbol{x}_1)$ and $\boldsymbol{B}_2(\boldsymbol{x}_2)$ are of dimensions $n \times c_1$, and $n \times c_2$ respectively. Then, the spatial B-spline basis $\boldsymbol{B}_s$ is of dimension $n \times c_s$, where $c_s = c_1 c_2$, with regression coefficients $\boldsymbol{\theta}_s$ of length $c_s \times 1$, and $2d$ penalty:

(6) $\boldsymbol{P}_s = \lambda_1 \boldsymbol{D}_1'\boldsymbol{D}_1 \otimes \boldsymbol{I}_{c_2} + \lambda_2 \boldsymbol{I}_{c_1} \otimes \boldsymbol{D}_2'\boldsymbol{D}_2$,

where $\boldsymbol{D}_i$, $i = 1, 2$, is a q order difference matrix (in general, we consider a $q = 2$). The penalty $\boldsymbol{P}_s$ in (6), allows for *anisotropic* smoothing by considering a different amount of smoothing for longitude and latitude coordinates ($\lambda_1 \neq \lambda_2$). Note that, scattered data are not in a array structure (or regular grids), hence we are not in the GLAM context, an then array algorithms are not applicable.

However, in the spatio-temporal setting, we can consider a 3-dimensional space-time interaction model in a GLAM context by constructing a full basis from the kronecker product of spatial and temporal B-spline bases, i.e. for model (3), we use:

(7) $\boldsymbol{B} = \boldsymbol{B}_s \otimes \boldsymbol{B}_t$, of dimension $nt \times c_s c_t$,

where $\boldsymbol{B}_t$ is the $t \times c_t$ marginal B-spline basis for time. Now, smoothness is imposed via the penalty matrix $\boldsymbol{P}_{st}$, based on second order difference matrices $\boldsymbol{D}_1$, $\boldsymbol{D}_2$ and $\boldsymbol{D}_t$. The penalty term in 3-dimensions is:

(8) $\boldsymbol{P}_{st} = \tau_1 \boldsymbol{D}_1'\boldsymbol{D}_1 \otimes \boldsymbol{I}_{c_2} \otimes \boldsymbol{I}_{c_3} + \tau_2 \boldsymbol{I}_{c_1} \otimes \boldsymbol{D}_2'\boldsymbol{D}_2 \otimes \boldsymbol{I}_{c_3} + \tau_t \boldsymbol{I}_{c_1} \otimes \boldsymbol{I}_{c_2} \otimes \boldsymbol{D}_t'\boldsymbol{D}_t$,

and for the temporal component (λ_t). Once, we are in the GLAM context, we can use the array methods to compute efficiently linear and inner products as: $\boldsymbol{B}\boldsymbol{\theta}$, $\boldsymbol{B}'\boldsymbol{W}_\delta\boldsymbol{y}$, and $\boldsymbol{B}'\boldsymbol{W}_\delta\boldsymbol{B}$ as detailed in Currie, et al. (2006).

Lee and Durbán (2011) proposed a mixed model representation of the spatio-temporal interaction model (2), where $\boldsymbol{B}\boldsymbol{\theta}$ is reparameterized into $\boldsymbol{X}\boldsymbol{\theta} + \boldsymbol{Z}\boldsymbol{\alpha}$, where $\boldsymbol{\alpha}$ can be viewed as a random effect

$\boldsymbol{\alpha} \sim \mathcal{N}(0, \boldsymbol{G})$, where $\boldsymbol{G} = \sigma^2 \boldsymbol{F}$, with some diagonal matrix $\boldsymbol{F}$, that penalizes the $\boldsymbol{\alpha}$ coefficients. Then, model can be estimated using standard mixed models equations and restricted or residual maximum likelihood (REML) for variance components. The mixed model is then formulated as:

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

where the new bases maintain the kronecker product structure:

(10) $\quad \boldsymbol{X} = \boldsymbol{X}_s \otimes \boldsymbol{X}_t, \text{ and } \boldsymbol{Z} = [\boldsymbol{X}_s \otimes \boldsymbol{Z}_t : \boldsymbol{Z}_s \otimes \boldsymbol{X}_t : \boldsymbol{Z}_s \otimes \boldsymbol{Z}_t],$

where $\boldsymbol{X}_s = [\mathbf{1}_n : \boldsymbol{x}_1] \otimes [\mathbf{1}_n : \boldsymbol{x}_2]$ and $\boldsymbol{X}_t = [\mathbf{1}_t : \boldsymbol{x}_t]$. Hence, the array methods can also be used for computational efficiency (see Durbán (2011) for details).

2.1 Spatio-temporal Smooth-ANOVA model

In most situations, the interpretation of a space-time interaction may result complex, hence we could be interested in a more interpretable model that identifies the spatial and temporal additive components, as well as the space-time interaction. Lee and Durbán (2011) proposed a model with functional form given by:

(11) $\quad \boldsymbol{y} = f_s(\boldsymbol{x}_1, \boldsymbol{x}_2) + f_t(\boldsymbol{x}_t) + f_{s,t}(\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{x}_t) + \boldsymbol{\epsilon},$

where main additive effects of space and time and space-time interaction are modelled and estimated explicitly. Then, an adequate basis for model (11) would be:

(12) $\quad \boldsymbol{B} = [\ \boldsymbol{B}_s \otimes \mathbf{1}_t : \mathbf{1}_n \otimes \boldsymbol{B}_t : \boldsymbol{B}_s \otimes \boldsymbol{B}_t\ ],$

with $\mathbf{1}_n$ and $\mathbf{1}_t$ are column vectors of ones' of length n and t respectively, where each block of (12) corresponds to each of the smooth functions defined in (11). And block-diagonal penalty:

(13) $\quad \boldsymbol{P} = \text{blockdiag}(\boldsymbol{P}_s, \boldsymbol{P}_t, \boldsymbol{P}_{st}),$

where the first two blocks ($\boldsymbol{P}_s$ and $\boldsymbol{P}_t$) correspond to the spatial and temporal penalty terms respectively, and $\boldsymbol{P}_{st}$ is the $3d$ penalty defined in (8) for the space-time interaction. The smoothness is control by a set of smoothing parameters: $\lambda_s = (\lambda_1, \lambda_2)$ for space, λ_t for time, and $\tau_{st} = (\tau_1, \tau_2, \tau_t)$ for space-time interaction.

Model (11) can be seen as a particular case of the Smoothing Splines ANOVA (SS-ANOVA) models proposed by Chen (1993) and Gu (2002). SS-ANOVA models are functional analogous to classical ANOVA models, however, their use are constrained to the dimension of the model basis, and the identifiability constraints on the functional terms of the decomposition.

The model proposed in Lee and Durbán (2011), has important advantages:

1. It is based on low-rank P-splines smoothers.

2. Identifiability problem is avoided using the mixed model representation. In fact, their reparameterization gives the linear dependent terms on the model basis, so that the repeated terms are removed from the model bases. The mixed model bases are exactly equivalent to those in the space-time interaction model in (10), but reordered according to the spatio-temporal ANOVA model functional form in (11).

3. They also demonstrated that, this reparameterization is exactly equivalent to impose the usual constraints on the original P-spline coefficients.

4. Models are also fitted taking advantage of the array structure of the model bases using the GLAM algorithms for the estimation of the variance components and coefficients by REML.

3. Application to air pollution data in Europe

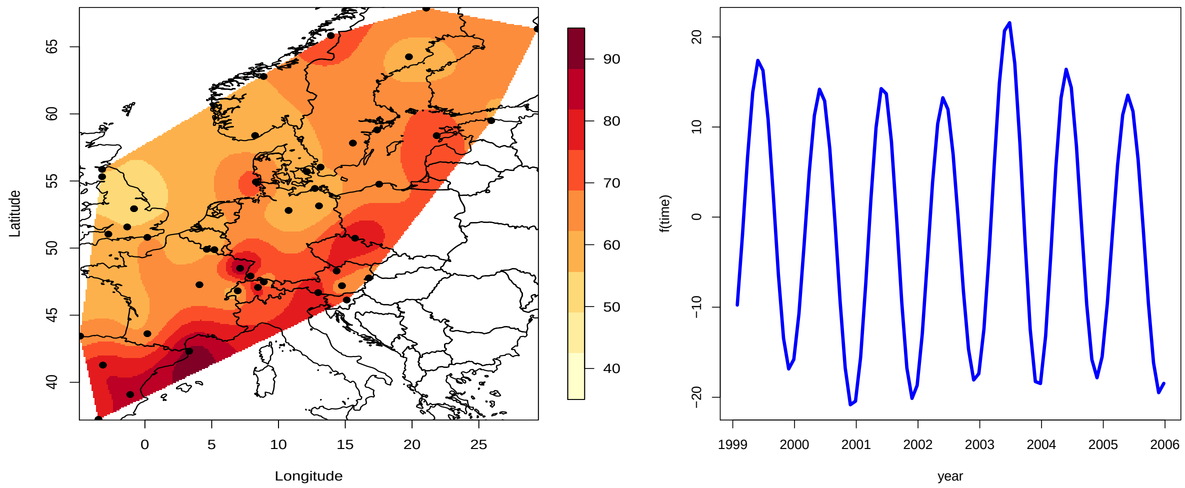
A repeated exposure to ozone pollution ground-level may cause important damages to human health (including asthma, reduced lung capacity or susceptibility to respiratory illnesses), ecosystems and agricultural crops. The formation of ozone is increased by hot weather and in urban industrial areas, and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

The harmful effects of ozone have become an important issue the development of new policies. The European Environment Agency (EEA) has established a program to monitor changes in ozone levels in the last decade. The EEA presents annual evaluation reports of ground-level ozone pollution in Europe from April-September, based on information submitted to the European Commission on ozone in ambient air. Further information is available at the Web site `http://www.eea.europa.eu/`. We analyzed monthly averages of air pollution by ground-level ozone (in $\mu g/m^3$ units) over Europe from January 1999 to December 2005. The data were collected in 43 monitoring stations in 15 european countries. Following the methodology described in previous sections, we fitted 3 models to the data: (i) *spatio-temporal S-ANOVA model*; (ii) 3*d interaction model* and (iii) *space-time additive model*. The three models formulation are then:

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

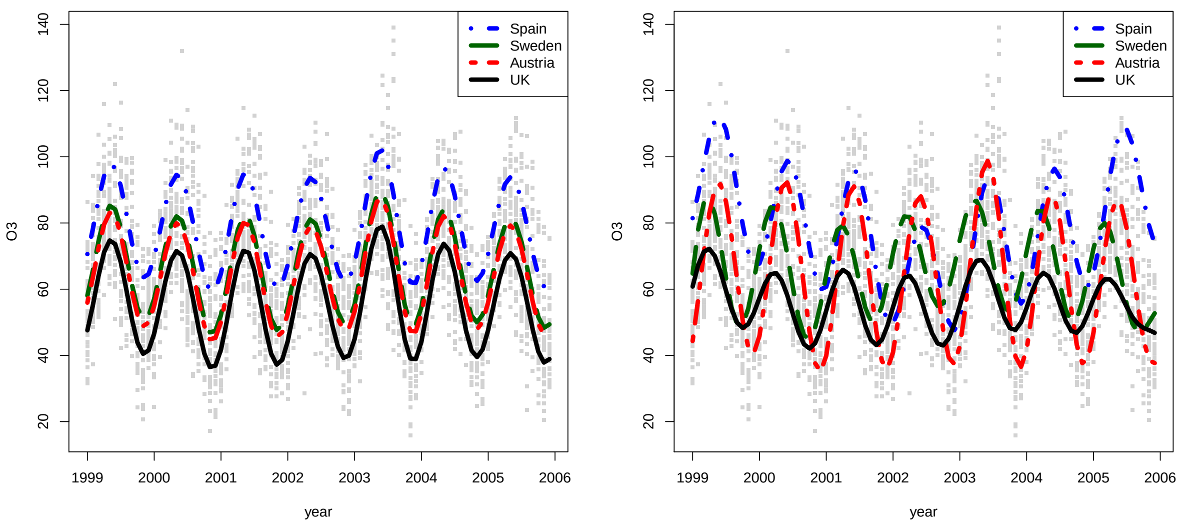
Lee and Durbán (2011) showed the superior performance of S-ANOVA and interaction models with respect to the additive model in terms of model residuals and Akaike Information Criteria (AIC). This could be expected since it is unrealistic to force the spatial pattern of ozone concentrations to increase and decrease similarly in all locations. The interaction model, although giving a better fit, uses a large amount of effective degrees of freedom. This is due to the fact that model has a single smoothing parameter for the temporal component.

Figure 2a shows the smoothed spatial surface for the ozone levels of the S-ANOVA model. The estimated spatial trend surface reflects a non-uniform picture across Europe, since the highest concentrations are observed in Southern Europe in Mediterranean countries as Spain, France and Italy, and the lowest levels are in North West Europe and the UK. The seasonal cycle of ozone levels is captured by the temporal trend shown in Figure 2b, where the highest levels are recorded during spring and summer months (April-September). The spatio-temporal S-ANOVA model also allows the explicit modelling of the space-time interaction in addition to the spatial and temporal trends. Figure 3 shows the fitted values of additive and S-ANOVA models plotted along the years. The additive model, ignores the interaction and assumes a spatial smooth surface over all monitoring stations



(a) Smoothed spatial surface, $f_s(\boldsymbol{x}_1, \boldsymbol{x}_2)$ in S-ANOVA model (b) Smoothed time trend, $f_t(\boldsymbol{x}_t)$ in S-ANOVA model

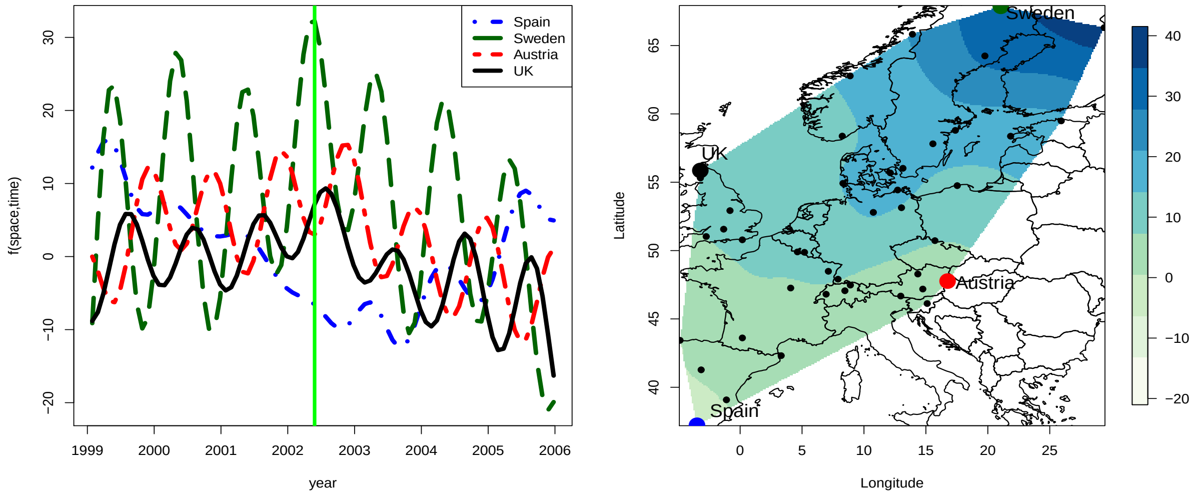
Figure 2: Smoothed main effects for space and time.



(a) Fitted values of additive model (b) Fitted values of S-ANOVA model

Figure 3: Comparison of fitted values of additive (no space-time interaction) and S-ANOVA model (with space-time interaction) in four selected countries.

that remains constant over time. The fitted values vary smoothly according to a seasonal pattern, but maintain the same differences among locations (Figure 3a). In contrast, the spatio-temporal S-ANOVA model fit, is able to capture the individual characteristics of the stations throughout time. Figure 3b shows the particular phase and amplitude given the geographic and seasonal inter-annual



(a) Space-time interaction plotted along time. (b) Space-time interaction plotted across space.

Figure 4: Space-time interaction of S-ANOVA model at May 2002.

variations of four monitoring stations. The space-time interaction is shown in Figure 4, we plotted the space-time interaction term $f_{s,t}(\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{x}_t)$ along time and across space. It can be seen that the patterns are very different. Figure 4b shows, the interaction at May 2002.

4. Discussion

We have presented a flexible modelling methodology for spatio-temporal data smoothing, that can be fitted using array methods. This methodology allowed us to construct ANOVA-type models. In practice, it is also easy to extend the model by the incorporation of other relevant covariates as smooth additive terms or as interactions. One of the main benefits of the spatio-temporal S-ANOVA model proposed is the interpretation of the smoothing and the ability of visualize each of the terms of the decomposition in descriptive plots. The S-ANOVA model also gives a direct interpretation in terms of their smoothing parameters and regression coefficients, since we set independent and separate penalties and coefficients for each smooth term.

If a larger sample of monitoring stations would have been considered in the study during a larger time period, the number of parameters in the interaction $\boldsymbol{B}_s \otimes \boldsymbol{B}_t$ could easily be of the order of thousands, and the computational burden prohibitive. Nevertheless, the GLAM methods also have an important role in the algorithms implementation, since they allow us to store the data and model matrices more efficiently and speed up the calculations. This computational aspect is a topic of current research.

Acknowledgements

The research of Dae-Jin Lee and María Durbán was supported by the Spanish Ministry of Science and Innovation (project MTM 2008-02901).

References

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

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. (2011). Generalized Linear Array Methods as mixed models. *58th Congress of the International Statistical Institute, Dublin.*

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

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

Gu, C. (2002). *Smoothing Spline ANOVA Models*. Springer Series in Statistics. Springer.

Lee, D.-J. and Durbán, M. (2011). P-splines ANOVA-type interaction models for spatio-temporal smoothing. *Statistical Modelling, Vol. 11, Issue 1, pages 49-69.*