Generalized Linear Array methods as Mixed Models

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

Linear mixed effects models, or simply mixed models, are an extension of regression models which incorporate random effects. Since the 90's, several authors have addressed the connection between smoothing splines and mixed models. The interest on this representation is due to the possibility of including smoothing in a large class of models (from correlated data to longitudinal studies and survival analysis), and the use of the methodology and software already developed for mixed models for estimation and inference. In the P-spline context, several authors have extended the model formulation into a mixed model. However, less work has been done in the multidimensional case. In this paper, we address the use of Genelarized Linear Array Methods (GLAM) as mixed models, and show how the reparameterization we propose has some interesting properties.

1. P-splines as mixed models

For simplicity, let us suppose the case of a univariate Gaussian data, with response variable $\boldsymbol{y}$ and regressor $\boldsymbol{x}$. The smooth model is of the form:

$$(1) \quad \boldsymbol{y} = f(\boldsymbol{x}) + \boldsymbol{\epsilon},\ \boldsymbol{\epsilon} \sim \mathcal{N}(0, \sigma^2 \boldsymbol{I}).$$

Eilers and Marx (19996) considered $f(\boldsymbol{x}) = \boldsymbol{B}\boldsymbol{\theta}$, where $\boldsymbol{B}$ is a B-spline regression basis constructed from the covariate $\boldsymbol{x}$, such that, $\boldsymbol{B} = \boldsymbol{B}(\boldsymbol{x})$, and $\boldsymbol{\theta}$, is the vector of regression coefficients. The coefficients $\boldsymbol{\theta}$, can be obtained solving the least squares problem by minimizing the penalized sum of squares:

$$(2) \quad S = (\boldsymbol{y} - \boldsymbol{B}\boldsymbol{\theta})'(\boldsymbol{y} - \boldsymbol{B}\boldsymbol{\theta}) + \boldsymbol{\theta}'\boldsymbol{P}\boldsymbol{\theta},$$

and obtain the explicit solution:

$$(3) \quad \widehat{\boldsymbol{\theta}} = (\boldsymbol{B}'\boldsymbol{B} + \boldsymbol{P})^{-1}\boldsymbol{B}'\boldsymbol{y}.$$

The term $\boldsymbol{P}$ is a difference penalty on the coefficients of the B-spline functions controlled by a smoothing parameter λ. In matrix form, $\boldsymbol{P} = \lambda \boldsymbol{D}_q' \boldsymbol{D}_q$, where q denotes the order of the penalty (in general we choose a second order penalty, i.e. $q = 2$).

The mixed model representation of a P-spline consists of reformulating the model (1) into a standard mixed model given by:

$$(4) \quad \boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{Z}\boldsymbol{\alpha} + \boldsymbol{\epsilon},\ \text{with } \boldsymbol{\alpha} \sim \mathcal{N}(0, \boldsymbol{G}) \text{ and } \boldsymbol{\epsilon} \sim \mathcal{N}(0, \sigma^2 \boldsymbol{I}),$$

where $\boldsymbol{X}$ 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 covariance matrix $\boldsymbol{G}$, which depends on a variance of the random effects σ_α^2.

This reformulation can be viewed as a reparameterization of the original non-parametric model, for which we transform the model B-spline basis into a new model basis, i.e: $\boldsymbol{B} \rightarrow [\boldsymbol{X} : \boldsymbol{Z}]$, and coefficients $\boldsymbol{\theta} \rightarrow (\boldsymbol{\beta}, \boldsymbol{\alpha})'$. This representation decomposes the fitted values as the sum of a polynomial/unpenalized part ($\boldsymbol{X\beta}$) and a non-linear/penalized ($\boldsymbol{Z\alpha}$) smooth term. There are several alternatives depending on the bases and the penalty used. The aim is to find a transformation matrix $\boldsymbol{T}$ to achieve this reparameterization. We consider the singular value decomposition (SVD) of the penalty matrix $\boldsymbol{P}$, i.e. $\boldsymbol{D}'\boldsymbol{D} = \boldsymbol{U\Sigma U}'$, where the matrix of eigenvectors $\boldsymbol{U}$ has two parts: $\boldsymbol{U}_n$ of dimension $c \times q$ containing the null-part eigenvectors and $\boldsymbol{U}_s$ of dimension $c \times (c-q)$ with the non-null eigenvectors. The diagonal matrix $\boldsymbol{\Sigma}$ contains the eigenvalues of the SVD of $\boldsymbol{D}'\boldsymbol{D}$, with q null eigenvectors. For $q = 2$, $\boldsymbol{\Sigma} = \text{diag}(0, 0, \widetilde{\boldsymbol{\Sigma}})$. Then, a suitable transformation matrix $\boldsymbol{T}$ is defined as $\boldsymbol{T} = [\boldsymbol{U}_n : \boldsymbol{U}_s]$, hence we obtain a transformation such that: $\boldsymbol{BT} = [\boldsymbol{X} : \boldsymbol{Z}]$, where the fixed effect matrix can be taken as $\boldsymbol{X} = [\boldsymbol{1} : \boldsymbol{x}]$ and $\boldsymbol{Z} = \boldsymbol{BU}_s$. Finally, the penalty $\boldsymbol{\theta}'\boldsymbol{P\theta}$ becomes $\boldsymbol{\alpha}'\boldsymbol{F\alpha}$, for a new diagonal penalty $\boldsymbol{F} = \lambda\widetilde{\boldsymbol{\Sigma}}$. Since $\boldsymbol{T}$ is orthogonal, we have $(\boldsymbol{\beta}, \boldsymbol{\alpha})' = \boldsymbol{T}'\boldsymbol{\theta}$. Estimates of $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$ follow from the standard mixed model theory:

(5) $\hat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{y}$, and

(6) $\hat{\boldsymbol{\alpha}} = \boldsymbol{GZ}'\boldsymbol{V}^{-1}(\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}})$, where $\boldsymbol{G} = \sigma^2\boldsymbol{F}^{-1}$.

One of the many attractive features of the mixed model formulation of a spline model, is that the smoothing parameter, becomes the ratio between the variance of the residuals and the variance of the random effects, i.e. $\lambda = \sigma^2/\sigma^2_\alpha$. And therefore, the selection of the smoothing parameter can be estimated using *restricted or residual maximum likelihood* (REML):

(7) $\mathcal{L}_R(\lambda) = -\frac{1}{2}\log|\boldsymbol{V}| - \frac{1}{2}\log|\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{X}| - \frac{1}{2}\boldsymbol{y}'(\boldsymbol{V}^{-1} - \boldsymbol{V}^{-1}\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{V}^{-1})\boldsymbol{y}$,

where $\boldsymbol{V} = \sigma^2\boldsymbol{I} + \boldsymbol{ZGZ}'$.

For computational efficiency, it is possible to avoid the direct calculation the determinant and inverse of the variance component matrix $\boldsymbol{V}$, of dimension $n \times n$. Given that $\boldsymbol{V}$ is an example of a Schur complement, it can be shown that:

(8) $|\boldsymbol{V}| = \sigma^{2n}\,|\boldsymbol{G}|\,|\boldsymbol{G}^{-1} + \frac{1}{\sigma^2}\boldsymbol{Z}'\boldsymbol{Z}|$,

(9) $\boldsymbol{V}^{-1} = \frac{1}{\sigma^2}\left(\boldsymbol{I} - \boldsymbol{Z}(\sigma^2\boldsymbol{G}^{-1} + \boldsymbol{Z}'\boldsymbol{Z})^{-1}\boldsymbol{Z}'\right)$

Note that, both expression (8) and (9), involves the inverses of $m \times m$ matrices, where m in this case denotes the number of random effects coefficients, and it is smaller than n.

2. Multidimensional P-splines as mixed models

Now, we consider the bivariate case, with data arrange in a $2d$ array or matrix $\boldsymbol{Y}$, of dimensions $n_1 \times n_2$. The model is now:

(10) $\mathbb{E}[\boldsymbol{Y}] = f(\boldsymbol{x}_1, \boldsymbol{x}_2)$, with covariates $\boldsymbol{x}_1 = (x_{11}, ..., x_{1n_1})'$ and $\boldsymbol{x}_2 = (x_{21}, ..., x_{2n_2})'$.

The regression basis $\boldsymbol{B}$, $n_1n_2 \times c_1c_2$, for model (10) is now the Tensor product of marginal B-splines bases, defined by the Kronecker product:

(11) $\boldsymbol{B}_2 \otimes \boldsymbol{B}_1$ with $\boldsymbol{B}_1 = \boldsymbol{B}(\boldsymbol{x}_1)$, $n_1 \times c_1$ and $\boldsymbol{B}_2 = \boldsymbol{B}(\boldsymbol{x}_2)$, $n_2 \times c_2$.

The regression coefficients are arranged into a matrix $\boldsymbol{\Theta}$, $c_1 \times c_2$, such that vec$\boldsymbol{\Theta} = \boldsymbol{\theta}$, of length $c_1c_2 \times 1$. The penalization of the rows and columns of $\boldsymbol{\Theta}$ is given by

$$(12) \quad \boldsymbol{P} = \lambda_2 \boldsymbol{D}_2'\boldsymbol{D}_2 \otimes \boldsymbol{I}_{c_1} + \lambda_1 \boldsymbol{I}_{c_2} \otimes \boldsymbol{D}_1'\boldsymbol{D}_1.$$

As we showed in the previous Section, we use the SVD on the penalty matrix $\boldsymbol{P}$ to obtain the mixed model transformation matrix $\boldsymbol{T}$. In $2d$, this transformation is defined as the Kronecker product of the marginal transformation matrices:

$$(13) \quad \boldsymbol{T} = \boldsymbol{T}_2 \otimes \boldsymbol{T}_1 = [\boldsymbol{U}_{2n} : \boldsymbol{U}_{2s}] \otimes [\boldsymbol{U}_{1n} : \boldsymbol{U}_{1s}],$$

where $\boldsymbol{U}_{in}$ and $\boldsymbol{U}_{is}$ are obtained from the SVD on $\boldsymbol{D}_i'\boldsymbol{D}_i$, for $i = 1, 2$. The mixed model matrices becomes:

$$(14) \quad \boldsymbol{X} = \boldsymbol{X}_2 \otimes \boldsymbol{X}_1, \text{ and } \boldsymbol{Z} = [\boldsymbol{X}_2 \otimes \boldsymbol{Z}_1 : \boldsymbol{Z}_2 \otimes \boldsymbol{X}_1 : \boldsymbol{Z}_2 \otimes \boldsymbol{Z}_1],$$

and diagonal penalty:

$$(15) \quad \boldsymbol{F} = \begin{pmatrix} \lambda_2 \widetilde{\boldsymbol{\Sigma}}_2 \otimes \boldsymbol{I}_{q_1} & & \\ & \lambda_1 \boldsymbol{I}_{q_2} \otimes \widetilde{\boldsymbol{\Sigma}}_1 & \\ & & \lambda_1 \boldsymbol{I}_{c_2-q_2} \otimes \widetilde{\boldsymbol{\Sigma}}_1 + \lambda_2 \widetilde{\boldsymbol{\Sigma}}_2 \otimes \boldsymbol{I}_{c_1-q_1} \end{pmatrix},$$

where $\widetilde{\boldsymbol{\Sigma}}_1$, and $\widetilde{\boldsymbol{\Sigma}}_2$, of dimensions $(c_1 - q_1) \times (c_1 - q_1)$ and $(c_2 - q_2) \times (c_2 - q_2)$, are the diagonal matrices of positive eigenvalues of $\boldsymbol{D}_1'\boldsymbol{D}_1$ and $\boldsymbol{D}_2'\boldsymbol{D}_2$.

Given the Kronecker structure of the model matrices, the use of array methods is straightforward.

2.1 Array methods

The main idea is the efficient computation of the matrix products involved in the REML expression in (7) and definitions of $|\boldsymbol{V}|$ and $\boldsymbol{V}^{-1}$, in (8), and (9). Taking advantage of the array structure of $\boldsymbol{X}$ and $\boldsymbol{Z}$ in (14), we may use the GLAM algorithms for a fast and efficient computation of linear and inner products. Let $\boldsymbol{X}_1$, $n_1 \times q_1$, $\boldsymbol{X}_2$, $n_2 \times q_2$, we use two basic array computations:

- **Linear functions:** the elements of $\boldsymbol{X}'\boldsymbol{y}$ are computed as

  $$(16) \quad (\boldsymbol{X}_2 \otimes \boldsymbol{X}_1)'\boldsymbol{y}, q_1q_2 \times 1 \equiv (\boldsymbol{X}_1'\boldsymbol{Y}\boldsymbol{X}_2), q_1 \times q_2,$$

  where symbol $\equiv$ indicates that both sides have the same elements but ordered in different dimensions.

- **Inner products:** elements of $\boldsymbol{X}'\boldsymbol{X}$ are computed as

  $$(17) \quad (\boldsymbol{X}_2 \otimes \boldsymbol{X}_1)'(\boldsymbol{X}_2 \otimes \boldsymbol{X}_1), q_1q_2 \times q_1q_2 \equiv \mathcal{G}(\boldsymbol{X}_1)'\boldsymbol{W}\mathcal{G}(\boldsymbol{X}_2), q_1^2 \times q_2^2,$$

  where $\boldsymbol{W} = \boldsymbol{1}\boldsymbol{1}'$, $n_1 \times n_2$, and $\mathcal{G}(\boldsymbol{X})$ is the *row-tensor a matrix*, defined as,

  $$\mathcal{G}(\boldsymbol{X}) = (\boldsymbol{X} \otimes \boldsymbol{1}_q') * (\boldsymbol{1}_q' \otimes \boldsymbol{X}), n \times q^2$$

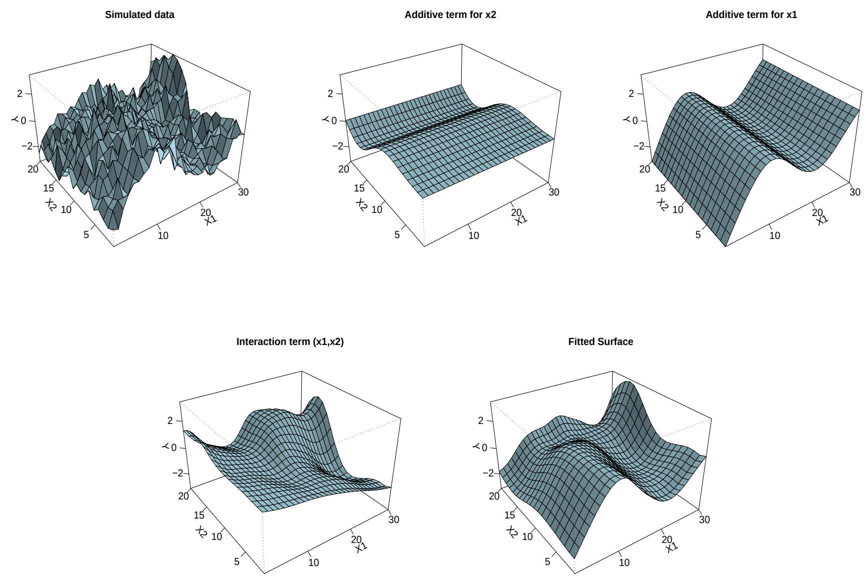


Figure 1: Decomposition of the two dimensional surface into additive terms and interactions.

Both linear and inner products are also used to compute: $\boldsymbol{Z}'\boldsymbol{y}$, $\boldsymbol{Z}'\boldsymbol{Z}$, and $\boldsymbol{X}'\boldsymbol{Z}$, but the operations are repeated block-wise given the block structure of matrix $\boldsymbol{Z}$ in (14).

Figure 1 shows data simulated on a 30×20 grid. We take $\boldsymbol{B}_1$, 30×13, and $\boldsymbol{B}_2$, 20×10 which gives a full regression matrix $\boldsymbol{B}$, 600×130. The bivariate model is reparameterized into a mixed model and fitted by REML, with smoothing parameters λ_1 and λ_2.

Top left panel of Figure 1 also shows the decomposition of the two dimensional surface in terms of a sum of additive and non-additive terms. Using our decomposition $\boldsymbol{X}$ and $\boldsymbol{Z}$ matrices in (14) can be expanded and reorganized as:

(18) $\boldsymbol{X} = [\mathbf{1}_{n_2} \otimes \mathbf{1}_{n_1} : \mathbf{1}_{n_2} \otimes \boldsymbol{x}_1 : \boldsymbol{x}_2 \otimes \mathbf{1}_{n_1} : \boldsymbol{x}_2 \otimes \boldsymbol{x}_1],$

(19) $\boldsymbol{Z} = [\mathbf{1}_n \otimes \boldsymbol{Z}_2 : \boldsymbol{Z}_2 \otimes \mathbf{1}_{n_1} : \boldsymbol{x}_2 \otimes \boldsymbol{Z}_1 : \boldsymbol{Z}_2 \otimes \boldsymbol{x}_1 : \boldsymbol{Z}_2 \otimes \boldsymbol{Z}_1]$

This partition facilitates two things:

- We can represent the fitted surface as an overall mean plus the sum of three components: (i) a term for the first covariate (i.e. $\boldsymbol{x}_1$ and $\boldsymbol{Z}_1$); (ii) a term for the second covariate (i.e. $\boldsymbol{x}_2$ and $\boldsymbol{Z}_2$); and (iii) those terms depending on the interactions of $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ (i.e. $[\boldsymbol{x}_1 : \boldsymbol{Z}_1] \otimes [\boldsymbol{x}_2 : \boldsymbol{Z}_2]$).

- We can consider a decomposition of the two-dimensional surface as:

  $$f(\boldsymbol{x}_1, \boldsymbol{x}_2) = f_1(\boldsymbol{x}_1) + f_2(\boldsymbol{x}_2) + f_{1,2}(\boldsymbol{x}_1, \boldsymbol{x}_2),$$

  where functions f_1 and f_2 can be interpreted as an ANOVA-type model as *main effects* of the covariates $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$, and $f_{1,2}$ is a two-dimensional interaction surface or *interaction effect* between $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$. This decomposition is strongly related to the work proposed by Gu (2002) and the *Smoothing Spline Analysis of Variance* (or SS-ANOVA) *models*. For this Smooth-

ANOVA model, the new mixed model penalty becomes:

$$(20)\quad \boldsymbol{F} = \begin{pmatrix} \lambda_1\widetilde{\boldsymbol{\Sigma}}_1 & & & & \\ & \lambda_2\widetilde{\boldsymbol{\Sigma}}_2 & & & \\ & & \lambda_3\widetilde{\boldsymbol{\Sigma}}_1 & & \\ & & & \lambda_4\widetilde{\boldsymbol{\Sigma}}_2 & \\ & & & & \lambda_3\boldsymbol{I}_{c_2-q_2}\otimes\widetilde{\boldsymbol{\Sigma}}_1 + \lambda_4\widetilde{\boldsymbol{\Sigma}}_2\otimes\boldsymbol{I}_{c_1-q_1} \end{pmatrix},$$

where we consider 4 smoothing parameters: λ_1 and λ_2 for the additive main effects for $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$, and λ_3 and λ_3 for the interaction between $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$. Lee and Durbán (2011) showed that this ANOVA decomposition avoids the identifiability problems in this type of models, and demonstrated how this procedure based on the mixed model reparameterization is exactlty equivalent to impose the usual linear constraints on the regression coefficients of the original P-spline model.

3. A generalized linear mixed model approach

We can extend the methodology to non-Gaussian responses in the context of generalized linear mixed models (GLMM). This approach consists of two stages: (i) reparameterize the linear predictor and (ii) estimate the model parameters. We consider the Poisson case with log link function:

- **Model:** $\boldsymbol{\mu} = \mathbb{E}[\boldsymbol{y}]$, $\log\boldsymbol{\mu} = \boldsymbol{B\theta}$ is reparameterized into $\log\boldsymbol{\mu} = \boldsymbol{X\beta} + \boldsymbol{Z\alpha}$ as we showed in the previous Section.
- **Linear predictor:** $\boldsymbol{\eta} = \exp(\boldsymbol{X\beta} + \boldsymbol{Z\alpha})$.
- **Estimation:** for Poisson data we use the penalized Quasi-Likelihood (PQL) approach of Breslow and Clayton (1993), where the scoring algorithm is

$$(21)\quad \begin{bmatrix} \boldsymbol{X}'\boldsymbol{W}_\delta\boldsymbol{X} & \boldsymbol{X}'\boldsymbol{W}_\delta\boldsymbol{Z} \\ \boldsymbol{Z}'\boldsymbol{W}_\delta\boldsymbol{X} & \boldsymbol{X}'\boldsymbol{W}_\delta\boldsymbol{Z} + \boldsymbol{G}^{-1} \end{bmatrix} \begin{bmatrix} \widehat{\boldsymbol{\beta}} \\ \widehat{\boldsymbol{\alpha}} \end{bmatrix} = \begin{bmatrix} \boldsymbol{X}'\boldsymbol{W}_\delta \\ \boldsymbol{Z}'\boldsymbol{W}_\delta \end{bmatrix} \boldsymbol{z},$$

where $\boldsymbol{z} = \boldsymbol{\eta} + \boldsymbol{W}_\delta^{-1}(\boldsymbol{y} - \boldsymbol{\mu})$ is the working vector, and $\boldsymbol{W}_\delta$ is a diagonal matrix of weights equal to diag(exp($\boldsymbol{X\beta} + \boldsymbol{Z\alpha}$)). Now, the variance $\boldsymbol{V}$ is given by $\boldsymbol{V} = \boldsymbol{W}_\delta^{-1} + \boldsymbol{ZGZ}'$, where as in the Gaussian case computational more efficient expressions for $\boldsymbol{V}^{-1}$ and $|\boldsymbol{V}|$ can be used. The PQL solution is obtained by iteration between (21) and REML estimates in (7) until convergence.

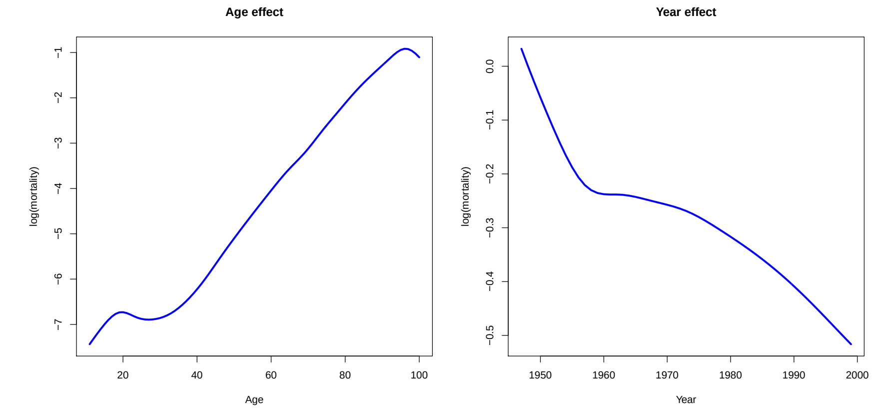
Now, the array methods are used for efficient estimation and updated in every iteration of the algorithm to compute: $\boldsymbol{X}'\boldsymbol{W}_\delta\boldsymbol{X}$, $\boldsymbol{X}'\boldsymbol{W}_\delta\boldsymbol{Z}$, $\boldsymbol{Z}'\boldsymbol{W}_\delta\boldsymbol{Z}$, $\boldsymbol{X}'\boldsymbol{W}_\delta\boldsymbol{z}$, and $\boldsymbol{Z}'\boldsymbol{W}_\delta\boldsymbol{z}$.

To illustrate the methodology for Poisson $2d$ data, we considered the mortality data analyzed in Currie, et al. (2004). The data consists of the number of policy claims (or mortality $\boldsymbol{y}$) and number of years lived (or exposures $\boldsymbol{e}$) for each calendar year (1947–1999) and each age (11-100). The total number of observations is $53 \times 90 = 4770$. We considered $\boldsymbol{B}_a = \boldsymbol{B}$(Age), 90×23 and $\boldsymbol{B}_y = \boldsymbol{B}$(Year), 53×13, hence the full regression basis if $\boldsymbol{B}_a \otimes \boldsymbol{B}_y$, 4770×299.

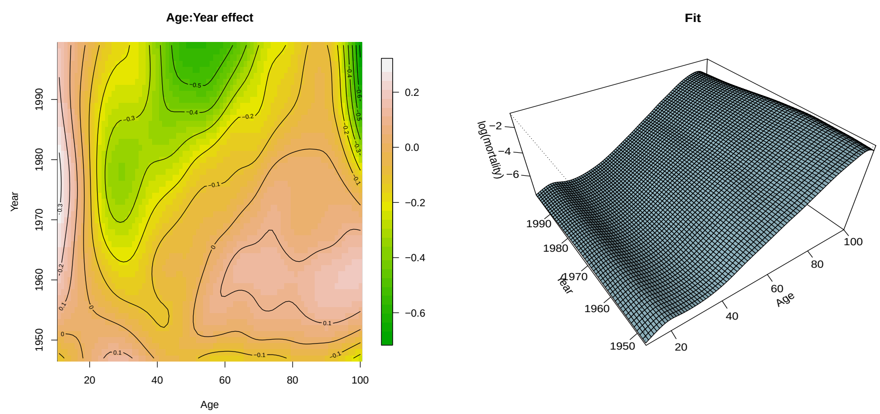
For this example, we fitted a ANOVA model:

$$(22)\quad \log\boldsymbol{\mu} = f(\text{Age}) + f(\text{Year}) + f(\text{Age,Year}),$$

using the mixed model representation showed in this paper. Figure 2 shows the components of ANOVA model fitted to this data. The two top panels of Figure 2 shows the overall main effects of `Age` and `Year`. Bottom panel shows the interaction effect, and the fitted surface.



(a) Additive `Age` main effect (b) Additive `Year` main effect



(c) `Age`-by-`Year` interaction effect (d) Fitted mortality surface

Figure 2: ANOVA decomposition of mortality data.

The main advantage of the ANOVA model, is in terms of interpretability and inference. Although both $2d$ and ANOVA models would give similar fitted values, the fact that we are considering different and independent penalties: λ_a and λ_y for main `Age` and `Year` effects, and $\breve{\lambda}_a$ and $\breve{\lambda}_y$ for the `Age`*`Year` interaction.

Then, we can compute separately effective degrees of freedom for each component and test for additive terms versus the interaction effect, for instance considering a single isotropic smoothing parameter, $\breve{\lambda} = \breve{\lambda}_a = \breve{\lambda}_y$, and then test for $\mathrm{H}_0 : \breve{\lambda}^{-1} = 0$.

4. Discussion

GLAM as mixed models give a fast and compact method of 2-dimensional smoothing. The method can be generalized to higher dimensions where the advantages of array methods in terms of storage and speed are even greater. The mixed model formulation presented in this paper, enables us not only to estimate the smoothing parameter/s maximizing REML rather than using an information criterion as AIC or BIC, but also to represent the fitted surface to be expressed as a sum of additive and interaction terms. The new bases also allows going further and clarify the role of the penalty by relating the penalty to each of the parts of the ANOVA-type decomposition (with its correspondent smoothing parameters). This allow us to build and identifiable model in a very simple and intuitive way and hence a number of possibilities can be considered as: model selection, or testing additive models versus interaction.

Acknowledgements

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

References

Breslow, N. E. and Clayton, D. G. (2006). Approximate inference in generalized linear mixed models. *Journal of the American Statistical Association*, 88,9–25.

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

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.*