Durban, M. & Aguilera-Morillo, M. C. (2017). On the estimation of functional random effects. *Statistical Modelling*, 17(1–2), pp. 50–58.

DOI: 10.1177/1471082x16681333

On the estimation of functional random effects

Maria Durban [1] and M. Carmen Aguilera-Morillo [1]

[1] Department of Statistics, Universidad Carlos III de Madrid, Spain

---

Address for correspondence: María Durban, Department of Statistics, Universidad Carlos III de Madrid, Avda. Universidad 30, CP–28911 Leganés – Madrid, Spain.

E-mail: `mdurban@est-econ.uc3m.es`.

Abstract: Functional regression modelling has become one of the most vibrant areas of research in the last years. This discussion provides some alternative approaches to one of the key issues of functional data analysis: the basis representation of curves, and in particular, of functional random effects. First, we propose the estimation of functional principal components by penalizing the norm, and as an alternative, we provide a an efficient and unified approach based on B-spline basis and quadratic penalties

---

Key words: Functional PCA; Subject-specific curves; Identifiability; SOP algo-

rithm

1 Introduction

It is a pleasure for us to contribute in the discussion of this paper. The authors present a complete review of the main functional regression models (scalar-on-function, function-on-scalar and function-on-function regression) developed in the context of mixed models. It is a very important piece of work indeed, since all models are expressed under a unified approach, and the authors should be congratulated for their contribution.

Our discussion is mainly focused on the estimation of functional random effects. Our intention is, on one hand, to bring the attention of the authors to an alternative method for the estimation of functional principal components. This method yields more parsimonious basis representation without compromising the amount of covariance explained. On the other hand, we present a fast estimation method when individual curves are fitted via B-spline basis and roughness penalties. The method proposed is a competitive alternative, both in terms of fit and efficiency.

2 Functional principal component estimation by penalizing the norm: a P-spline approach

In this section we will focused on the basis representation of the model parameters. In the paper, authors have considered their representation in terms of two different types of basis functions, that is, splines and functional principal components (FPCs).

The authors propose to estimate the FPC functions based on a smoothed version of the sample covariance. We were surprised since, in some cases, the number of FPCs needed to get a good representation is quite high. For example, in section 3.5 (Application Example) 34 FPCs were considered. In the estimation of model (1.1) 8 FPCs were needed to explain the 95% of variability. Both cases are based on the functional variable FA-CCA (fractional anisotropy values along CCA tract).

Usually, FPCA is a good technique to reduce the dimension and only a reduced number of FPCs is required to explain a hight proportion of variability. Because of this, the results in section 3.5. surprise us.

It is very well known that in the presence of noisy data a roughness penalty must be considered during the estimation of the FPCs. In that case, a good and smooth estimation of the FPCs is obtained, and also more variability is explained using less components.

There are different ways to penalized the FPCs algorithm: the one proposed by the authors or others (more usual in functional data context) based on the orthonormality constraints (Silverman, 1996; Reiss and Ogden, 2007; Aguilera and Aguilera-Morillo, 2013). In particular, our proposal is based on a penalized estimation of FPCs as follows (see Aguilera and Aguilera-Morillo, 2013, for more details).

Let us consider a functional random variable X whose observations are realizations of a second order stochastic process $X = \{X(t), \quad t \in \tau\}$, continuous in quadratic mean, and whose sample functions $\{x_i(t) : \ t \in \tau, \ i = 1, \ldots, n\}$ belong to the Hilbert space of square integrable functions $L_2(\tau)$, with the usual inner product given by

$$\langle f, g\rangle = \int_{\tau} f(t)g(t)dt, \quad \forall f, g \in L_2(\tau).$$

In practice, sample curves are observed in a finite set of points, $\{t_{i,0}, \ldots, t_{i,m_i} \in \tau\}$ $\forall i = 1, \ldots, n$, so that our first step is to get the functions from the raw data. To do this, the basis representation of the sample curves in terms of cubic B-splines basis is considered, so that

$$x_i(t) = \sum_{l=1}^{K} a_{il}\phi_l(t), \quad i = 1, \ldots, n,$$

with $a_i = (a_{i1}, \ldots, a_{iK})'$ being the vector of basis coefficients estimated for the i-th sample curve.

After that, the FPCs are obtained as generalized linear combinations with maximum variance, so that the l-th principal component scores are given by

$$\xi_{il} = \int_{\tau} x_i(t) f_l(t) dt, \quad i = 1, \ldots, n,$$

where the eigenfunctions $f_l(t)$ can be obtained as solution to the following maximization problem

$$\max_{f \in L_2(\tau)} \frac{Var\left[\int_{\tau} X\left(t\right) f\left(t\right) dt\right]}{\|f\|^2 + \lambda Pen_d\left(f\right)},$$

with $\|.\|$ being the norm associated to the usual inner product $\langle ., . \rangle$, λ being the smoothing parameter and $Pen_d\left(f\right) = f' P_d f$, where f denotes the vector of basis coefficients of $f(t)$ and $P_d = \left(\triangle^d\right)' \triangle^d$, with $\triangle^d$ being the matrix of d-order differences (Eilers and Marx, 1996).

Taking as example the FA-CCA variable, we have compared our version of penalized FPCA (PEN-FPCA, Aguilera and Aguilera-Morillo, 2013) with the results provided by the authors (FPCA by smoothed covariance). Regarding the previous basis representation of the sample paths, Scheipl and Greven (2016) recommended the use a sufficiently large number of basis functions if a pre-processing is required. In that

Method	PC1	PC2	PC3	PC4	PC5	PC6	PC 7	PC8	Accumulated
I	60.42%	10.52%	7.85%	6.46%	4.50%	2.51%	1.70%	1.40%	95.36%
II	65.35%	11.11%	7.41%	5.23%	4.41%	2.54%	1.45%	0.90%	98.41%

Table 1: Proportion of variability explained by the first eight FPCs. FPCs were estimated by Method I (FPCA by smoothed covariance) and Method II (penalized FPCA by Aguilera and Aguilera-Morillo, 2013).

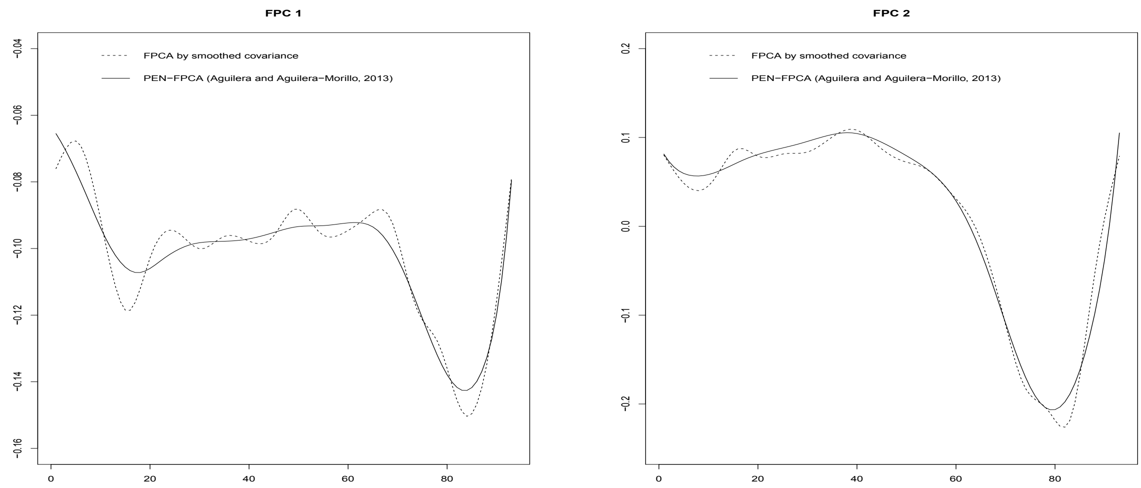

Figure 1: First two eigenfunctions estimated by FPCA by smoothed covariance (dashed line) and Penalized FPCA (Aguilera and Aguilera-Morillo, 2013) (solid line).

sense, twenty B-spline basis were considered to get the functional form of the residual curves. Also, a second order difference penalty has been considered in the penalized FPCA algorithm.

After computing the two FPCA algorithms, the results in terms of the proportion of variability explained by the first eight FPC's can be seen in Table 1. Moreover, in Figure 1 the first two eigenfunction estimated by FPCA by smoothed covariance and PEN-FPCA have been represented in dashed and solid lines, respectively.

From these results, it can be seen that PEN-FPCA achieves smoother eigenfunctions than FPCA by smoothed covariance, which should be easier to interpret. Besides, this method increases the proportion of explained variability by the first principal components, so that less components are required to get a god data representation. In this particular example the difference between both approaches is not dramatic, however, in other situations PEN-FPCA may result in a more parsimonious basis representation.

3 Functional random effects: a P-spline mixed model approach

Another possibility for the for the representation of functional random effects /smooth residuals is the use of B-spline basis and a roughness penalty. However, when the number of individuals in the dataset is large, and the shape of the functions is complicated, it might be necessary to use a basis of moderate size. This can increase greatly the computational burden, and so, approaches based on FPCs (like the one in the paper or in the previous section) are preferred. The aim of this section is to show that B-spline basis and derivative/difference penalties can be a competitive alternative, both in terms of fit and efficiency. Furthermore, they can easily deal with correlated error curves (Aguilera-Morillo et al., 2016) and other important issues such as identifiability. Furthermore, they avoid a two-stage estimation procedure which might seems unsatisfactory.

We will focus on the simplest situation: we will use model (1.1) in the paper, and assume that the functional intercept is zero and that all , i.e.:

$$Y_i(t) = B_i(t) + \epsilon_{it}, i = i, \ldots, n \tag{3.1}$$

where the functional random intercept (or smooth residual) is expanded in terms of B-spline basis:

$$B_i(t) = \sum_{l=1}^{K} \phi_l(t) a_{il} = \mathbf{\Phi a}_i \quad \mathbf{a_i} \sim N(\mathbf{0}, \sigma^2 \lambda^{-1} \mathbf{S}^-),$$

$N(\mathbf{0}, \sigma^2 \lambda \mathbf{S}^-)$ is a partially improper Gaussian distribution with positive semi-definite covariance matrix $\mathbf{S}$ (since roughness penalties are, in general , rank deficient), and λ is the smoothing parameter controlling the amount of smoothness of the curves. When the curves are observed in a common dense grid to all functions, model (3.1) can be written in matrix form as:

$$\mathbf{Y} = (\mathbf{I}_n \otimes \Phi)\mathbf{a} + \epsilon \quad \mathbf{a} \sim N(\mathbf{0}, \sigma^2(\mathbf{I}_n \otimes \lambda^{-1}\mathbf{S}^-)). \tag{3.2}$$

Estimation now can be carried out using the mixed model approach presented in the paper. However, if other terms are included in the model some identifiability issues arise, and the estimation of the smoothing parameters may become a burden. We propose to use the approach introduced by Durban et al. (2005) which makes use of one of the possible reparameterizations of penalized splines as mixed models. The proposed reparameterization separates the random functional effect into a penalized and unpenalized term which has now a proper distribution. In particular, using the singular value decomposition of $\mathbf{S}$, $\mathbf{S} = \mathbf{U\Omega U}$, model (3.2) becomes:

$$\mathbf{Y} = (\mathbf{I}_n \otimes \mathbf{X})\beta + (\mathbf{I}_n \otimes \mathbf{Z})\alpha + \epsilon \quad \alpha \sim N(\mathbf{0}, \tau^2(\mathbf{I}_n \otimes \mathbf{I}_K)) \tag{3.3}$$

with $\mathbf{X} = \Phi\mathbf{U_n}\widetilde{\mathbf{\Omega}}^{-1/2}$ and $\mathbf{Z} = \Phi\mathbf{U_s}\widetilde{\mathbf{\Omega}}^{-1/2}$ and $\lambda = \tau^2/\sigma^2$. $\mathbf{U_n}$ and $\mathbf{U_s}$ are matrices of eigenvectors corresponding to the null and non-null spaced spanned by the penalty matrix, while $\widetilde{\mathbf{\Omega}}$ is a diagonal matrix with the non-zero eigenvalues of $\mathbf{S}$.

Brumback and Rice (1998) already used this approach, but they pointed out severe computational problems when the number of individuals was large, since the size of the vector of parameters β is $n \times rank(\mathbf{X})$. A possible solution, is to assume that all terms are random, i.e.

$$\mathbf{Y} = (\mathbf{I}_n \otimes \mathbf{X})\beta + (\mathbf{I}_n \otimes \mathbf{Z})\alpha + \epsilon \quad \beta \sim N(\mathbf{0}, \mathbf{I}_n \otimes \mathbf{\Sigma}) \quad \alpha \sim N(\mathbf{0}, \tau_3^2(\mathbf{I}_n \otimes \mathbf{I}_K)). \quad (3.4)$$

We will assume here that $\mathbf{\Sigma}$ is a diagonal matrix, however, a more general covariance matrix can be used. If, for example, $rank(\mathbf{X}) = 2$, individual curves would be the sum of random linear and non-linear departures from the population mean, and $\mathbf{\Sigma} = diag(\tau_1^2, \tau_2^2)$.

This approach has several advantages: No further identifiability constraints need to be imposed, there is no need to be limited to difference penalties of low order (which might not be appropriate in some situations), and more flexibility is achieved. We have used this methodology to fit model (1.1) in the paper. We used 25 B-spline basis functions for group means, 20 for individual curves, and second order penalties.

Figure 2 shows the estimated smooths residuals for two control and two MS patients, using the functional principal component methods proposed by the authors, and the penalized spline approach described above. In all cases the use of individual curves estimated via penalized splines gave a better estimation of the smooth residuals. The use of this latter approach not only did not increasing the computational burden, but was significantly more efficient. We are not sure why the shape of the estimated random effects using functional principal components were not so close to the residuals in some cases. Maybe due to the covariance smoothing, if the size of the basis is too small or the smoothing method is not appropriate.

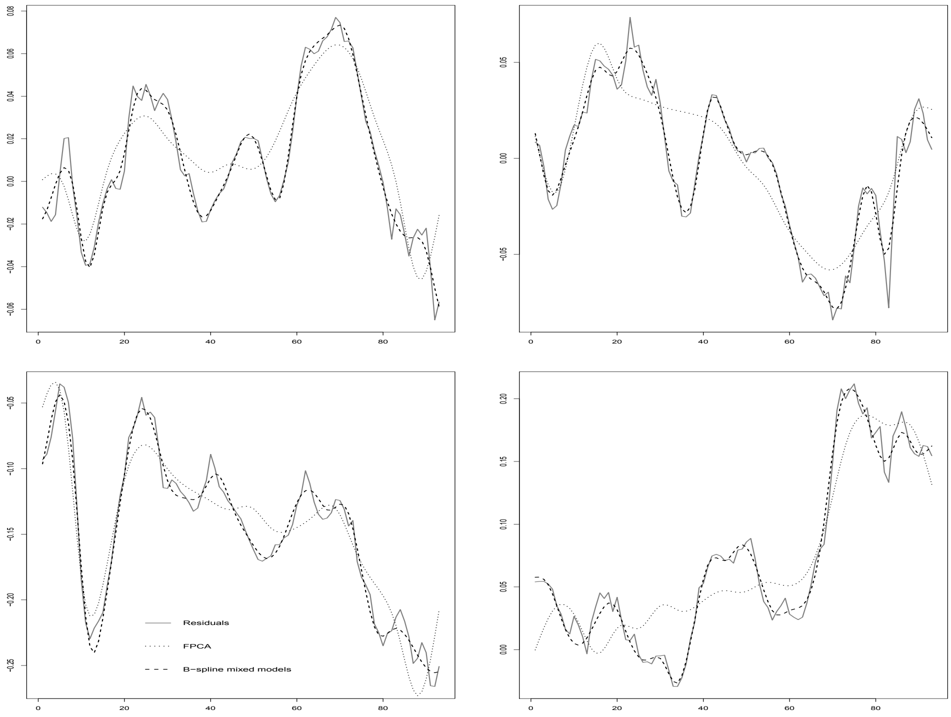

Figure 2: Plot of residuals (grey) after estimation of mean structure and centering, and estimated random effects using functional principal components (dotted lines) and penalized splines (dashed lines). Top row corresponds to two control cases and bottom row to patients affected with MS.

3.1 Fast estimation of functional random effects

As previously pointed out, one of the possible advantages of the use of functional principal components over penalized splines is the fact that no smoothing parameter/variance component has to be estimated within the large-overall-model. The intention of this section is to give some hints on how to fit efficiently the approach presented above. Here we focus on data with a full array structure, but the methods introduced here can be adapted to sparse arrays.

We propose the use of the algorithm introduce by Rodriguez-Alvarez et al. (2016)

called *Separation of Overlapping Penalties* (SOP). This is a generalization of an earlier algorithm: *Separation of Anisotropic Penalties* (Rodriguez-Alvarez et al., 2015) which was designed for the estimation of smoothing parameters of multidimensional P-splines with anisotropic penalties. The new algorithm, SOP, can be applied to any mixed model (or models in which mixed model based inference is used) in which the precision matrix of random effects can be expressed as a linear combination over the variance components. For example, model (3.4) can be written as:

$$\mathbf{Y} = (\mathbf{I}_n \otimes \widetilde{\mathbf{Z}})\gamma + \epsilon \quad \gamma \sim N(\mathbf{0}, \mathbf{G}),$$

with $\widetilde{\mathbf{Z}} = [\mathbf{X} : \mathbf{Z}]$, $\gamma^T = [\beta : \alpha]$. Then, the precision matrix $\mathbf{G}$ can be expressed as:

$$\mathbf{G^{-1}} = \sum_{h=1}^{3} \frac{1}{\tau_k^2} \mathbf{\Lambda_k}, \quad \mathbf{\Lambda_1} = \mathbf{\Lambda_2} = \mathbf{I_n} \otimes 1, \mathbf{\Lambda_3} = \mathbf{I_n} \otimes \mathbf{I_K},$$

and so, the SOP algorithm can be used. This algorithm is much faster than other existing approaches, and can cope with the estimation of hundreds of variance components in a matter of seconds (Rodriguez-Alvarez et al., 2016). We could, therefore, make the model even more flexible and allow a specific smoothing parameter per subject. Then we would have a different variance component per individual, and each $\mathbf{\Lambda_k}$ would become $\mathbf{\Lambda_k} = \mathbf{E_k} \otimes \mathbf{I_K}$ (where $\mathbf{E_k}$ is an indicator matrix for each subject).

This algorithm combined with the use of *Generalized Linear Array Models* (GLAM) introduced by Currie et al. (2006) (in which the calculation of Kronecker products is avoided) yields a very efficient estimation procedure, even when the basis used for individual curves are of moderate size. Of course, the SOP algorithm can be also applied when the model is expressed in the original parameterizacion, i.e., $\alpha \sim N(\mathbf{0}, \mathbf{P}^{-})$, since in all cases provided in the paper, $\mathbf{P}$ is a linear combination of smoothing parameters.

4 Identifiability via penalties

The authors mention two possible sources for non-identifiability: i) arising in all models with functional responses, and ii) in the case of function-on-function regression. In the first case, the authors propose a sum-to-zero constraint for each term in the model. The second case is more complicated and some countermeasures (such as use of first order penalties) are proposed.

We bring to their attention a different approach based on Djeundje and Currie (2010). They propose a novel methods to address smoothness and identifiability simultaneously via penalties. The idea is to double penalize the coefficients, one with the usual derivative/difference penalty for smoothness, and a second ridge penalty that ensures identifiability (by shrinking the coefficient to zero). For example, in model (3.2), the distribution of the random effects would be

$$\mathbf{a} \sim N(\mathbf{0}, \sigma^2(\mathbf{I}_n \otimes (\lambda_1 \mathbf{S} + \lambda_2 \mathbf{I}_K)^{-1})$$

.

In the function on function case, the penalty on the coefficients surface would become:

$$\mathbf{P} = \widetilde{\mathbf{P}}_s \otimes \mathbf{I}_{K_Y} + \mathbf{I}_{K_X} \otimes \mathbf{P}_Y, \quad \widetilde{\mathbf{P}}_s = \lambda_s \mathbf{P}_s + \lambda \mathbf{I}_{K_X}.$$

This approach is, in fact, closely related to the one proposed in Marra and Wood (2012): In that paper, the authors modify the null space of the penalty so that it has full rank. However, in their proposal, the zero eigenvalues of the penalty matrix are modified by a small number previously chosen. Our proposal allows for the estimation of the smoothing parameter in the ridge penalty at the same time as the rest of variance components in the model. Again, this can be achieved easily using the SOP algorithm, since it can be applied to penalized regression with more than one penalty acting on the same vector of coefficients.

5 Conclusions

The aim of our discussion was to provide a small contribution to one important topic within functional regression models: the estimation of functional random effects. We have presented different alternatives with the goal of finding parsimonious basis representations and/or faster estimation methods. One approach is based on a penalized estimation of the functional principal components, and the other on a the usual B-spline representation, but formulated in such a way that several goals are achieved simultaneously: efficient and accurate estimation, compact representation and a solution to the identifiability problems, embedded withing the estimation algorithm. The application this approach to one of the examples in the paper improved significantly the methods proposed there, both in goodness of fit and computational efficiency. One again we want to remark the importance of the paper by Sonja Greven and Fabian Scheipl. Their effort to provide a unified approach to functional regression modelling constitute a nice bridge between multivariate regression and functional data modelling. This, indeed, will attract a large group of researchers which, somehow, are still reluctant to see the potential of functional data modelling.

References

Aguilera, A. M. and Aguilera-Morillo, M. C. (2013). Penalized PCA approaches for B-spline expansions of smooth functional data. *Applied Mathematics and Computation*, **219**, 7805–7819.

Aguilera-Morillo, M., Dutran, M., and Aguilera, A. M. (2016). Prediction of functional data with spatial dependence: a penalized approach. *Stochastic Environmental Research and Risk Assessment*, **doi:10.1007/s00477-016-1216-8**.

Brumback, B. and Rice, J. (1998). Smoothing spline models for the analysis of nested and crossed samples of curves. *Journal of the American Statistical Association*, **93**, 961–994.

Currie, I., Durban, M., and Eilers, P. (2006). Generalized linear array models with applications to multidimensional smoothing. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **68**, 259–280.

Djeundje, V. and Currie, I. (2010). Appropriate covariance-specification via penalties for penalized splines in mixed models for longitudinal data. *Electronic Journal of Statistics*, **2**, 1202–1224.

Durban, M., Harezlak, J., Wand, M., and Carroll, R. (2005). Simple fitting of subject-specific curves for longitudinal data. *Statistics in Medicine*, **24**, 1153–1167.

Eilers, P. H. C. and Marx, B. D. (1996). Flexible smoothing with B-splines and penalties. *Statistical Science*, **11**, 107–121.

Marra, G. and Wood, S. (2012). Coverage properties of confidence intervals for generalized additive model components. *Scandinavian Journal of Statistics*, **39**, 53–74.

Reiss, P. T. and Ogden, R. T. (2007). Functional principal component regression and functional partial least squares. *Journal of the American Statistical Association*, **102**, 984–996.

Rodriguez-Alvarez, M.X.a nd Durban, M., Lee, D.-J.-J., and Eilers, P. H. C. (2016). Fast estimation of multidimensional adaptive p-spline models. Technical report, arXiv preprint.

Rodriguez-Alvarez, M., Lee, D.-J., Kneib, T., Durban, M., and Eilers, P. H. C. (2015). Fast algorithm for smoothing parameter selection in multidimensional generalized p-splines. *Statistics and Computing*, **25**, 241–257.

Scheipl, F. and Greven, S. (2016). Identifiability in penalized function-on-function regression models. *Electronic Journal of Statistics*, **10**, 495–526.

Silverman, B. W. (1996). Smoothed Functional Principal Component Analysis by Choice of Norm. *Annal Statistics*, **24**, 1–24.