Predicting Pregnancy Outcomes Using Longitudinal Information: A Penalized Splines Mixed–Effects Model Approach

February 9, 2017

Rolando De la Cruz[1*], Claudio Fuentes[2], Cristian Meza[3], Dae-Jin Lee[4] and Ana Arribas–Gil[5]

[1]Instituto de Estadística, Pontificia Universidad Católica de Valparaíso, Valparaíso, Chile.
[2]Department of Statistics, Oregon State University, Corvallis, USA.
[3]CIMFAV - Facultad de Ingeniería, Universidad de Valparaíso, Valparaíso, Chile.
[4]BCAM - Basque Centre for Applied Mathematics, Bilbao, Basque Country, Spain.
[5]Departamento de Estadística, Universidad Carlos III de Madrid, Getafe, Spain.

[*] **Corresponding author:** `rolando.delacruz@pucv.cl`

Abstract

We propose a semiparametric mixed–effects model (SNMM) using penalized splines to classify longitudinal data and improve the prediction of a binary outcome. The work is motivated by a study in which different hormone levels were measured during the early stages of pregnancy, and the challenge is using this information to predict *normal* versus *abnormal* pregnancy outcomes. The aim of this paper is to compare models and estimation strategies based on alternative formulations of SNMMs depending on the characteristics of the data set under consideration. For our motivating example, we address the classification problem using a particular case of the SNMM in which the parameter space has a finite dimensional component (fixed effects and variance components) and an infinite dimensional component (unknown function) that need to be estimated. The nonparametric component of the model is estimated using penalized splines. For the parametric component, we compare the advantages of using random effects versus direct modeling of the correlation structure of the errors. Numerical studies show that our approach improves over other existing methods for the analysis of this type of data. Furthermore, the results obtained using our method support the idea that explicit modeling of the serial correlation of the error term improves the prediction accuracy with respect to a model with random effects, but independent errors.

Keywords: Classification models; Correlated observations; Longitudinal data; Mixed–effects models; P–splines.

1 Introduction

The human chorionic gonadotropin beta subunit, β–HCG, increases its concentration levels during the first stages of pregnancy. It is generally used in obstetrics as it is a powerful indicator of pregnancy evolution, particularly in the context of *in vitro* fertilization, where it is used as a prognostic marker to detect ectopic pregnancies and other possible complications [30, 3]. In fact, β–HCG is known as the "pregnant hormone" since it is also used to detect pregnancy, as its level increases fast after fertilization, being detectable by blood and urine tests between 10 and 15 days after conception. The main increment is produced right after the implantation of the fertilized egg to the uterine wall, although the levels of β–HCG will continue rising during first trimester of gestation. Abnormally high or low concentration levels may be an indicator of an anomalous situation. For this reason, it is important to follow the evolution of the concentration levels over time, that is, its longitudinal profile. In a normal pregnancy, the level of β–HCG will double its value every 1.5 days up to 5 weeks after the last menstrual period, and then every 3.5 days starting on the 7th week [15]. After the first trimester, levels should gradually decrease over time and quickly decrease to zero after the pregnancy is ended. Nonetheless, ectopic pregnancies, miscarriages or spontaneous abortions often follow a sudden decrease in β–HCG levels during the first trimester. Other complications, however, are usually preceded by an abrupt rise of hormone concentration.

In this paper, we consider the problem of detection of pathologies during pregnancy based on longitudinal information of hormone concentrations in pregnant women. It seems clear that any attempt to predict an adverse pregnancy condition based on β–HCG levels needs to take into account the dynamics of the concentration over time and not only the absolute values at a given moment. Unfortunately, hormone measurements during the first weeks of pregnancy are only recorded occasionally, and not always at the same stage of pregnancy for every woman, so available records often consist of very sparse and irregularly designed longitudinal data (see Figure 1). In this context, the use of mixed–effects models, in which individual trajectories can borrow information from each other, are extremely useful to model individual and group behaviors, which is the keystone to a successful classification. Here, we approach the classification problem by first considering separate models for each class, and then establishing a classification rule based on the underlying density estimates, using a Bayes classifier. Other approaches for the classification of longitudinal data, in which individuals in all groups are modeled jointly, and estimation and classification are performed simultaneously have been discussed in the literature using a functional data analysis (FDA) approach [33, 34]. The unbalanced longitudinal data can be seen as a set of sparse functional data (i.e., curves observed in few time points that are different among the sample individuals). For instance, [24] considered a functional binary regression model for sparse functional data, and [21] proposed a least squares support vector machine classifier for longitudinal data. Recently, [35] considered the projection of the sparse functional data onto the most effective directions associated with a functional index model using a weighted support vector machine and proposed a cumulative slicing approach to borrow information across individuals.

The pregnancy data set we introduce in Section 2 has been the catalyst for the development of a mixed model–based collection of powerful classification methods for sparse longitudinal data. For instance, [23] used a nonlinear mixed effects (NLME) model approach to describe the evolution in the different groups and produce an optimal allocation rule. The authors showed the necessity of modeling the interaction of time with fixed and random effects in a nonlinear way in order to capture the dynamics of the data set. In this direction, [6] proposed a Bayesian framework for the classification of longitudinal profiles, when the underlying structure in each group or populations can be expressed by nonlinear hierarchical models. In [7], the authors extended these ideas and developed a semiparametric Bayesian approach, in which the distribution of the random effects was estimated using a Dirichlet process mixture prior.

More recently, [2] proposed a semiparametric linear mixed–effects models (SLMM) for the longitudinal trajectories of both groups and the use of a Bayes classifier to predict pregnancy outcomes. For the semiparametric component of the model, a LASSO approach was considered to estimate the function capturing the temporal trend of the data. In contrast to previous approaches, the random effects entered the function linearly, which allowed for the development of an exact type–EM estimation algorithm. In fact, when modeling complex longitudinal data sets such as the one at hand, the use of sophisticated methods for some of the components of the model may condition the way in which the rest of the components are dealt with. However, the flexibility used to estimate the function compensated for the lack of nonlinear interaction, providing a better fit and lower classification errors. It turns out that a balanced combination of simple methods for the different components of a model may deliver better overall results. Although the use of a LASSO–type estimator was partially motivated by the irregularities in the trend of the abnormal pregnancy group, the resulting estimated function was fairly smooth, suggesting that the use of more conventional methods, such as low-rank penalized splines, might be appropriate. Indeed, penalized splines provide a simple and accurate way of fitting smooth functions for longitudinal data. Moreover, because the penalized spline fitting can be written in the form of a linear mixed model (see Section 3), they can be easily incorporated into a mixed–effects model facilitating the relaxation of some unrealistic assumptions, such as the independence of the individual error terms.

The main objectives of this study are: (i) to propose an efficient and parsimonious method for the classification of sparse longitudinal profiles with semiparametric mixed effects models via low-rank penalized splines; (ii) to find the best possible mixed–model for classification of the pregnancy data set; and (iii) to provide guidelines for the choice of mixed models in the context of sparse longitudinal profiles. Specifically, we study the use of the nonparametric methods and the modeling individual effects via random effects and correlated errors.

The rest of the paper is organized as follows: in Section 2 we present a detailed description of the data set that motivates this paper. Section 3 describes the different semiparametric mixed–effects models under consideration and the classification procedure using the specific models studied here. In Section 4 we apply our methodology to the data set of β–HCG concentration during the early stages of pregnancy. We conclude the paper in Section 6 with a short discussion of the methodology and results obtained in this study.

2 Description of the data

The data used in this study were collected during a clinical trial in a private assisted reproduction center in Santiago, Chile. The data set consists of repeated measures of β–HCG concentration levels taken over a period of two years on 173 pregnant women and is divided in two groups: one corresponding to pregnancies with a normal development that went to term without important complications, and a second group including records of abnormal pregnancies with serious anomalies that ended up with the lost of the fetus. From the 173 pregnancies, 124 belong to the first group and the remaining 49 to the second one. The measurements were recorded at different times for each women during the first trimester of pregnancy (first 80 days), and it is well known that the β–HCG concentration levels in the two groups follow different patterns. Figure 1, shows the logarithm of these hormone concentrations. Each curve represents the evolution over time of β–HCG for each woman, and each profile is displayed in its corresponding group of normal or abnormal pregnancies.

The data set is unbalanced, making the analysis particularly challenging. Moreover, the number of observations per subject is very small and the measurement time grid is very irregular. The number of measurements per woman ranges between one and six, with a median of two. In the group of normal pregnancies 28% of the women have only one measurement and almost 98%

of the women have three or less measurements. These percentages are 35% and 86% respectively in the group of abnormal pregnancies. Furthermore, the times at which the measurement were taken show a large variability among subjects; and the time between two consecutive measurements for the same subject exhibits a variability that goes from 2 up to 51 days. See Figure 2 for a summary of these features.

Looking at the profiles, we observe that the group of abnormal pregnancies show an erratic behaviour difficult to model. Even using adaptive techniques for function approximation (such as LASSO), the fitting is poor and the classification does not improve (significantly) with respect to other conventional techniques [2]. Finally, unlike other studies of similar nature, no covariates are available for this data. In that sense, we address a pure longitudinal classification problem, and our methods could be directly applied or extended to frameworks in which only longitudinal profiles are available.

3 Model formulation

Suppose that a training data set consists of N individuals and observations $\{(y_i, \omega_i),\ i = 1, \ldots, N\}$, where $y_i = (y_{i1}, \ldots, y_{in_i})' \in \mathbb{R}^{n_i}$ is the response vector for the ith individual, taken at arbitrary times $t_i = (t_{i1}, \ldots, t_{in_i})'$ and $\omega_i \in \{1, 2, \ldots, G\}$ denotes the true (known) group or population label for the ith individual. For instance, in the pregnancy data set, $N = 173$, $1 \leq n_i \leq 6$ and $G = 2$, with $\omega_i = 1$ for the normal pregnancy group, and $\omega_i = 2$ for the abnormal group. Notice that the label ω_i is known for some women with already reported delivery, but unknown for women with partial data before delivery. Then, without any loss of generality, we assume that ω_i, $i = 1, \ldots, N$, is known, and ω_{N+1} is unknown. Finally, denote by $y^N = (y_1 \ldots, y_N, \omega_1, \ldots, \omega_N)$ the training data set, including the recorded class memberships ω_i up to the Nth subject.

Let $q_g(y_i; \theta_g)$ be the probability density function (pdf) for the response vector y_i in group g. That is, $y_i|\omega_i = g \sim q_g(y_i; \theta_g)$, where θ_g is a corresponding set of parameters. Then, the classification problem can be approached either from a within–sample or a predictive perspective. In the first case, once we obtain an estimate $\hat{\theta}$ of θ, we aim at estimating posterior probabilities $\{p(\omega_i = g|y^N, \theta);\ g = 1, \ldots, G\}$, where $\theta = (\theta_1, \theta_2, \ldots, \theta_G)$. For the predictive classification approach, we assume that a new subject $N + 1$ with unknown label ω_{N+1} is recorded, so that the problem of interest is the prediction of ω_{N+1}. In this context, we are interested in $\{p(\omega_{N+1}|y^N, y_{N+1}, \theta)\}$, where y_{N+1} is the currently available partial response vector for the new individual $N + 1$.

Most of the models for serial measurements can be described as full multivariate models or multi–stage mixed–effects models. For the full multivariate model, the vector of responses y_i, within the gth group, is assumed to be multivariate normal with mean μ_{ig} $(n_i \times 1)$ and an arbitrary $n_i \times n_i$ dispersion matrix V_{ig}. The mean vector may depend upon the pattern of observations and also upon covariates.

If the longitudinal design is balanced (i.e., $n_1 = n_2 = \cdots = n_N$), but observations are missing at random, traditional multivariate discriminant analysis based on the full multivariate model can be easily applied using multivariate methods for missing observations [8, 29]. However, when the individuals are measured at arbitrary or unique times, or when the dimension of V_{ig} is large, this approach becomes unattractive, since a full multivariate model with unrestricted dispersion matrix requires a proliferation of variance parameters, many of which will be poorly estimated. In addition, the full multivariate model does not permit the definition and estimation of (random) individual–specific characteristics [19].

Two–stage mixed–effects models are based on explicit identification of individual–specific and group or population characteristics, and their form extends naturally to the unbalanced situation. Most of the methods for longitudinal data focus on data that can be represented using

a function for the mean that is linear in its parameters [19]. However, in many situations, we are concerned with data for which the assumption of normal errors is plausible, but the proposed mean function is nonlinear. When parametric mixed–effects models are not flexible enough for complex longitudinal data, semiparametric mixed–effects model offer an interesting extension, since they are a good compromise of both a parametric and nonparametric approach. In the next section we describe the mixed–effects models considered to analyze complex longitudinal data like the pregnant women data set.

3.1 Semiparametric nonlinear mixed–effects model

For each group g, we can consider the following general semiparametric nonlinear mixed–effects model (SNMM) proposed by [17]

$$y_{ij} \quad = \quad \eta(\mathrm{x}_{ij}, \phi_{ig}, f_g) + \epsilon_{ijg}, \quad i = 1, \ldots, N, \quad j = 1, \ldots, n_i, \quad g = 1, \ldots, G, \tag{1}$$

where $y_{ij} \in \mathbb{R}$ is the jth observation in the ith individual, $\mathrm{x}_{ij} \in \mathbb{R}^d$ is a vector of known regression variables, η is a common known function governing within–individual behavior and f_g is an unknown nonlinear function to estimate using the training data set. They assume that η is linear in f_g conditional on ϕ_{ig},

$$\eta(\mathrm{x}_{ij}, \phi_{ig}, f_g) = a(\phi_{ig}; \mathrm{x}_{ij}) + b(\phi_{ig}; \mathrm{x}_{ij}) f_g(c(\phi_{ig}; \mathrm{x}_{ij})), \tag{2}$$

where a, b and c are known parametric functions which may depend on i. For the error term, we assume $\epsilon_{ig} = (\epsilon_{i1g}, \ldots, \epsilon_{in_ig})' \sim \mathrm{MVN}(0, \sigma_g^2 I_{n_i})$, a multivariate normal with mean vector 0 and covariance matrix $\sigma_g^2 I_{n_i}$, where I_a denotes the identity matrix of size a. The vector of random effects ϕ_{ig} satisfies

$$\phi_{ig} = A_i \beta_g + U_{ig}, \quad U_{ig} \sim \mathrm{MVN}_p(0, \Gamma_g), \tag{3}$$

where, $A_i \in \mathcal{M}_{p,q}$ is a known design matrix, $\beta_g \in \mathbb{R}^q$ is the unknown vector of fixed effects, and ϵ_{ig}, U_{ig} are mutually independent. Here, q represents the number of fixed effects parameters.

The model parameter is the pair (θ_g, f_g), where $\theta_g = (\beta_g, \Gamma_g, \sigma_g^2)$ belongs to a finite dimensional space, whereas f_g lies in an infinite dimensional space of functions $\mathcal{H}$. Notice that if f_g is known, we fall into the classical NLME model.

From the general formulation of the SNMM, we can consider different submodels. If estimation is the main objective, it is important to analyze the smoothing (or modeling) of f_g, the way in which the random effects interact with f_g, and the error structure. Moreover, these three components cannot be addressed independently, since, for instance, a given smoothing method may not be able to cope with nonlinear interaction of the random effects or correlated error terms. It is then important to look for a compromise between the properties of the models for each component and the overall properties of the global model. Next, we analyze each of these components and their interaction, and propose a model for the pregnancy data set.

Semiparametric linear mixed–effects model Recently, a special case of model (1) was studied in [2], leading to the following semiparametric linear mixed–effects model (SLMM)

$$y_{ij} = f_g(t_{ij}) + \mathrm{x}_{ij}' \beta_g + U_{ig} + \epsilon_{ijg}, \quad i = 1, \ldots, N; \quad j = 1, \ldots, n_i, \quad g = 1, \ldots, G, \tag{4}$$

where $t_{ij} \in \mathbb{R}$ is a predictor with a possibly nonlinear effect and $\mathrm{x}_{ij} \in \mathbb{R}^d$ is a vector of predictors with a linear effect, with corresponding coefficient vector β_g. The authors assume $U_{ig} \sim$ iid $N(0, \tau_g^2)$ and $\epsilon_{ijg} \sim$ iid $N(0, \sigma_g^2)$, independent of the U_{ig}. Note that in (4) U_{ig} is a random intercept. In [2], the main argument to prefer this model to (1) was that the linear interaction between f_g and the random effects allowed for the implementation of an exact EM–type algorithm for the estimation of all the parameters of the model, when f_g is estimated through LASSO with a dictionary approach. This model also allows the implementation of accurate estimation algorithms when other smoothing methods are chosen.

Correlated errors Typically, the error terms ϵ_{ig} in the gth group or population are assumed to be independent over time [17, 1, 2]. However, in longitudinal data, measurements taken over time on individuals usually show a highly unbalanced structure (i.e., measurement times may be unequally spaced within an individual and may differ across individuals) and may be serially related. This can be considered explicitly by taking $\epsilon_{ig} \sim \text{MVN}_{n_i}(0, \Sigma_{ig})$, with $\Sigma_{ig} = \Sigma_i(\sigma_g^2, \rho_g)$, where σ_g^2 is a scalar parameter, and ρ_g a vector of parameters describing the correlation structure. Depending on the context, different assumptions about the matrix $\Sigma_i(\sigma_g^2, \rho_g)$ can be considered [31].

3.2 Low-rank penalized splines as mixed–effects models for longitudinal data

As an alternative to the semiparametric nonlinear mixed–effects model in Section 3.1, in this section we propose to use low-rank penalized splines and its mixed model representation to estimate the unknown function f_g in (4). The main advantage of low-rank penalized splines is that the number of basis functions does not grow with the sample size. Moreover, the mixed–effects model representation also allows for the extension to multilevel models, longitudinal data and correlated errors, where the estimation of the amount of smoothness and the correlation in the errors can be done simultaneously [4, 9]. In this paper, we consider two popular approaches: i) penalized splines as mixed models using truncated power basis functions (TPF–Splines) with ridge penalties for the coefficients [27] and ii) penalized splines (P–splines) coined by [12] based on B–spline bases with discrete difference penalties on adjacent coefficients using the mixed model reparameterization in [4, 20]. The main difference of both approaches are: i) the regression basis chosen ii) the type of penalization on the regression coefficients. The performance of both approaches was discussed in [13, 14]. Here, we compare both approaches for unbalanced longitudinal data modelling as in [10] with particular focus on predicting binary outcomes. Let us consider the function f_g in (4) can be written as a penalized spline, i.e.

$$f_g(t_{ij}) = \alpha_{0g} + \alpha_{1g}t_{ij} + \cdots + \alpha_{pg}t_{ij}^p + \sum_{k=1}^{K} u_{kg}z_{kg}(t_{ij})^p, \tag{5}$$

where α_{0g} is an intercept term, $\alpha_{1g}t_{ij} + \cdots + \alpha_{pg}t_{ij}^p$ is a polynomial of degree p over time t_{ij} and z_{kg}, $1 \leq k \leq K$, is an appropriate spline basis. In compact matrix form, for each group g model (4) can be written as

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{Z}\boldsymbol{u} + \boldsymbol{\epsilon}\ , \text{ with } \boldsymbol{u} \sim \text{N}(0, \boldsymbol{G}) \text{ and } \boldsymbol{\epsilon} \sim \text{N}(0, \boldsymbol{R}), \tag{6}$$

where $\boldsymbol{G}$ is the covariance of the random effects $\boldsymbol{u}$, and $\boldsymbol{R} = \Sigma(\sigma^2, \rho)$ a correlation matrix for the errors $\boldsymbol{\epsilon}$. Hence, given N individuals in group g, the rest is defined as $\boldsymbol{y} = (y_1', y_2', \ldots, y_m')$, with $y_i = (y_{i1}, y_{i2}, \ldots, y_{in_i})'$, with fixed effects matrix $\boldsymbol{X}$ defined as:

$$\boldsymbol{X} = \begin{pmatrix} X_1 \\ \vdots \\ X_N \end{pmatrix}, \text{with } X_i = \begin{pmatrix} 1 & t_{i1} & \cdots & t_{i1}^p \\ 1 & t_{i2} & \cdots & t_{i2}^p \\ \vdots & \cdots & \ddots & \vdots \\ 1 & t_{iN} & \cdots & t_{iN}^p \end{pmatrix}, \text{of size } n_i \times (p+1)$$

and random effects matrix $\boldsymbol{Z}$ given by

$$\boldsymbol{Z} = \begin{pmatrix} Z_1 & \mathbf{1}_1 & 0 & \cdots & 0 \\ Z_2 & 0 & \mathbf{1}_2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ Z_N & 0 & 0 & \cdots & \mathbf{1}_N \end{pmatrix}, \text{with } Z_i = \begin{pmatrix} z_{1,i_1} & \cdots & z_{K,i_1} \\ z_{1,i_2} & \cdots & z_{K,i_2} \\ \vdots & \ddots & \vdots \\ z_{1,i_{n_i}} & \cdots & z_{K,i_{n_i}} \end{pmatrix}, \text{ of size } n_i \times K$$

where $\mathbf{1}_i$ is a column vector of ones of length n_i, for $i = 1, ..., m$. The fixed effects are $\boldsymbol{\beta} = (\beta_0, \beta_1, ..., \beta_p)'$ and the vector of random effects $\boldsymbol{u} = (u_1, ..., u_k, U_1, ..., U_m)'$ has covariance matrix

$$\boldsymbol{G} = \text{Cov}(\boldsymbol{u}) \sim \text{MVN} \begin{pmatrix} \sigma_u^2 \boldsymbol{I}_K & 0 \\ 0 & \sigma_U^2 \boldsymbol{I}_N \end{pmatrix}.$$

The penalized spline smoother of model (4) corresponds to the optimal predictor in a mixed model framework assuming a given smoothing parameter $\lambda = \sigma^2/\sigma_u^2$, where $0 < \lambda < \infty$. Hence, for a small λ the solution will be equivalent to ordinary least squares, and for $\lambda \to \infty$ the strong penalty forces the fit to the polynomial (fixed effect) term $\boldsymbol{X}\hat{\boldsymbol{\beta}}$. Estimation is usually done using restricted maximum likelihood (REML) [25], which provides an advantage when the correlation is unknown [18].

TPF–Splines For each g group, let be $\kappa_1, ..., \kappa_K$ be a set of distinct knots in the range of t_{ij} and $t_+ = \max(0, t)$. The number K of knots is chosen and fixed to ensure enough flexibility of the fitted curve. Usually, a simple default choice of K based only in the sample size is $\min(1/4 \times \text{number of unique} t_i, 35)$. The usual criteria to choose the position of the knots is by given fixed quantiles for the covariate t. However, there are times where one needs a more sophisticated algorithm that uses the data to the number K of knots and its position. TPF–Splines choice for the random effects matrix $\boldsymbol{Z}$ is $z_k(v)^p = (v - \kappa_k)_+^p$. The function $(v - \kappa_k)_+^p$ has $p - 1$ continuous derivatives. Hence, higher values of p lead to smoother spline functions. Smoothness is controlled by a ridge penalty on the size of the u_k coefficients, i.e. via $\lambda \boldsymbol{u}'\boldsymbol{u}$. The book by [27] popularized this low-rank penalized splines (for a complete review see [28]). The function `lme` in the `R` library `nlme` allows for the estimation of the model specifying the argument `method="REML"` (see [26]).

P–Splines P-splines assumes a given smooth function f of a covariate t as $f(t_i) = \sum_{l=1}^{L} a_l B_l(t_i)$ for $i = 1, ..., m$, and $l = 1, ..., L$. The regression matrix is formed by a low-rank matrix of B-splines, i.e. $\boldsymbol{B} = B_l(t_i)$ for $l = 1, ..., L$, where L are the number of regression coefficients $\boldsymbol{a} = (a_1, ..., a_L)'$ much smaller than the number of observations. Smoothness is controlled by a discrete difference penalty on the regression coefficients a, i.e. $\Delta a_l = a_l - a_{l-1}$, $\Delta^2 a_l = \Delta(\Delta a_l) = a_l - 2a_{l-1} + a_{l-2}$. In general, a second order penalty is used (i.e. $d = 2$), hence the straight line is fitted for a large smoothing parameter λ. In matrix form D_d is a difference matrix operator such that $\boldsymbol{D}_d\boldsymbol{a} = \Delta^d\boldsymbol{a}$. The penalty is then of the form $\lambda \boldsymbol{a}'\boldsymbol{P}\boldsymbol{a}$, where $\boldsymbol{P} = \boldsymbol{D}'\boldsymbol{D}$, of size $L \times L$. The penalty matrix $\boldsymbol{P}$ is rank deficient, i.e. $\text{rank}(\boldsymbol{P}) = L - d$, indeed d eigenvectors of $\boldsymbol{P}$ are zero. A solution is to use mixed model representation and rewrite $\boldsymbol{B}\boldsymbol{a}$ as $\boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{Z}\boldsymbol{u}$, such that $\boldsymbol{X}$ has d columns that expand the polynomial null space of $\boldsymbol{P}$ (the unpenalized part) and the $(L - d)$ columns of $\boldsymbol{Z}$ (penalized part) expands its complementary space. This reparameterization is done in different ways, for instance [11] proposed $\boldsymbol{Z} = \boldsymbol{B}\boldsymbol{D}'(\boldsymbol{D}\boldsymbol{D}')^{-1}$ and [20] a reparameterization based on the singular value decomposition of the penalty matrix $\boldsymbol{P}$, $\boldsymbol{D}'\boldsymbol{D} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}'$, such that $\boldsymbol{Z} = \boldsymbol{B}\boldsymbol{U}_s\tilde{\boldsymbol{\Sigma}}^{-1/2}$, where $\boldsymbol{U}_s$ are the column of $\boldsymbol{U}$ containing the non-null eigenvectors, and $\tilde{\boldsymbol{\Sigma}}$ is a diagonal matrix of the non-null eigenvalues of $\boldsymbol{\Sigma}$. In both reparameterizations the random effects $\boldsymbol{u}$ have covariance $\boldsymbol{G} = \sigma_u^2 \boldsymbol{I}_K$, and hence the penalty on the random effects also becomes $\lambda \boldsymbol{u}'\boldsymbol{u}$. The model can be fitted by constructing the $\boldsymbol{X}$ and $\boldsymbol{Z}$ matrices as shown or equivalently by means of the function `gamm` with the arguments `bs="ps"`, and `method="REML"` in the `R` library `mgcv` (see [32]).

Penalized splines with correlated errors The mixed model representation of a penalized splines has the great advantage of handling correlated errors very naturally [9, 18]. Indeed, a correlation structure can be considered in the matrix $\boldsymbol{R}$ defined in (6). In our case, for the

within–individual error, we propose to use a continuous time first order process, CAR(1), in order to address the non equally spaced measurements. A CAR(1) process is similar to a discrete time AR(1) process in that its correlation function decreases exponentially with increasing time separation. For a time interval of δt, the correlation is

$$\rho(\delta t) = \rho^{\delta t},$$

where $0 \leq \rho \leq 1$ and $\delta = |i - j|$ is the difference between ϵ_i and ϵ_j. The `R` library `nlme` allows for this type of correlation structure through the function `corCAR1()` (see [26, p. 239]).

In the rest of the paper, we refer as "TPF–splines" to the model with truncated power functions, and "P–splines" to the model with basis from the mixed models reparameterization of B-splines respectively. Both approaches are combined to include a random intercept (U_i) and/or a first order continuous autoregressive process CAR(1).

3.3 The Classification Problem

In this section we consider a generic individual with response vector $y = (y_1, y_2, \ldots, y_n)'$ taken at arbitrary times $t = (t_1, t_2, \ldots, t_n)'$ and thus we simplify the notation by dropping the subindex i. The goal of the classification problem is to allocate an individual into one of g groups or populations on the basis of the observations y, the time of these observations t, a vector of covariates x and the distribution of y in the G groups or populations. Assume that the vector of responses y in the l group or population has pdf $q_l(y; \theta_l)$, where θ_l is the set of parameters associated with this distribution. If we assume that $\pi_1, \pi_2, \ldots, \pi_G$ are the prior probabilities of group or population membership, the Bayes rule for allocation classifies y to population g if

$$\log \pi_g + \log q_g(y; \theta_g) = \max_l \{\log \pi_l + \log q_l(y; \theta_l)\}, \quad l = 1, 2, \ldots, G, \tag{7}$$

where $\pi_g q_g(y; \theta_g)$ is proportional to the posterior distribution of membership in group or population g, i.e,

$$g = \arg\max_{l=1,..,G} \{\log \pi_l + \log q_l(y; \theta)\}.$$

From (6) the conditional distribution of the response vector $y_i = (y_{i1}, y_{i2}, \ldots, y_{in_i})'$ given the random effects u_i in group or population g is given by

$$y_i | u_{kg} \sim \text{MVN}_{n_i}(X_i\beta_g + Z_{ig}u_{ig}, V_{ig}), \text{ where } V_{ig} = \tau_g^2 \mathbf{1}\mathbf{1}_i' + \sigma_g^2 \Sigma_i(\rho_g), \tag{8}$$

that is, q_g has a multivariate normal distribution. Let μ_l and Ψ_l denote the vector mean and the covariance matrix of the multivariate normal distribution (8) respectively. In classical discriminant analysis the Mahalanobis distance plays a central role in both the conceptual framework and the allocation rules. The squared Mahalanobis distance between the response vector y and the mean of the distribution of group or population l, μ_l, with respect to Ψ_l is $D_l(y) = (y - \mu_l)'\Psi_l^{-1}(y - \mu_l)$, and the classification rule is to allocate y to group or population g if $\varphi_{lg}(y) \leq 0$ for $l = 1, \ldots, G$ and $l \neq g$, where

$$\varphi_{lg}(y) = D_g^*(y) - D_l^*(y) + 2 \log \frac{\pi_l}{\pi_g}, \tag{9}$$

with $D_c^*(y) = D_c(y) + \log |\Psi_c|$. Note that, although the mixed-effects model in (6) specifies the response vector, $\boldsymbol{y}$, conditionally on a vector $\boldsymbol{u}$ of random effects, classification is based on the marginal distribution obtained from integrating over the random effects. In practice, the quantity in (9) is replaced by an estimate based on the parameter estimates obtained in the mixed-effects model.

4 Analysis of the Pregnancy Data

Let $y_i = (y_{i1}, \ldots, y_{in_i})'$ be the observed $\log_{10}(\beta$–HCG) measurements at time $t_i = (t_{i1}, \ldots, t_{in_i})'$ for woman $i = 1, \ldots, N = 173$. We compare the following models with both independent and correlated errors:

- nonlinear mixed effects model with a parametric logistic function (NLME) as in [23], i.e.

$$y_{ij} = \frac{\phi_{i1g}}{1 + \exp\{(\beta_{2g} - t_{ij})/\beta_{3g}\}} + \epsilon_{ijg}, \quad i = 1, \ldots, N, \quad 1 \leq j \leq n_i, \quad g = 1, 2, \qquad (10)$$

  where $\phi_{i1g} \sim N(\beta_{1g}, \tau_g^2)$. With independent errors ("NLME") or CAR(1) errors ("NLME+CAR1").
- Penalized splines models in (6):
  - TPF–splines and P-splines.
  - TPF–splines/P–splines + Individual Random intercept (U_{ig}), where $U_{ig} \sim N(0, \sigma_{ug}^2)$ ("TPF–splines+Ui" and "P–splines+Ui").
  - TPF–splines/P–splines + Individual Random intercept (U_{ig}) with correlated CAR(1) errors ϵ_{jig} ("TPF–splines+Ui+CAR1" and "P–splines+Ui+CAR1").

For the correlated errors models we consider a CAR(1) structure, as described in Section 3. Additionally, we fit the nonlinear model (10) with uncorrelated and correlated errors, but without random effects. We call them "NLM" and "NLM + CAR1", respectively.

The ten resulting models are fitted to both groups of the pregnancy data set and a model selection criterion is used to compare them in terms of fitting accuracy. Table 1 shows the log likelihood (logLik), the Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) for the all models with independent and correlated errors. We observe that for both groups, model selection criteria consistently favor those models that explicitly consider a correlation structure for the errors. Among those models fitted using penalized splines, P–splines have a better performance than TPF–splines in all comparisons. This is due to the fact that for TPF–splines, we used the position of the knots based on the quantiles of the unique values of the time measurement (days of pregnancy). Figure 3 illustrates the difference in the basis functions design for TPF and P–splines for both normal and abnormal groups.

We can also observe that the incorporation of a random intercept to TPF/P–splines models does not improve their performance, in fact, the estimated variance is very close to zero ($\hat{\sigma}_U^2 \approx 10^{-5}$) when correlation errors are considered. Indeed, the within-individuals correlation parameter estimate $\hat{\rho}$ is very high and greater than $0,90$ when CAR(1) structures are included. Figure 4 shows the fitted smooth curves by "P–splines + CAR1" model for both groups in the same plot. The differences between the two groups mean functions estimates are evident. The corresponding 95% confidence intervals of the fitted curves $\hat{f}$ were computed as in [27] (for further details see also [32, Chapter 4] and [22]). It is also interesting noticing that Figure 4 shows that both mean curves differs from approximately day 21 (see vertical line) and how the confidence intervals are wider in the abnormal group (due to the higher variability).

With respect to classification problem, Figure 5 shows the Receiver Operating Characteristic (ROC) curves and the area under the curve (AUC) measures for the penalized splines (TPF/P–splines) and NLME models with independent and correlated errors. We see how the "P–splines + CAR1" model outperforms the other models (with an AUC of 0.932). As a result, this classifier has a very good performance, as the true positive rate increases very quickly. In addition, it is noticeable that TPF–splines has no much better performance with respect to NLME models. Additionally, Figure 6 shows estimated classification probabilities in the normal group for all 173 women, arranged by true group, and within each group sorted in decreasing order. We see how the proposed "P–splines + CAR1" model dominates the parametric model with correlated errors ("NLME+CAR1") in the sense of obtaining higher probabilities for abnormal pregnancies (i.e., a lower area below the abnormal group probability of classification in the normal group).

5 Simulation study

In order to compare the performance of the penalized splines models (TPF–splines versus P–splines), we conducted a simulation study based on data sets generated to mimic longitudinal pregnancy data for both normal and abnormal groups. We consider 100 samples for each normal and abnormal groups in two scenarios:

- **Scenario 1.** *("Dense design")* using the same time points as in the original data set and simulating a mean smooth function estimated with a CAR(1) correlation with $\phi = 0.90$. Figure 7 shows one sample of this design for both normal and abnormal groups.
- **Scenario 2.** *("Sparse design")*, with randomly chosen time points of the original data set and simulating a mean smooth function estimated with a CAR(1) correlation with $\phi = 0.90$. Figure 8 shows one sample of this design, where the simulated response is more similar to the real pregnancy data in comparison to Scenario 1.

For both scenarios, we fitted penalized spline models with correlated CAR(1) errors, for TPF–splines we use the default criteria for choosing the number K of knots (i.e. $\min(1/4\times\text{Number of unique time points}, 35)$, which in this case is chosen as $K = 14$), for P–splines we choose a relatively large number of B–spline basis functions ($L = 40$). Table 2, show the average classification results obtained with each model considering for the classification of each individual a leave-one-out training set. Note that for both scenarios, P–splines performance is better than TPF–splines, in particular for sparse designs (Scenario 2), the misclassification performance is much lower with P–splines. The main reason is that for sparse designs, the use of TPF–splines with K knots chosen by sample quantiles of the unique time points leads to poor coverage of the full range of the time covariate (as illustrated in Figure 3), even if increasing the number K, the knots will concentrate in the dense part of the time measurements (not shown). Algorithms for the optimal choice of the number and positions of knots in TPF–splines are discussed in [27, Section 5.6]. In contrast, using equally spaced knots with difference order penalties (P–splines) results for both sparse and dense designs are much better (we also decreased the number of basis functions to $L = 20$ leading to similar results). As expected, the difference between TPF–splines and P–splines is not very dramatic in Scenario 1 where dense designs were generated (knots positions using TPF—splines are evenly spaced). But, when the design is sparse (particularly as in pregnancy data for days > 51), the effect of the selection of the knots in TPF–splines for prediction performance is at least remarkable in comparison to P–splines.

6 Discussion

In this paper we discuss the classification problem of unbalanced (or sparse) longitudinal data, and we focus on the modeling of each class behavior, using a Bayes classifier for the determination of the discriminant rule between classes. The proposed method uses a penalized splines as mixed–effects model representation for the nonlinear mean function of each group and both random intercept per individual and correlated errors are estimated simultaneously. We compared two popular penalized splines as mixed–effects approaches: TPF–splines (with truncated polynomials as basis functions and ridge penalties on the random effects) as in [27] and P–splines (with B–splines basis function with discrete order difference penalties on the regression coefficients) reparameterized as a mixed model [4, 20].

We analyze a data set that exhibits highly unbalanced longitudinal measurements, with many individuals having only of one observation, and serves as motivation for other case studies in which the availability of dense and regularly spaced observations is not possible. We compare the combination of different components within the mixed–effects model described including nonlinear mixed–effects with a parametric logistic function. We conclude that the effect of

adding a random intercept can be coped with by considering time correlation between individual errors. In fact, for those models with correlated error terms, adding an random effect to model individual intercept did not affect the classification performance and showed little improvement in terms of model selection criteria. The question of the confounding effect of serial correlation and random effects is not a new one. The author in [16] already advised to take this phenomenon into account, especially when dealing with unequally spaced observations. In our example, the variability captured by the error structure is enough to model individual deviations with respect to a common mean function that is well captured through penalized splines smoothing. This simple model provides better classification results on this particular data set than other alternatives that do not allow for the incorporation of correlated structure in the errors. We also performed a simulation study, to show that for sparse designs P–splines outperforms TPF–splines when default criteria for knots selection is used. Our results suggests that the use of models for unbalanced longitudinal in which the different components can be combined and embedded into a general flexible model using P–splines are more appropriate. Further extensions of this approach would consider avoiding the two-step procedure in a joint model as proposed in [5].

An interesting problem in this type of analysis consists of investigating the effects of informative censoring on the parameter estimates and classification. Although the data from our motivating example did not appear to exhibit informative censoring, the general setting from which the study arose could have led to shorter follow-up times on average for the women with abnormal pregnancy outcomes. This topic, however, is beyond the scope of the present study and is left for future research.

Acknowledgements

The first author was supported by: Proyecto DI Regular 37.0/2016 VRIEA PUCV; FONDECYT grants 1120739, 1130233 and 1141256; Ministerio de Economía y Competitividad and FEDER, Spain, grants MTM2013–40941–P and MTM2016–74931–P; FONDAP grant 15130011; and grant 16–MATH–03 SIDRE from CONICYT MATH–AmSud Program. The third author was supported by: FONDECYT grant 1141256; and grant 16–MATH–03 SIDRE from CONICYT MATH–AmSud Program. The fourth author was supported by the Basque Government through the BERC 360 2014-2017 and the Department of Education, Language Policy and Culture of the Basque Government IT-620-13 programs and Basque Government Industry Department under the ELKARTEK Program, by the Spanish Ministry of Economy and Competitiveness MINECO: BCAM Severo Ochoa excellence accreditation SEV-2013-0323 and MTM2014-52184-P, and by the RISE Horizon 2020 European Project GEAGAM (644602). The fifth author was funded by grants MTM2014–56535–R and ECO2015–66593–P of the Ministerio de Economía y Competitividad, Spain. Contributions of the second, fourth and fifth authors were partially supported by grants FONDECYT 1120739 and 1141256 while visiting PUCV and CIMFAV–UV.

References

[1] Arribas-Gil, A., Bertin, K., Meza, C., and Rivoirard, V. (2014). Lasso-type estimators for semiparametric nonlinear mixed-effects models estimation. *Statistics and Computing*, 24(3):443–460.

[2] Arribas-Gil, A., De la Cruz, R., Lebarbier, E., and Meza, C. (2015). Classification of longitudinal data through a semiparametric mixed-effects model based on lasso-type estimators. *Biometrics*, 71(2):333–343.

[3] Confino, E., Demir, R. H., Friberg, J., and Gleicher, N. (1986). The predictive value of hcg beta subunit levels in pregnancies achieved by in vitro fertilization and embryo transfer: an international collaborative study. *Fertility and Sterility*, 45:526–531.

[4] Currie, I. D. and Durban, M. (2002). Flexible smoothing with P-splines: a unified approach. *Statistical Modelling*, 2(4):333–349.

[5] De la Cruz, R., Meza, C., Arribas-Gil, A., and Carroll, R. J. (2016). Bayesian regression analysis of data with random effects covariates from nonlinear longitudinal measurements. *Journal of Multivariate Analysis*, 143:94–106.

[6] De la Cruz-Mesía, R. and Quintana, F. A. (2007). A model–based approach to Bayesian classification with applications to predicting pregnancy outcomes from longitudinal β–hCG profiles. *Biostatistics*, 8:228–238.

[7] De la Cruz-Mesía, R., Quintana, F. A., and Müller, P. (2007). Semiparametric Bayesian classification with longitudinal markers. *Journal of the Royal Statitical Society, Series C*, 56:119–137.

[8] Dempster, A. E., Laird, N. M., and Rubin, D. B. (1977). Maximum likelihood with incomplete data via the E-M algorithm. *Journal of the Royal Statistical Society, Series B*, 39:1–38.

[9] Durbán, M. and Currie, I. D. (2003). A note on P-spline additive models with correlated errors. *Computational Statistics*, 18:251–262.

[10] Durbán, M., Harezlak, J., Wand, M., and Carroll, R. J. (2005). Simple fitting of subject-specific curves for longitudinal data. *Statist. Med.*, 24(8):1153–67.

[11] Eilers, P. H. C. (1999). Discussion on: the analysis of designed experiments and longitudinal data by using smoothing splines. *Journal of the Royal Statistical Society, C - Applied Statistics*, 48:307–308.

[12] Eilers, P. H. C. and Marx, B. D. (1996). Flexible smoothing with B-splines and penalties. *Statist. Sci.*, 11(2):89–121.

[13] Eilers, P. H. C. and Marx, B. D. (2010). Splines, knots and penalties. *Wiley Interdisciplinary Reviews: Computational Statistics*, 2:637–653.

[14] Eilers, P. H. C., Marx, B. D., and Durbán, M. (2015). Twenty years of P-splines. *SORT*, 39(2):149–186.

[15] Frits, M. A. and Guo, S. M. (1987). Doubling time of human chorionic gonadotropin (hcg) in early normal pregnancy: relationship to hcg concentration and gestational age. *Fertility and Sterility*, 47:584–589.

[16] Jones, R. H. (1990). Serial correlation or random subject effects. *Communications in Statistics - Simulation and Computation*, 19(3):1105–1123.

[17] Ke, C. and Wang, Y. (2001). Semiparametric nonlinear mixed-effects models and their applications (with discussion). *Journal of the American Statistical Association*, 96(456):1272–1298.

[18] Krivovoboka, T. and Kauermann, G. (2007). A note on penalized spline smoothing with correlated errors. *Journal of American Statistical Association*, 102(480):1328–1337.

[19] Laird, N. M. and Ware, J. H. (1982). Random-effects models for longitudinal data. *Biometrics*, 38:963–974.

[20] Lee, D.-J. and Durbán, M. (2009). Smooth-car mixed models for spatial count data. *Computational Statistics and Data Analysis*, 53(8):2958–2979.

[21] Luts, J., Molenberghs, G., Verbeke, G., Van Huffel, S., and Suykens, J. A. K. (2012). A mixed effects least squares support vector machine model for classification of longitudinal data. *Computational Statistics and Data Analysis*, 56:611–628.

[22] Marra, G. and Wood, S. N. (2011). Coverage properties of confidence intervals for generalized additive model components. *Scandinavian Journal of Statistics*, 39(1):53–74.

[23] Marshall, G. and Barón, A. E. (2000). Linear discriminant models for unbalanced longitudinal data. *Statistics in Medicine*, 19:1969–1981.

[24] Müller, H.-G. (2005). Functional modelling and classification of longitudinal data. *Scandinavian Journal of Statistics*, 32:223–240.

[25] Patterson, H. D. and Thompson, R. (1971). Recovery of inter-block information when block sizes are unequal. *Biometrika*, 58(545–554).

[26] Pinheiro, J. C. and Bates, D. M. (2000). *Mixed-effects models in S and S-PLUS*. Springer, New York.

[27] 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.

[28] Ruppert, D., Wand, M. P., and Carroll, R. J. (2009). Semiparametric regression during 2003-2007. *Electronic Journal of Statistics*, 3:1193–1256.

[29] Schafer, J. L. (1997). *Analysis of Incomplete Multivariate Data*. Chapman and Hall, London.

[30] Shepherd, R. W., Patton, P. E., Novy, M. J., and A., B. K. (1990). Serial beta-hcg measurements in the early detection of ectopic pregnancy. *Obstetrics and Gynecology*, 75:417–420.

[31] Vonesh, E. F. and Chinchilli, V. M. (1997). *Linear and Nonlinear Models for the Analysis of Repeated Measurements*. Marcel Dekker, New York.

[32] Wood, S. N. (2006). *Generalized Additive Models - An introduction with R*. Texts in Statistical Science. Chapman & Hall.

[33] Yao, F., Müller, H. G., and Wang, J. L. (2005a). Functional data analysis for sparse longitudinal data. *J. Am. Statist. Assoc.*, 100:577–590.

[34] Yao, F., Müller, H. G., and Wang, J. L. (2005b). Functional linear regression analysis for longitudinal data. *Annals of Statistics*, 33:2873–2903.

[35] Yao, F., Wu, Y., and Zou, J. (2016). Probability-enhanced effective dimension reduction for classifying sparse functional data. *Test*, 25(1):1–22.

List of Figures

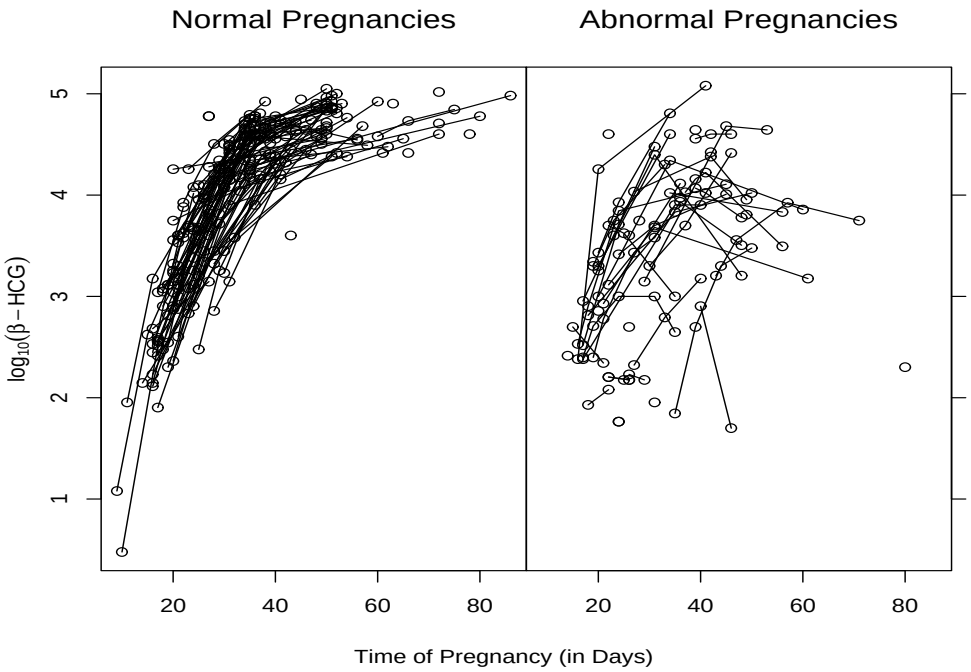

Figure 1: Observed profiles of $\log_{10}(\beta - \text{HCG})$ for normal (left panel) and abnormal groups (right panel).

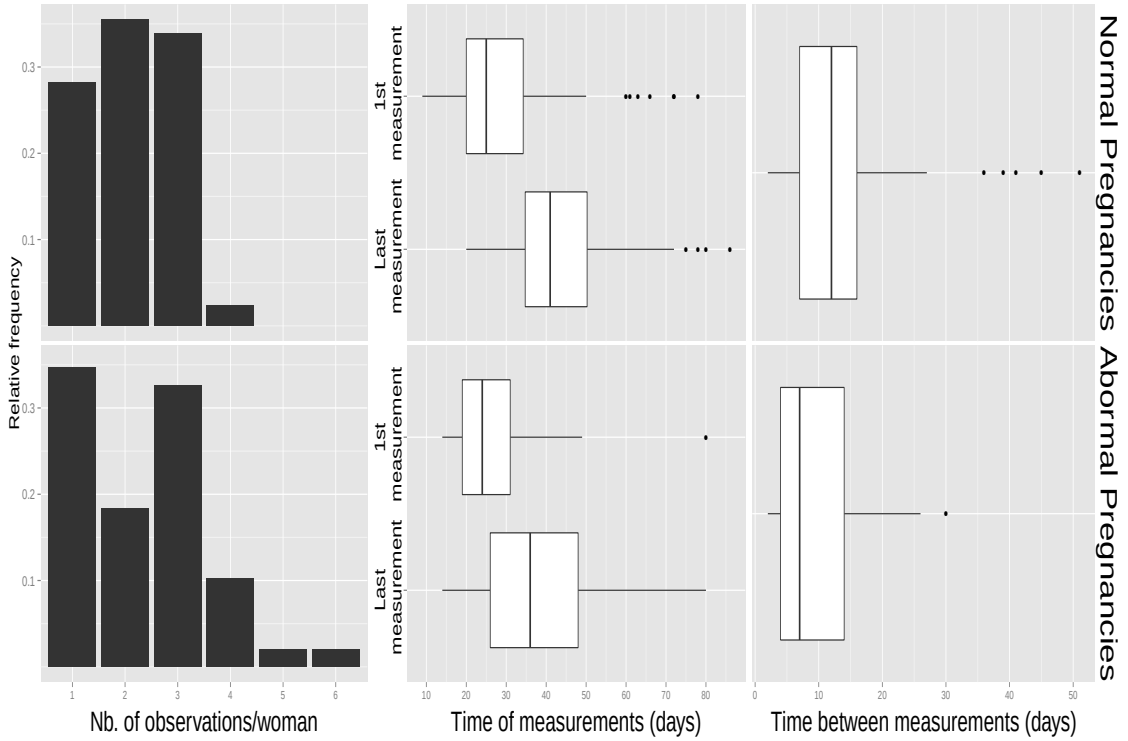

Figure 2: Characteristics of the measurements contained in the pregnancy data set. The top row corresponds to the group of normal pregnancies and the bottom one to the group of abnormal pregnancies.

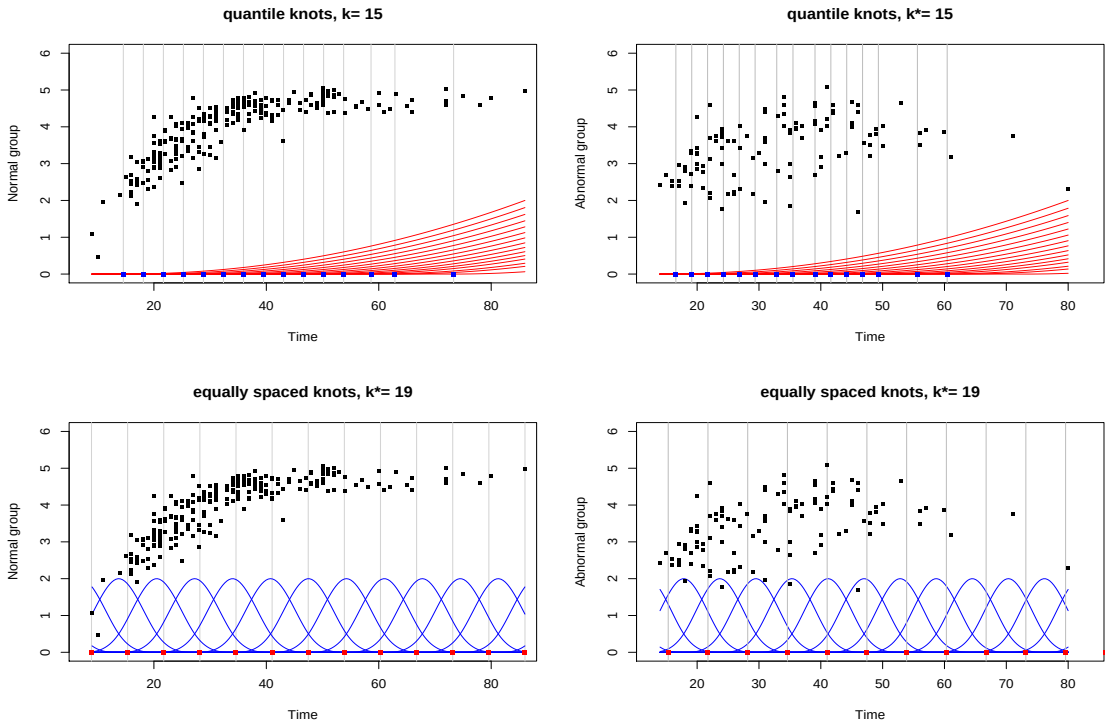

Figure 3: Illustrative representation of regression basis functions for normal and abnormal pregnancy groups with TPF–splines (top) and P–splines (bottom). The positions of the knots are indicated as vertical grey lines. Note that, for TPF–splines we considered the usual quantile-based criteria and hence position of the knots are more dense between 15 and 50 days, and scattered from day 50 in both normal and abnormal groups. Using P–splines, the knots are chosen equally spaced, by construction 2 additional knots at the beginning and at the end are considered, denoted as k^*.

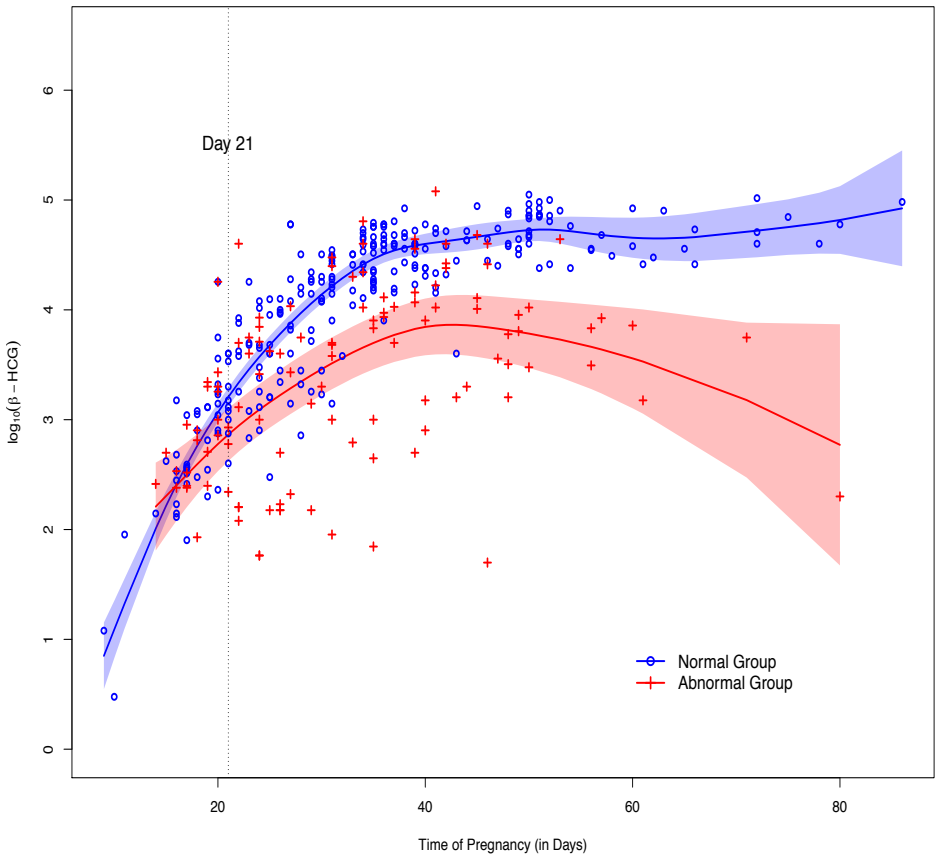

Figure 4: Mean functions estimated by fitting the P–splines + CAR1 model and actual observed data for the pregnancies data set. The individuals trajectories were omitted in order to facilitate the interpretation of the plot. Vertical line at day 21 indicates where mean curves and confidence bands do not cross, indicating the significant difference between both groups starts approximately from that day.

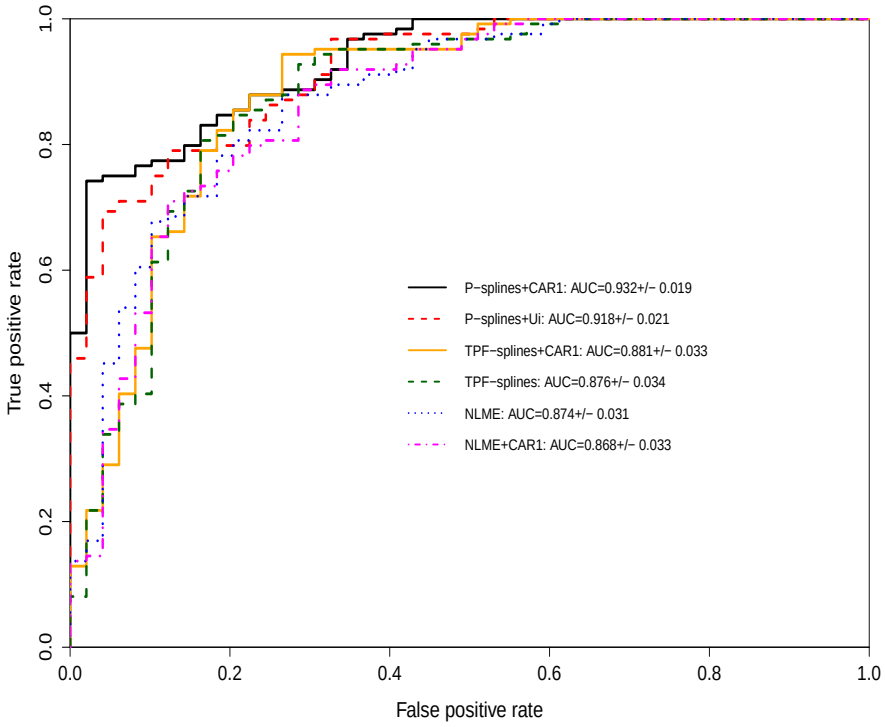

Figure 5: Receiver operating characteristic (ROC) curves for classification in the pregnancies data set under the penalized splines models (TPF–splines/P–splines) and NLME models, with independent errors and correlated errors. The standard deviation of the AUC of a ROC curve is also indicated with "+/-".

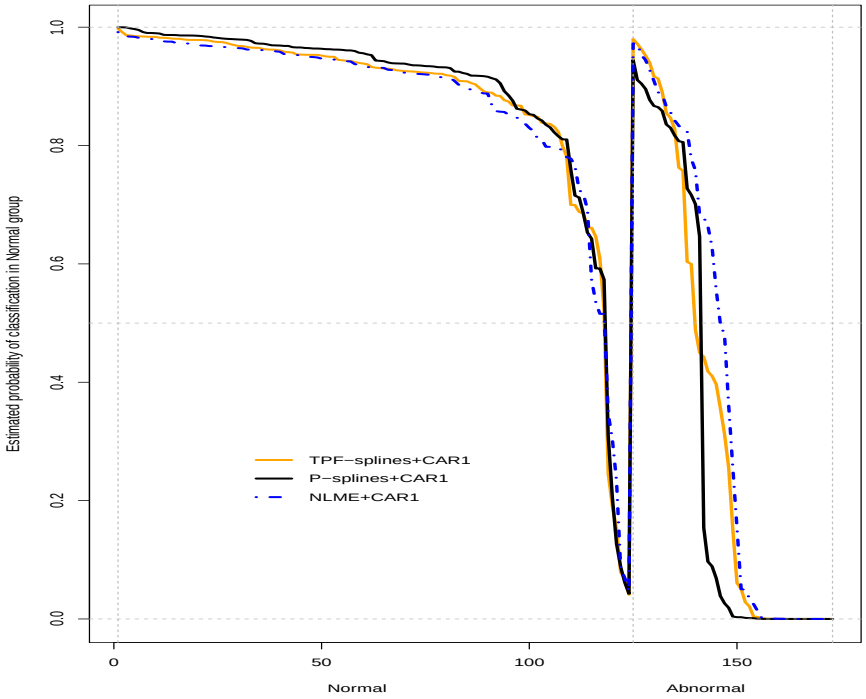

Figure 6: Estimated probabilities of classification in the normal group for all women in decreasing order within normal and abnormal groups using leave-one-out CV. The plot shows that for the normal group, the estimated probabilities have a similar performance for TPF and P–splines with correlated errors and "NLME+CAR1", but the main differences are in the abnormal group where the "P–splines+CAR1" model has a lower misclassification probability (it decays faster than the other models) which means that it obtains higher probabilities for abnormal pregnancies.

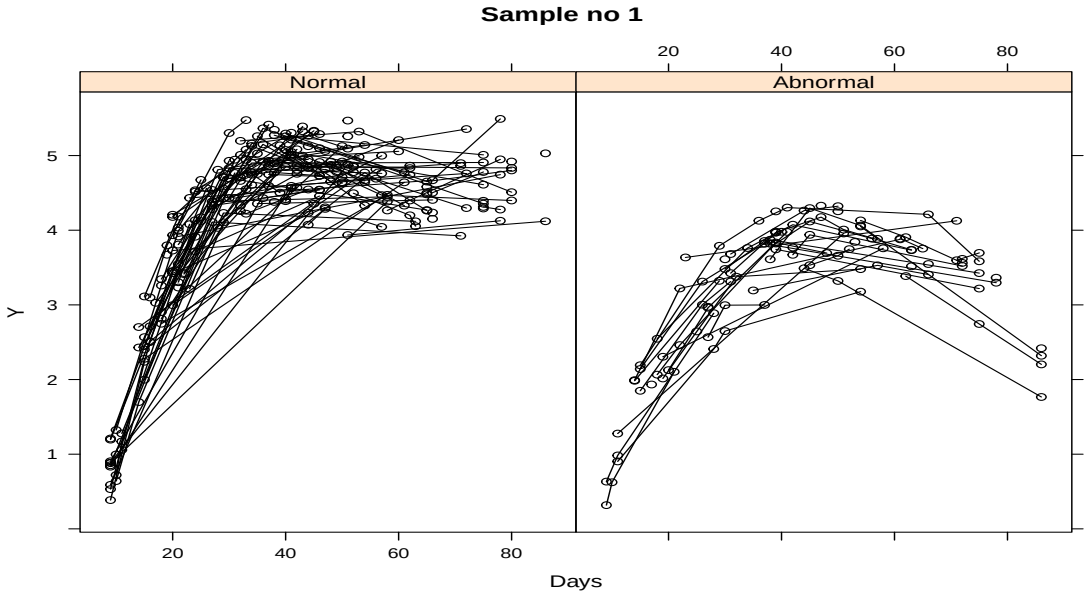

Figure 7: Sample no. 1 for Scenario 1.

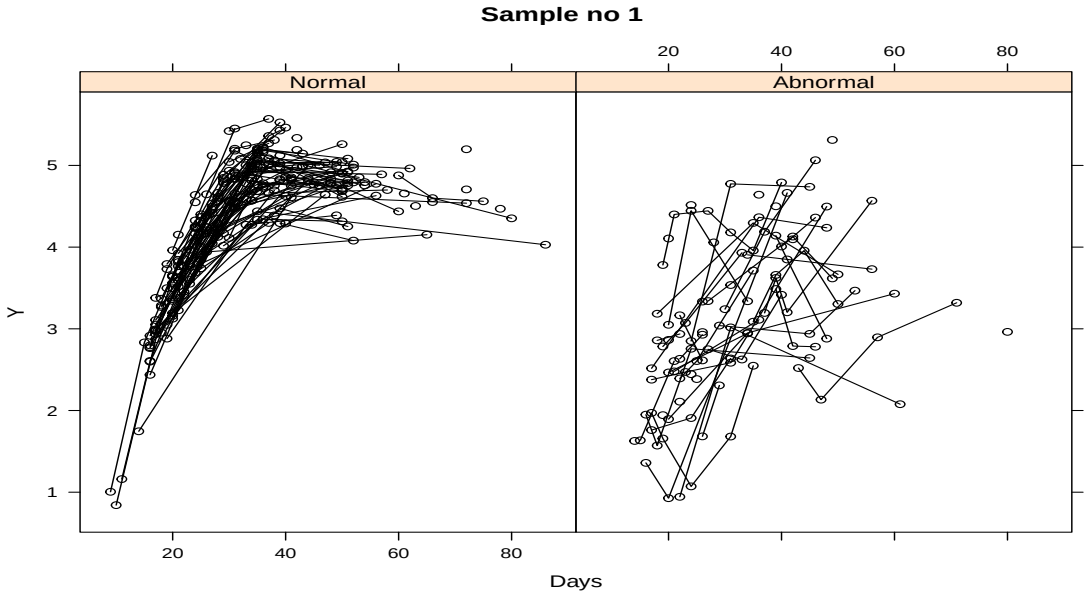

Figure 8: Sample no. 1 for Scenario 2.

List of Tables

Table 1: Model selection criteria for the different models under consideration.

Model	logLik	AIC	BIC
Normal group			
TPF–splines	-102.578	213.155	227.382
P–splines	-97.751	203.503	217.730
TPF–splines + CAR1	-71.374	152.748	170.532
P–splines + CAR1	-67.390	144.780	162.564
TPF–splines + Random intercept	-91.300	192.600	210.384
P–splines + Random intercept	-86.735	183.471	201.255
TPF–splines + Random intercept + CAR1	-71.374	154.748	176.089
P–splines + Random intercept + CAR1	-67.376	146.752	168.093
NLME	-83.730	177.460	195.282
NLME + CAR1	-60.168	132.336	153.723
NLM	-88.214	184.428	198.686
NLM + CAR1	-60.168	130.336	148.159
Abnormal group			
TPF–splines	-128.562	265.125	275.999
P–splines	-123.362	254.724	265.598
TPF–splines + CAR1	-104.874	219.748	233.340
P–splines + CAR1	-99.435	208.870	222.462
TPF–splines + Random intercept	-114.436	238.871	252.464
P–splines + Random intercept	-108.771	227.542	241.135
TPF–splines + Random intercept + CAR1	-104.874	222.748	238.059
P–splines + Random intercept + CAR1	-99.435	210.870	227.180
NLME	-104.289	218.579	232.260
NLME + CAR1	-96.391	204.783	221.200
NLM	-122.637	253.274	264.219
NLM + CAR 1	-97.437	204.874	218.555

Table 2: Average classification women in the pregnancies over 100 simulated data sets by using TPF–splines and P–spline models in Scenario 1 and Scenario 2.

Scenario 1

Group	*Normal*	*Abnormal*	*Normal*	*Abnormal*	*Total (173)*
	TPF–splines + CAR1		*P–splines + CAR1*		
Normal	120.28	3.72	122.96	1.04	124
Abnormal	7.13	41.87	6.24	42.76	49

Scenario 2

Group	*Normal*	*Abnormal*	*Normal*	*Abnormal*	*Total (173)*
	TPF–splines + CAR1		*P–splines + CAR1*		
Normal	113.09	10.91	121.20	2.80	124
Abnormal	3.22	45.78	3.16	45.84	49