T E C H N I C A L

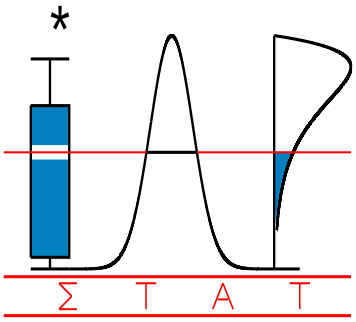
R E P O R T

06107

FLEXIBLE ESTIMATION OF SERIAL CORRELATION IN LINEAR MIXED MODELS

SERROYEN, J., MOLENBERGHS, G., AERTS, M.,

VLOEBERGHS, E., DE DEYN, P. P. and G. VERBEKE



I A P S T A T I S T I C S

N E T W O R K

INTERUNIVERSITY ATTRACTION POLE

http://www.stat.ucl.ac.be/IAP

Flexible Estimation of Serial Correlation in Linear Mixed Models

Jan Serroyen, Geert Molenberghs, Marc Aerts

Center for Statistics, Hasselt University,
Agoralaan 1, Diepenbeek, Belgium

Ellen Vloeberghs, Peter Paul De Deyn

Laboratory of Neurochemistry & Behaviour, University of Antwerp,
Universiteitsplein 1, Wilrijk, Belgium

Geert Verbeke

Biostatistical Centre, Katholieke Universiteit Leuven,
Kapucijnenvoer 35, Leuven, Belgium

Abstract

The linear mixed effects model has, arguably, become the most commonly used tool for analyzing continuous, normally distributed longitudinal data. In its general model formulation four structures can be distinguished: fixed effects, random effects, measurement error and serial correlation. Broadly speaking, serial correlation captures the phenomenon that the correlation structure within a subject depends on the time lag between two measurements. While the general linear mixed model is rather flexible, the need has arisen to further increase flexibility. In response, quite some work has been done to relax the model assumptions and/or to extend the model. For example, the normality assumption for the random effects has been generalized in several ways. Comparatively less work has been devoted to more flexible serial correlation structures. Therefore, we propose the use of spline-based modeling of the serial correlation function. The approach is applied to data from a pre-clinical experiment in dementia which studied the eating and drinking behavior in mice.

Some Keywords: Alzheimer's Disease; Dementia; Generalized estimating equations; Ordinary least squares; Mixed-effects model; Random effect.

Running Title: Flexible Serial Correlation.

1 Introduction

Arguably, the linear mixed effects model (Laird and Ware 1982, Verbeke and Molenberghs 2000, Diggle *et al* 2002) has become the most commonly used tool for analyzing continuous, normally distributed longitudinal data, arising from measuring a subject's response repeatedly over time. In its general formulation, based on Diggle's (1988) model, four structures can be distinguished. First, so-called *fixed regression effects* describe population-averaged relationships between covariates and the outcome of interest. Second, between-subject variability is captured by means of subject-specific parameters. For example, one often assumes a subject-specific intercept and slope in time. Such effects are then assumed to follow from a stochastic distribution, usually of the normal type. This gives rise to the term *random effects*. Third, the outcome is often measured with error, adding a second stochastic component to the model, the *measurement error*, usually assumed to follow a normal distribution. When model specification concludes here, the so-called conditional independence model is obtained, meaning that all within-subject correlation is captured by the random-effects structure. In case this is deemed less plausible, the fourth structure, termed *serial correlation* can be included into the model. Broadly speaking, serial correlation captures the phenomenon that the correlation structure within a subject depends on the time lag between two measurements. Often, indeed, measurements taken closer in time will exhibit a larger correlation than when they are further apart. Diggle (1988) assumed the serial correlation to arise from a normally distributed stochastic process. Combining these four components leads to the general linear mixed-effects model, which has been implemented in a good number of standard statistical software packages, such as the SAS procedure MIXED. Standard fitting methods are based on maximum likelihood and variations there upon. Diggle (1988) proposed the *semi-variogram* as a convenient graphical tool to study the overall variance-covariance structure and to separate it into its three constituents. For this tool to be applicable, one has to assume a constant variance over time and restrict the random-effects structure to a random intercept only.

While the above model is rather flexible, the need has arisen for further flexibility. In response, quite some work has been done to relax the model assumptions and/or to extend the model.

One strand of research is directed towards flexible covariance-structure modeling (Pan and Mackenzie 2003). Furthermore, the normality assumption for the random effects has been generalized in several ways. Lee and Nelder (1996) proposed *hierarchical generalized linear models* in which, combined with an alternative fitting method, both the outcome as well as the random effects can, but do not have to be normally distributed. Verbeke and Lesaffre (1996) presented what they termed the *heterogeneity linear mixed model*, allowing the random effect to follow a mixture of normal distributions. Aitkin (1999) assumed a discrete random effects distribution, based on a finite number of support points, and suggests to use non-parametric maximum likelihood as a convenient fitting tool. Various strands of research have considered spline-based formulations for the random-effects structure (Verbyla *et al* 1999, Ruppert *et al* 2003). Such spline-based models can be implemented, on a routine basis, in the SAS procedure GLIMMIX. Ruppert *et al* (2003) present the necessary SPlus code to fit their model.

Comparatively less work has been devoted to more flexible serial correlation structures. Verbeke, Lesaffre, and Brant (1998) presented an extension of the semi-variogram, allowing for random effects other than merely a random intercept. While elegant in concept, the method is not invariant to the choice of transformation on which it is based. Lesaffre *et al* (2000) used *fractional polynomials* (Royston and Altman 1994) to obtain a flexible yet still fully parametric description of the serial correlation function. This is an appealing idea, worth of further refinement. Consequently, it is taken up in this paper. Next to this, we also propose the use of spline-based modeling of the serial correlation function.

When one is not directly interested in the correlation structure as such, but merely needs to correct for it, the generalized estimating equations (GEE) approach of Liang and Zeger (1986) can be adopted. Even in this situation, however, there are reasons to prefer a mixed model approach. First, this is the case when subject-specific predictions are needed. Second, the full likelihood-based mixed models are preferable when one is confronted with missing data and the assumption of missing completely at random (MCAR, Little and Rubin 2002) is considered too restrictive and one needs to revert to missing at random (MAR).

Section 2 introduces the motivating case study, of which the analysis is taken up in Section 5.

An overview of existing methodology, the linear mixed model and relevant extensions to be found in the literature, is the topic of Section 3. Our proposals for flexible serial correlation methodology are described in Section 4.

2 Motivating Case Study

Alzheimer's disease (AD) and other dementias have been defined by cognitive and non-cognitive symptomatology. These neuropsychological characteristics are referred to as *Behavioral and Psychological Signs and symptoms of Dementia (BPSD)*. Besides these behavioral disturbances and psychological symptoms described by Reisberg *et al* (1987), demented patients develop changes in eating and drinking behavior. The data introduced in this section were obtained from a study which was set up to investigate behavioral changes in genetically modified mice. These so-called transgenic APP23 mice were genetically engineered based on an animal model for dementia (Vloeberghs *et al* 2004). The specific aim of the study was to investigate whether this valuable mouse model develops eating and drinking disturbances. The APP23 mice were compared with wild-type (WT) control littermates. The total sample size was 85, of which 44 were transgenic mice and 41 were controls.

Eating and drinking behavior were simultaneously recorded for one week by employing so-called Skinner boxes placed inside ventilated isolation compartments. Each mouse cubicle was equipped with a pellet feeder and a water bottle (optical lickometer) to provide 20 mg dustless precision pellets of the rodent grain-based formula and tap water. Photocell sensors were used to detect pellet removal, i.e., the number of pellets taken, and the number of licks at the drinking tube. Registration periods typically started Wednesday at 10 am and ended exactly 167 hours later on Wednesday at 9 am. During this 1-week recording period, the *12-hour light—12-hour dark* cycle was continued in the same way as in the facility where mice were previously housed (i.e., lights off at 8 pm).

Figure 1 presents the average evolutions in number of licks and pellets over time for the WT and APP23 group. A circadian pattern can clearly be observed: the mice show more activity at night (e.g., after 12 hours) compared to during the day (e.g., after 24 hours). The

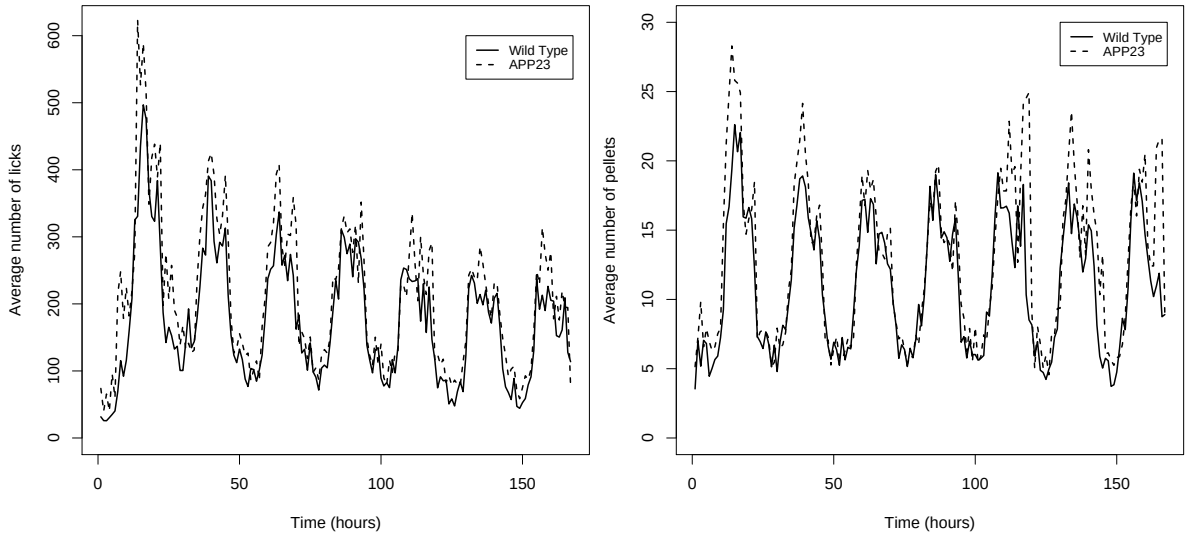


Figure 1: *Average evolutions for number of licks and pellets over time.*

individual ordinary least squares (OLS) residual profiles of the log-transformed number of licks of 5 randomly selected mice are shown in Figure 2. The variability is not constant and the circadian pattern also seems to be present in the variance structure.

3 Existing Methodology

After briefly describing the well-known linear mixed model, we turn to existing proposals for flexible random-effects modeling, where the main focus will be placed on spline-based methods.

3.1 The Linear Mixed Model

As mentioned in the introduction, the linear mixed-effects model (Laird and Ware 1982, Verbeke and Molenberghs 2000) is very commonly used with continuous longitudinal data. The model will be introduced and briefly discussed.

Let $\boldsymbol{Y_i}$ denote the n_i-dimensional vector of measurements available for subject $i = 1, \ldots, N$. A general linear mixed model decomposes $\boldsymbol{Y_i}$ as:

$$\boldsymbol{Y_i} = X_i\boldsymbol{\beta} + Z_i\boldsymbol{b_i} + \boldsymbol{\varepsilon_i}, \tag{1}$$

in which $\boldsymbol{\beta}$ is a vector of population-average regression coefficients called fixed effects, and

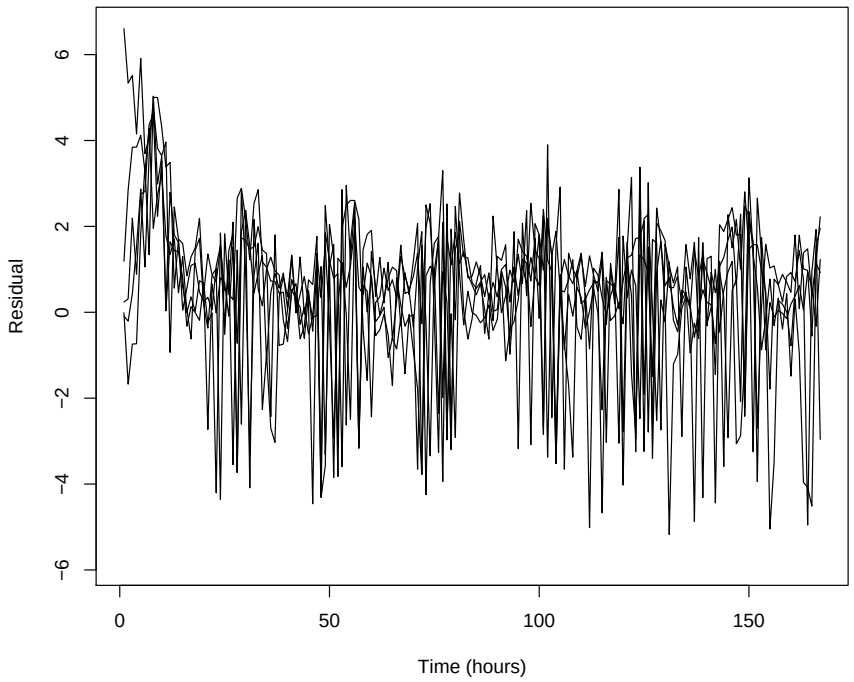


Figure 2: *Ordinary least squares (OLS) residual profiles for log(licks) of 5 randomly selected mice.*

where $\boldsymbol{b_i}$ is a vector of subject-specific regression coefficients. The $\boldsymbol{b_i}$ describe how the evolution of the ith subject deviates from the average evolution in the population. The matrices X_i and Z_i are $(n_i \times p)$ and $(n_i \times q)$ matrices of known covariates. The random effects $\boldsymbol{b_i}$ and residual components $\boldsymbol{\varepsilon_i}$ are assumed to be independent with distributions $N(\mathbf{0}, D)$, and $N(\mathbf{0}, \Sigma_i)$, respectively. Note that Σ_i depends on i only dimension-wise, i.e., through the number of measurements available for a particular subject. In other words, the parameters governing Σ_i are generally common to all subjects. Thus, in summary,

$$\boldsymbol{Y_i}|\boldsymbol{b}_i \sim N(X_i\boldsymbol{\beta} + Z_i\boldsymbol{b}_i, \Sigma_i), \qquad \boldsymbol{b}_i \sim N(0, D). \tag{2}$$

Let $f(\boldsymbol{y_i}|\boldsymbol{b_i})$ and $f(\boldsymbol{b_i})$ be the density functions of $\boldsymbol{Y_i}$ conditional on $\boldsymbol{b_i}$, and of $\boldsymbol{b_i}$, respectively. The marginal density function of $\boldsymbol{Y_i}$ is then given by

$$f(\boldsymbol{y_i}) = \int f(\boldsymbol{y_i}|\boldsymbol{b_i})\ f(\boldsymbol{b_i})\ d\boldsymbol{b_i}, \tag{3}$$

the density of the n_i-dimensional normal distribution $N(X_i\boldsymbol{\beta}, Z_iDZ_i' + \Sigma_i)$. Further, let $\boldsymbol{\alpha}$ denote the vector of all variance and covariance parameters (usually called variance compo-

nents) found in $V_i = Z_i D Z_i' + \Sigma_i$, that is, $\boldsymbol{\alpha}$ consists of the $q(q+1)/2$ different elements in D and of all parameters in Σ_i. Finally, let $\boldsymbol{\theta} = (\boldsymbol{\beta}', \boldsymbol{\alpha}')$ be the s-dimensional vector of all parameters in the marginal model for $\boldsymbol{Y_i}$.

Oftentimes, Σ_i in model (2) is chosen to be equal to $\sigma^2 I_{n_i}$ where I_{n_i} denotes the identity matrix of dimension n_i. We then call this model the conditional independence model, since it implies that the n_i responses on individual i are independent, conditional on $\boldsymbol{b_i}$ and $\boldsymbol{\beta}$. This model may imply unrealistically simple covariance structures for the response vector $\boldsymbol{Y_i}$, especially for models with few random effects. The covariance assumptions can often be relaxed by allowing an appropriate, more general, residual covariance structure Σ_i for the vector $\boldsymbol{\varepsilon_i}$ of subject-specific error components.

Diggle *et al* (2002), based on Diggle (1988), proposed such a general model. They assume that $\boldsymbol{\varepsilon_i}$ has constant variance and can be decomposed as $\boldsymbol{\varepsilon_i} = \boldsymbol{\varepsilon_{(1)i}} + \boldsymbol{\varepsilon_{(2)i}}$ in which $\boldsymbol{\varepsilon_{(2)i}}$ is a component of serial correlation, suggesting that at least part of an individual's observed profile is a response to time-varying stochastic processes operating within that individual. This type of random variation results in a correlation between serial measurements, which is usually, and quite sensibly, a decreasing function of the time separation between these measurements. Further, $\boldsymbol{\varepsilon_{(1)i}}$ is an extra component of measurement error reflecting variation added by the measurement process itself, and assumed to be independent of $\boldsymbol{\varepsilon_{(2)i}}$.

The resulting linear mixed model can now be written as

$$\begin{cases} \boldsymbol{Y_i} = X_i\boldsymbol{\beta} + Z_i\boldsymbol{b_i} + \boldsymbol{\varepsilon_{(1)i}} + \boldsymbol{\varepsilon_{(2)i}} \\ \boldsymbol{b_i} \sim N(\mathbf{0}, D), \\ \boldsymbol{\varepsilon_{(1)i}} \sim N(\mathbf{0}, \sigma^2 I_{n_i}), \\ \boldsymbol{\varepsilon_{(2)i}} \sim N(\mathbf{0}, \tau^2 H_i), \\ \boldsymbol{b_1}, \ldots, \boldsymbol{b_N}, \boldsymbol{\varepsilon_{(1)1}}, \ldots, \boldsymbol{\varepsilon_{(1)N}}, \boldsymbol{\varepsilon_{(2)1}}, \ldots, \boldsymbol{\varepsilon_{(2)N}} \text{ independent}, \end{cases} \tag{4}$$

and the model is completed by assuming a specific structure for the $(n_i \times n_i)$ correlation matrix H_i. One usually assumes that the serial effect $\boldsymbol{\varepsilon_{(2)i}}$ is a population phenomenon, independent of the individual. The serial correlation matrix H_i then only depends on i through the number

of n_i observations and the time points t_{ij} at which measurements were taken. Further, it is assumed that the (j, k) element h_{ijk} of H_i is modeled as

$$h_{ijk} = g(|t_{ij} - t_{ik}|) \tag{5}$$

for some decreasing function $g(\cdot)$ with $g(0) = 1$. This means that the correlation between the measurements $\varepsilon_{(1)ij}$ and $\varepsilon_{(2)ik}$ only depends on the time interval between the measurements y_{ij} and y_{ik}, and decreases if the length of this interval increases.

Two frequently used $g(\cdot)$ functions are the exponential and Gaussian serial correlation functions defined as $g(u) = \exp(-\phi u)$ and $g(u) = \exp(-\phi u^2)$, respectively ($\phi > 0$).

The marginal covariance matrix is then of the form

$$V_i = Z_i D Z_i' + \tau^2 H_i + \sigma^2 I_{n_i}. \tag{6}$$

A classical inferential approach is based on maximizing the marginal likelihood function

$$L_{\mathsf{ML}}(\boldsymbol{\theta}) = \prod_{i=1}^{N} \left\{ (2\pi)^{-n_i/2} |V_i(\boldsymbol{\alpha})|^{-1/2} \times \mathsf{exp}\left(-\frac{1}{2} (\boldsymbol{Y_i} - X_i\beta)' \, V_i^{-1}(\boldsymbol{\alpha}) \, (\boldsymbol{Y_i} - X_i\beta) \right) \right\} \tag{7}$$

with respect to $\boldsymbol{\theta}$. Alternatively, and with an eye on reducing small-sample likelihood-type bias, restricted maximum likelihood (REML, Harville 1974, 1977, Molenberghs and Verbeke 2000) can be used, which comes down to the maximization of the so-called REML likelihood

$$L_{\mathsf{REML}}(\boldsymbol{\theta}) = \left| \sum_{i=1}^{N} X_i' \, V_i^{-1}(\boldsymbol{\alpha}) \, X_i \right|^{-1/2} L_{\mathsf{ML}}(\boldsymbol{\theta}). \tag{8}$$

The proposals made further in this paper can be implemented with both ML and REML according to taste.

3.2 Finite Mixture

Verbeke and Lesaffre (1996) relaxed the assumption of normal random effects to a mixture of normal distributions. Beunckens, Molenberghs, and Verbeke (2006) broadened this type of model for the incomplete data setting. Verbeke and Lesaffre (1996) used the form

$$\boldsymbol{Y}_i | q_{ik} = 1, \boldsymbol{b}_i \sim N(X_i\boldsymbol{\beta}_k + Z_i\boldsymbol{b}_i, \Sigma_i^{(k)}),$$

where X_i and Z_i are design matrices, $\boldsymbol{\beta}_k$ are fixed effects, possibly depending on the group components, $\boldsymbol{b}_i$ denote the random-effect parameters, following a mixture of g normal distributions with indicator variables q_{ik}, mean vectors $\boldsymbol{\mu}_k$, and covariance matrices $\boldsymbol{D}_k$, i.e.,

$$\boldsymbol{b}_i | q_{ik} = 1 \sim N(\boldsymbol{\mu}_k, D_k),$$

and thus

$$\boldsymbol{b}_i \sim \sum_{k=1}^{g} \pi_k N(\boldsymbol{\mu}_k, D_k).$$

The measurement error terms $\boldsymbol{\varepsilon}_i$ follow a normal distribution with mean zero and covariance matrix $\Sigma_i^{(k)}$. The mean and the variance of $\boldsymbol{Y}_i$ take the form:

$$\begin{aligned} E(\boldsymbol{Y}_i) &= X_i \sum_{k=1}^{g} \pi_k \boldsymbol{\beta}_k + Z_i \sum_{k=1}^{g} \pi_k \boldsymbol{\mu}_k, \\ \mathsf{Var}(\boldsymbol{Y}_i) &= Z_i' \left[\sum_{k=1}^{g} \pi_k \boldsymbol{\mu}_k^2 - \left(\sum_{k=1}^{g} \pi_k \boldsymbol{\mu}_k \right)^2 + \sum_{k=1}^{g} \pi_k D_k \right] Z_i + \sum_{k=1}^{g} \pi_k \Sigma_i^{(k)}. \end{aligned}$$

To ensure identifiability, a restriction has to be placed on the $\boldsymbol{\mu}_k$ parameters. A common choice is to impose a zero sum constraint.

3.3 Smoothing Splines

Another flexible way for obtaining a smooth fit to one's data is through *splines*, which are piecewise polynomials with components smoothly spliced together. The joining points of the polynomial pieces are called *knots*, that do not have to be evenly spaced. A spline is of degree p when the highest degree of the polynomial segments is p. Ruppert *et al* (2003) define a pth-degree spline model with knots at $\kappa_1, \ldots, \kappa_K$ as

$$f(x) = \beta_0 + \beta_1 x + \ldots + \beta_p x^p + \sum_{k=1}^{K} \beta_{p+k} (x - \kappa_k)_+^p, \tag{9}$$

where $(x - \kappa_k)_+$ is the truncated power basis function, i.e., the *positive part* of the function $(x - \kappa_k)$. Other possible basis functions include the B-spline (Dierckx 1993), P-spline (Eilers and Marx 1996), natural cubic spline, and radial basis.

A simple and straightforward way to fit splines is by using ordinary least squares to estimate the (unrestricted) knot point coefficients β_{p+k}. This essentially means that the coefficient at

each knot point is considered a fixed effect. Thus, each knot point is associated with a single parameter, hence the term *parametric spline*. However, this approach usually tends to overfit the data, leading to too coarse a regression curve, unless the number of knot points is small and their location carefully chosen.

Owing to the aforementioned coarseness of the parametric spline, various methods have been developed to constrain the knots' influence. Classically, the amount of smoothing is controlled by adding a term to the likelihood function, penalizing large coefficients at the knot points, which amounts to counterbalancing such coefficients' contribution to the raggedness of the curves. An candidate penalty term is $\sum_{k=1}^{K} \beta_{p+k}^2$, but there are many more. Roughly speaking, they can be subdivided into three classes. First, the *fixed smoothing parameter* approach controls the amount of smoothing by fixing the degrees of freedom of the fit, df_{fit}, which is the trace of the so-called smoother matrix (Ruppert *et al* 2003). Another way to refer to this concept is by *equivalent number of parameters*. However, this approach is not easily translated to the setting of smoothing serial correlation functions, the goal of this manuscript, since constructing the smoother matrix is less than straightforward, owing to the fact that this method can be cast into the form of a ridge regression when applied to conventional regression or the fixed-effects structure in a linear mixed model, whereas such a reformulation is not possible in the serial correlation case. Second, we can fit models with different numbers of knot points, combined with the use of some information criterion, such as, for example, the *Akaike Information Criterion* (AIC). However, for this method to work, adequately fitting a large number of knot points is crucial, and this can be problematic in the context of covariance modeling. Third, penalized splines can also be represented in mixed-model form (Verbyla *et al* 1999, Ruppert *et al* 2003), meaning that each knot point coefficient acts as a random effect. This results in a multivariate normal density entering the marginal likelihood, which then needs to be integrated out. The variance component governing these additional random effects is usually set equal for all knot points. This variance component controls and describes the degree of flexibility and smoothness. The fitted curve can be constructed by means of the empirical Bayes estimates. The aforementioned integration can be carried out using conventional numerical integration (e.g., Gaussian quadrature, Laplace approximation)

or sampling based (e.g., Monte Carlo Markov chain) methods.

The linear mixed model representation can be set up by considering the following random-spline design matrix:

$$Z_i = \begin{bmatrix} (x_1 - \kappa_1)_+ & \cdots & (x_1 - \kappa_K)_+ \\ \vdots & \ddots & \vdots \\ (x_n - \kappa_1)_+ & \cdots & (x_n - \kappa_K)_+ \end{bmatrix}.$$

Of course, such additional random effects can be combined with random effects already present in (1). Other modeling assumptions expressed in conjunction with (1) are left unaltered. Let the sole variance component governing the smoothing process be σ_u^2 and assume the residual error structure is of the conditional independence type with variance component σ_2^ε, then the smoothing parameter λ^2 can be shown to take the form $\lambda^2 = \sigma_\varepsilon^2/\sigma_u^2$ (Verbyla *et al* 1999, Ruppert *et al* 2003).

4 Flexible Serial Correlation Structures

In analogy with choosing flexible functions and modeling concepts for the fixed and random effects structures, it would also be desirable to dispose of flexible tools for the serial structure. Verbeke, Lesaffre, and Brant (1998) proposed an extended semi-variogram to flexibly explore this structure, which, together with some issues surrounding it, will be reviewed briefly in Section 4.1. Subsequent sections deal with using smoothing spline ideas, the concept of which was introduced before, when describing the serial association. All of these methods are rooted in studying the function $g(\cdot)$ in (5).

4.1 An Extension of the Variogram

Borrowing ideas from spatial statistics, the empirical semi-variogram was introduced by Diggle (1998) into the field of longitudinal data. For a linear mixed model with random intercepts, a time-stationary serial correlation process, and constant-variance measurement error, he was able to (graphically) separate these three components of variability so as to enable convenient assessment of, amongst others, the relative importance of each of the three components. In its original form, the variogram is restrictive in the variance-covariance structure allowed; in

particular, the random-effects structure is confined to a random intercept only. Subsequently, Verbeke, Lesaffre, and Brant (VLB, 1998) extended the technique to allow for models containing additional random effects.

The main idea of VLB is to project the ordinary least squares residuals $\boldsymbol{r_i} = \boldsymbol{y_i} - X_i \widehat{\boldsymbol{\beta}}_{OLS}$ orthogonal to the columns of Z_i, allowing one to directly study the variability in the data not explained by the random effects. It follows from the theory of generalized estimating equations (GEE) that this OLS estimator is consistent for $\boldsymbol{\beta}$ (Liang and Zeger 1986, Verbeke and Molenberghs 2000, p. 125). This justifies the use of the OLS residuals $\boldsymbol{r_i}$ for studying the dependence among the repeated measures. For each i, $i = 1, \ldots, N$, let A_i be an $n_i \times (n_i - q)$ matrix such that $A_i' Z_i = 0$ and $A_i' A_i = I_{n_i - q}$. The $(n_i - q)$-dimensional transformed OLS residuals are then defined as $\boldsymbol{\Re_i} = A_i' \boldsymbol{r_i}$. Similar to the logic developed by Diggle (1998) the relative importance of the variance components can then be assessed by means of the quantities $(\boldsymbol{\Re_{ij}} - \boldsymbol{\Re_{ik}})^2$.

However, there are several pitfalls associated with the VLB approach. Using simulations, Verbeke (1995) has shown that the method yields estimates for the $g(\cdot)$ function too instable to be useful for residual serial correlation detection, which is caused by the high degree of scatter in the values $(\boldsymbol{\Re_{ij}} - \boldsymbol{\Re_{ik}})^2$ and, at the same time, by the high degree of multicollinearity in the approximate regression model that is needed by VLB to calculate these residuals. Furthermore, the choice of the matrix A_i is non-unique and leads to different residuals. Notwithstanding the fact that VLB empirically found that different choices for A_i yield only slightly different non-parametric estimates for $g(\cdot)$, leaving conclusions about presence and type of serial correlation untouched, this feature subtracts some appeal from the method and suggests consideration of alternative approaches.

4.2 Smoothing Splines

Returning to the smoothing spline ideas laid out in Section 3.3, we are now in a position to formulate a spline model for the serial process:

$$\tau^2 g(u) = \frac{\exp(\phi_0)}{1 + \exp\{\phi_1 + \phi_2 \, \log(u) + \sum_{k=1}^{K} \phi_{k+2} [\log(u) - \log(\kappa_k)]_+\}}. \qquad (10)$$

This means that ϕ_0 acts as a (strictly positive) intercept, capturing the variance of the serial correlation component, τ^2. Further, ϕ_1 acts as an intercept, ϕ_2 is the linear slope and $\phi_3, \ldots, \phi_{K+2}$ are the spline coefficients associated with the serial correlation function $g(\cdot)$. The logistic link ensures that the estimated $g(\cdot)$ function stays within the $[0, 1]$ interval.

A penalty term is added to (7) to obtain a smoother fit, leading to the following marginal likelihood function:

$$l(\boldsymbol{\theta}) = L_{\mathsf{ML}}(\boldsymbol{\theta}) + \lambda \sum_{k=1}^{K} \phi_{k+2}^2, \tag{11}$$

where the fixed, user-defined, λ controls the amount of smoothing. In principle, it is conceivable to develop methods for data-driven selection of λ. However, this falls outside of the scope of this paper. Moreover, we have observed in our case study and limited simulations (the latter not reported here), that the presence of λ has a favorable since stabilizing effect on the algorithm's convergence.

5 Analysis of Case Study

The approach as described above in Section 4.2 has been applied to the case study data introduced in Section 2. For the reasons mentioned in Section 4.1, the analyzes were performed on the OLS residual profiles for the log-transformed number of licks. A small smoothing parameter value ($\lambda = 0.01$) was chosen since this improved convergency, while it only had a small impact on the actual fit. No random effects were included in the model, since they did not lead to a significant improvement in likelihood value. A classical exponential fit together with spline fits at two different sets of knot points of the serial correlation function $g(.)$ for number of licks is shown in Figure 3.

The spline fits indicate that the serial correlation function is non-monotone. This fact would go entirely unnoticed with a conventional serial correlation approach. For example, we now see that the serial correlation is substantially lower for a 12-hour time lag than for one of 24 hours. Very likely, this can be ascribed to the circadian rythm. This 24-hours pattern could also be observed in Figure 1. It therefore plays a role in the mean structure and the correlation structure simultaneously. Since a classical serial correlation model would not allow for this,

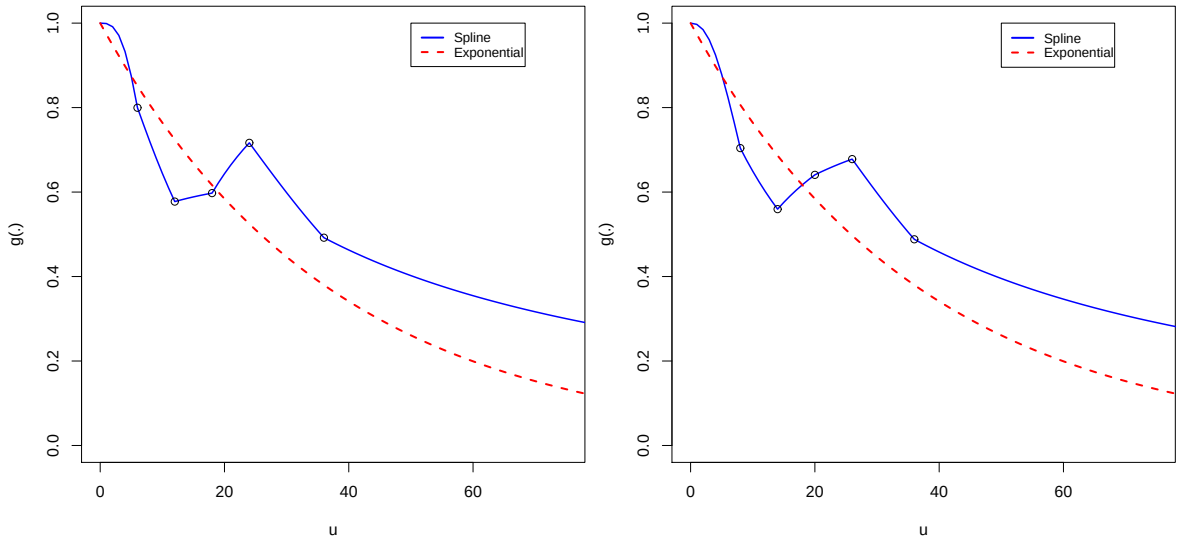


Figure 3: *Exponential and spline fit of the serial correlation function* $g(.)$ *for number of licks. Left graph: knot points at* $u = 6, 12, 18, 24, 36$. *Right graph: knot points at* $u = 8, 14, 20, 26, 36$.

it is conceivable that in such a model the mean structure fit would be distorted, rendering associated inferences less reliable.

The fact that the one-parameter exponential function cannot detect this type of serial correlation also shows through the difference in log-likelihood between the exponential, and the spline model with knots located at $u = 6, 12, 18, 24, 36$. Precisely, the test statistic equals $2(27395.0 - 27306.9) = 176.2$, which under the null follows a χ^2_6. This represents a considerable improvement in fit of the spline model compared to the exponential model.

Note that all knot points were positioned for time lags below 36 hours, confirming that this is the time frame were quite a lot is happening, in contrast to larger lags. Of course, there is more information in a set of data about shorter lags, since relatively more pairs corresponding to such lags can be formed.

6 Concluding Remarks

Flexible serial correlation structures, just like flexible random-effects modeling, are necessary when modeling complex longitudinal profiles, especially with a long period of follow up and/or

a large number of measurements within subjects. To this effect, we have proposed a spline-based approach. Such a parametric spline approach works acceptably well, as long as the number of knot points is chosen to be relatively small compared to the number of time points. In our case study, we essentially used 5 knot points for 167 follow-up occasions. The choice of the knot points' position, too, is important, both for the quality of the fit as well as for convergence of the updating algorithm.

Convergency can be problematic when fitting an elaborate covariance structure. However, in the analysis of the case study the proposed spline approach actually performed better than some of the simpler serial correlation based models, such as one featuring Gaussian serial correlation. Arguably, the added flexibility allows for a better fit and ultimately therefore better convergence.

The choice of the smoothing parameter λ is rather subjective by nature. In our analysis, we chose a small value since this improved convergence, without smoothing out the non-monotone trend in the fitted serial correlation function and without adversely impacting the model's fit.

Admittedly, observations made in a case study are always a bit ad hoc. Therefore, we performed a small simulation study (details not reported), which largely confirmed the finding that adding knot points can improve the fit, while at the same time causing rapid variance increases and having a detrimental influence on convergence.

References

Aitkin, M. (1999) A general maximum likelihood analysis of variance components in generalized linear models. *Biometrics*, **55**, 218–234.

Dierckx, P. (1993) *Curve and Surface Fitting with Splines*. Clarendon, Oxford.

Diggle, P.J. (1988) An approach to the analysis of repeated measures. *Biometrics*, **44**, 959–971.

Diggle, P.J., Heagerty, P.J., Liang, K.-Y., and Zeger, S.L. (2002) *Analysis of Longitudinal Data (2nd ed.)*. Oxford Science Publications. Oxford: Clarendon Press.

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

Harville, D.A. (1974) Bayesian inference for variance components using only error contrasts. *Biometrika*, **61**, 383–385.

Harville, D.A. (1977) Maximum likelihood approaches to variance component estimation and to related problems. *Journal of the American Statistical Association*, **72**, 320–340.

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

Lee, Y. and Nelder, J.A. (1996) Hierarchical generalized linear models (with discussion). *Journal of the Royal Statistical Society, Series B*, **58**, 619–678.

Lesaffre, E., Todem, D., Verbeke, G. and Kenward, M. (2000) Flexible modelling of the covariance matrix in a linear random effects model. *Biometrical Journal*, **42**, 807–822.

Liang, K.-Y. and Zeger, S.L. (1986) Longitudinal data analysis using generalized linear models. *Biometrika*, **73**, 13–22.

Little, R.J.A. and Rubin, D.B. (2002) *Statistical Analysis with Missing Data* (2nd ed.). New York: John Wiley & Sons.

Pan, J. and Mackenzie, G. (2003) On modelling mean-covariance structures in longitudinal studies. *Biometrika*, **90**, 239–244.

Reisberg, B., Borenstein, J., Salob, S.P., Ferris, S.H., Franssen, E., Georgotas, A. (1987) Behavioral symptoms in Alzheimer's disease: phenomenology and treatment. *Journal of Clinical Psychiatry*, **48**, 9–15.

Royston, P. and Altman, D.G. (1994) Regression using fractional polynomials of continuous covariates: parsimonious parametric modelling. *Applied Statistics*, **43**, 429–468.

Ruppert, D., Wand, M.P. and Carroll, R.J. (2003) *Semiparametric Regression*. New York: Cambridge University Press.

Verbeke, G. (1995) *The linear mixed model. A critical investigation in the context of longitudinal data analysis.* Ph.D. thesis, Catholic University of Leuven, Faculty of Science,

Department of Mathematics.

Verbeke, G. and Lesaffre, E. (1996) A linear mixed-effects model with heterogeneity in the random-effects population. *Journal of the American Statistical Association*, **91**, 217–221.

Verbeke, G., Lesaffre, E. and Brant, L.J. (1998) The detection of residual serial correlation in linear mixed models. *Statistics in Medicine*, **17**, 1391–1402.

Verbeke, G. and Molenberghs, G. (2000). *Linear mixed models for longitudinal data*, Springer Series in Statistics, Springer-Verlag, New-York.

Verbyla, A.P., Cullis, B.R., Kenward, M.G. and Welham, S.J. (1999) The analysis of designed experiments and longitudinal data by using smoothing splines. *Applied Statistics*, **48**, 269–311.

Vloeberghs, E., Van Dam, D., Engelborghs, S., Nagels, G., Staufenbiel, M. and De Deyn, P.P. (2004) Altered circadian locomotor activity in APP23 mice: A model for BPSD disturbances. *European Journal of Neuroscience*, **20**, 2757–2766.