Flexible estimation of serial correlation in nonlinear mixed models

Peer-reviewed author version

Flexible Estimation of Serial Correlation in Nonlinear Mixed Models

Jan Serroyen [1], **Geert Molenberghs** [2,3], **Marc Aerts** [2]

Ellen Vloeberghs [4], **Peter Paul De Deyn** [4], **Geert Verbeke** [3,2]

[1] Methodology and Statistics, University Maastricht,
Peter Debyeplein 1, 6229 HA Maastricht, the Netherlands

[2] I-BioStat, Hasselt University,
Agoralaan 1, 3590 Diepenbeek, Belgium

[3] I-BioStat, Katholieke Universiteit Leuven,
Kapucijnenvoer 35, 3000 Leuven, Belgium

[4] Laboratory of Neurochemistry & Behaviour, University of
Antwerp, Universiteitsplein 1, 2170 Wilrijk, Belgium

Abstract

In the conventional linear mixed-effects model, four structures can be distinguished: fixed effects, random effects, measurement error and serial correlation. The latter 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 addition to work done in the area, we we propose the use of spline-based modeling of the serial correlation function, so as to allow for additional flexibility. 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; Ordinary least squares; Random effect.

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. Such an effect is then assumed to follow from a stochastic distribution, usually of the normal type. This gives rise to the term *random effect*. 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 would conclude 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 so-called general linear mixed-effects model, which has been implemented in a good number of standard statistical software packages, such as the SAS procedure MIXED or similar procedures in S-PLUS and R. 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), while another strand of research has 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 S-Plus code to fit their model.

Comparatively less work has been devoted to more flexible serial correlation structures. Diggle and Verbyla (1998) proposed kernel smoothing to provide a nom-parametric estimator for the covariance structure without assuming stationarity. As mentioned by these authors, in practice it often becomes essential to impose structural restrictions on the covariance matrix either by smoothing or by fitting a parametric model. In their approach they clearly opt for (non-parametric) smoothing, while we believe that the spline-based approach, which we will introduce further on, strikes a good balance between the parametric and smoothing ideas. 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 $20\,\mathrm{mg}$ 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 st 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).

The response variables were defined as the total number of licks and pellets per hour. Since these responses showed severe right-tailed skewness, they were transformed to log(response $+$ 1). Figure 1 presents the average evolutions in the log-transformed 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). Let $\log(\ell_{ij} + 1)$ be the log-transformed number of licks for mouse i at time point j. The observed individual profiles for $\log(\ell_{ij} + 1)$ of 5 randomly selected mice are shown in
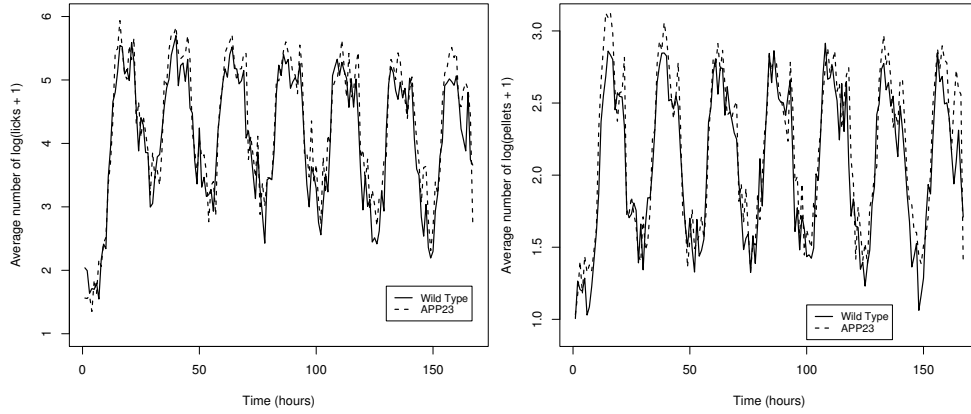
Figure 1: *Average evolutions for log-transformed number of licks and pellets over time.*

Figure 2, while the corresponding observed variance function is portrayed in the left hand panel of Figure 3. The variability is not constant and the circadian pattern also seems to be present in the variance structure. The circadian rhythm thus appears to be a dominant biological factor in this experiment. The right hand panel of Figure 3 presents the semi-variogram for $\log(\ell_{ij}+1)$, based on a random-intercept model with an unstructured model for the mean. However, we should be careful in interpreting this semi-variogram, since the non-constant variance is a clear indication of non-stationarity, thereby rendering the semi-variogram less than trustworthy. It is therefore prudent to consider the semi-variogram for exploratory purposes only. Nevertheless, there appears to be signs of a circadian trend in this graph as well. This was also the direct motivation for studying whether the circadian pattern was also present on the level of serial correlation. To corroborate this finding, we have also constructed a sample semi-variogram, in terms of both mean and median values, supplemented with the interquartile ranges. These can be found in Figure 4. Note that there is both a large spread as well as apparent asymmetry (given the difference between mean and median). Both of these phenomena result from the quadratic nature of the variogram contributions. One can see an apparent discrepancy between the variograms in Figures 3 and 4. This is because the semi-variogram in Figure 4 is based on standardised residuals. Thus, when the variance function is non-constant, but is assumed to be constant nevertheless when computing semi-variogram, it may be less than trustworthy.
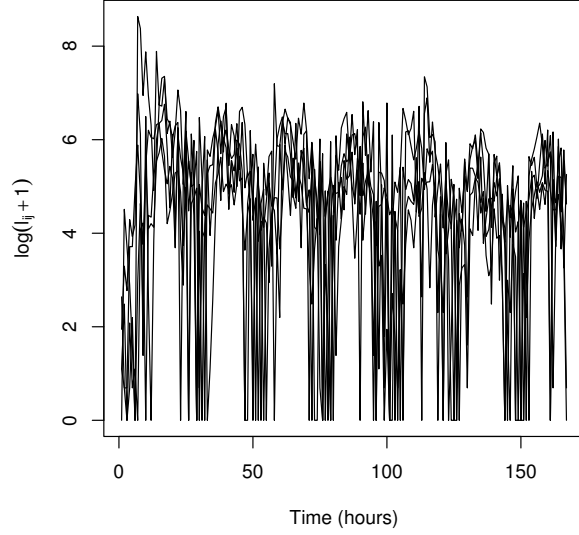
Figure 2: *Observed profiles for $\log(\ell_{ij} + 1)$ of 5 randomly selected mice.*

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

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 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 $N(\boldsymbol{0}, D)$, and $N(\boldsymbol{0}, \Sigma_i)$, 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. Let $f(\boldsymbol{Y_i}|\boldsymbol{b_i})$ and $f(\boldsymbol{b_i})$ be the density
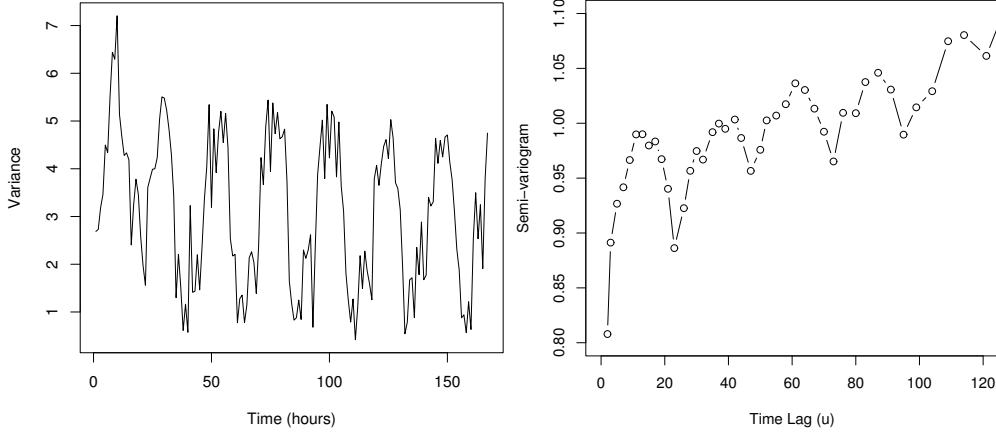
6

Figure 3: *Left hand panel: observed variance function for* $\log(\ell_{ij} + 1)$. *Right hand panel: Semi-variogram for* $\log(\ell_{ij} + 1)$.

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 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 components) found in $V_i = Z_iDZ_i' + \Sigma_i$, 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 $\boldsymbol{Y_i}$.

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 corresponding distributional assumptions are $\boldsymbol{\varepsilon}_{(1)i} \sim N(\boldsymbol{0}, \sigma^2 I_{n_i})$ and $\boldsymbol{\varepsilon}_{(2)i} \sim N(\boldsymbol{0}, \tau^2 H_i)$. 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
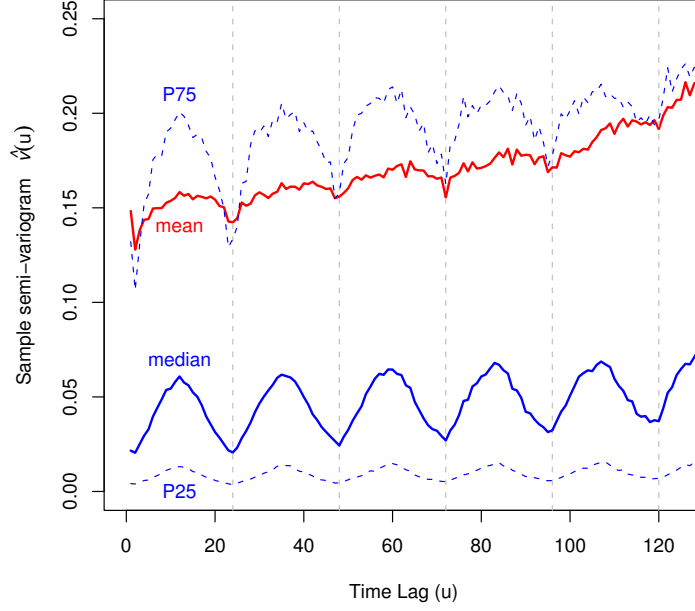
Figure 4: *Sample semi-variogram: mean and median values, supplemented with interquartile ranges, plotted over time.*

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

$$h_{ijk} = g(|t_{ij} - t_{ik}|) \tag{2}$$

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 $V_i = Z_i D Z_i' + \tau^2 H_i + \sigma^2 I_{n_i}$. Inference usually proceeds via maximum likelihood (ML) or restricted maximum likelihood (REML) (Verbeke and Molenberghs 2000).

3.2 Fractional Polynomials

Fractional polynomials have been used in several contexts as a flexible regression device. Royston and Altman (1994) define a fractional polynomial as any function of the form

$$f(u) = \sum_{j=0}^{m} \phi_j \, H_j(X),$$

where the degree m is a positive integer, and $\phi_0, \ldots, \phi_m$ are real-valued unknown regression coefficients. Further, $H_j(X)$ is defined as

$$H_j(X) = \begin{cases} x^{p_j} & \text{if } p_j \neq p_{j-1}, \\ H_{j-1}(X) \ln(X) & \text{if } p_j = p_{j-1}. \end{cases} \tag{3}$$

where $H_0(x) = 1$, $p_0 = 0$, and $p_1 \leq \ldots \leq p_m$ are real-valued prespecified powers.

In the context of linear and logistic regression analyses, Royston and Altman (1994) have shown that the family of fractional polynomials is very flexible and that models with degree m larger than 2 are rarely required. In practice, several values for the powers $p_1, \ldots, p_m$ can be tried and the model with the best selected.

In the context of the linear mixed model, fractional polynomials have been used both in the fixed-effects structure (Verbeke and Molenberghs 2000, Ch. 24.5) as well as in the serial correlation structure. We will return to the latter application in Section 4.1.

3.3 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*, which 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

$$f(x) = \beta_0 + \beta_1 \, x + \ldots + \beta_p \, x^p + \sum_{k=1}^{K} \beta_{p+k} \, (x - \kappa_k)_+^p, \tag{4}$$

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), natural cubic spline (Eubank 1988), and radial basis (Green and Silverman 1994).

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 and this is usually referred to as a *regression 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, as in adaptive splines (Zhang 2004).

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. A candidate penalty term, though not the only one, is $\lambda \sum_{k=1}^{K} \beta_{p+k}^2$. There is a vast amount of literature on the selection of the optimal smoothing parameter λ. Roughly speaking, there are three (related) ways to determine the smoothing parameter λ. A first option is cross-validation, where for a grid of λ-values the squared error loss criterion is minimized in a leave-one-out cross-validation procedure. A second approach is the use of some model selection criterion, such as *Akaike's Information Criterion* (AIC, Hurvich, Simonoff, and Tsai 1998). This method, however, requires a so-called equivalent number of parameters or the effective degrees of freedom (Ruppert *et al* 2003) to define the AIC appropriately. Thirdly, penalized splines can also be represented in mixed-model form (Verbyla *et al* 1999, Ruppert *et al* 2003), in the sense 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 kept constant across knot points; it controls and describes the degree of flexibility and smoothness. The fitted curve can be constructed by means of empirical Bayes estimation.

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_ε^2, 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, it would be desirable to have available flexible tools for the serial structure. Lesaffre *et al* (2000) proposed fractional polynomials to flexibly model this structure, which, jointly with some issues surrounding it, will be reviewed briefly in Section 4.1. The subsequent section deals with penalized splines, the concept of which was introduced in Section 3.3, when describing the serial association. All methods are rooted in studying the function $g(\cdot)$ in (2).

4.1 Fractional Polynomials

Lesaffre *et al* (2000) applied fractional polynomials to model the serial correlation function $g(\cdot)$. Their model is of the form

$$\tau^2 g(u) = \exp\left\{\phi_0 + \sum_{j=1}^{m} \phi_j\, u^{(p_j)}\right\}. \tag{5}$$

This parameterization does not *a priori* ensure that $g(\cdot)$ is a decreasing function, nor that it is has an upper bound of 1. Furthermore, these authors construct a fractional polynomial of degree 4, using the power set $\{0, 0.5, 1, 2\}$, in clear contrast with the recommended degree of $m = 2$. As a result, concerns of multicollinearity and model stability can be raised.

The fractional polynomial approach, using the recommended degree of $m = 2$, will be applied to the case study data and the results will be presented in Section 5.

4.2 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:

$$\tau^2 g(u) = \frac{\exp(\phi_0)}{1 + \exp\{\phi_1 + \phi_2 \log(u) + \sum_{k=1}^{K} \phi_{k+2} [\log(u) - \log(\kappa_k)]_+\}}. \tag{6}$$

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. Defining the spline on the log-scale, using argument $\log(u)$, and constraining ϕ_2 to be strictly positive, the serial correlation function satisfies the natural assumption that $\lim_{u \to 0} g(u) = 1$. Additionally, the use of the log-scale improved convergence considerably.

For a particular, rich enough set of knots, a penalty term is added to the marginal likelihood to obtain a smooth fit, leading to the following marginal log-likelihood function:

$$\ell(\boldsymbol{\theta}) = \ell_{\mathsf{ML}}(\boldsymbol{\theta}) + \lambda \sum_{k=1}^{K} \phi_{k+2}^2, \tag{7}$$

where ℓ denotes the natural logarithm of the corresponding likelihood L and, again, the smoothing parameter λ controls the amount of smoothing. In principle, it is conceivable to develop methods for an optimal, data-driven selection of λ, as those briefly discussed in Section 3.3. The main difficulty, however, is that criteria such as cross-validation and the effective degrees-of-freedom, being the trace of the so-called smoother matrix in the classical setting, are defined on the scale of the data, and it is no means clear how to translate these concepts to serial correlation. The same holds for the mixed model representation. The aforementioned integration can be done analytically in case of normal random spline effects, but this is no longer true for spline effects in the serial correlation function. In this case the 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.

A data-driven optimal selection of the smoothing parameter falls outside of the scope of this paper. Instead, we limit the number of (well-chosen) knots and try out some fixed values for

the smoothing parameter. The number and position of knot points for the spline function has to be chosen wisely. This is generally the case in smoothing when using only a limited number of knots, but maybe even more so when modeling the covariance matrix, where parsimonious modeling is very important. Covariance modeling can in general be considered as computationally heavy, and this is particularly true for our case study, where we have 167 measurement occasions. In practice, splines with different sets of knot points can be fitted to the data. This then gradually gives an idea of the shape of the serial correlation function. Since most pairs of data can be formed for shorter time lags, it is also most sensible to focus at least some knot points in this range.

Note that, in principle, spline functions allow for almost unlimited flexibility. Evidently, the more complex the model, the more data would be needed to stably estimate the corresponding parameters. Too complex a model could easily result in convergence issues. For this reason, we have chosen to restrict periodicity to less than 36 hours, thereby capturing the main component of periodicity.

All analyzes were performed using the statistical software package R (R Development Core Team 2007). The marginal log-likelihood expression for (7) was constructed and then maximized using a general purpose numerical optimizer based on a quasi-Newton method.

5 Analysis of Case Study

First, a model for the mean number of licks will be presented. Then, the approaches as described above in Section 4 will be applied to the case study data introduced in Section 2.

The mean structure was modeled using the six-parameter function:

$$\log(\ell_{ij} + 1) = \beta_0 + \beta_1 \, t_{ij} + \beta_2 \, t_{ij}^2 + \beta_3 \, \log(t_{ij})^2 + \beta_4 \, \sin(\beta_5 \, t_{ij}). \tag{8}$$

Note that this model includes one non-linear parameter, i.e., the sine frequency parameter β_5. This parameter, together with β_4, is included to accommodate the clear circadian pattern present in the data. However, since we opted for a fully marginal approach, capturing the within-subject association through the serial correlation function, no random effect enters the

13

Table 1: Parameter estimates and standard errors for the spline model with knots located at $u = 6, 12, 18, 24, 36$

Effect	Parameter	Estimate (s.e.)
Mean function:		
Intercept	β_0	4.6670 (0.0655)
Linear slope	β_1	-0.3580 (0.0559)
Quadratic slope	β_2	-0.0251 (0.0366)
Log quadr. slope	β_3	-0.1529 (0.0082)
Amplitude sine	β_4	-1.2128 (0.0282)
Frequency sine	β_5	26.6225 (0.0232)
Variance function:		
Intercept	σ_0	1.8990 (0.0062)
Linear slope	σ_1	-0.1514 (0.0235)
Amplitude sine	σ_2	1.6004 (0.0295)
Frequency sine	σ_3	26.9143 (0.0217)
Serial correlation function:		
Variance (τ^2)	ϕ_0	-0.1370 (0.0275)
Intercept	ϕ_1	-3.7514 (0.0962)
Slope	ϕ_2	-1.6284 (0.0450)
Knot ($u = 6$)	ϕ_3	0.8145 (0.0613)
Knot ($u = 12$)	ϕ_4	0.4875 (0.1082)
Knot ($u = 18$)	ϕ_5	2.9516 (0.1838)
Knot ($u = 24$)	ϕ_6	-5.2902 (0.2273)
Knot ($u = 36$)	ϕ_7	1.5939 (0.2637)

likelihood in a non-linear way. This means that no random effect needs to be numerically or approximately integrated out from the likelihood function and therefore, we can still use the marginal likelihood to fit (8). No treatment effect is included in (8). It is obviously virtually absent from the model from graphical inspection (Figure 1). Also, initially fitting the model with treatment effect included produced a highly non-significant p-value. Hence, we start further modeling efforts from mean-model description (8).

The resulting model fit for the mean number of licks is shown in Table 1 and in Figure 5.

Let us now turn to the main topic of interest, namely the modeling of the serial correlation function. After fitting models over a range of smoothing parameter values, a small value ($\lambda = 0.01$) was chosen since this improved convergency, while it only had a small impact on
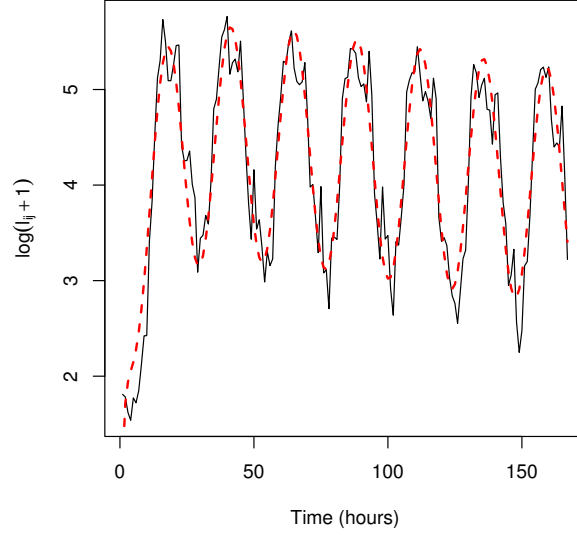
Figure 5: *Observed mean profile (solid) and fitted mean model (dashed) for* $\log(\ell_{ij} + 1)$.

the actual fit. A classical exponential function, a fractional polynomial fit, together with spline fits at two different sets of knot points of the serial correlation function $g(.)$ for the number of licks is shown in Figure 6. The combination of fractional powers that provided the best fit was $\{0.5, 1\}$. The choice of knot point locations did not seem to have a large impact on convergence, as long as they were not positioned too close to one other. For this reason, we chose to set the knot points at 6 hourly intervals. The results obtained with different sets of knot points were all consistent.

It may be surprising at first sight that the serial correlation function in Figure 6 predicts high values, whereas the observed correlations in Table 2 are relatively low. However, they cannot be compared directly. The observed correlations result from the interplay between three variance components: measurement error, serial correlation, and random effects. The latter two contribute to the correlation, whereas the first one does now. Figure 6 singles out the serial correlation component. Suppose, for example, that the measurement error contribution is very high. Then the observed correlations may be low, even when computed between two instantaneous replications (two measurements taken at the same time). However, between two such instantaneous replicates, the serial correlation would always be 1, because for every choice of $g(\cdot)$, we have that $g(0) = 1$. Likewise, $g(+\infty) = 0$, even though the observed correlation

15

Table 2: Estimated marginal correlation matrix based upon the spline model presented in Table 1.

						k					
		1	**7**	**13**	**19**	**25**	**31**	**37**	**43**	**49**	**55**
	1	1.0000	0.1533	0.1621	0.1831	0.1711	0.1280	0.1362	0.1382	0.0890	0.0794
	7	0.1533	1.0000	0.1769	0.1725	0.1142	0.1405	0.1531	0.1347	0.0861	0.0763
	13	0.1621	0.1769	1.0000	0.2759	0.1577	0.1375	0.2462	0.2220	0.1230	0.1082
	19	0.1831	0.1725	0.2759	1.0000	0.2326	0.1750	0.2222	0.3293	0.1869	0.1425
j	**25**	0.1711	0.1142	0.1577	0.2326	1.0000	0.1513	0.1658	0.1742	0.1625	0.1270
	31	0.1280	0.1405	0.1375	0.1750	0.1513	1.0000	0.1888	0.1711	0.1132	0.1453
	37	0.1362	0.1531	0.2462	0.2222	0.1658	0.1888	1.0000	0.2838	0.1620	0.1474
	43	0.1382	0.1347	0.2220	0.3293	0.1742	0.1711	0.2838	1.0000	0.2222	0.1745
	49	0.0890	0.0861	0.1230	0.1869	0.1625	0.1132	0.1620	0.2222	1.0000	0.1507
	55	0.0794	0.0763	0.1082	0.1425	0.1270	0.1453	0.1474	0.1745	0.1507	1.0000

function may not fully phase out, owing to the presence of random effects. Thus, one must not loose sight of the fact that the serial correlation is one component of the variance-covariance structure.

A 10×10 subset, using time intervals of 6 hours, of the in principle 167×167 fitted correlation matrix, is displayed in Table 2.

Note that unlike in the fully parametric case, the fractional-polynomial and spline-based models can be tuned as far as smoothness is concerned, resulting in more or less smoothness. Related to this, by choosing this particular level of smoothness, we are able to follow trends in the data, which are less smooth than the parametric model can possibly support.

The spline fits indicate that the serial correlation function is non-monotone. The (point-wise) 95% confidence bands for this model fit is presented in the left hand graph in Figure 7. The non-monotone trend is indeed confirmed by this graph. This fact would go entirely unnoticed with a conventional serial correlation approach and is also missed by the fractional polynomial fit. 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 rhythm.
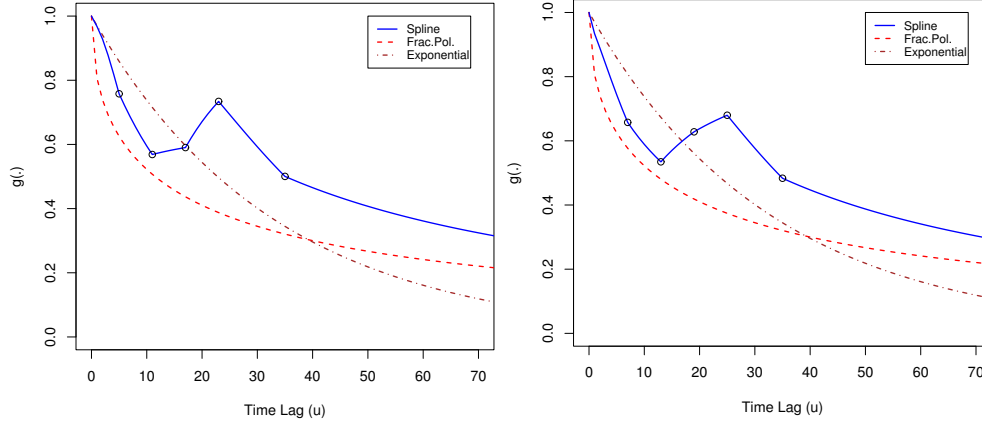
Figure 6: *Exponential, fractional polynomial and spline fit of the serial correlation function $g(.)$ for number of licks. Left hand graph: knot points at $u = 6, 12, 18, 24, 36$. Right hand graph: knot points at $u = 8, 14, 20, 26, 36$.*

This 24-hours pattern can also be observed in Figures 1 and 2. It therefore plays a role in the mean structure, variance structure, and the correlation structure simultaneously. Since a classical serial correlation model would not allow for this, it is conceivable that in such a model, for this type of applications, the mean structure fit would be distorted, rendering associated inferences less reliable. To check whether the estimation of the serial correlation function was not distorted by the non-constant variance, the variance itself was modeled with a sine function. The right hand side graph in Figure 7 shows the observed and fitted variance function. Although modeling the variance function improved the likelihood considerably, the estimated serial correlation function remained virtually unchanged.

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, and 36. Precisely, the difference equals $2(27395.0 - 27306.9) = 176.2$. The corresponding value of Akaike's Information Criterion provides strong evidence of an improved fit by the spline model, when compared to the exponential model.
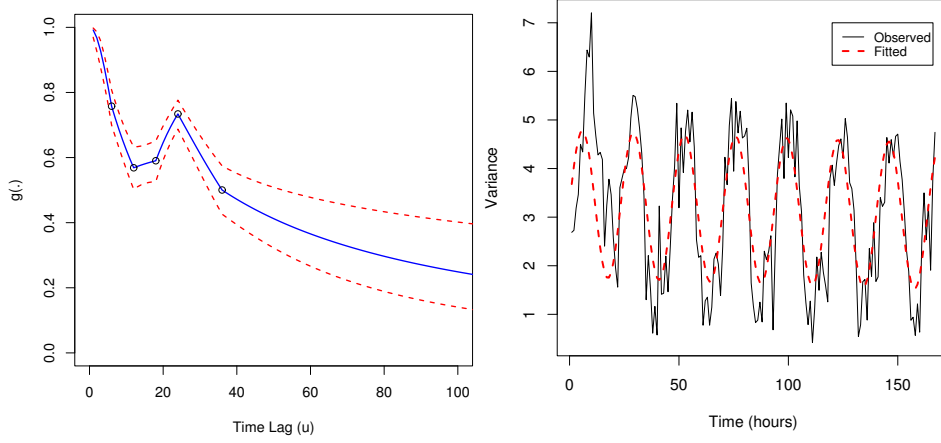
Figure 7: *Spline model with knot points at $u = 6, 12, 18, 24, 36$. Left hand graph: spline fit with 95% confidence bands. Right hand graph: observed and fitted variance function.*

6 Concluding Remarks

Flexible serial correlation structures, in agreement with 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 end, 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 specific parameterization in combination with the added flexibility allows for a better fit and ultimately therefore better convergence.

Since model (8) includes a non-linear parameter, i.e., frequency of the sine wave, we opted for a fully marginal approach, thus omitting random effects. Although it is possible to include both random effects and serial correlation in a linear mixed-effects model, this is not advisable

in many applications. Diggle *et al* (2002) and Verbeke and Molenberghs (2000) argue that, in applications, the effects of random effects and serial correlation are often intertwined in the sense of high correlation between the corresponding parameter estimates. In practice, this is often reflected by way of estimation problems. Therefore, restricting the model's covariance structure to either serial correlation or random effects, often combined with measurement error, may be pragmatically a sensible choice. A situation where all aspects of (co)variability could successfully be included is where: (1) the time lag between subsequent measurements is sufficiently large to separate instantaneous measurement error from serial correlation on the one hand; and (2) there are a sufficient number of measurement occasions so as to separate serial correlation from random-effects variability on the other hand.

The presented spline approach provides a flexible alternative to the simple classical models, such as, for example, the exponential model, which is useful when interest lies directly in the shape of the serial correlation function. The approach can also be used for sensitivity analysis purposes, i.e., checking if an assumed simple serial correlation function is adequate.

For the case study data, the standard errors of the fixed-effects parameter estimates did not change substantially when comparing the relatively simple exponential serial correlation model with the more elaborate spline model. Thus, this particular analysis could not provide evidence for a gain in efficiency related to inference about the fixed-effects parameters. However, a simulation study might shed some light as to whether this conclusion would be more broadly tenable.

The choice of the smoothing parameter λ is rather subjective in 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

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. and Verbyla, A.P. (1998) Nonparametric estimation of covariance structure in longitudinal data. *Biometrics*, **54**, 401–415.

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.

Eubank, R.L. (1988) *Spline Smoothing and Nonparametric Regression*. New York: Marcel Dekker.

Green, P.J., and Silverman, B.W. (1994) *Nonparametric Regression and Generalized Linear Models*. London: Chapman & Hall.

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.

Hurvich, C.M, Simonoff, J.S., and Tsai, C.-L. (1998) Smoothing parameter selection in nonparametric regression using an improved Akaike information criterion. *Journal of the Royal Statistical Society, Series B*, **60**, 271–293.

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.

R Development Core Team (2007) *R: A language and environment for statistical computing.* R Foundation for Statistical Computing, Vienna, Austria.

Reisberg, B., Borenstein, J., Salob, S.P., Ferris, S.H., Franssen, E., and 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., 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.

Zhang, H. (2004) Mixed effects multivariate adaptive splines model for the analysis of longitudinal and growth curve data. *Statistical Methods in Medical Research*, **13**, 63–82.