Statistics in Medicine

Research Article

Received 13 July 2014, Accepted 11 February 2015 Published online 12 March 2015 in Wiley Online Library

(wileyonlinelibrary.com) DOI: 10.1002/sim.6465

Bayesian modeling of the covariance structure for irregular longitudinal data using the partial autocorrelation function

Li Su[a] and Michael J. Daniels[b*†]

In long-term follow-up studies, irregular longitudinal data are observed when individuals are assessed repeatedly over time but at uncommon and irregularly spaced time points. Modeling the covariance structure for this type of data is challenging, as it requires specification of a covariance function that is positive definite. Moreover, in certain settings, careful modeling of the covariance structure for irregular longitudinal data can be crucial in order to ensure no bias arises in the mean structure. Two common settings where this occurs are studies with 'outcome-dependent follow-up' and studies with 'ignorable missing data'. 'Outcome-dependent follow-up' occurs when individuals with a history of poor health outcomes had more follow-up measurements, and the intervals between the repeated measurements were shorter. When the follow-up time process only depends on previous outcomes, likelihood-based methods can still provide consistent estimates of the regression parameters, given that both the mean and covariance structures of the irregular longitudinal data are correctly specified and no model for the follow-up time process is required. For 'ignorable missing data', the missing data mechanism does not need to be specified, but valid likelihood-based inference requires correct specification of the covariance structure. In both cases, flexible modeling approaches for the covariance structure are essential. In this paper, we develop a flexible approach to modeling the covariance structure for irregular continuous longitudinal data using the partial autocorrelation function and the variance function. In particular, we propose semiparametric non-stationary partial autocorrelation function models, which do not suffer from complex positive definiteness restrictions like the autocorrelation function. We describe a Bayesian approach, discuss computational issues, and apply the proposed methods to CD4 count data from a pediatric AIDS clinical trial. © 2015 The Authors. *Statistics in Medicine* Published by John Wiley & Sons Ltd.

Keywords: missing data; non-stationary covariance function; outcome-dependent follow-up; penalized splines; semiparametric covariance function

1. Introduction

In long-term follow-up studies, irregular longitudinal data are observed when individuals are assessed repeatedly over time but at uncommon and irregularly spaced time points. Modeling the covariance structure for irregular longitudinal data requires the specification of a covariance function. This specification can be challenging due to the need to keep the covariance function positive definite. Moreover, careful modeling of the covariance structure for the irregular longitudinal data can be crucial for valid inferences on the mean structure in certain settings (e.g., outcome-dependent follow-up and ignorable missing data). In such settings, flexible models for the covariance structure are certainly desirable. In this paper, we describe a flexible approach to modeling the covariance structure for irregular continuous longitudinal data. We will start by providing more details on settings where correctly modeling the covariance structure is essential. We then review the relevant literature on covariance modeling and provide details on our motivating example.

[a]*MRC Biostatistics Unit, Cambridge Institute of Public Health, Robinson Way, Cambridge CB2 0SR,U.K.*
[b]*Department of Statistics & Data Sciences,Department of Integrative Biology, University of Texas at Austin, Austin, TX 78712, U.S.A.*
[*]*Correspondence to: Michael J. Daniels, Department of Statistics & Data Sciences, Department of Integrative Biology, University of Texas at Austin, Austin, TX 78712, U.S.A.*
[†]*E-mail: mjdaniels@austin.utexas.edu*
This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

1.1. Settings where incorrect covariance modeling results in bias in the mean structure

Irregularly measured longitudinal data can be a result of 'outcome-dependent follow-up', where individuals with a history of poor health outcomes are being assessed with greater frequency and regularity. Lipsitz *et al*. [1], Fitzmaurice *et al.* [2], and Ryu *et al.* [3] discussed assumptions that the follow-up time process can be ignored when making likelihood-based inferences about the longitudinal outcome process. Basically, when the follow-up time process only depends on previous outcomes, likelihood-based methods can still provide consistent estimates for the regression parameters in the mean structure of the longitudinal outcome process, and no modeling of the follow-up time process is needed. However, this requires the correct specification of the whole joint distribution of the longitudinal measures for each individual, which of course includes the covariance structure. A similar problem is that of 'ignorable missing data'. As with the 'outcome-dependent follow-up' case, the missing data mechanism does not need to be explicitly specified, but proper inference about the mean structure requires correct specification of the covariance structure [4, 5].

1.2. Modeling the covariance function

Parametric stationary models for the correlation function have been explored in [6]. Qian [7] proposed flexible stationary models that allows the correlation function to decay with lag. Pan and MacKenzie [8] proposed flexible models for irregularly spaced observations using the modified Cholesky decomposition [9] of a covariance matrix, but again, the focus is on 'stationary-type' models, where the covariance structure depends on the time lag but not on time. More recently, Zhang and Leng [10] proposed similar ('stationary-type') models based on a moving average Cholesky factorization.

Parametric non-stationary models for the covariance function have been proposed in [11] and [12]. Diggle and Verbyla [13] proposed nonparametric approaches for the covariance function using kernel weighted local linear regression, but there is no guarantee that their estimator results in a positive definite covariance function. Fan *et al.* [14] proposed semiparametric models, allowing for nonparametric estimation of the variance function, but parametric estimation of the correlation function. Yao *et al.* [15] used a functional principal components analysis approach to modeling the non-stationary covariance structure, but their main purpose was to characterize the time trend and variation of the irregular longitudinal data in a functional data setting. Here, instead, we focus on regression settings, where covariate effects are of interest.

In this paper, we develop a flexible approach to modeling the covariance structure of the irregular continuous longitudinal data by focusing on the partial autocorrelation function (PACF). The advantage of modeling using a PACF is that the only restriction to maintain positive definiteness is to ensure its values are in the interval $(-1, 1)$. Therefore, in our approach, the positive definiteness obstacle is removed.

The PACF has been well explored for the stationary setting in the time series literature (e.g., [16]) and is often used to determine the order of a stationary autoregressive model in the time series models [17] but appears to have not been used for longitudinal data. Parameterizing a correlation matrix using partial autocorrelations has been explored in [18] and [19] and for a spherical parameterization in [20] but only for cases of *common, equally* spaced follow-up times across units. Zimmerman and Nunez-Anton [21] proposed structured (parametric) antedependence models for the correlation matrix based on partial autocorrelations.

To handle irregularly spaced longitudinal data and accommodate non-stationarity in the correlation function, we develop a new class of semiparametric models for the PACF. To our knowledge, such models have not yet been developed. Together with a model for the variance function, our approach offers flexibility in modeling the covariance structure in challenging situations, such as the 'outcome-dependent follow-up' and 'ignorable missing data' problems described earlier.

1.3. Motivating example

This work was motivated by a randomized double-blinded equivalency trial of high-dose (180 mg per square meter body surface area, six times daily) versus low-dose (90 mg) zidovudine (ZDV) for HIV-infected children (Protocol 128 of the AIDS Clinical Trial Group) [22]. The study enrolled 426 children who were randomized to receive one of the two doses and scheduled for measurement of CD4 count before enrollment and every 12 weeks up to 5 years. The analysis objective is to compare the treatment-specific longitudinal trajectories of CD4 counts.

© 2015 The Authors. *Statistics in Medicine* Published by John Wiley & Sons Ltd.

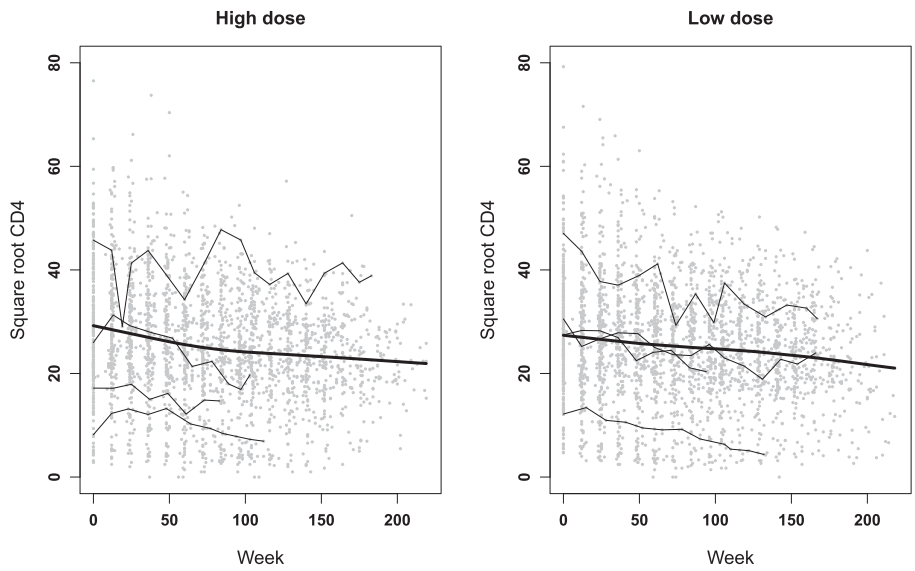

Figure 1. Observed (square root) CD4 count data over time by the dose groups, with local regression fits to the pooled sample (dark lines) and profiles from 4 selected participants in each group highlighted.

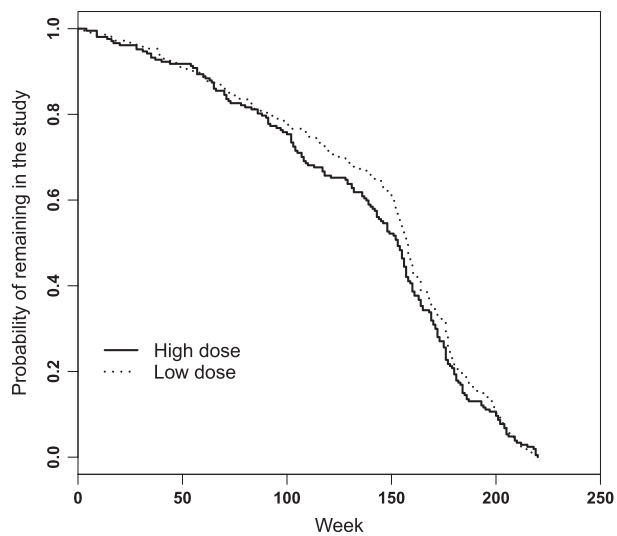

Figure 2. Kaplan-Meier curves for observed dropout times by the dose groups in the AIDS example.

However, the actual measurement times were irregular and varied considerably across children. Figure 1 presents the observed CD4 count data over time by the dose groups with local regression fits to the pooled sample and four individual profiles highlighted. Note that a square root transformation is used to reduce the right skewness in these data. The total number of measurements was 4999, while the number of measurements per child varied from 1 to 21. The observed maximum follow-up time was 219 weeks, and there were 214 unique measurement times following enrollment. In addition, only about half of the children completed 3 years of follow-up (Figure 2). Previous analyses [5, 23] suggest that the dropout was possibly informative in the sense that children with a more rapid decline in CD4 count were more likely to drop out. Figure 3 presents the estimated individual ordinary least-squares intercepts and slopes of the square root of CD4 counts against the observed dropout times, and it appears that lower intercepts in both dose groups and lower slopes in the low-dose group are associated with early dropout. In Section 4, we will demonstrate how to accommodate the irregular measurement times in modeling the covariance structure of these CD4 data while dealing with informative dropout at the same time.

The remainder of this paper is organized as follows. In Section 2, we formally define partial autocorrelations and the PACF, then describe a flexible class of semiparametric non-stationary models for the

© 2015 The Authors. *Statistics in Medicine* Published by John Wiley & Sons Ltd.

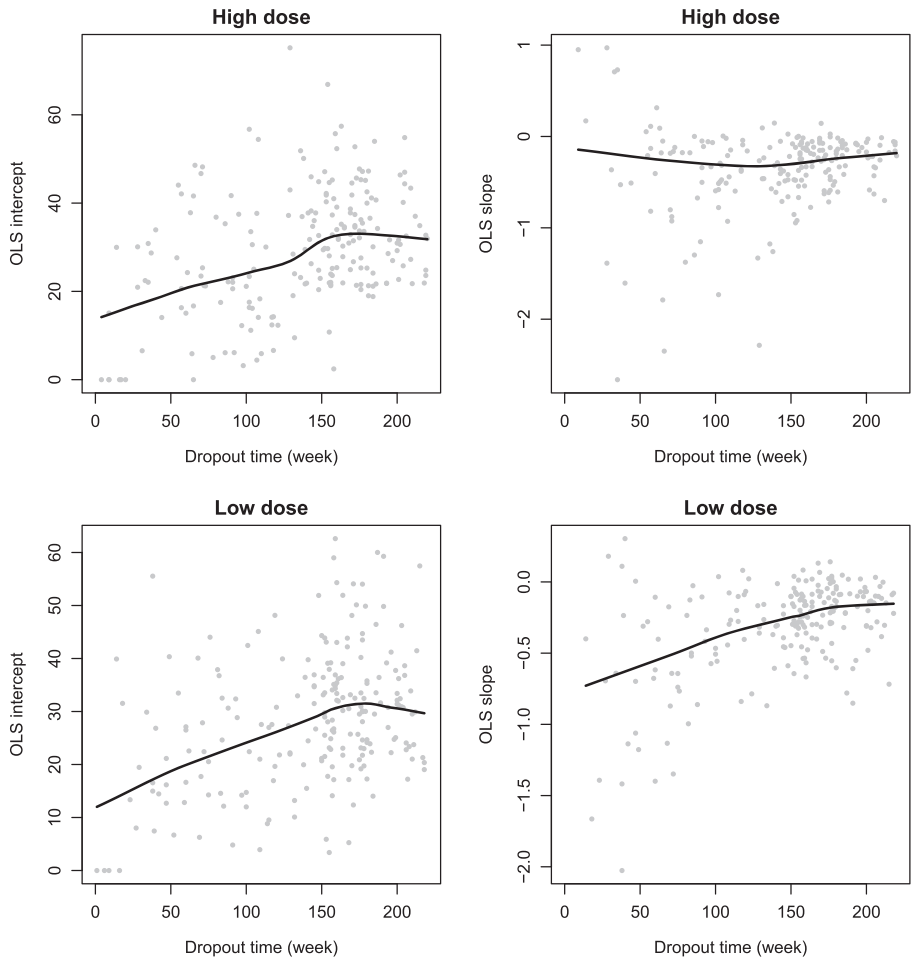

Figure 3. Individual OLS intercepts and slopes of square root CD4 count as functions of the dropout time by the dose groups, with local regression fits highlighted.

covariance function. Section 3 discusses the computational issues and 'tricks' that can be used to perform Bayesian (or likelihood) inference efficiently. In Section 4, we apply the proposed methods to the pediatric CD4 count data. We offer conclusions and extensions in Section 5.

2. Definitions and models

Suppose that N independent individuals are to be followed up intermittently at discrete time $t = 1, 2, \ldots, T$. Here, t could be months, weeks, days or hours (in the AIDS example in Section 4, it is 'months'). The constant T is determined by the potential maximum follow-up time where a longitudinal measurement can be taken in the study; and the value of T can be large depending on the chosen time unit. We assume a discrete time Gaussian process for the longitudinal outcome from the ith individual ($i = 1, \ldots, N$), $\{Y_i(t) : t \in 1, \ldots, T\}$, with a mean function $\mu_i(t)$ and a covariance function, $\mathrm{Cov}\{Y_i(t), Y_i(t')\} = \sigma_i(t)\sigma_i(t')\rho_i(t, t')$ $(t, t' \in 1, \ldots, T)$, where $\sigma_i^2(t)$ is the variance function and $\rho_i(t, t') = \mathrm{Cor}\{Y_i(t), Y_i(t')\}$ is the autocorrelation function.

We assume that the mean function $\mu_i(t)$ can be described by a linear model

$$\mu_i(t) = \mathbf{X}_i(t)\boldsymbol{\beta}, \tag{1}$$

where $\mathbf{X}_i(t)$ is a p-dimensional covariate process that can include both time-invariant and exogenous time-varying covariates, and $\boldsymbol{\beta}$ is the $p \times 1$ vector of corresponding regression coefficients. Note that random effects can be added (we will do this in the data analysis in Section 4) or more flexible structures can be specified for the mean function, for example, using semiparametric regression approaches [24]. In this

© 2015 The Authors. *Statistics in Medicine* Published by John Wiley & Sons Ltd.

paper, however, we focus on developing flexible models for the covariance function $\mathrm{Cov}\{Y_i(t), Y_i(t')\}$ and assume that the mean function can be appropriately modeled by the linear model in (1), possibly with random effects.

At time points $t_{i1}, \dots, t_{in_i}$, the continuous longitudinal outcome measurements $\mathbf{Y}_i = (Y_{i1}, \dots, Y_{in_i})^{\mathrm{T}}$ for the ith individual are taken. Thus, the covariance matrix of $\mathbf{Y}_i$, Σ_i is n_i-dimensional. This matrix can be decomposed as $\mathbf{\Sigma}_i = \mathbf{S}_i\mathbf{R}_i\mathbf{S}_i$, where $\mathbf{S}_i = \mathrm{diag}\{\sigma(t_{i1}), \dots, \sigma(t_{n_i})\}$ is the diagonal matrix of standard deviations, and $\mathbf{R}_i = \{\rho(t_{ik}, t_{il})\}_{kl}$ $(k, l \in 1, \dots, n_i)$ is the correlation matrix. Note that $t_{i1}, \dots, t_{in_i}$ can vary across individuals and be unequally spaced, and in this set-up, irregular longitudinal data are accommodated.

The elements of the correlation matrix, $\mathbf{R}_i$, the marginal correlations, can be expressed in terms of the PACF. In this paper, we model the autocorrelation function $\rho_i(t, t') = \mathrm{Cor}\{Y_i(t), Y_i(t')\}$ by parameterizing the PACF. For the remainder of this section, we drop the subscript i and introduce the PACF as well as our semiparametric PACF models.

2.1. *The partial autocorrelation function*

The PACF, $\pi(t, t+j)$ is defined as

$$\pi(t, t+j) = \mathrm{Cor}(Y(t), Y(t+j) | Y(k), k \in \{t+1, \dots, t+j-1\}),$$

the correlation between $Y(t)$ and $Y(t+j)$ conditional on the intervening values, $\{Y(t+1), \dots, Y(t+j-1)\}$, where j is the time lag. These correlations can also be represented as the correlation of the residuals for $Y(t)$ and $Y(t+j)$ from regressing each on the intervening values.

In other words, this is the remaining correlation between $Y(t)$ and $Y(t+j)$ that cannot be explained by all the intervening variables. Therefore, in settings with decaying (serial) correlation, we expect partial autocorrelations to be zero after a certain lag (i.e., with quite a few intervening variables, there is little remaining correlation between $Y(t)$ and $Y(t+j)$), unlike marginal correlations.

In addition, the domain of the set of partial autocorrelations induced by the PACF is a $[T(T-1)/2]-$dimensional hypercube, so each can vary independently in $(-1, 1)$. This is a major advantage over the autocorrelation function, which is highly restricted. For further intuition on this, we recommend reading about partial correlation vines in the linear algebra literature (references can be found in [18]).

In principle, it is easy to move between the PACF, $\pi(t, t+j)$, and the autocorrelation function, $\rho(t, t+j) = \mathrm{Cor}\{Y(t), Y(t+j)\}$. In particular, a partial autocorrelation, $\pi(t, t+j)$, is a function of the correlation matrix corresponding to components $(Y(t), \dots, Y(t+j))$. The following is the expression for computing the marginal correlations from the partial autocorrelations,

$$\rho(t, t+j) = \pi(t, t+j)D_{t,t+j} + \mathbf{r}_1(t+1, t+j-1)^{\mathrm{T}}\mathbf{R}(t+1, t+j-1)^{-1}\mathbf{r}_3(t+1, t+j-1),$$

where $\mathbf{R}(t+1, t+j-1)$ is the correlation matrix of the vector $\{Y(t+1), \dots, Y(t+j-1)\}$, $\mathbf{r}_1(t+1, t+j-1)^{\mathrm{T}} = (\rho(t, t+1), \dots, \rho(t, t+j-1))$, and $\mathbf{r}_3(t+1, t+j-1)^{\mathrm{T}} = (\rho(t+j, t+1), \dots, \rho(t+j, t+j-1))$. $D_{t,t+j} = \left\{1 - \mathbf{r}_1(t+1, t+j-1)^{\mathrm{T}}\mathbf{R}(t+1, t+j-1)^{-1}\mathbf{r}_1(t+1, t+j-1)\right\}^{1/2} \{1 - \mathbf{r}_3(t+1, t+j-1)^{\mathrm{T}}\mathbf{R}(t+1, t+j-1)^{-1}\mathbf{r}_3(t+1, t+j-1)\}^{1/2}$ is the product of the partial standard deviations [18].

Because T can be large in the irregular longitudinal data setting, moving between $\pi(t, t+j)$ and $\rho(t, t+j)$ can be challenging computationally. For example, inverting $\mathbf{R}(t+1, t+j-1)$ can be time-consuming if j is large. We discuss solutions to these important computational issues in Section 3.

2.2. *PACF models*

2.2.1. Stationary PACF models. We can construct unrestricted PACF models, using Fisher's z-transform, $z(x) = \frac{1+x}{1-x}$ as a link function. To start, we consider a stationary model,

$$z\{\pi(t, t+j)\} = g(j)\mathrm{I}(j \leqslant a), \tag{2}$$

where g is a monotonically non-increasing function of j (time lag), and the indicator function denotes when the lag j PACF is zero. The function $g(j)$ can be formulated as $\mathbf{W}(j)\boldsymbol{\gamma}$, where $\mathbf{W}(j)$ is a design matrix that is a function of the time lag, j; and $\boldsymbol{\gamma}$ is a vector of regression coefficients, which does not depend on t. This is a stationary PACF model, as $\pi(t, t+j)$ only depends on the time lag j, not t.

When $j > a$, $\pi(t, t+j) = 0$, as we expect that $\pi(t, t+j)$ will decrease to zero after a certain number of lags, a, which is generally expected to not be very large in the serial correlation setting. As a result,

© 2015 The Authors. *Statistics in Medicine* Published by John Wiley & Sons Ltd.

this model structure bands the partial autocorrelation matrix, defined as a $T \times T$ matrix with the ones on the main diagonal, and kl and lkth elements set to $\pi(k,l)$ $(k,l \in 1,\dots,T)$, where a is the number of bands.

This set-up is important for computations, as it avoids the need to invert large dimensional matrices (when T is large) and only requires inversion of at most $(a+1)-$dimensional matrices when moving between $\pi(t,t+j)$ and $\rho(t,t+j)$ (see Section 3 and Wang and Daniels [25] for further details). Any correlation matrix of the components of $\{Y_1,\dots,Y_T\}$ will be positive definite under this model.

2.2.2. Semiparametric non-stationary PACF models. We also consider a related non-stationary model,

$$z\{\pi(t,t+j)\} = g_t(j)\mathrm{I}(j \leqslant a), \tag{3}$$

where now the function g is indexed by time t and allows a different rate of decay in partial autocorrelations at different times (thus, non-stationary). The function $g_t(j)$ can be formulated as $\mathbf{W}(j)\boldsymbol{\gamma}(t)$, where $\mathbf{W}(j)$ is a design matrix that is a function of the time lag, j; and $\boldsymbol{\gamma}(t)$ is a vector of smooth functions of time, t.

In the AIDS example presented in Section 4, we specify a non-stationary PACF model as follows:

$$\begin{aligned} z\{\pi(t,t+j)\} &= \{g_{t0} + g_{t1} \times (j-1)\}\mathrm{I}\{j \leqslant a\} \\ &= \{\mathbf{B}(t)\boldsymbol{\gamma}_0 + \mathbf{B}(t)\boldsymbol{\gamma}_1 \times (j-1)\}\mathrm{I}\{j \leqslant a\}, \end{aligned} \tag{4}$$

where $\forall t\ g_{t1} \leqslant 0$ and g_{t0}, g_{t1} are smooth functions of t that are modelled nonparametrically using penalized splines with a low-rank thin-plate basis $\mathbf{B}(t)$, that is, $g_{t0} = \mathbf{B}(t)\boldsymbol{\gamma}_0$, $g_{t1} = \mathbf{B}(t)\boldsymbol{\gamma}_1$ (see details in Section 4). Note that other bases for penalized splines, such as truncated polynomial basis or cubic B-spline basis, can be used. We follow [26] to use the low-rank thin-plate basis because of its better Markov Chain Monte Carlo (MCMC) mixing compared with truncated polynomial basis. In this model, we assume that the partial autocorrelations within band a are decaying linearly as a function of j by restricting $g_{t1} \leqslant 0\ \forall t$. Other more flexible structures in the linear model framework can be specified, for example, by adding a quadratic term of j. However, because the partial autocorrelation matrix under our model is banded at a, two-dimensional surface estimation for $g_t(j)$ in (3) is not very practical.

We will choose the number of bands a using the Deviance Information Criterion (DIC) [27] in the AIDS example, but it is also possible to put a prior on a and obtain its posterior distribution (and integrate over its uncertainty).

2.3. Model for the variance function

In some models (e.g., multivariate probit models), the covariance structure is characterized by a correlation matrix (for identifiability) that would be completely specified with the PACF. In our setting with a covariance function, we also need a model for the marginal variance at each time, $\sigma^2(t)$. This could be modeled as a smooth function of time t by again using penalized splines on the log scale,

$$\log\{\sigma^2(t)\} = \mathbf{B}(t)\boldsymbol{\alpha}. \tag{5}$$

In the AIDS example, we use a parametric model for the marginal variance, as the variance over time reveals a simple linear pattern on the log scale (see details in Section 4).

3. Overview of posterior computations

For the ith individual, we observe the response vector, $\mathbf{Y}_i = (Y_{i1},\dots,Y_{in_i})^{\mathrm{T}}$, at time points $t_{i1},\dots,t_{in_i}$. It follows that

$$\mathbf{Y}_i \sim N(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i), \tag{6}$$

where $\boldsymbol{\mu}_i$ is the mean vector that is determined by the model for $\mu_i(t)$ as in (1).

© 2015 The Authors. *Statistics in Medicine* Published by John Wiley & Sons Ltd.

We decompose the marginal covariance matrix $\mathbf{\Sigma}_i$ such that $\mathbf{\Sigma}_i = \mathbf{S}_i\mathbf{R}_i\mathbf{S}_i$, where $\mathbf{R}_i$ is the marginal correlation matrix, and $\mathbf{S}_i$ is a diagonal matrix with marginal standard deviations along the diagonal. $\mathbf{R}_i$ is determined by the model for $\pi(t, t+j)$ as in (2) or (3), and $\mathbf{S}_i$ is determined by the model for $\sigma^2(t)$ as in (5).

The corresponding log-likelihood from the ith individual is

$$l_i = -n_i \log(2\pi)/2 - \log(|\mathbf{\Sigma}_i|)/2 - (\mathbf{Y}_i - \boldsymbol{\mu}_i)^{\mathrm{T}}\mathbf{\Sigma}_i^{-1}(\mathbf{Y}_i - \boldsymbol{\mu}_i)/2. \tag{7}$$

Posterior sampling via Gibbs sampling is relatively straightforward. The details of the MCMC algorithm for the AIDS example can be found in the Supplementary Materials.

However, the key to the proposed approach involves efficient ways to compute and invert the covariance matrix for each subject, $\mathbf{\Sigma}_i$. We now review some results for correlation matrices based on banded partial autocorrelation matrices that make the computation of $\mathbf{\Sigma}_i^{-1}$ efficient.

Because $\mathbf{R}_i$ can be written directly as a function of the marginal correlation function, $\rho(\cdot, \cdot)$, we need an efficient algorithm to move between $\pi(\cdot, \cdot)$ and $\rho(\cdot, \cdot)$, as T can be quite large. The recursive algorithm proceeds as follows:

(1) $\pi(t, t+1) \rightarrow \rho(t, t+1)$
(2) $\rho(t, t+1), \pi(t, t+2) \rightarrow \rho(t, t+2)$
(3) $\rho(t, t+1), \rho(t, t+2), \pi(t, t+3) \rightarrow \rho(t, t+3)$
$\vdots$
(j) $\rho(t, t+1), \ldots, \rho(t, t+j-1), \pi(t, t+j) \rightarrow \rho(t, t+j)$,

where '$\rightarrow$' corresponds to the parameters to the left of the arrow being needed to compute the parameters to the right of the arrow based on the formula given in Section 2.1. Moving between $\pi(t, t+j)$ and $\rho(t, t+j)$ also involves inverting a $(j-1)$-dimensional correlation matrix for the sub-vector $(Y(t+1), \ldots, Y(t+j-1))$ $(j = 1, \ldots, T-1)$ as was seen in Section 2.1. However, the classes of models proposed in Section 2.2 reduces the computational burden by limiting the dimension of matrices that need to be inverted.

Wang and Daniels [25] provide a result for inversion of a-band PACF matrices.

Result 1 [25]: Inverting the correlation matrix constructed from an a-band partial autocorrelation matrix only requires the inversion of $(a+1)$-dimensional matrices, and its precision matrix is also an a-band matrix.

This result illustrates that these matrix inversions only require inversion of at most $(a+1)$-dimensional matrices (even for $j > a$), which is essential when T is large. The expression for inverting these matrices and the proof of this result can be found in the Supplementary Materials to [25].

4. Application to the AIDS pediatric trial data

In this section, we use the proposed methods to analyze the AIDS pediatric trial data introduced in Section 1. Recall that the analysis objective is to compare the treatment-specific longitudinal trajectories of CD4 counts. To deal with the informative dropout problem, we use conditional linear models (CLMs) from [28] in the mean structure because dropout occurred in continuous time (Figure 2), and pattern mixture models [29] for discrete dropout times are not suitable here. Similar to pattern mixture models, the joint distribution of the outcome and dropout time in CLMs is factorized as the marginal distribution of the dropout time times the conditional distribution of the outcome given the dropout time. To obtain marginal covariate effects in the mean structure, we do not model the dropout time distribution but use the Rubin's Bayesian bootstrap [30, 31] for averaging over the dropout time distribution.

For the covariance structure, we fit several models to compare their fits to these CD4 data, while the same CLM is assumed for the mean structure. The first is a linear mixed effects model (LMM) with random intercepts and random slopes, which is a simple but often reasonable approach. The marginal correlation structure induced by this model is non-stationary but follows a specific parametric form depending on the covariance matrix of the random effects (see page 133 in [32]). Given the large number of unique measurement times and the resulting lack of replications at many times and lags in these CD4 count data, directly estimating an unstructured marginal covariance matrix is not feasible. In addition, we would rather not make strong parametric assumptions about the covariance structure. Thus, we apply our proposed semiparametric PACF model to these CD4 data. In preliminary analyses, we found that the partial autocorrelations were effectively non-zero even at a relatively large number of time lags,

© 2015 The Authors. *Statistics in Medicine* Published by John Wiley & Sons Ltd.

which suggests that there was a non-diminishing correlation over time in these data. In addition, we compared model fits for the models with and without random intercepts (to allow for long-term correlation) using DIC (with the random intercepts integrated out) and found that the model with random intercepts provided a better fit to the observed data. We therefore include a random intercept to account for the non-diminishing correlation and use our PACF models to model the serial correlation. We fit both semiparametric and parametric PACF models with different numbers of bands and compare their fits to the observed data using DIC.

Because the maximum of the observed measurement times was 219 weeks, in order to simplify the computations to a more manageable level, we round the actual measurement times to the nearest 4 weeks in the following analyses. Therefore, the maximum follow-up time (i.e., T) based on the measurement times is 56. In Section 5, we provide more discussions of the computations when the dimension increases.

4.1. Conditional linear model for the mean structure

Let $\mathbf{Y}_i(t)$ represent the square root of CD4 count at time t for the ith child. In our proposed PACF model, we assume that $\mathbf{Y}_i(t)$ is a discrete-time Gaussian process with a mean function $\mu_i^{(1)}(t)$,

$$\mu_i^{(1)}(t) = \beta_0(d_i) + \beta_1(d_i) \cdot (t-1)/13 + \beta_2(d_i) \cdot \text{dose}_i + \beta_3(d_i) \cdot \text{dose}_i \cdot (t-1)/13 + b_{i0}, \tag{8}$$

where $b_{i0} \sim N\left(0, \sigma_b^2\right)$ is a random intercept, dose_i is an indicator variable for the low-dose group, and d_i is the observed dropout time in weeks (rescaled by taking $(d_i - \bar{d})/\text{range}(d_i)$, where $\bar{d}$ is the sample mean of the observed dropout times). The unit for time t is 4 weeks, and $t = 1, \ldots, 56$, where $t = 1$ corresponds to enrollment. We scale the time t by 13 such that $\beta_1(d_i)$ represents the change rate of CD4 counts over a year, given d_i in the high-dose group.

Following the approach in Wu and Bailey [28], we assume that regression coefficients $\beta_0(d_i)$, $\beta_1(d_i)$, $\beta_2(d_i)$, and $\beta_3(d_i)$ in (8) are functions of the observed dropout time d_i. Because Figure 3 shows the local regression fits to the individual ordinary least-squares intercepts and slopes that are not far apart from linear patterns, we assume that $\beta_0(d_i) = \theta_{00} + \theta_{01}d_i$, $\beta_1(d_i) = \theta_{10} + \theta_{11}d_i$, $\beta_2(d_i) = \theta_{20} + \theta_{21}d_i$, and $\beta_3(d_i) = \theta_{30} + \theta_{31}d_i$.

In the LMM, we have the same conditional linear model for the mean as in (8) but add a random slope as follows

$$\mu_i^{(2)}(t) = \mu_i^{(1)}(t) + b_{i1} \cdot (t-1)/13, \tag{9}$$

where

$$\begin{bmatrix} b_{i0} \\ b_{i1} \end{bmatrix} \sim N\left(\mathbf{0}, \mathbf{G} = \begin{bmatrix} \sigma_0^2 & \sigma_{01} \\ \sigma_{01} & \sigma_1^2 \end{bmatrix}\right)$$

are random effects for the intercept and time slope for the ith child.

4.2. Models for the covariance structure

For the LMM, we assume that the (residual) covariance function is $\text{Cov}\{Y_i(t), Y_i(t')\} = \sigma_\epsilon^2 \text{I}(t = t')$.

In the semiparametric PACF model, we assume that

$$\begin{aligned} z\{\pi_i(t, t+j)\} &= \{g_{t0} + g_{t1} \times (j-1)\}\text{I}(j \leqslant a) \\ &= \{\mathbf{B}(t/50)\boldsymbol{\gamma}_0 + \mathbf{B}(t/50)\boldsymbol{\gamma}_1 \times (j-1)\}\text{I}(j \leqslant a), \end{aligned}$$

where $\forall t\ g_{t1} \leqslant 0$, g_{t0} and g_{t1} are smooth functions that are modelled nonparametrically using penalized splines with a low-rank thin-plate basis $\{B(t)\} = \{1, t, |t - \nu_1|^3, \ldots, |t - \nu_{10}|^3\}$, and $\nu_1 < \ldots < \nu_{10}$ are 10 equally spaced fixed knots that are set at $t = 5, 10, \ldots, 50$. For example, the low-rank thin-plate spline representation of g_{t0} is

$$g_{t0} = \gamma_0^0 + \gamma_1^0 \cdot t/50 + \sum_{k=1}^{10} \gamma_{k+2}^0 \cdot |t/50 - \nu_k/50|^3, \tag{10}$$

© 2015 The Authors. *Statistics in Medicine* Published by John Wiley & Sons Ltd.

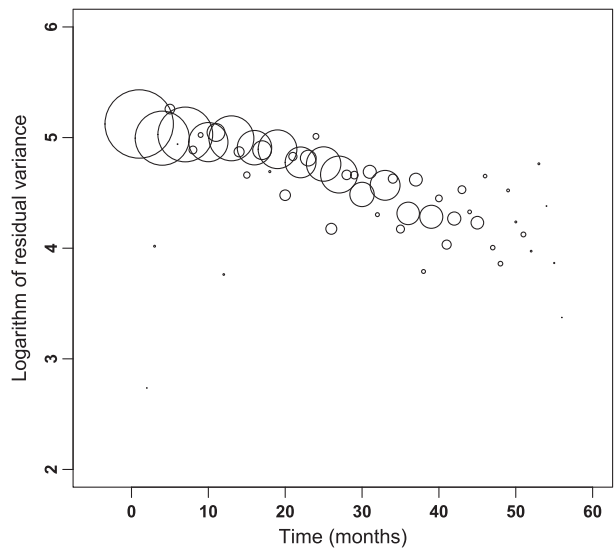

Figure 4. Logarithm of the sample residual variance (from a simple linear model fit to the square root CD4 data) for each time point $t = 1, \ldots, 56$ in the AIDS example. The size of the circles is proportional to the number of observations at that time.

where $\boldsymbol{\gamma}_0 = \left(\gamma_0^0, \ldots, \gamma_{12}^0\right)^{\mathrm{T}}$ is a vector of regression coefficients. Similarly, we have $g_{t1} = \mathbf{B}(t/50)\boldsymbol{\gamma}_1$ and $\boldsymbol{\gamma}_1 = \left(\gamma_0^1, \ldots, \gamma_{12}^1\right)^{\mathrm{T}}$. Note that, to facilitate the numerical computations, we scale t by 50 when constructing the penalized spline basis.

For comparison, we also fit a stationary PACF model of the following form:

$$\begin{aligned} z\{\pi_i(t, t+j)\} &= g(j)\mathrm{I}(j \leqslant a) \\ &= \{\gamma_0 + \gamma_1 \times (j-1)\}\mathrm{I}(j \leqslant a). \end{aligned}$$

For the variance function $\sigma_i^2(t)$, we use a parametric model,

$$\log\left\{\sigma_i^2(t)\right\} = \alpha_0 + \alpha_1 t/50, \tag{11}$$

because the logarithm of the residual variances (from a simple linear model fit to these data), over time reveals a linear pattern (Figure 4).

4.3. Summarizing marginal covariate effects

Following [31], we leave the dropout-time distribution completely unspecified and use the Rubin's Bayesian bootstrap [30] to obtain the posterior for $\mathrm{P}(D = d_i)$ (see details in [31]), where D is the dropout time. Basically, at each iteration of the MCMC, we simulate $\mathrm{P}(D = d_i)$ from Dirichlet$(1, \ldots, 1)$ for both dose groups. The marginal covariate effects can be approximated by $\sum_{i=1}^{N} \mathrm{P}(D = d_i)\boldsymbol{\beta}(d_i)$, where $\boldsymbol{\beta}(d_i) = (\beta_0(d_i), \beta_1(d_i), \beta_2(d_i), \beta_3(d_i))^{\mathrm{T}}$. Details can be found in the Supplementary Materials.

4.4. Prior specification and MCMC algorithm

For the mean structure, we specify independent Normal priors $N(0, 10^3)$ for θ_{00}, θ_{01}, θ_{10}, θ_{11}, θ_{20}, θ_{21}, θ_{30}, θ_{31}.

In the LMM, we use the modified Cholesky decomposition in [33] for modeling the random effect covariance structure, that is, we assume that $b_{i1} = \lambda b_{i0} + e_i$ in (9), where $e_i \sim N(0, \sigma_e^2)$. The reparameterization through λ, σ_0, and σ_e will guarantee that the covariance matrix of the random effects is positive definite. We assign Uniform$(0, 20)$ prior to the standard deviations σ_0 and σ_e, and $N(0, 10^3)$ prior to λ. Note that $\sigma_1^2 = \lambda^2\sigma_0^2 + \sigma_e^2$, and we can obtain $\sigma_{01} = \lambda\sigma_0^2$. Further, for the (residual) error variance, we assign the prior $\sigma_\epsilon^2 \sim$ inverse-Gamma$(0.01, 0.01)$.

© 2015 The Authors. *Statistics in Medicine* Published by John Wiley & Sons Ltd.

For both non-stationary and stationary PACF models, we specify independent Normal priors $N(0, 10^3)$ for the parameters in the variance function α_0 and α_1 and assign a Uniform(0, 20) prior to the random intercept standard deviation σ_b.

In addition, for the stationary PACF model, we assign the following priors for γ_0 and γ_1: $\gamma_0 \sim N(0, 10^3)$, $\gamma_1 \sim N(0, 10^3)$.

Let $\boldsymbol{\xi}_0 = \left(\gamma_0^0, \gamma_1^0\right)^{\mathrm{T}}$, $\boldsymbol{\psi}_0 = \left(\gamma_3^0, \ldots, \gamma_{12}^0\right)^{\mathrm{T}}$, $\mathrm{T}_1 = (1, t/50)$, $\mathrm{T}_2 = (|t/50 - v_1/50|^3, \ldots, |t/50 - v_{10}/50|^3)$, and $\boldsymbol{\Omega}$ be a 10×10 penalty matrix whose (l, k)th entry is $|v_l/50 - v_k/50|^3$. Using the reparameterization $\tilde{\boldsymbol{\psi}}_0 = \boldsymbol{\Omega}^{1/2}\boldsymbol{\psi}_0$ and $\tilde{\mathrm{T}}_2 = \mathrm{T}_2\boldsymbol{\Omega}^{-1/2}$, (10) can be rewritten as $g_{t0} = \mathrm{T}_1\boldsymbol{\xi}_0 + \tilde{\mathrm{T}}_2\tilde{\boldsymbol{\psi}}_0$. Note that this reparameterization is useful because we can then assign an independent Normal prior to $\tilde{\boldsymbol{\psi}}_0$ [26]. Similarly, we can define $\boldsymbol{\xi}_1 = \left(\gamma_0^1, \gamma_1^1\right)^{\mathrm{T}}$, $\boldsymbol{\psi}_1 = \left(\gamma_3^1, \ldots, \gamma_{12}^1\right)^{\mathrm{T}}$, then $\tilde{\boldsymbol{\psi}}_1 = \boldsymbol{\Omega}^{1/2}\boldsymbol{\psi}_1$ and we have $g_{t1} = \mathrm{T}_1\boldsymbol{\xi}_1 + \tilde{\mathrm{T}}_2\tilde{\boldsymbol{\psi}}_1$.

We assign to $\boldsymbol{\xi}_0$, $\boldsymbol{\xi}_1$ independent Normal priors with a mean zero and a large variance and to $\tilde{\boldsymbol{\psi}}_0$ and $\tilde{\boldsymbol{\psi}}_1$ the prior $N\left(\mathbf{0}, \sigma_{\gamma_0}^2 \cdot \mathrm{I}_{10\times10}\right)$ and $N\left(\mathbf{0}, \sigma_{\gamma_1}^2 \cdot \mathrm{I}_{10\times10}\right)$, respectively. Estimating the smoothing parameters $\sigma_{\gamma_0}^2$ and $\sigma_{\gamma_1}^2$ is similar to estimating variance components in Bayesian hierarchical models [34], and the curve estimation by penalized splines can be sensitive to the choice of prior for $\sigma_{\gamma_0}^2$ and $\sigma_{\gamma_1}^2$. Crainiceanu *et al.* [35] discussed this issue and found that inverse-Gamma priors can be used in practice when certain conditions are met, such that the posterior inference of $\sigma_{\gamma_0}^2$ and $\sigma_{\gamma_1}^2$ is insensitive to the hyper-parameters in the prior for $\sigma_{\gamma_0}^2$ and $\sigma_{\gamma_1}^2$. In our application, we use Uniform(0, 20) priors for σ_{γ_0} and σ_{γ_1}.

Because the MCMC algorithm used in the AIDS example involves numerous matrix operations, we decided to use MATLAB (MathWorks, Natick, MA, USA) (instead of R) due to its greater efficiency in matrix operations. The MATLAB code can be run in an open-source alternative QtOctave under Linux. The MATLAB code for fitting the non-stationary PACF models in the AIDS example is available at http://www.mrc-bsu.cam.ac.uk/software/miscellaneous-software/.

Two separate chains were run for each model. Convergence (checked using trace plots) was reached at about 2000 iterations; and pooled samples of 20,000 after convergence were used for inference.

4.5. Results

4.5.1. Model selection and assessment. In Table I, we compare fitted models using DIC based on marginal likelihood with the random effects integrated out. The parameterization used for non-stationary PACF models is $\theta_{00}, \theta_{01}, \theta_{10}, \theta_{11}, \theta_{20}, \theta_{21}, \theta_{30}, \theta_{31}$ $\boldsymbol{\xi}_0, \boldsymbol{\xi}_1, \tilde{\boldsymbol{\psi}}_0, \tilde{\boldsymbol{\psi}}_1, \alpha_0, \alpha_1, 1/\sigma_b^2$. For parametric stationary PACF model, we replace $\boldsymbol{\xi}_0, \boldsymbol{\xi}_1, \tilde{\boldsymbol{\psi}}_0, \tilde{\boldsymbol{\psi}}_1$ with γ_0 and γ_1. For the LMM, the parameterization is $\theta_{01}, \theta_{10}, \theta_{11}, \theta_{20}, \theta_{21}, \theta_{30}, \theta_{31}, \mathbf{G}^{-1}, \sigma_\epsilon^2$.

Table I. DIC values from the fitted models in the AIDS example. $\bar{D}$ is the posterior mean of the deviance ($-2\log L$), $\hat{D}$ is the deviance by substituting in the posterior means of parameters, $p_D = \bar{D} - \hat{D}$ is the effective number of parameters and DIC $= \bar{D} + p_D$.

		$\bar{D}$	$\hat{D}$	p_D	DIC
Linear mixed model		30345	30333	12	30357
Stationary PACF model	Band (a)				
	1	30437	30425	12	30449
	2	30187	30175	13	30200
	3	29997	29984	13	30010
	4	29951	29938	13	29964
	5	29955	29942	13	29968
	6	29986	29973	13	29999
Non-stationary PACF model	Band (a)				
	1	30405	30387	18	30424
	2	30148	30127	21	30169
	3	29968	29948	20	29988
	4	29911	29890	21	29933
	5	29903	29880	22	**29925**
	6	29922	29899	23	29945

© 2015 The Authors. *Statistics in Medicine* Published by John Wiley & Sons Ltd.

All PACF models except those with one band had smaller DIC values than the LMM fit. The non-stationary PACF model with five bands gives the smallest DIC; therefore, we will present the results based on this model. Note that for each number of bands considered, the non-stationary PACF model results in a smaller DIC than the corresponding stationary PACF model.

To assess the fit of the best model chosen via DIC, we use posterior predictive checks based on replicated observed data as recommended in Daniels *et al.* [36] and a χ^2 discrepancy described in Gelman *et al.* [37]. Specifically, the steps are the following:

(1) Sample a replicated dropout time, D^{rep}, from the empirical distribution of the observed dropout times.
(2) Sample from the empirical distribution of the gap times between irregular measurement times and construct the replicated measurement times up to D^{rep}.
(3) Sample a set of responses $\mathbf{Y}_i^{rep}$ at those measurement times given D^{rep} and the current posterior sample of the parameters.
(4) Repeat Steps 1–3 for all $N = 421$ subjects.

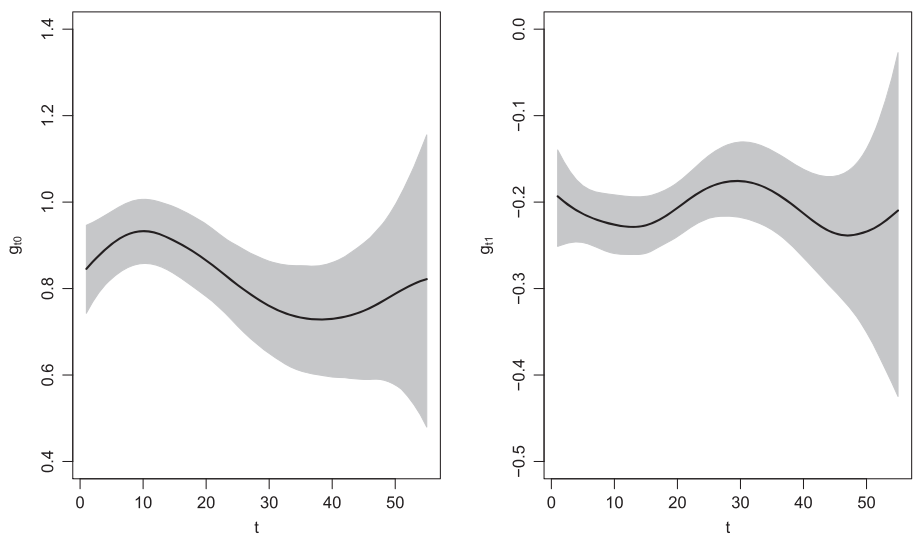

Figure 5. Estimated g_{t0} and g_{t1} functions (posterior median and pointwise 95% credible band) from the fitted non-stationary 5-band PACF model for the AIDS example.

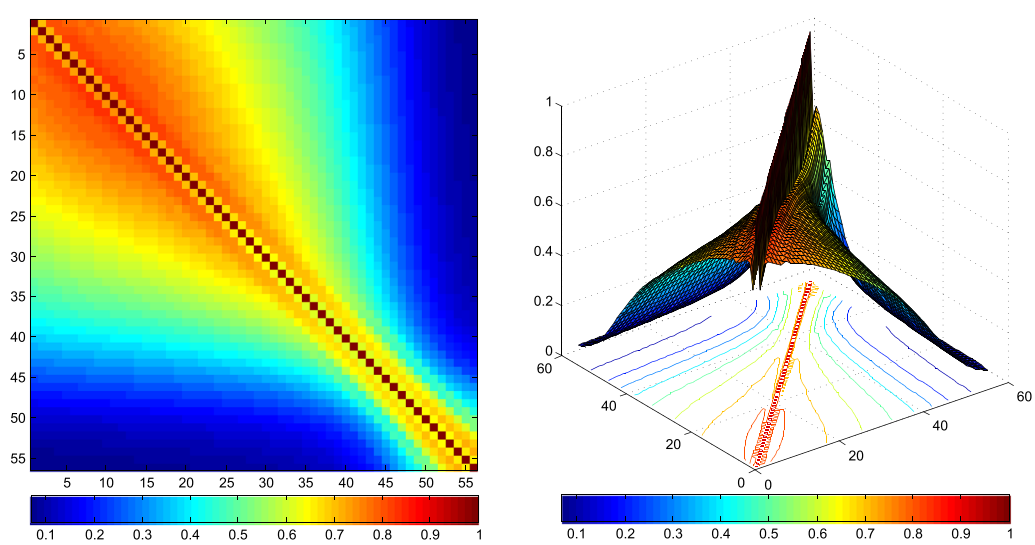

Figure 6. Image plot and surface plot of the estimated 56×56 marginal correlation matrix (based on posterior medians of the parameters) from the fitted non-stationary 5-band PACF model for the AIDS example.

© 2015 The Authors. *Statistics in Medicine* Published by John Wiley & Sons Ltd.

(5) Compute the χ^2 discrepancy for the replicated data,

$$\sum_{i=1}^{421} \left(\mathbf{Y}_i^{rep} - \boldsymbol{\mu}_i\right)^{\mathrm{T}} \boldsymbol{\Sigma}^{-1} \left(\mathbf{Y}_i^{rep} - \boldsymbol{\mu}_i\right) / n^{rep},$$

where n^{rep} is the total number of replicated responses, μ_i is the mean given in (8) with the random effects integrated out, and $\boldsymbol{\Sigma}$ is the marginal covariance structure after integrating out the random effects.

(6) Compute the χ^2 discrepancy for the observed data.

(7) Repeat Steps 1–6 for each posterior sample and compute the posterior predictive probability that the replicated χ^2 statistic is larger than the observed χ^2 statistic.

The posterior probability that the replicated χ^2 statistic is larger than the observed χ^2 statistic is 0.50, which provides no evidence for the lack of fit of the non-stationary PACF model with five bands to the observed data.

4.5.2. Covariance structure. Figure 5 shows the estimated functions g_{t0} and g_{t1} that determine the non-stationarity of the correlation structure in the fitted five-band non-stationary PACF model. A stationary structure would correspond to both these functions being constant over time. The deviations from stationarity are apparent when $t \in [0, 30]$. Figure 6 displays image and surface plots of the correlation structure induced by our PACF model and demonstrates the lack of stationarity as well. In particular, we can see the larger short lag correlations and slower decay at the earlier times versus the later times.

The large estimate for random intercept variance σ_b^2 (Table II) indicates large heterogeneity in terms of the overall CD4 count level across children. For the variance function, there is a significant decrease

Table II. Posterior medians and 95% credible intervals for the parameters of the fitted LMM and non-stationary 5-band PACF model as well as the marginal covariate effects inferred from these models.

	Linear mixed model			**5-band PACF model**		
	Median	2.5%	97.5%	Median	2.5%	97.5%
θ_{00}	30.50	28.96	32.02	30.38	28.69	32.07
θ_{01}	12.87	6.48	19.42	13.12	6.33	19.81
θ_{10}	−4.97	−5.48	−4.45	−4.82	−5.33	−4.31
θ_{11}	8.18	5.55	10.80	7.41	5.05	9.82
θ_{20}	−1.75	−3.86	0.48	−1.56	−3.88	0.79
θ_{21}	1.26	−8.24	10.63	2.24	−7.17	11.79
θ_{30}	1.02	0.28	1.76	1.03	0.30	1.76
θ_{31}	−1.11	−4.86	2.59	−2.21	−5.70	1.18
σ_ϵ	3.94	3.86	4.02			
σ_0	11.56	10.80	12.42			
σ_{01}	−22.62	−27.86	−18.22			
σ_1	3.16	2.89	3.46			
α_0				4.65	4.48	4.81
α_1				−1.76	−2.10	−1.38
σ_b				8.01	7.12	8.91
	Marginal covariate effects					
	Linear mixed model			**5-band PACF model**		
	Median	2.5%	97.5%	Median	2.5%	97.5%
Intercept	30.28	28.67	31.85	30.15	28.42	31.90
Time	−5.11	−5.71	−4.52	−4.95	−5.55	−4.38
Dose	−1.39	−3.62	0.94	−1.19	−3.62	1.23
Interaction	1.23	0.40	2.06	1.21	0.41	2.01

© 2015 The Authors. *Statistics in Medicine* Published by John Wiley & Sons Ltd.

over time (on the log scale) by noticing that the 95% credible interval for the slope, $(-2.10, -1.38)$, does not contain zero (also see Figure 4).

4.5.3. Mean structure. Although the same CLM is applied to both the LMM and the non-stationary five-band PACF model, the parameter estimates in the mean structure differ in these fits (Table II). For example, the parameter θ_{11} quantifies the degree of association of CD4 change rate from baseline to the maximum follow-up with the dropout time in the high-dose group. The large positive estimates from both model fits suggest that later dropouts were associated with a slower decline rate of CD4 count over time. However, in the LMM, this estimated association (8.18 vs. 7.41) is stronger than in the non-stationary five-band PACF model. As a result, the LMM imposes a larger adjustment for selection bias due to dropout in the marginal time slope estimate of the high-dose group (-5.11 vs. -4.95) than the non-stationary PACF model.

On the other hand, θ_{31} quantifies the degree of association of the interaction between time and dose effects with the dropout time. Both model fits show no evidence of this association although the point estimates are quite different (-1.11 vs. -2.21). Therefore, we expect that in both model fits, the adjustments for selection bias in marginal time slopes will be similar for both dose groups. Consequently, the marginal interactions between time and dose effects are similar in both model fits (1.23 vs 1.21).

Based on the non-stationary five-band PACF model fit (the best fitting model by DIC), there is a significant difference in the slopes between the two treatments (doses) with the 95% credible interval for the difference of $(0.41, 2.01)$, indicating that the rate of decline in CD4 counts is less severe under the low-dose arm.

Overall, our analyses of these CD4 data demonstrate that models for the covariance structure can influence the inference for the mean structure under complex scenarios with informative dropout and irregular measurement times.

5. Discussion

We have proposed a flexible class of semiparametric, non-stationary models for the covariance structure in a Gaussian process for irregular continuous longitudinal data. The analysis of the CD4 count data showed that modeling covariance structures can impact the inference of the mean structure in complex situations with informative dropout and irregular measurement times. Our models will be useful when careful modeling of the covariance structure is required to ensure valid inference of the mean structure.

In the AIDS data, to make the MCMC algorithm quicker, we aggregated the weekly data to months to change the dimension of T from 220 to 56. Note that the issue is not matrix inversions as for an a-band PACF, only $(a+1)$-dimensional matrices need to be inverted. However, as the dimension increases, there are more matrices that need to be inverted to move from the PACF to the autocorrelation function. For example, in our computations, converting a three-band 56×56 partial autocorrelation matrix to a marginal correlation matrix took 0.25 s in MATLAB 7.1 (2.59GHz CPU, 32GB RAM on PC). However, to convert a three-band 220×220 matrix, it took 40 s. This is less of an issue with more powerful High Performance Computing Clusters, as much of this conversion can be parallelized. Therefore, actual recorded measurement times can be used when applying the proposed PACF models in High Performance Computing Clusters environments.

As with other approaches for dealing with informative dropout, sensitivity analysis is required to assess the unverifiable assumption used in our CLM for the mean structure. In particular, the CLM assumes that the slope before dropout is the same as the slope after dropout; the latter is not identifiable from the observed data. Strategies such as in [31] can be adopted for sensitivity analysis.

The proposed PACF models can be extended to categorical and count data by specifying them for a Gaussian process in the mean function (e.g., [38]). For example,

$$h\left[\mathrm{E}\{\mathbf{Y}_i(t) \mid \mathbf{X}_i(t), S(t)\}\right] = \mathbf{X}_i(t)\boldsymbol{\beta} + S(t),$$

where h is a link function, $\{S(t)\}$ is a zero-mean Gaussian process with a covariance function $\mathrm{Cov}(S(t), S(t+j)) = \sigma(t)\sigma(t+j)\rho(t, t+j)$, and the proposed PACF models can be applied to model $\rho(\cdot,\cdot)$. Covariates can be incorporated into the proposed PACF models; we are currently working on this extension.

© 2015 The Authors. *Statistics in Medicine* Published by John Wiley & Sons Ltd.

Acknowledgements

This work was supported by the Medical Research Council [Unit Programme number U105261167] and U.S. NIH grants R01-CA85295 and R01-CA183854.

References

1. Lipsitz S, Fitzmaurice G, Ibrahim J, Gelber R, Lipshultz S. Parameter estimation in longitudinal studies with outcome-dependent follow-up. *Biometrics* 2002; **58**:621–630.
2. Fitzmaurice GM, Lipsitz SR, Ibrahim JG, Gelber R, Lipshultz S. Estimation in regression models for longitudinal binary data with outcome-dependent follow-up. *Biostatistics* 2006; **7**(3):469–485.
3. Ryu D, Sinha D, Mallick B, Lipsitz SL, Lipshultz S. Longitudinal studies with outcome-dependent follow-up: models and Bayesian regression. *Journal of the American Statistical Association* 2007; **102**:952–967.
4. Little R, Rubin D. *Statistical Analysis with Missing Data* (2nd edition). John Wiley & Sons: New York, 2002.
5. Daniels M, Hogan J. *Missing Data in Longitudinal Studies: Strategies for Bayesian Modeling and Sensitivity Analysis*, Monographs on Statistics and Applied Probability, vol. 101. Chapman & Hall/CRC: New York, 2008.
6. Muñoz A, Carey V, Schouten J, Segal M, Rosner B. A parametric family of correlation structures for the analysis of longitudinal data. *Biometrics* 1992; **48**:733–742.
7. Qian L. Bayesian semiparametric correlation models for longitudinal data with application to an HIV/AIDS biomarker study. *PhD Thesis*, UCLA 2009.
8. Pan J, MacKenzie G. On modelling mean-covariance structures in longitudinal studies. *Biometrika* 2003; **90**(1):239–244.
9. Pourahmadi M. Joint mean-covariance models with applications to longitudinal data: unconstrained parameterisation. *Biometrika* 1999; **86**(3):677–690.
10. Zhang W, Leng C. A moving average Cholesky factor model in covariance modelling for longitudinal data. *Biometrika* 2012; **99**(1):141–150.
11. Taylor J, Cumberland W, Sy J. A stochastic model for analysis of longitudinal AIDS data. *Journal of the American Statistical Association* 1994; **89**(427):727–736.
12. Wang Y, Taylor J. Jointly modeling longitudinal and event time data with application to acquired immunodeficiency syndrome. *Journal of the American Statistical Association* 2001; **96**(455):895–905.
13. Diggle P, Verbyla A. Nonparametric estimation of covariance structure in longitudinal data. *Biometrics* 1998; **54**:401–415.
14. Fan J, Huang T, Li R. Analysis of longitudinal data with semiparametric estimation of covariance function. *Journal of the American Statistical Association* 2007; **102**(478):632–641.
15. Yao F, Müller HG, Wang JL. Functional data analysis for sparse longitudinal data. *Journal of the American Statistical Association* 2005; **100**(470):577–590.
16. Pourahmadi M. *Foundations of Time Series Analysis and Prediction Theory*. Wiley: New York, 2001.
17. Box G, Jenkins G, Reinsel G. *Time Series Analysis, Forecasting and Control* (4th edition). Wiley: Hoboken, N.J, 2008.
18. Daniels M, Pourahmadi M. Modeling covariance matrices via partial autocorrelations. *Journal of Multivariate Analysis* 2009; **100**(10):2352–2363.
19. Wang Y, Daniels M. Bayesian modeling of the dependence in longitudinal data via partial autocorrelations and marginal variances. *Journal of Multivariate Analysis* 2013; **116**:130–140.
20. Zhang W, Leng C, Tang C. A joint modelling approach for longitudinal studies. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 2015; **10**(1):219–238.
21. Zimmerman D, Nunez-Anton V. *Antedependence Models for Longitudinal Data*, Vol. 112. CRC Press: Hoboken, N.J, 2010.
22. Brady MT, McGrath N, Brouwers P, Gelber R, Fowler MG, Yogev R, Hutton N, Bryson YJ, Mitchell CD, Fikrig S, Borkowsky W, Jimenez E, McSherry E, Rubinstein A, Wilfert CM, McIntosh K, Elkins MM, Weintrub PS. Randomized study of the tolerance and efficacy of high-versus low-dose zidovudine in human immunodeficiency virus-infected children with mild to moderate symptoms (AIDS Clinical Trials Group 128). *Journal of Infectious Diseases* 1996; **173**(5): 1097–1106.
23. Hogan J, Lin X, Herman B. Mixtures of varying coefficient models for longitudinal data with discrete or continuous nonignorable dropout. *Biometrics* 2004; **60**:854–864.
24. Ruppert D, Wand M, Carroll R. *Semiparametric Regression*. Cambridge University Press: Cambridge, 2003.
25. Wang Y, Daniels M. Computationally efficient banding of large covariance matrices for ordered data and connections to banding the inverse Cholesky factor. *Journal of Multivariate Analysis* 2014; **130**:21–26.
26. Crainiceanu C, Ruppert D, Wand M. Bayesian analysis for penalized spline regression using WinBUGS. *Journal of Statistical Software* 2005; **14**(14):1–24.
27. Spiegelhalter D, Best N, Carlin B, van der Linde A. Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society. Series B* 2002; **64**:583–639.
28. Wu M, Bailey K. Estimation and comparison of changes in the presence of informative right censoring: conditional linear model (corr:v46 p889). *Biometrics* 1989; **45**:939–955.
29. Little R. Pattern-mixture models for multivariate incomplete data. *Journal of the American Statistical Association* 1993; **88**:125–134.
30. Rubin D. The Bayesian bootstrap. *Annals of Statistics* 1981; **9**:130–134.
31. Su L, Hogan J. Varying-coefficient models for longitudinal processes with continuous-time informative dropout. *Biostatistics* 2010; **11**(1):93–110.
32. Diggle P, Heagerty P, Liang KY, Zeger S. *Analysis of Longitudinal Data*. Oxford University Press: New York, 2002.
33. Daniels M, Zhao Y. Modelling the random effects covariance matrix in longitudinal data. *Statistics in Medicine* 2003; **22**:1631–1647.

© 2015 The Authors. *Statistics in Medicine* Published by John Wiley & Sons Ltd.

34. Gelman A. Prior distributions for variance parameters in hierarchical models. *Bayesian Analysis* 2005; **1**(2):1–19.
35. Crainiceanu C, Ruppert D, Carroll RJ. Spatially adaptive penalized splines with heteroscedastic errors. *Journal of Computational and Graphical Statistics* 2007; **16**:265–288.
36. Daniels M, Chatterjee A, Wang C. Bayesian model selection for incomplete data using the posterior predictive distribution. *Biometrics* 2012; **68**:1055–1063.
37. Gelman A, Meng X, Stern H. Posterior predictive assessment of model fitness via realized discrepancies. *Statistica Sinica* 1996; **6**:733–807.
38. Zeger S, Diggle P. Semiparametric models for longitudinal data with application to CD4 cell numbers in HIV seroconverters. *Biometrics* 1994; **50**:689–699.

Supporting information

Additional supporting information may be found in the online version of this article at the publisher's web site.

© 2015 The Authors. *Statistics in Medicine* Published by John Wiley & Sons Ltd.