Censored mixed-effects models for irregularly observed repeated measures with applications to HIV viral loads

Larissa A. Matos[a], Luis M. Castro[b] and Victor H. Lachos[a*]

[a] Department of Statistics, Campinas State University, Campinas, São Paulo, Brazil

[b] Department of Statistics, Universidad de Concepción, Concepción, Chile

Abstract

In some AIDS clinical trials, the HIV-1 RNA measurements are collected irregularly over time and are often subject to some upper and lower detection limits, depending on the quantification assays. Linear and nonlinear mixed-effects models, with modifications to accommodate censored observations, are routinely used to analyze this type of data Vaida & Liu (2009); Matos *et al.* (2013a). This paper presents a framework for fitting LMEC/NLMEC with response variables recorded at irregular intervals. To address the serial correlation among the within-subject errors, a damped exponential correlation structure is considered in the random error and an EM-type algorithm is developed for computing the maximum likelihood estimates, obtaining as a byproduct the standard errors of the fixed effects and the likelihood value. The proposed methods are illustrated with simulations and the analysis of two real AIDS case studies.

Key words: Censored data; EM Algorithm; HIV viral load; Influential observations; Linear/nonlinear mixed models

1. Introduction

Nowadays, study of acquired immunodeficiency syndrome (AIDS) and understanding of the dynamic of the human immunodeficiency virus (HIV) have become the focus of biomedical and biostatistical research. As mentioned by many researchers, HIV is an extremely dynamic and variable virus having new subtypes

* *Address for correspondence*: Víctor Hugo Lachos Dávila, Departamento de Estatística, IMECC, Universidade Estadual de Campinas, CEP 13083-859, Campinas, São Paulo, Brazil. E-mail: hlachos@ime.unicamp.br.

and recombinant forms, about which the scientific community knows little or nothing.

HIV/AIDS clinical trails are research studies with the aim to find new ways to prevent, detect and/or treat AIDS by determining whether a new anti-retroviral (ARV) agent/therapy is safe and effective in people. Most of these clinical trials assess the quantitative rates/changes of HIV-1 ribonucleic acid (RNA) levels in plasma (or simply HIV-1 viral load), since is an important surrogate marker to assess the risk of disease progression and to monitor response to ARV therapy in routine medical care of infected patients.

However, modeling HIV-1 viral load presents many challenges from the statistical point of view. Three are of particular importance. First, the viral load measurements are often left or right censored (undetected) due to a lower and/or upper detection limit of quantification (100-500 copies/ml). This is because some quantification assays cannot accurately quantify HIV-1 RNA above/below a specific level. Second, as a result of unscheduled follow-up visits of patients and/or some missing preplanned responses, the viral load are usually recorded at irregular intervals. Finally, since the viral load is measured longitudinally over time, the between-subject and within-subject variations have to be taken into account.

Recently, some alternatives for modeling the irregular observation responses and correlations induced by longitudinal data have been proposed in the statistical literature. These proposals consider not only the correlation structure induced by the random effects term but also by other types of correlation in the error term. Particularly, Wang (2013) proposed a multivariate Student-t linear mixed model (t-LMM) for outcome variables recorded on irregular occasions considering a damping exponential correlation (DEC) structure Muñoz *et al.* (1992). This correlation structure takes into account the autocorrelation generated by the within-subject dependence among irregular occasions. Moreover, Wang & Fan (2011) considered the multivariate Student-t linear mixed with autoregressive of order p (AR(p)) dependence structure for the within-subject errors in the case of multiple outcomes.

In the case of censored responses, there are several alternatives proposed in the literature to deal with them in the context of linear/nonlinear mixed-effects (LME/NLME) models. For example, Hughes (1999) proposed a likelihood-based Monte Carlo EM algorithm (MCEM) for LME with censored responses (LMEC). In turn, Vaida *et al.* (2007); Vaida & Liu (2009) extended the work of Hughes, proposing a more efficient EM algorithm than Hughes's algorithm. Recently, Matos

et al. (2013a), Matos *et al.* (2013b) and Bandyopadhyay *et al.* (2014) proposed a likelihood-based estimation and influence analysis for LMEC/NLMEC models, respectively. However, to our knowledge there is no work considering irregular observations, damping exponential correlation and censored longitudinal responses simultaneously in the context of LMEC/NLMEC models. Consequently, the aim of this paper is to study the impact of censoring and irregularly timed observed responses under Gaussian LMEC and NLMEC models.

For this purpose, we consider the analysis of two AIDS case studies. The first one investigated the effect of a highly active antiretroviral therapy (HAART) in persons with moderately advanced HIV-1 infection. This case study presented 11% of observations below (left-censored) the detection limits. The second case study evaluated the immune responses to HIV during acute infection, presenting about 22% of measurements lying above (right-censored) the limits of assay quantifications. Moreover, in both case studies, the viral loads were irregularly measured over time.

The rest of the paper is organized as follows. In Section 2 we describe the AIDS case studies that motivate our paper. Section 3 introduces the model (DEC-LMEC) and the likelihood function. In Section 4, the related likelihood-based inference is presented, including estimation of the random effects and the expected information matrix. The method for the prediction of future observations is presented in Section 5. Section 6 presents the extension to the nonlinear case (DEC-NLMEC). The application of the proposed method is presented in Sections 7 and 8 through a simulation study and the analysis of two AIDS case studies of HIV viral load. Finally, Section 9 concludes with a short discussion of issues raised by our study and some possible directions for a future research.

2. Case studies

In this section, we present the two motivating datasets, which will be analyzed next.

2.1. ACTG 315 data

The ACTG 315 protocol considers 46 HIV-1 infected patients treated with a potent antiretroviral drug cocktail based on protease inhibitor ritonavir and reverse transcriptase inhibitor drugs (zidovudine and lamivudine). Before initiating the antiretroviral therapy, all patients discontinued their own antiretroviral regimen for five weeks as a "washout" period. The aim of this antiretroviral regimen is to show

that immunity can be partially restored in people with moderately advance HIV disease.

The viral load was quantified on days 0, 2, 7, 10, 14, 21, 28, 56, 84, 168 and 196 after starting treatment. The dataset includes 361 observations. An immunologic marker known as CD4+ cell count was also measured along with viral load and 72 out of 361 (20%) CD4 values were missing due to a mismatch of the CD4 and the viral load measurement schedules. The number of measurements per subject varied from 4 to 10. Viral load measurements below the detectable threshold of 100 copies/mL (40 out of 361, 11%) were considered left-censored, and the censoring process assumed independence of the complete data. The individual profiles are shown in Figure 1.

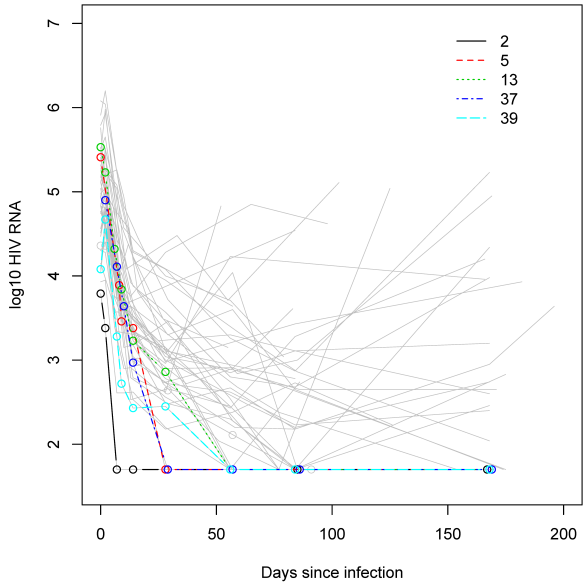

Figure 1: ACTG 315 data. Individual profiles (in $\log_{10}$ scale) for HIV viral load at different follow-up times. Trajectories for some censored individuals are indicated in different colors.

Table 1 shows the observed correlation for a selected individual at different time points. It is important to note that the correlation structure does not seem to be symmetric or even to have equal covariances across time, as considered by Vaida & Liu (2009).

Table 1: ACTG 315 data. Observed correlation of $\log_{10}$RNA for a single response over different times.

		$\log_{10}$RNA								
		t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8	t_9
$\log_{10}$RNA	t_1		0.6697	0.7779	0.7753	0.6181	0.4294	0.3268	0.1703	0.4121
	t_2	0.6697		0.7577	0.6969	0.6563	0.4486	0.2763	0.0464	0.3324
	t_3	0.7779	0.7577		0.8397	0.7904	0.4513	0.3352	0.1645	0.3023
	t_4	0.7753	0.6969	0.8397		0.8479	0.4855	0.3846	0.2488	0.2597
	t_5	0.6181	0.6563	0.7904	0.8479		0.5422	0.3943	0.3694	0.1784
	t_6	0.4294	0.4486	0.4513	0.4855	0.5422		0.7276	0.3715	0.0090
	t_7	0.3268	0.2763	0.3352	0.3846	0.3943	0.7276		0.5446	-0.1219
	t_8	0.1703	0.0464	0.1645	0.2488	0.3694	0.3715	0.5446		0.1159
	t_9	0.4121	0.3324	0.3023	0.2597	0.1784	0.0090	-0.1219	0.1159	

2.2. AIEDRP data

The second AIDS case study is from the AIEDRP program. This program, which is a large multicenter observational study of subjects with acute and early HIV infection, covers areas such as the evaluation of immune responses to HIV during acute infection, the assessment of thymic function and T-cell turnover during acute HIV infection and the assessment of transmission and prevalence of HIV resistance among treatment-naive subjects. The aim of this study was to help design future vaccines and to know the implications of new anti-HIV treatments.

We consider 320 untreated individuals with acute HIV infection (See Vaida & Liu (2009) for more details). Of the 830 recorded observations, 185 (22%) were above the limit of assay quantification. The individual profiles are shown in Figure 2.

Table 2: AIEDRP data. Observed correlation of $\log_{10}$RNA for a single response over different times.

		$\log_{10}$RNA									
		t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8	t_9	t_{10}
$\log_{10}$RNA	t_1		0.4980	0.4703	0.5708	0.5141	0.6023	0.4574	0.6992	0.7187	0.7593
	t_2	0.4980		0.7778	0.7136	0.5640	0.4990	0.3254	0.6484	0.7061	0.7775
	t_3	0.4703	0.7778		0.8513	0.7492	0.6585	0.4778	0.6394	0.6542	0.6069
	t_4	0.5708	0.7136	0.8513		0.8679	0.7662	0.6284	0.6796	0.9186	0.8844
	t_5	0.5141	0.5640	0.7492	0.8679		0.8223	0.7902	0.8626	0.9910	0.9881
	t_6	0.6023	0.4990	0.6585	0.7662	0.8223		0.8849	0.9395	0.9733	0.9915
	t_7	0.4574	0.3254	0.4778	0.6284	0.7902	0.8849		0.9646	0.9779	0.9944
	t_8	0.6992	0.6484	0.6394	0.6796	0.8626	0.9395	0.9646		0.9993	0.9960
	t_9	0.7187	0.7061	0.6542	0.9186	0.9910	0.9733	0.9779	0.9993		0.9970
	t_{10}	0.7593	0.7775	0.6069	0.8844	0.9881	0.9915	0.9944	0.9960	0.9970	

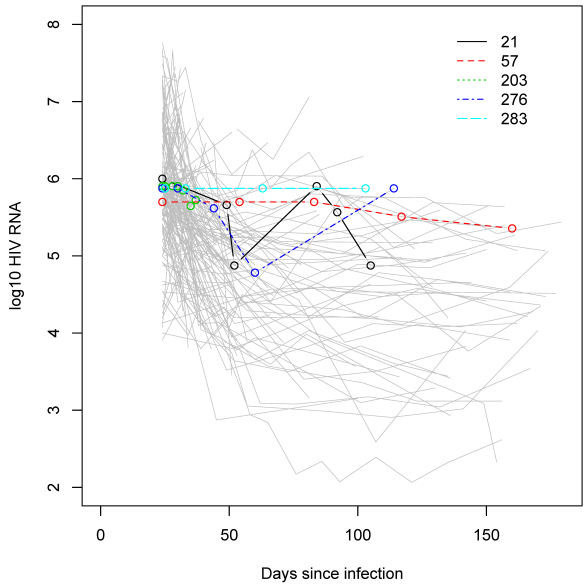

Figure 2: AIEDRP data. Individual profiles (in $\log_{10}$ scale) for HIV viral load at different follow-up times. Trajectories for some censored individuals are indicated in different colors.

Table 2 shows the observed correlation for a selected individual at different time points. Note that once again the correlation structure of the AIEDRP data does not seem to be symmetric or equal across time.

3. Model formulation

In the non-censored case, the classical Gaussian LME model is specified as follows Laird & Ware (1982):

$$\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{b}_i + \boldsymbol{\epsilon}_i, \tag{1}$$

where $\mathbf{b}_i \overset{iid}{\sim} \mathrm{N}_q(\mathbf{0}, \mathbf{D})$ is independent of $\boldsymbol{\epsilon}_i \overset{ind.}{\sim} \mathrm{N}_{n_i}(\mathbf{0}, \boldsymbol{\Omega}_i)$, $i = 1, \ldots, n$; the subscript i is the subject index; $\mathbf{y}_i = (y_{i1}, \ldots, y_{in_i})^\top$ is an $n_i \times 1$ vector of observed continuous responses for subject i measured at particular time points $\mathbf{t}_i = (t_{i1}, \ldots, t_{n_i})^\top$; $\mathbf{X}_i$ is the $n_i \times p$ design matrix corresponding to the fixed effects, $\boldsymbol{\beta}$, of dimension $p \times 1$; $\mathbf{Z}_i$ is the $n_i \times q$ design matrix corresponding to the $q \times 1$ vector of random effects $\mathbf{b}_i$; $\boldsymbol{\epsilon}_i$ of dimension $(n_i \times 1)$ is the vector of random errors; and the dispersion matrix $\mathbf{D} = \mathbf{D}(\boldsymbol{\alpha})$ depends on unknown and reduced parameters $\boldsymbol{\alpha}$. The correlation structure of the error vector is assumed to be $\boldsymbol{\Omega}_i = \sigma^2 \mathrm{E}_i$, where the $n_i \times n_i$ matrix E_i incorporates a time-dependence structure. Consequently, to capture the serial

correlation among irregularly observed longitudinal data, such as the ACTG 315 and AIEDRP datasets, it is necessary to consider a parsimonious parameterization of the matrix E_i. Following Muñoz *et al.* (1992), we adopt a DEC structure for E_i, which is defined as:

$$\mathrm{E}_i = \mathrm{E}_i(\boldsymbol{\phi}, \mathbf{t}_i) = \left[\phi_1^{|t_{ij}-t_{ik}|^{\phi_2}}\right], \; i = 1, \ldots, n, \;\; j, k = 1, \ldots, n_i, \tag{2}$$

where $\boldsymbol{\phi} = (\phi_1, \phi_2)^\top$, the parameter ϕ_1 describes the autocorrelation between observations separated by the absolute length of two time points, and the parameter ϕ_2 permits acceleration of the exponential decay of the autocorrelation function, defining a continuous-time autoregressive model.

For practical reasons, the parameter space of ϕ_1 and ϕ_2 is confined within $\boldsymbol{\Phi} = \{(\phi_1, \phi_1) : 0 < \phi_1 < 1, \, \phi_2 > 0\}$. It is important to stress that different values of the damping parameter ϕ_2 produce a variety of correlation structures for a given value of $\phi_1 > 0$, as follows: (a) if $\phi_2 = 0$, then E_i generates the compound symmetry correlation structure; (b) when $0 < \phi_2 < 1$, then E_i presents a decay rate between the compound symmetry structure and the first-order AR (AR (1)) model; (c) if $\phi_2 = 1$, then E_i generates an AR(1) structure; (d) when $\phi_2 > 1$, E_i presents a decay rate faster than the AR(1) structure; and (e) if $\phi_2 \to \infty$, then E_i represents the first-order moving average model, MA(1). A more detailed discussion of the DEC structure presenting more complex scenarios of the parameter space $\boldsymbol{\Phi}$ can be found in Muñoz *et al.* (1992).

As mentioned earlier, our model also considers censored observations, *i.e.*, we assume that the response Y_{ij} is not fully observed for all i, j. Let $(\mathbf{V}_i, \mathbf{C}_i)$ be the observed data for the i-th subject, where $\mathbf{V}_i$ represents the vector of uncensored reading or censoring level and $\mathbf{C}_i$ is the vector of censoring indicators, such that

$$\begin{aligned} y_{ij} &\leq V_{ij} \quad \text{if} \;\; C_{ij} = 1, \\ y_{ij} &= V_{ij} \quad \text{if} \;\; C_{ij} = 0. \end{aligned} \tag{3}$$

Note that since the observed response y_{ij} is defined over the real line, extensions to right censored data are straightforward. In fact, the right censored problem can be represented by a left censored problem by simultaneously transforming the response y_{ij} and censoring level V_{ij} to $-y_{ij}$ and $-V_{ij}$. The model defined in (1)-(3), will be called DEC-LMEC.

3.1. The log-likelihood function

Following Vaida & Liu (2009), classic inference on the parameter vector $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \boldsymbol{\alpha}^\top, \sigma^2, \boldsymbol{\phi}^\top)^\top$ is based on the marginal distribution of $\mathbf{y}_i$. For complete data, the marginal distribution of the vector $\mathbf{y}_i$, for $i = 1, \ldots, n$ is $\mathrm{N}_{n_i}(\mathbf{X}_i\boldsymbol{\beta}, \boldsymbol{\Sigma}_i)$, where $\boldsymbol{\Sigma}_i = \boldsymbol{\Omega}_i + \mathbf{Z}_i\mathbf{D}\mathbf{Z}_i^\top$. The responses with censoring pattern as in (3) follow a multivariate truncated normal distribution $\mathrm{TN}_{n_i}(\mathbf{X}_i\boldsymbol{\beta}, \boldsymbol{\Sigma}_i; \mathbf{A})$, where $\mathrm{TN}_{n_i}(\cdot; \mathbf{A})$ denotes the truncated normal distribution on the interval $\mathbf{A}$, where $\mathbf{A}_i = A_{i1} \times \ldots \times A_{ini}$, with A_{ij} being the interval $(-\infty, \infty)$ if $C_{ij} = 0$ and the interval $(-\infty, V_{ij}]$ if $C_{ij} = 1$. The strategy followed to compute the likelihood function associated with model (1) and (2) is to treat separately the observed and censored components of $\mathbf{y}_i$.

Let $\mathbf{y}_i^o$ be the n_i^o-vector of observed outcomes and $\mathbf{y}_i^c$ be the n_i^c-vector of censored observations for subject i with $(n_i = n_i^o + n_i^c)$ such that $C_{ij} = 0$ for all elements in $\mathbf{y}_i^o$, and $C_{ij} = 1$ for all elements in $\mathbf{y}_i^c$. After reordering, $\mathbf{y}_i$, $\mathbf{V}_i$, $\mathbf{X}_i$, and $\boldsymbol{\Sigma}_i$ can be partitioned as follows:

$$\mathbf{y}_i = vec(\mathbf{y}_i^o, \mathbf{y}_i^c),\ \mathbf{V}_i = vec(\mathbf{V}_i^o, \mathbf{V}_i^c),\ \mathbf{X}_i^\top = (\mathbf{X}_i^o, \mathbf{X}_i^c) \text{ and } \boldsymbol{\Sigma}_i = \begin{pmatrix} \boldsymbol{\Sigma}_i^{oo} & \boldsymbol{\Sigma}_i^{oc} \\ \boldsymbol{\Sigma}_i^{co} & \boldsymbol{\Sigma}_i^{cc} \end{pmatrix}.$$

In this setup, the operator $vec(\cdot)$ denotes the function which stack vectors or matrices of the same number of columns. Consequently, from the marginal-conditional decomposition of the multivariate normal distribution, $\mathbf{y}_i^o \sim \mathrm{N}_{n_i^o}(\mathbf{X}_i^o\boldsymbol{\beta}, \boldsymbol{\Sigma}_i^{oo})$ and $\mathbf{y}_i^c|\mathbf{y}_i^o \sim \mathrm{N}_{n_i^c}(\boldsymbol{\mu}_i, \mathbf{S}_i)$, where $\boldsymbol{\mu}_i = \mathbf{X}_i^c\boldsymbol{\beta} + \boldsymbol{\Sigma}_i^{co}(\boldsymbol{\Sigma}_i^{oo})^{-1}(\mathbf{y}_i^o - \mathbf{X}_i^o\boldsymbol{\beta})$ and $\mathbf{S}_i = \boldsymbol{\Sigma}_i^{cc} - \boldsymbol{\Sigma}_i^{co}(\boldsymbol{\Sigma}_i^{oo})^{-1}\boldsymbol{\Sigma}_i^{oc}$. Now, let $\Phi_{n_i}(\mathbf{u}; \mathbf{a}, \mathbf{A})$ and $\phi_{n_i}(\mathbf{u}; \mathbf{a}, \mathbf{A})$ be the *cdf* (left tail) and *pdf*, respectively, of $\mathrm{N}_{n_i}(\mathbf{a}, \mathbf{A})$ computed at vector $\mathbf{u}$. From Vaida & Liu (2009) and Matos *et al.* (2013a), the likelihood function for subject i (using conditional probability arguments) is given by:

$$\begin{aligned} L_i(\boldsymbol{\theta}) = f(\mathbf{y}_i|\boldsymbol{\theta}) &= P(\mathbf{V}_i|\mathbf{C}_i, \boldsymbol{\theta}) = P(\mathbf{y}_i^c \leq \mathbf{V}_i^c|\mathbf{y}_i^o = \mathbf{V}_i^o, \boldsymbol{\theta})P(\mathbf{y}_i^o = \mathbf{V}_i^o|\boldsymbol{\theta}), \\ &= P(\mathbf{y}_i^c \leq \mathbf{V}_i^c|\mathbf{y}_i^o, \boldsymbol{\theta})f(\mathbf{y}_i^o|\boldsymbol{\theta}) \\ &= \phi_{n_i^o}(\mathbf{y}_i^o; \mathbf{X}_i^o\boldsymbol{\beta}, \boldsymbol{\Sigma}_i^{oo})\Phi_{n_i^c}(\mathbf{V}_i^c; \boldsymbol{\mu}_i, \mathbf{S}_i), \end{aligned} \tag{4}$$

which can be easily evaluated computationally.

The log-likelihood function for the observed data, given by $\ell(\boldsymbol{\theta}|\mathbf{y}) = \sum_{i=1}^n \{\log L_i(\boldsymbol{\theta})\}$, is used to compute different model selection criteria, such as:

$$AIC = 2\,m - 2\,\ell_{max} \text{ and } BIC = m\,log\,N - 2\,\ell_{max},$$

where m is the number of model parameters, $N = \sum_{i=1}^n n_i$ and ℓ_{max} is the maximized log-likelihood value.

4. The EM algorithm

In this section, we describe in detail how the proposed model specified in (1)-(3) can be fitted by using the ECM algorithm. The EM algorithm (proposed originally by Dempster *et al.* (1977)) has several appealing features, such as stability of monotone convergence with each iteration, increasing the likelihood and simplicity of implementation. Due to the computational difficulty at the M-step, we use the ECM algorithm (an extension of the EM algorithm), which shares the appealing features of the EM and presents faster convergence than the original algorithm.

Let $\mathbf{y} = (\mathbf{y}_1^\top, \ldots, \mathbf{y}_n^\top)^\top$, $\mathbf{b} = (\mathbf{b}_1^\top, \ldots, \mathbf{b}_n^\top)^\top$, $\mathbf{V} = vec(\mathbf{V}_1, \ldots, \mathbf{V}_n)$ and $\mathbf{C} = vec(\mathbf{C}_1, \ldots, \mathbf{C}_n)$. Considering $\mathbf{b}$ as the hypothetical missing data, the complete data are denoted by $\mathbf{y}_c = (\mathbf{C}^\top, \mathbf{V}^\top, \mathbf{y}^\top, \mathbf{b}^\top)^\top$. Hence, the ECM algorithm is applied to the complete data log-likelihood function:

$$\begin{aligned}\ell_i(\boldsymbol{\theta}|\mathbf{y}_c) &= -\frac{1}{2}\Big[n_i \log\sigma^2 + \log(|\mathrm{E}_i|) + \frac{1}{\sigma^2}(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta} - \mathbf{Z}_i\mathbf{b}_i)^\top \mathrm{E}_i^{-1}(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta} - \mathbf{Z}_i\mathbf{b}_i) \\ &\quad + \log|\mathbf{D}| + \mathbf{b}_i^\top \mathbf{D}^{-1}\mathbf{b}_i\Big] + K, \end{aligned} \tag{5}$$

with K being a constant that does not depend on the parameter vector $\boldsymbol{\theta}$. Given the current estimate $\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}^{(k)}$, the E-step calculates the conditional expectation of the complete data log-likelihood function, given by:

$$\begin{aligned} Q\left(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}^{(k)}\right) &= E\left[\ell_c(\boldsymbol{\theta}|\mathbf{y}_c)|\mathbf{Q}, \mathbf{C}, \widehat{\boldsymbol{\theta}}^{(k)}\right] \\ &= \sum_{i=1}^{n} Q_{1i}\left(\boldsymbol{\beta}, \sigma^2|\widehat{\boldsymbol{\theta}}^{(k)}\right) + \sum_{i=1}^{n} Q_{2i}\left(\boldsymbol{\alpha}|\widehat{\boldsymbol{\theta}}^{(k)}\right), \end{aligned}$$

where

$$\begin{aligned} Q_{1i}\left(\boldsymbol{\beta}, \sigma^2, \boldsymbol{\phi}|\widehat{\boldsymbol{\theta}}^{(k)}\right) &= -\frac{n_i}{2}\log\sigma^2 - \frac{1}{2}\log(|\mathrm{E}_i|) - \frac{1}{2\sigma^2}\Big[\widehat{a}_i^{(k)} \\ &\quad -2\boldsymbol{\beta}^\top \mathbf{X}_i^\top \mathrm{E}_i^{-1}\left(\widehat{\mathbf{y}}_i^{(k)} - \mathbf{Z}_i\widehat{\mathbf{b}}_i^{(k)}\right) + \boldsymbol{\beta}^\top \mathbf{X}_i^\top \mathrm{E}_i^{-1}\mathbf{X}_i\boldsymbol{\beta}\Big], \end{aligned} \tag{6}$$

$$Q_{2i}\left(\boldsymbol{\alpha}|\widehat{\boldsymbol{\theta}}^{(k)}\right) = -\frac{1}{2}\log|\mathbf{D}| - \frac{1}{2}tr\left(\widehat{\mathbf{b}_i\mathbf{b}_i^\top}^{(k)}\mathbf{D}^{-1}\right), \tag{7}$$

and $\widehat{a}_i^{(k)} = tr\left(\widehat{(\mathbf{y}_i\mathbf{y}_i^\top}^{(k)}\mathrm{E}_i^{-1} - 2\widehat{\mathbf{y}_i\mathbf{b}_i^\top}^{(k)}\mathbf{Z}_i^\top\mathrm{E}_i^{-1} + \widehat{\mathbf{b}_i\mathbf{b}_i^\top}^{(k)}\mathbf{Z}_i^\top\mathrm{E}_i^{-1}\mathbf{Z}_i)\right)$, $\widehat{\mathbf{b}_i\mathbf{b}_i^\top}^{(k)} = E\left\{\mathbf{b}_i\mathbf{b}_i^\top|\mathbf{V}_i, \mathbf{C}_i, \widehat{\boldsymbol{\theta}}^{(k)}\right\}$ $= \widehat{\boldsymbol{\Lambda}}_i^{(k)} + \widehat{\boldsymbol{\varphi}}_i^{(k)}(\widehat{\mathbf{y}_i\mathbf{y}_i^\top}^{(k)} - \widehat{\mathbf{y}}_i^{(k)}\widehat{\boldsymbol{\beta}}^{(k)\top}\mathbf{X}_i^\top - \mathbf{X}_i\widehat{\boldsymbol{\beta}}^{(k)}\widehat{\mathbf{y}}_i^{(k)\top} + \mathbf{X}_i\widehat{\boldsymbol{\beta}}^{(k)}\widehat{\boldsymbol{\beta}}^{(k)\top}\mathbf{X}_i^\top)\widehat{\boldsymbol{\varphi}}_i^\top$, $\widehat{\mathbf{b}}_i^{(k)} =$ $E\left\{\mathbf{b}_i|\mathbf{V}_i, \mathbf{C}_i, \widehat{\boldsymbol{\theta}}^{(k)}\right\} = \widehat{\boldsymbol{\varphi}}_i^{(k)}(\widehat{\mathbf{y}}_i^{(k)} - \mathbf{X}_i\widehat{\boldsymbol{\beta}}^{(k)})$, $\widehat{\mathbf{y}_i\mathbf{b}_i^\top} = E\{\mathbf{y}_i\mathbf{b}_i^\top|\mathbf{V}_i, \mathbf{C}_i, \widehat{\boldsymbol{\theta}}^{(k)}\} = (\widehat{\mathbf{y}_i\mathbf{y}_i^\top}^{(k)} -$ $\widehat{\mathbf{y}}_i^{(k)}\widehat{\boldsymbol{\beta}}^{(k)\top}\mathbf{X}_i^\top)\widehat{\boldsymbol{\varphi}}_i^\top$, with $\widehat{\boldsymbol{\Lambda}}_i^{(k)} = (\widehat{\mathbf{D}}^{-1(k)} + \mathbf{Z}_i^\top\widehat{\mathrm{E}}_i^{-1(k)}\mathbf{Z}_i/\widehat{\sigma^2}^{(k)})^{-1}$ and $\widehat{\boldsymbol{\varphi}}_i^{(k)} = \widehat{\boldsymbol{\Lambda}}_i^{(k)}\mathbf{Z}_i^\top\widehat{\mathrm{E}}_i^{-1(k)}/\widehat{\sigma^2}^{(k)}$.

It is easy to see from (6) and (7) that the E-step reduces only to the computation of

$$\widehat{\mathbf{y}_i\mathbf{y}_i^\top} = E\{\mathbf{y}_i\mathbf{y}_i^\top|\mathbf{V}_i,\mathbf{C}_i,\widehat{\boldsymbol{\theta}}\} \text{ and } \widehat{\mathbf{y}}_i = E\{\mathbf{y}_i|\mathbf{V}_i,\mathbf{C}_i,\widehat{\boldsymbol{\theta}}\}.$$

These conditional expectations rely on the first and second moments of a multivariate truncated normal distribution and can be determined in closed-form (for more details on the computation of these moments see Vaida & Liu (2009)).

The conditional maximization step (CM) conditionally maximizes $Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}^{(k)})$ with respect to $\boldsymbol{\theta}$ obtaining a new estimate $\widehat{\boldsymbol{\theta}}^{(k+1)}$, as follows:

$$\widehat{\boldsymbol{\beta}}^{(k+1)} = \left(\sum_{i=1}^{n}\mathbf{X}_i^\top\widehat{\mathrm{E}}_i^{-1(k)}\mathbf{X}_i\right)^{-1}\sum_{i=1}^{n}\mathbf{X}_i^\top\widehat{\mathrm{E}}_i^{-1(k)}\left(\widehat{\mathbf{y}}_i^{(k)} - \mathbf{Z}_i\widehat{\mathbf{b}}_i^{(k)}\right), \tag{8}$$

$$\begin{aligned}\widehat{\sigma^2}^{(k+1)} = \frac{1}{N}\sum_{i=1}^{n}\Big[&\widehat{a}_i^{(k)} - 2\widehat{\boldsymbol{\beta}}^{(k+1)\top}\mathbf{X}_i^\top\widehat{\mathrm{E}}_i^{-1(k)}\left(\widehat{\mathbf{y}}_i^{(k)} - \mathbf{Z}_i\widehat{\mathbf{b}}_i^{(k)}\right)\\ &+\widehat{\boldsymbol{\beta}}^{(k+1)\top}\mathbf{X}_i^\top\widehat{\mathrm{E}}_i^{-1(k)}\mathbf{X}_i\widehat{\boldsymbol{\beta}}^{(k+1)}\Big],\end{aligned} \tag{9}$$

$$\widehat{\mathbf{D}}^{(k+1)} = \frac{1}{n}\sum_{i=1}^{n}\widehat{\mathbf{b}_i\mathbf{b}_i^\top}^{(k)}, \tag{10}$$

$$\begin{aligned}\boldsymbol{\phi}^{(k+1)} = \underset{\boldsymbol{\phi}\in(0,1)\times\mathbb{R}^+}{\operatorname{argmax}}\Bigg(&-\frac{1}{2}\log(|\mathrm{E}_i|) - \frac{1}{2\sigma^2}\Big[\widehat{a}_i^{(k)}\\ &-2\boldsymbol{\beta}^{(k+1)\top}\mathbf{X}_i^\top\mathrm{E}_i^{-1}\left(\widehat{\mathbf{y}}_i^{(k)} - \mathbf{Z}_i\widehat{\mathbf{b}}_i^{(k)}\right) + \boldsymbol{\beta}^{(k+1)\top}\mathbf{X}_i^\top\mathrm{E}_i^{-1}\mathbf{X}_i\boldsymbol{\beta}^{(k+1)}\Big]\Bigg).\end{aligned} \tag{11}$$

4.1. Estimation of random effects and the expected information matrix

For the estimation of random effects, we consider the conditional mean of $\mathbf{b}_i$ given the observed data $\mathbf{V}_i$ and $\mathbf{C}_i$, that is, $E\{\mathbf{b}_i|\mathbf{V}_i,\mathbf{C}_i\}$. Thus, for a given value of $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top,\sigma^2,\boldsymbol{\alpha}^\top,\boldsymbol{\phi}^\top)^\top$, the conditional mean of $\mathbf{b}_i$ given $\mathbf{V}_i$ and $\mathbf{C}_i$ is:

$$\widehat{\mathbf{b}}_i(\boldsymbol{\theta}) = E\{\mathbf{b}_i|\mathbf{V}_i,\mathbf{C}_i\} = \boldsymbol{\varphi}_i(\widehat{\mathbf{y}}_i - \mathbf{X}_i\boldsymbol{\beta}), \tag{12}$$

where $\boldsymbol{\varphi}_i = \boldsymbol{\Lambda}_i\mathbf{Z}_i^\top\mathrm{E}_i^{-1}/\sigma^2$ and $\boldsymbol{\Lambda}_i = (\mathbf{D}^{-1} + \mathbf{Z}_i^\top\mathrm{E}_i^{-1}\mathbf{Z}_i/\sigma^2)^{-1}$. Note that $\widehat{\mathbf{y}}_i = E\{\mathbf{y}_i|\mathbf{Q}_i,\mathbf{C}_i\}$ is given by the first moment of the $\mathrm{TN}_{n_i}(\mathbf{X}_i\boldsymbol{\beta},\boldsymbol{\Sigma}_i;\mathbb{A}_i)$. In practice, the estimator of $\mathbf{b}_i$, $\widehat{\mathbf{b}}_i$, can be obtained by substituting the ML estimate $\widehat{\boldsymbol{\theta}}$ into (12), leading to $\widehat{\mathbf{b}}_i = \widehat{\mathbf{b}}_i(\widehat{\boldsymbol{\theta}})$. On the other hand, the conditional covariance matrix of $\mathbf{b}_i$ given $\mathbf{V}_i$ and $\mathbf{C}_i$ is:

$$Var\{\mathbf{b}_i|\mathbf{Q}_i,\mathbf{C}_i\} = E\{\mathbf{b}_i\mathbf{b}_i^\top|\mathbf{Q}_i,\mathbf{C}_i\} - \widehat{\mathbf{b}}_i(\boldsymbol{\theta})\widehat{\mathbf{b}}_i(\boldsymbol{\theta})^\top = \boldsymbol{\Lambda}_i + \boldsymbol{\varphi}_i Var(\mathbf{y}_i|\mathbf{V}_i,\mathbf{C}_i)\boldsymbol{\varphi}_i^\top.$$

Note that $Var(\mathbf{y}_i|\mathbf{V}_i, \mathbf{C}_i)$ can be easily obtained as a byproduct of our proposed ECM algorithm developed in Section 4.

Louis (1982) proposed a technique for computing the observed information matrix within the EM algorithm framework. Using this method, and from the results given by Vaida & Liu (2009) (see also Hughes (1999)), we can find an asymptotic approximation for the variances of the fixed effects in the DEC-LMEC model. This approximation is given by:

$$\mathbf{J}_{\boldsymbol{\beta}\boldsymbol{\beta}} = Var(\widehat{\boldsymbol{\beta}}) = \left(\sum_{i=1}^{n} \left(\mathbf{X}_i^\top \boldsymbol{\Sigma}_i^{-1} \mathbf{X}_i - \mathbf{X}_i^\top \boldsymbol{\Sigma}_i^{-1} Var(\mathbf{y}_i|\mathbf{V}_i, \mathbf{C}_i) \boldsymbol{\Sigma}_i^{-1} \mathbf{X}_i \right) \right)^{-1}. \quad (13)$$

Asymptotic confidence intervals and hypothesis tests for the fixed effects are obtained assuming that the ML estimates $\widehat{\boldsymbol{\beta}}$ approximately have an $\mathrm{N}_p(\boldsymbol{\beta}, \mathbf{J}^{-1}_{\boldsymbol{\beta}\boldsymbol{\beta}})$ distribution. In practice, $\mathbf{J}_{\boldsymbol{\beta}\boldsymbol{\beta}}$ is usually unknown, so it needs to be replaced by its ML estimate $\mathbf{J}_{\widehat{\boldsymbol{\beta}}\widehat{\boldsymbol{\beta}}}$.

5. Prediction of future observations

The problem related to the prediction of future values has a great impact in many practical applications. Rao (1987) pointed out that the predictive accuracy of future observations can be taken as an alternative measure of "goodness-of-fit". In order to propose a strategy to generate predicted values from our t-MLC model, we used the approach proposed by Wang (2013). Thus, let $\mathbf{y}_{i,obs}$ be an observed response vector of dimension $n_{i,obs} \times 1$ for a new subject i over the first portion of time and $\mathbf{y}_{i,pred}$ the corresponding $n_{i,pred} \times 1$ response vector over the future portion of time. Let $\bar{\mathbf{X}}_i = (\mathbf{X}_{i,obs}, \mathbf{X}_{i,pred})$ be the $(n_{i,obs} + n_{i,pred}) \times p$ design matrix corresponding to $\bar{\mathbf{y}}_i = (\mathbf{y}_{i,obs}^\top, \mathbf{y}_{i,pred}^\top)$.

To deal with the censored values existing in $\mathbf{y}_{i,obs}$, we used the imputation procedure, by replacing the censored values by $\widehat{\mathbf{y}}_i = E\{\mathbf{y}_i|\mathbf{V}_i, \mathbf{C}_i, \widehat{\boldsymbol{\theta}}\}$ obtained from the EM algorithm. Therefore, when the censored values are imputed, a complete data set, denoted by $\mathbf{y}_{i,obs^*}$, is obtained. The reason to use the imputation procedure is that we avoid computing truncated conditional expectations of the multivariate normal distribution originated by the censoring scheme. Hence, we have that

$$\bar{\mathbf{y}}_i^* = \left(\mathbf{y}_{i,obs^*}^\top, \mathbf{y}_{i,pred}^\top\right)^\top \sim N_{n_{i,obs}+n_{i,pred}}\left(\mathbf{X}_i\boldsymbol{\beta}, \boldsymbol{\Sigma}_i\right),$$

where the matrix $\boldsymbol{\Sigma}_i$, can be represented by $\boldsymbol{\Sigma}_i = \begin{pmatrix} \boldsymbol{\Sigma}_i^{obs^*,obs^*} & \boldsymbol{\Sigma}_i^{obs^*,pred} \\ \boldsymbol{\Sigma}_i^{pred,obs^*} & \boldsymbol{\Sigma}_i^{pred,pred} \end{pmatrix}$. As mentioned in Wang (2013), the best linear predictor of $\mathbf{y}_{i,pred}$ with respect to the minimum mean squared error (MSE) criterion is the conditional expectation of $\mathbf{y}_{i,pred}$ given $\mathbf{y}_{i,obs^*}$, which is given by:

$$\widehat{\mathbf{y}}_{i,pred}(\boldsymbol{\theta}) = \mathbf{X}_{i,pred}\boldsymbol{\beta} + \boldsymbol{\Sigma}_i^{pred,obs^*}\boldsymbol{\Sigma}_i^{obs^*,obs^*-1}\left(\mathbf{y}_{i,obs^*} - \mathbf{X}_{i,obs^*}\boldsymbol{\beta}\right). \tag{14}$$

Therefore, $\mathbf{y}_{i,pred}$ can be estimated directly by substituting $\widehat{\boldsymbol{\theta}}$ into (14), leading to $\widehat{\widehat{\mathbf{y}}_{i,pred}} = \widehat{\mathbf{y}}_{i,pred}(\widehat{\boldsymbol{\theta}})$.

6. The nonlinear case

The NLME (without censoring) Pinheiro & Bates (2000) is defined as:

$$\mathbf{y}_i \;=\; \eta(\boldsymbol{\psi}_i, \mathbf{X}_i) + \boldsymbol{\epsilon}_i, \;\; \boldsymbol{\psi}_i = \mathbf{A}_i\boldsymbol{\beta} + B_i\mathbf{b}_i, \;\; i = 1, \ldots, n, \tag{15}$$

where $\mathbf{b}_i \overset{iid}{\sim} \mathrm{N}_q(0, \mathbf{D})$ and $\boldsymbol{\epsilon}_i \overset{ind}{\sim} \mathrm{N}_{n_i}(0, \sigma^2\mathbf{E}_i)$ are independent; $\mathbf{y}_i$ is an $(n_i \times 1)$ vector of observed responses for subject i; η is a nonlinear function of the individual random parameter $\boldsymbol{\psi}_i$; $\mathbf{A}_i$ and $\mathbf{B}_i$ are known design matrices of dimensions $r \times p$ and $r \times q$, respectively, possibly depending on some covariate values; $\boldsymbol{\beta}$ is the $(p \times 1)$ vector of fixed effects and $\mathbf{b}_i$ is the $(q \times 1)$ vector of random effects.

As mentioned by Vaida & Liu (2009), the linearization (L) procedure to obtain the approximate MLE of $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \sigma^2, \boldsymbol{\alpha}^\top, \boldsymbol{\phi}^\top)^\top$ involves taking the first-order Taylor expansion of η_i around the current parameter estimate $\widetilde{\boldsymbol{\beta}}$ and the random effect estimates $\widetilde{\mathbf{b}}_i$ (empirical predictors). This procedure is equivalent to iteratively solving the following LME model (L-step):

$$\widetilde{\mathbf{Y}}_i = \widetilde{\mathbf{W}}_i\boldsymbol{\beta} + \widetilde{\mathbf{H}}_i\mathbf{b}_i + \boldsymbol{\epsilon}_i, \;\; i = 1, \ldots, n, \tag{16}$$

where $\mathbf{b}_i \overset{iid}{\sim} \mathrm{N}_q(0, \mathbf{D})$ and $\boldsymbol{\epsilon}_i \overset{ind}{\sim} \mathrm{N}_{n_i}(\mathbf{0}, \sigma^2\mathrm{E}_i)$, and $\widetilde{\mathbf{Y}}_i = \mathbf{Y}_i - \eta(\boldsymbol{\psi}(\widetilde{\boldsymbol{\beta}}, \widetilde{\mathbf{b}}_i), \mathbf{X}_i)$, $\widetilde{\mathbf{H}}_i = \dfrac{\partial\eta(\mathbf{A}_i\boldsymbol{\beta} + \mathbf{B}_i\mathbf{b}_i, \mathbf{X}_i)}{\partial\mathbf{b}_i^\top}\Big|_{\mathbf{b}_i=\widetilde{\mathbf{b}}_i}$ and $\widetilde{\mathbf{W}}_i = \dfrac{\partial\eta(\mathbf{A}_i\boldsymbol{\beta} + \mathbf{B}_i\mathbf{b}_i, \mathbf{X}_i)}{\partial\boldsymbol{\beta}_i^\top}\Big|_{\boldsymbol{\beta}_i=\widetilde{\boldsymbol{\beta}}_i}$. Thus, in the censored case, the model in (16) is an LME with censored data that can be fitted using the strategy explained in Section 4. The model matrices in (16) depend on the current parameter value, and need to be recalculated at each iteration. The algorithm iterates between the L-, E- and CM-steps until convergence.

7. Analysis of case studies

We illustrate the performance of the proposed methods with the analysis of two HIV datasets, previously analyzed by Wu (2002) and Vaida & Liu (2009), respectively.

7.1. ACTG 315 data

We re-analyze the HIV viral load data from clinical trial ACTG 315 Wu (2002) considering four different correlation structures, namely the uncorrelated structure (UNC), damped exponential correlation (DEC), continuous-time autoregressive of order 1 (AR(1)) and compound symmetric structure (SYM). As mentioned in Section 2, the dataset consists of 46 HIV-1 infected patients treated with a potent ARV therapy. The viral load was repeatedly quantified on days 0, 2, 7, 10, 14, 21, 28, 56, 84, 168, and 196 after start of treatment, with a total of 361 observations. The viral load detectable limit is 100 copies/mL, and 40 out of 361 (11%) of all viral load measurements are below the detection limit. Wu & Ding (1999) proposed the use of a biphasic model:

$$V(t) = e^{p_1-\varphi_1 t} + e^{p_2-\varphi_2 t}, \tag{17}$$

where $V(t)$ is the viral load at time t. The parameters φ_1 and φ_2 are called the first and second phase viral decay rates, which can represent the minimum turnover rate of productively infected cells and that of latently or long-lived infected cells, respectively. The parameters p_1 and p_2 are macro-parameters and $e^{p_1} + e^{p_2}$ is the baseline viral load at time $t = 0$.

As was noted by Wu & Ding (1999), the inter-subject variation of observed viral loads motivates the use of a NLME model. The viral load trajectories initially exhibit rapid decay (known as first-phase decay), followed by a phase of slow decay for some (the second-phase) with the others rebounding back to the original levels Liu & Wu (2012). Therefore, following Wu (2002) we consider the following NLME model to model the dynamics of the HIV viral load:

$$\begin{aligned} y_{ij} &= \log_{10}(e^{\varphi_{1i}-\varphi_{2i}t_{ij}} + e^{\varphi_{3i}-\varphi_{4ij}t_{ij}}) + \epsilon_{ij}, && (18)\\ \beta_{1ij} &= \varphi_{1i} = \beta_1 + b_{1i}, \;\; \beta_{3ij} = \varphi_{3i} = \beta_3 + b_{3i}, && (19)\\ \beta_{2ij} &= \varphi_{2i} = \beta_2 + b_{2i}, \;\; \beta_{4_{ij}} = \varphi_{4ij} = \beta_4 + \beta_5 \mathrm{CD4}_{ij} + b_{4i}, && (20) \end{aligned}$$

where y_{ij} is the $\log_{10}$-transformation of the viral load for the ith subject at time t_{ij} $(i = 1, 2, \ldots, n, j = 1, 2, \ldots, n_i)$ and $\boldsymbol{\epsilon}_i = (\epsilon_{i1}, \ldots, \epsilon_{in_i})^\top$ represents the vector of

within-individual random errors; CD4_{ij} indicates the observed CD4 values up to time t_{ij}; $\boldsymbol{\beta}_{ij} = (\beta_{1ij}, \beta_{2ij}, \beta_{3ij}, \beta_{4ij})^\top$ and $\boldsymbol{\beta} = (\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_5)^\top$ are individual parameters for the i-th subject at time t_{ij} and population parameters, respectively and $\mathbf{b}_i = (b_{1i}, \ldots, b_{4i})^\top$ is the random effects vector for subject i.

The values of ℓ_{max}, AIC and BIC for the four considered models are presented in Table 3. Note that, based on these criteria, the model presenting the best fit corresponds to the model with a damped exponential correlation structure (DEC).

ML estimates corresponding to the best model are presented in Table 4. Using these estimates, we can quantify the population decay rates and viral load parameters. The first- and second-phase decay rates can be approximated as $\widehat{\varphi}_1 = 31.598$ and $\widehat{\varphi}_2(t) = -0.990 + 0.616\,\text{CD4}$. The population viral load process can be represented as $\widehat{V}(t) = \exp\{11.555 - \widehat{\varphi}_1(t)t\} + \exp\{6.865 - \widehat{\varphi}_2(t)t\}$.

Table 3: ACTG 315 data. Model selection criteria for the NLMEC model under different correlation structures.

	NLMEC			
Criteria	UNC	DEC	AR(1)	SYM
ℓ_{max}	-281.305	**-255.828**	-264.986	-279.330
AIC	594.611	**547.655**	563.971	592.660
AIC corr	596.192	**549.655**	565.756	594.445
BIC	656.833	**617.655**	630.082	658.771

Table 4: ACTG 315 data. ML estimates with standard errors for the NLMEC model under DEC structure.

Fixed effects			Between-subject variances		Within-subject variances	
Parameter	Estimative	SE	Parameter	Estimative	Parameter	Estimative
β_1	11.555	0.221	α_{11}	0.184	σ^2	0.397
β_2	31.598	1.823	α_{12}	-0.969	ϕ_1	0.183
β_3	6.865	0.284	α_{22}	8.568	ϕ_2	0.644
β_4	-0.990	0.932	α_{13}	0.078		
β_5	0.616	0.273	α_{23}	0.287		
			α_{33}	0.412		
			α_{14}	-0.602		
			α_{24}	5.589		
			α_{34}	0.752		
			α_{44}	5.845		

7.2. AIEDRP data

The second case study is taken from the AIEDRP program, a large multicenter observational study of subjects with acute and early HIV infection. We consider 320 untreated individuals with acute HIV infection. Of the 830 recorded observations, 185 (22%) were above the limit of assay quantification. Therefore, in the spirit of Vaida & Liu (2009), we consider a right-censored five-parameter NLME model (inverted S-shaped curve) as follows:

$$y_{ij} = \lambda_{1i} + \frac{\lambda_2}{1 + \exp((t_{ij} - \lambda_3)/\lambda_4)} + \lambda_{5i}(t_{ij} - 50) + \epsilon_{ij}, \tag{21}$$

where y_{ij} is the $\log_{10}$ of the viral load for subject i at time t_{ij}. The parameters λ_{1i} and λ_2 represent the subject-specific set-point value and decrease from the maximum HIV-1 RNA. The location parameter λ_3 indicates the time point at which half of the change in HIV-1 RNA is attained, λ_4 is a scale parameter modeling the rate of decline and λ_{5i} allows us to increase the HIV-1 RNA trajectory after day 50. The reparameterization given by $\beta_{1i} = \log(\lambda_{1i}) = \beta_1 + b_{1i}$; $\beta_k = \log(\lambda_k)$, $k = 2, 3, 4$, and $\lambda_{5i} = \beta_5 + b_{2i}$ is adopted to assure positive values for the model parameters.

As in Section 7.1, we considered the correlation structures UNC, DEC, AR(1) and SYM. Table 5 summarizes the values of ℓ_{max}, AIC and BIC for all considered models. Note that the values of ℓ_{max} for the DEC and $AR(1)$ models are close. This is explained because the estimated values of ϕ_1 and ϕ_2 under the DEC model are 0.83 and 1.15 respectively. Based on this observation and the criteria, the best (parsimonious) fit is obtained using the continuous-time autoregressive of order 1 correlation (AR(1)). Moreover, the model fit of the AR(1) (and DEC) model is slightly better than the SYM model, with the smooth mean residual curve in Figure 3 (b) always being closer to zero. The ML estimates under this model are presented in Table 6. We can use the AR(1) model with reasonable confidence for predictions of viral load. For example, at 6 months since infection, the average viral load is 4.537 $\log_{10}$ units. The individual 6-month viral load estimates vary between 1.794 and 6.469, with 5th and 95th quantiles at 3.466 and 5.549. The average slope after day 50 is negative, $\beta_{5i} = -0.004$ $\log_{10}$HIV/day, with 95% CI(-0.006,-0.002). And, for the individual slopes α_{5i} the 5th and 95th quantiles are -0.0061 and -0.0015.

8. Simulation Studies

In order to examine the performance of our proposed method, we conducted two simulation studies to investigate the consequences for parameter estimation and

Table 5: AIEDRP data. Model selection criteria for the NLMEC model under different correlation structures.

	LMEC			
Criteria	UNC	DEC	AR(1)	SYM
ℓ_{max}	-783.794	-769.814	**-770.097**	-775.624
AIC	1585.589	1561.629	**1560.194**	1571.248
AIC corr	1585.808	1561.952	**1560.462**	1571.517
BIC	1628.082	1613.564	**1607.408**	1618.463

Table 6: AIEDRP data. ML estimates with standard errors for the NLMEC model under AR(1) structure.

Fixed effects			Between-subject variances		Within-subject variances	
Parameter	Estimative	SE	Parameter	Estimative	Parameter	Estimative
β_1	1.614	0.014	α_{11}	0.01658	σ^2	0.308
β_2	0.128	0.107	α_{12}	0.00020	ϕ_1	0.808
β_3	3.516	0.027	α_{22}	0.00003		
β_4	1.118	0.280				
β_5	-0.004	0.001				

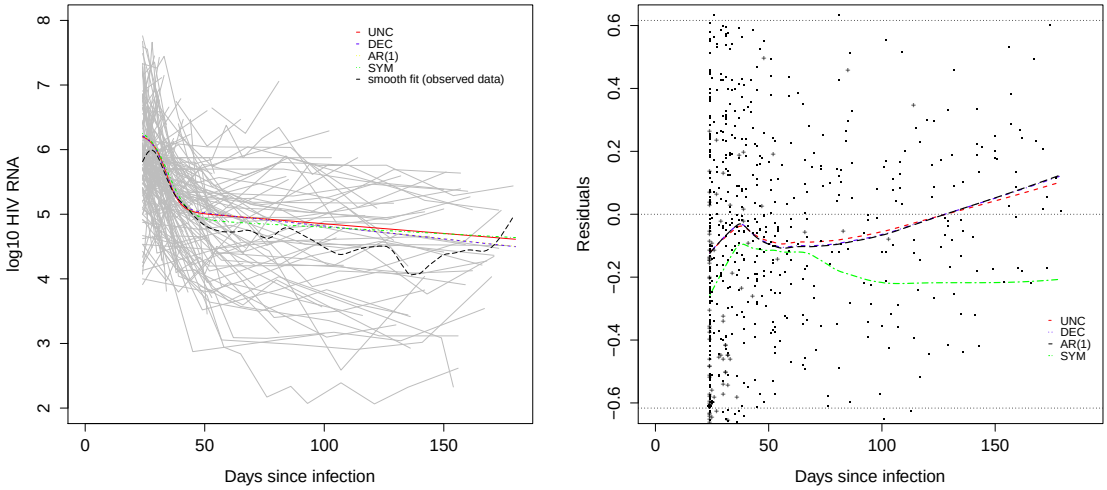

Figure 3: AIEDRP data. (Left panel) Individual profiles (in log10 scale) for HIV viral load at different follow-up times with the model fits. (Right panel) Smooth means of residuals from model fits. The residuals from the model with autoregressive of order 1 correlation appear as points.

the behavior of the prediction when the correlation structure of the error term is misspecified. For this purpose and simplicity reasons, we consider a logistic model

similar to that studied in Section 7.2, with random set-points λ_{1i} and random decline rates λ_{4i}, as follows:

$$y_{ij} = \lambda_{1i} + \frac{\lambda_2}{1 + \exp((t_{ij} - \lambda_3)/\lambda_{4i})} + \epsilon_{ij}, \tag{22}$$

with $i = 1, \ldots, 100$, $j = 1, \ldots, 10$, $\alpha_{1i} = \exp(\beta_1 + b_{1i})$, $\beta_k = \log(\lambda_k)$, $k = 2, 3$, $\lambda_{4i} = \exp(\beta_4 + b_{2i})$, $(b_{1i}, b_{2i}) \overset{ind.}{\sim} \mathrm{N}_2(\mathbf{0}, \mathbf{D})$, and $\epsilon_{ij} \overset{ind.}{\sim} \mathrm{N}_{n_i}(\mathbf{0}, \mathbf{\Omega}_i)$. We set $\boldsymbol{\beta} = (1.6094, 0.6931, 3.8067, 2.3026)^\top$, $\sigma^2 = 0.55$, and $\mathbf{D}$ with elements $\alpha_{11} = 0.05$ and $\alpha_{22} = 0.1$.

For the first study, we simulated several datasets considering different values of the parameter ϕ_1 under the correlation structure AR(1) with the aim to study the effect of the correlation level on the estimation. For each value of ϕ_1, we simulated 100 datasets. In addition, we considered 5% and 20% of censored observations for each value of ϕ_1. Once the simulated datasets were generated, we fitted the proposed model assuming the uncorrelated (UNC) and AR(1) structures. The model selection criteria (AIC and BIC) as well as the estimates of the model parameters were stored for each simulation. Summary statistics such as the mean estimate (MC mean), coverage probability (MC CP) and the mean of the approximate standard error obtained through the information-based method described in Section 4.1 (IM SE), are presented in Tables 7 and 8.

From the results shown in Tables 7 and 8, we can observe that when the AR(1) is chosen as the true model, the MC CP values are higher than those obtained under the uncorrelated model, even when the correlation parameter ϕ_1 is small (0.3). Moreover, the biases of fixed effects estimates under the AR(1) structure are lower than the obtained ones under the uncorrelated structure (see Figures 4 and 5) for different values of the ϕ_1 parameter. The model selection criteria chose the true model (AR(1)) for moderate values of the ϕ_1 parameter (greater than 0.5) for the two levels of censoring considered.

The second simulation study analyzes the performance of the prediction of future values described in Section 5. For this purpose, we compared the prediction of the NLMEC model in (22) under the UNC and AR(1) structures. As in the first study, we generated 100 datasets of size $n = 100$ under AR(1) structure with parameter $\phi_1 = 0.9$, considering two different settings of censoring proportions, say 5% and 20%. For the prediction, we excluded the last two measurements of each simulated individual in the datasets. For comparing the performance of the prediction, we considered two empirical discrepancy measures, namely the MAE (mean absolute

Table 7: 5% censored. Summary statistics based on 100 simulated AR(1) samples.

			Parameter estimates					Criteria	
ϕ_1	Corr. Structure		β_1	β_2	β_3	β_4	σ^2	MC AIC	MC BIC
0.3	UNC	MC Mean	1.665	0.529	3.735	2.106	0.551	3020.406	3056.037
		IM SE	0.037	0.114	0.045	0.199			
		MC CP	79%	78%	66%	89%			
	AR(1)	MC Mean	1.607	0.713	3.833	2.270	0.553	3024.403	3065.124
		IM SE	0.118	0.224	0.096	0.256			
		MC CP	84%	88%	94%	91%			
0.5	UNC	MC Mean	1.663	0.540	3.735	2.125	0.547	3014.613	3050.244
		IM SE	0.038	0.114	0.046	0.198			
		MC CP	82%	80%	67%	91%			
	AR(1)	MC Mean	1.605	0.714	3.834	2.269	0.551	3017.634	3058.355
		IM SE	0.124	0.227	0.098	0.258			
		MC CP	84%	88%	93%	91%			
0.6	UNC	MC Mean	1.659	0.561	3.736	2.160	0.539	3003.867	3039.497
		IM SE	0.038	0.114	0.046	0.196			
		MC CP	82%	86%	69%	92%			
	AR(1)	MC Mean	1.603	0.715	3.834	2.268	0.551	3003.228	3043.948
		IM SE	0.131	0.233	0.101	0.264			
		MC CP	84%	88%	93%	91%			
0.7	UNC	MC Mean	1.647	0.624	3.735	2.267	0.520	2978.063	3013.694
		IM SE	0.039	0.114	0.046	0.189			
		MC CP	90%	91%	68%	94%			
	AR(1)	MC Mean	1.594	0.721	3.837	2.269	0.550	2961.560	3002.281
		IM SE	0.198	0.251	0.109	0.275			
		MC CP	84%	88%	91%	92%			
0.8	UNC	MC Mean	1.621	0.750	3.735	2.513	0.474	2912.026	2947.656
		IM SE	0.038	0.099	0.039	0.171			
		MC CP	99%	96%	50%	76%			
	AR(1)	MC Mean	1.597	0.720	3.837	2.268	0.548	2840.346	2881.066
		IM SE	0.170	0.254	0.112	0.285			
		MC CP	84%	88%	93%	91%			
0.9	UNC	MC Mean	1.605	0.899	3.725	2.770	0.359	2672.775	2708.406
		IM SE	0.035	0.070	0.026	0.144			
		MC CP	95%	17%	13%	12%			
	AR(1)	MC Mean	1.612	0.699	3.826	2.256	0.530	2452.579	2493.300
		IM SE	0.117	0.211	0.094	0.263			
		MC CP	83%	88%	94%	91%			

error) and MSE (mean square error), given by:

$$\text{MAE} = \frac{1}{200}\sum_{i,j}|y_{ij} - y_{ij}^{\star}| \quad \text{and} \quad \text{MSE} = \frac{1}{200}\sum_{i,j}(y_{ij} - y_{ij}^{\star})^2, \tag{23}$$

where y_{ij} is the original value and $y_{ij}^{\star}$ is the predicted value, for $i = 1, \ldots, 100$ and $j = 1, \ldots, 2$. Table 9 shows the comparison between the predicted values and real ones under the NLMEC model considering the UNC and AR(1) structures. We can see from these results that the model with AR(1) structure generates predictive values close to the real ones.

Table 8: 20% censored. Summary statistics based on 100 simulated AR(1) samples.

			Parameter estimates					Criteria	
ϕ_1	Corr. Structure		β_1	β_2	β_3	β_4	σ^2	MC AIC	MC BIC
0.3	UNC	MC Mean	1.672	0.499	3.724	2.077	0.547	2796.411	2832.041
		IM SE	0.037	0.117	0.047	0.209			
		MC CP	68%	69%	63%	87%			
	AR(1)	MC Mean	1.587	0.724	3.836	2.279	0.553	2799.969	2840.690
		IM SE	0.233	0.262	0.114	0.273			
		MC CP	87%	90%	92%	91%			
0.5	UNC	MC Mean	1.671	0.507	3.724	2.093	0.544	2791.306	2826.936
		IM SE	0.037	0.117	0.047	0.208			
		MC CP	76%	73%	63%	87%			
	AR(1)	MC Mean	1.599	0.712	3.830	2.270	0.551	2794.133	2834.853
		IM SE	0.170	0.249	0.108	0.271			
		MC CP	87%	90%	92%	91%			
0.6	UNC	MC Mean	1.668	0.522	3.724	2.121	0.536	2780.670	2816.301
		IM SE	0.038	0.118	0.048	0.207			
		MC CP	78%	77%	64%	88%			
	AR(1)	MC Mean	1.596	0.715	3.832	2.270	0.550	2780.178	2820.899
		IM SE	0.189	0.257	0.111	0.277			
		MC CP	87%	90%	91%	92%			
0.7	UNC	MC Mean	1.658	0.581	3.724	2.224	0.518	2756.906	2792.536
		IM SE	0.039	0.119	0.048	0.202			
		MC CP	85%	87%	67%	92%			
	AR(1)	MC Mean	1.593	0.719	3.833	2.271	0.549	2742.308	2783.028
		IM SE	0.203	0.266	0.116	0.287			
		MC CP	86%	90%	91%	94%			
0.8	UNC	MC Mean	1.628	0.709	3.725	2.477	0.476	2703.285	2738.915
		IM SE	0.038	0.105	0.041	0.183			
		MC CP	96%	99%	45%	80%			
	AR(1)	MC Mean	1.617	0.685	3.818	2.237	0.549	2636.814	2677.535
		IM SE	0.115	0.232	0.100	0.285			
		MC CP	86%	89%	93%	94%			
0.9	UNC	MC Mean	1.615	0.853	3.724	2.735	0.359	2484.058	2519.689
		IM SE	0.034	0.072	0.025	0.156			
		MC CP	98%	43%	24%	26%			
	AR(1)	MC Mean	1.625	0.668	3.811	2.214	0.532	2289.859	2330.580
		IM SE	0.093	0.200	0.087	0.264			
		MC CP	86%	89%	95%	94%			

Table 9: Evaluation of the prediction accuracy for the NLMEC model with different correlation structures.

	5% censored		20% censored	
Corr. Structure	MAE	MSE	MAE	MSE
UNC	0.5507	0.4739	0.6418	0.6746
AR(1)	0.5169	0.4299	0.6073	0.6165

9. Conclusions

In this paper, we proposed a mixed effects model with censored observations based on the multivariate normal distribution. We adopted a damped exponential correlation structure as proposed by Muñoz *et al.* (1992) to model the autocorrelation existing among irregularly observed measures. An ECM algorithm to obtain

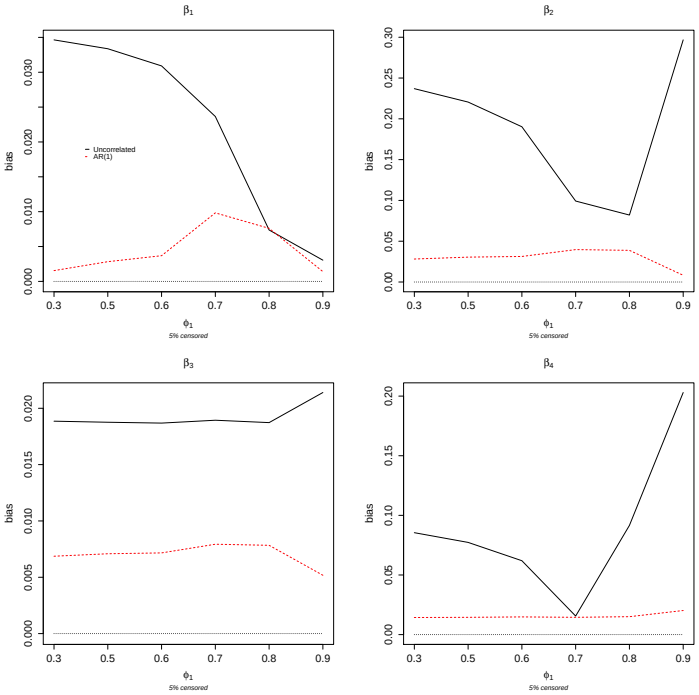

Figure 4: Simulation study. 5% censored. Bias of $\boldsymbol{\beta}$ estimates under the uncorrelated and AR(1) models for 6 different values of ϕ_1.

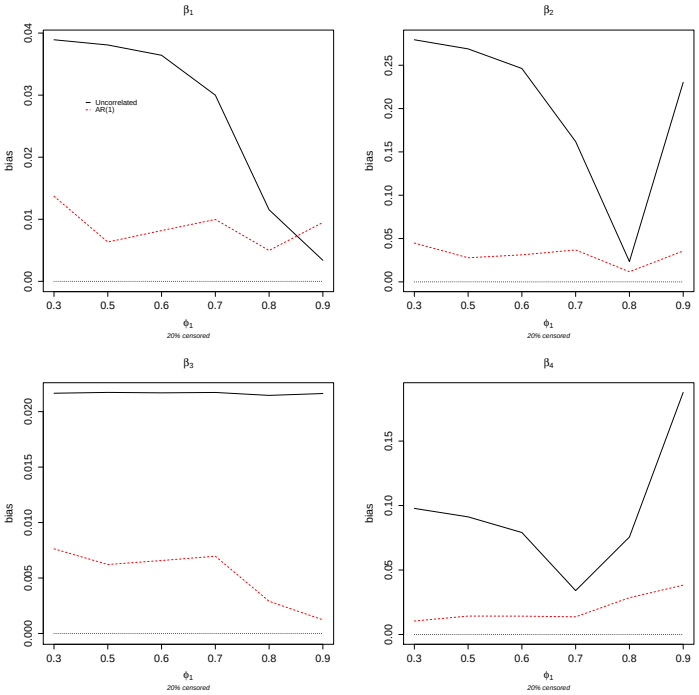

Figure 5: Simulation study. 20% censored. Bias of $\boldsymbol{\beta}$ estimates under the uncorrelated and AR(1) models for 6 different values of ϕ_1.

the ML estimates was developed by using the statistical properties of the multivariate truncated normal distribution. Our proposed algorithm has a closed-form expression for the E-step, based on the first two moments of the truncated normal distribution. We applied our methods to two AIDS case studies and performed a simulation study, showing the effects of misspecification on the correlation structure over the fixed effects estimates.

Although our LMEC/NLMEC models showed great flexibility for modeling symmetric data, they can be seriously affected by the presence of outliers. Recently, Garay *et al.* (2014) proposed a remedy to accommodate outliers using a Student's t regression model with DEC structure. We believe that our method can be extended by considering the Student's t in the context of LMEC/NLMEC models, providing satisfactory results at the expense of additional complexity in implementation. Further, it is also of interest to develop an effective Markov chain Monte Carlo algorithm for the DEC-LMEC/NLMEC in a fully Bayesian treatment.

Acknowledgements

L. Matos acknowledges support from FAPESP-Brazil. The research of L.M. Castro was partially supported by Grant FONDECYT 1130233 from the Chilean government and Grant 2012/19445-0 from FAPESP-Brazil. V.H. acknowledges support from CNPq-Brazil (Grant 305054/2011-2) and FAPESP-Brazil (Grant 2011/17400-6).

References

Bandyopadhyay, D., Matos, L., Castro, L. & Lachos, V. (2014). Diagnostics for censored mixed-effects models using the multivariate t-distribution. Technical report, Universidad Estadual de Campinas.

Dempster, A., Laird, N. & Rubin, D. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, Series B,*, **39**, 1–38.

Garay, A., Castro, L., Leskow, J. & Lachos, V. H. (2014). Censored linear regression models for irregulary observed longitudinal data using the multivariate-t distribution. *Statistical Methods in Medical Research*, **Forthcoming**.

Hughes, J. (1999). Mixed effects models with censored data with application to HIV RNA levels. *Biometrics*, **55**, 625–629.

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

Liu, W. & Wu, L. (2012). Two-step and likelihood methods for HIV viral dynamic models with covariate measurement errors and missing data. *Journal of Applied Statistics*, **39**(5), 963–978.

Louis, T. A. (1982). Finding the observed information matrix when using the em algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 226–233.

Matos, L., Lachos, V., Balakrishnan, N. & Labra, F. (2013a). Influence diagnostics in linear and nonlinear mixed-effects models with censored data. *Computational Statistical & Data Analysis*, **57**(1), 450–464.

Matos, L. A., Prates, M. O., Chen, M. H. & Lachos, V. H. (2013b). Likelihood-based inference for mixed-effects models with censored response using the multivariate-t distribution. *Statistica Sinica*, **23**, 1323–1342.

Muñoz, A., Carey, V., Schouten, J. P., Segal, M. & Rosner, B. (1992). A parametric family of correlation structures for the analysis of longitudinal data. *Biometrics*, **48**, 733–742.

Pinheiro, J. C. & Bates, D. M. (2000). *Mixed-Effects Models in S and S-PLUS*. Springer, New York, NY.

Rao, C. R. (1987). Prediction of future observations in growth curve models. *Statistical Science*, **2**, 434–447.

Vaida, F. & Liu, L. (2009). Fast implementation for normal mixed Effects models with censored response. *Journal of Computational and Graphical Statistics*, **18**, 797–817.

Vaida, F., Fitzgerald, A. & DeGruttola, V. (2007). Efficient hybrid EM for linear and nonlinear mixed effects models with censored response. *Computational Statistics & Data Analysis*, **51**, 5718–5730.

Wang, W.-L. (2013). Multivariate t linear mixed models for irregularly observed multiple repeated measures with missing outcomes. *Biometrical Journal*, **55**(4), 554–571.

Wang, W. L. & Fan, T. H. (2011). Estimation in multivariate t linear mixed models for multivariate longitudinal data. *Statistica Sinica*, **21**, 1857–1880.

Wu, H. & Ding, A. (1999). Population HIV-1 dynamics in vivo: applicable models and inferential tools for virological data from AIDS clinical trials. *Biometrics*, **55**(2), 410–418.

Wu, L. (2002). A joint model for nonlinear mixed-effects models with censoring and covariates measured with error, with application to AIDS studies. *Journal of the American Statistical Association*, **97**(460), 955–964.