



HAL
open science

Individual prediction regions for multivariate longitudinal data with small samples

Didier Concordet, Rémi Servien

► **To cite this version:**

Didier Concordet, Rémi Servien. Individual prediction regions for multivariate longitudinal data with small samples. *Biometrics*, 2014, 70 (3), pp.629-638. 10.1111/biom.12201 . hal-02639518

HAL Id: hal-02639518

<https://hal.inrae.fr/hal-02639518>

Submitted on 28 May 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Individual prediction regions for multivariate longitudinal data with small samples

Didier Concordet, Rémi Servien

► To cite this version:

Didier Concordet, Rémi Servien. Individual prediction regions for multivariate longitudinal data with small samples. 2013. <hal-00856736v3>

HAL Id: hal-00856736

<https://hal.archives-ouvertes.fr/hal-00856736v3>

Submitted on 12 May 2014

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Comment citer ce document :

Concordet, D., Servien, R. (2014). Individual prediction regions for multivariate longitudinal data with small samples. *Biometrics*, 70 (3), 629-638. DOI : 10.1111/biom.12201

Individual prediction regions for multivariate longitudinal data with small samples

DIDIER CONCORDET & RÉMI SERVIEN

INRA-ENVT, Université de Toulouse, UMR1331 Toxalim
Research Centre in Food Toxicology, F-31027 Toulouse

remi.servien@toulouse.inra.fr

Abstract

Follow-up is more and more used in medicine/doping control to identify abnormal results in an individual. Currently, follow-ups are mostly carried out variable by variable using “reference intervals” that contain the values observable in $100(1 - \alpha)\%$ of healthy/undoped individuals. Observations of the evolution of the variables over time in a sample of N healthy/undoped individuals, allows these reference intervals to be individualized by taking into account the possible effect of covariables and some previous observations of these variables obtained when the individual was healthy/undoped. For each variable these individualized intervals should contain $100(1 - \alpha)\%$ of observable values compatible with previous observed values in this individual. General methods to build these intervals are available, but they allow only a variable by variable follow-up whatever the possible correlations over time between them. In this article, we propose a general method to jointly follow-up several correlated variables over time. This methodology relies on a multivariate linear mixed effects model. We first provide a method to estimate the model’s parameters. In an asymptotic framework (N large enough), we then derive a $(1 - \alpha)$ individualized prediction region. Sometimes, the sample size N is not large enough for the asymptotic framework to give a reasonable prediction region. It is for this reason we propose and compare three different prediction regions that should behave better for small N . Finally, the whole methodology is illustrated by the follow-up of kidney insufficiency in cats.

Keywords : Corrected coverage rate; Longitudinal follow-up; Multivariate mixed Gaussian model; Plug-in estimator; Prediction region; Reference intervals.

1 Introduction

Longitudinal follow-up of biological variables is more and more used in preventive medicine. It consists of monitoring the markers of important functions for the early detection of slowly progressive diseases with a subclinical phase. For example, the prostate specific antigen (PSA) is used to detect prostate cancer in men. The same kind of follow-up is systematically done with teenagers using their weight and height to detect the beginning of obesity. In sport, like cycling or athletics, anti-doping control authorities try to generalize the use of a biological passport which consists of a longitudinal follow-up of some endogenous substances of interest in order to detect abnormal variations in an individual (Sottas et al., 2007; Zorzoli and Rossi, 2010).

A standard method of doing these follow-ups is to use the so-called reference intervals (CLSI, 2008). These intervals contain a fixed percentage (usually 95%) of measurements that can be observed in healthy individuals. However, this method suffers from several flaws. First, it does not use individual information i.e. a healthy individual can have extreme values, outside the reference interval, while for some other individuals values inside the reference interval are pathologic. Second, these intervals are built in an univariate framework (i.e. variable by variable) without taking into account the possible correlations between them. Finally, it does not account for their evolution over time within a given individual.

The individual reference intervals (or prediction intervals) mitigate this flaw by allowing the construction of a reference individual based on the observed values in a healthy individual and taking into account some covariables (such as sex, age). The literature on this subject is plentiful and the usual methodology is to use linear/nonlinear mixed effects models (Chi and Reinsel, 1989; Verbeke and Molenberghs, 2000; Davidian and Giltinan, 1995). In these models the observations are usually assumed to be independent conditional to the individual specific parameters (compound symmetry assumption). To our knowledge, the development of reliable methods to detect abnormal variations of longitudinal variables has remained limited. Sottas et al. (2007) proposed a Bayesian approach to combine population-derived limits and individual-based thresholds. Nevertheless, this method is built in an univariate framework whereas a follow-up is usually performed on several markers. By consequence, the individual reference intervals are built on each marker independently. Intuition suggests that building regions using simultaneous information on correlated variables could help to better detect abnormal values. Recently, Wang and Fan (2010) proposed a method to build prediction regions. They used a p order autoregressive process to model the autocorrelation of a variable with time while the correlations between different variables is assumed to be fixed over time. These regions are built using the asymptotic distribution of a statistic. In other words, the distribution of the statistics is computed assuming that the model parameters are known while estimates are used to compute it. While this plug-in method is easy to use, its very nature does not guarantee an exact coverage rate for the prediction region because it does not account for the imprecision of the parameter estimates. This can be a real problem when the sample size is small (Barndorff-Nielsen and Cox, 1996). In this case, assuming that the asymptotic framework holds could lead to considerable approximation errors. Therefore, special attention has to be paid to this problem to control the real coverage rate of the built

prediction region.

In this article, we propose to build an individual prediction region from previous observations of these variables carried out in the same individual and model parameter estimates. The observations obtained in an individual are assumed to be correlated over time. The correlation between a variable X_1 at time t_1 and a variable X_2 at time t_2 is not assumed to be equal to the correlation between X_1 at time t_2 and X_2 at time t_1 . This leads to highly structured autocorrelations that cannot be directly estimated by conventional methods (the NLMIXED procedure in SAS or the R nlme package). Therefore, we proposed a specific estimation method.

The model used to build these prediction regions is detailed on Section 2 and the estimation of these parameters is performed in Section 3. The prediction regions are then built in Section 4 using a plug-in approach. Three different corrections of the asymptotic confidence region are proposed and compared in Section 5. These corrections aim at correcting the plug-in estimation of the prediction region. The first two come from Hall et al. (1999); Ueki and Fueda (2007); Vidoni (2009); Fonseca et al. (2012), while the third is inspired by the work of Beran (1990) and Fonseca et al. (2014). A real dataset on kidney insufficiency follow-up in cats is then treated in Section 6. Finally, a discussion is provided in Section 7.

2 The model

Let us denote $\mathbf{X}_i = [\mathbf{X}_{i1} : \dots : \mathbf{X}_{ir}]$ the measurements performed in the i^{th} individual of a sample of size N . The vector \mathbf{X}_{ij} contains the n_i observations carried out over time for the j^{th} variable. More precisely, X_{ijk} is the value observed for the i^{th} individual for the j^{th} variable at time t_{ik} . Without loss of generality, we can assume that $t_{i1} \leq t_{i2} \leq \dots \leq t_{in_i}$. Note that all the variables are supposed to be measured at the same time for an individual, but time measures may differ from one individual to another. We assume that, up to a monotonic transformation

$$\mathbf{X}_i = \mathbf{B}_i \boldsymbol{\beta} + \mathbf{T}_i \boldsymbol{\Phi}_i + \boldsymbol{\zeta}_i \quad (1)$$

where \mathbf{B}_i and \mathbf{T}_i are known full-rank covariate matrices of dimensions $n_i \times p$ and $n_i \times q$ respectively, $\boldsymbol{\beta} = [\beta_1 : \dots : \beta_r]$ is a $p \times r$ matrix of parameters used to describe the population mean, $\boldsymbol{\Phi}_i = [\boldsymbol{\Phi}_{i1} : \dots : \boldsymbol{\Phi}_{ir}]$ and $\boldsymbol{\zeta}_i = [\zeta_{i1} : \dots : \zeta_{ir}]$ are respectively $q \times r$ and $n_i \times r$ matrices of unobserved Gaussian random effects. The variance of the components of the random matrix $\boldsymbol{\zeta}_i$ is assumed to be highly structured:

$$\text{cov}(\zeta_{ijk}, \zeta_{ij'k'}) = \sum_{jj'} \rho_{jj'}^{t_{ik} - t_{ik'}} \text{ if } k > k' \text{ and } \text{cov}(\zeta_{ijk}, \zeta_{ij'k'}) = \sum_{jj'} \rho_{jj'}^{t_{ik'} - t_{ik}} \text{ if } k < k'$$

where $\rho_{jj'} \in [0, 1]$ and $\sum_{jj'} = \alpha_{jj'} \sigma_j \sigma_{j'}$. The numbers $\alpha_{jj} = 1, \forall j \in \{1, \dots, r\}$, $\alpha_{jk} = \alpha_{kj} \in [-1, 1] \forall j \neq k \in \{1, \dots, r\}$ and $\rho_{jk} \in [0, 1] \forall j, k \in \{1, \dots, r\}$, σ_j represents the standard deviation of the j^{th} variable at each measurement time. The correlation between ζ_{ijk} and $\zeta_{ij'k'}$ is assumed to be $\rho_{jj'}^{t_{ik} - t_{ik'}}$ for $k > k'$ and $\rho_{jj'}^{t_{ik'} - t_{ik}}$ for $k < k'$. This means that we do not assume that the correlation between the j^{th} variable in $\boldsymbol{\zeta}_i$ measured at time k and the j^{th} variable in

ζ_i measured at time k' is the same as the correlation between the j^{th} variable in ζ_i measured at time k' and the j^{th} variable in ζ_i measured at time k . The major difference with the paper of Wang and Fan (2010) is that they assume that the observation times t_{ik} are equally spaced integer numbers, and that for all j and j' , $\text{cov}(\zeta_{ijk}, \zeta_{ij'k'}) = \sum_{jj'} \rho_{|t-t'|}$ where $\rho_{|t-t'|}$ is the correlation of an auto-regressive process of order p .

The matrix of the covariance of the ζ_i is a variance/covariance matrix because it is a symmetric and positive-definite matrix as the Kronecker and Shur products of two positives matrices (Bhatia, 2009). If this model writing is easy to understand, its multidimensional nature does not facilitate the estimation of parameters and the distribution definition of Φ_i and ζ_i . Thus, we rewrite this model in a vectorial framework to facilitate further estimations. Let us define $\psi_i = \text{vec}(\Phi_i)$ the vector obtained by stacking the columns of Φ_i columnwise. We assume that $\psi_i \stackrel{iid}{\sim} N(0, \Omega)$. The variance matrix $\Omega = [\omega_{jm}]_{jm}$ is block-partitioned with $q \times q$ variance matrices $\omega_{jm} = \text{cov}(\Phi_{ij}, \Phi_{im})$.

Similarly, the within subject error ζ_i can be stored columnwise into a vector $\varepsilon_i = \text{vec}(\zeta_i) \sim N(0, \Lambda_i(\rho, \Sigma))$. The matrix $\Lambda_i(\rho, \Sigma)$ can be written as $\mathbf{D}_i^{-1} \mathbf{R}_i^{-1} \mathbf{D}_i^{-1}$ where $\mathbf{D}_i^{-1} = \text{diag}(\sigma_1, \dots, \sigma_1, \sigma_2, \dots, \sigma_2, \sigma_r, \dots, \sigma_r)$ with each σ_j repeated n_i times. Thus, it is assumed to be constant over time. The matrix $\mathbf{R}_i^{-1}(\rho)$ is block-partitioned with $n_i \times n_i$ matrices ω_{ijk} with

$$\omega_{ijk}(\rho) = (\text{corr}(\zeta_{ijl}, \zeta_{ikf}))_{l,f \in \{1, \dots, n_i\}} = \alpha_{jk} \rho_{jk}^{|t_{if} - t_{il}|}$$

and $\alpha_{jj} = 1$. The matrix $\omega_{ijk}(\rho)$ contains the correlation between the j^{th} and k^{th} variable at the different sampling times.

The ε_i 's are assumed mutually independent and independent of the ψ_i 's. Let us define $\mathbf{Y}_i = \text{vec}(\mathbf{X}_i)$, $\mathbf{A}_i = \mathbf{1}_r \otimes \mathbf{B}_i$ and $\mathbf{Z}_i = \mathbf{1}_r \otimes \mathbf{T}_i$. Using these notations, the model (1) can be re-written as

$$\mathbf{Y}_i = \mathbf{A}_i \boldsymbol{\theta} + \mathbf{Z}_i \psi_i + \varepsilon_i \quad (2)$$

where $\boldsymbol{\theta} = \text{vec}(\boldsymbol{\beta})$. This model may appear to be a standard linear mixed effect model whose parameter $\boldsymbol{\xi} = (\boldsymbol{\theta}, \Omega, \Sigma, \rho) \in \Xi$ can be easily estimated using standard statistical software. However, the covariance matrices of this model are highly structured and their estimation needs careful development, which is done in Section 3.

Assume that n_w observations of the r variables are available at times (t_1, \dots, t_{n_w}) in a new individual. Let us denote $\mathbf{U} \in \mathbb{R}^{r \times 1}$ the future values that will be observed at time $t_u > t_{n_w}$ for the r variables in this new individual. We assume that

$$(\mathbf{W}' \mathbf{U})' = \mathbf{A} \boldsymbol{\theta} + \mathbf{Z} \psi + \varepsilon$$

where $\mathbf{Z} = (\mathbf{Z}'_w \mathbf{Z}'_u)'$, $\mathbf{A} = (\mathbf{A}'_w \mathbf{A}'_u)'$ are known matrices and $\varepsilon = (\varepsilon'_w \varepsilon'_u)'$. The random matrix $(\mathbf{W}' \mathbf{U})'$ is assumed to be independent of the \mathbf{Y}_i 's. We are looking for a region $\mathcal{R}_\xi^\alpha(\mathbf{W})$ so that $P(\mathbf{U} \in \mathcal{R}_\xi^\alpha(\mathbf{W}) | \mathbf{W}) = 1 - \alpha$.

To build such a region, we need two things: a random sample of individuals $(\mathbf{Y}_i)_{i \in \{1, \dots, N\}}$ that enables the population parameters $\boldsymbol{\xi}$ to be estimated and some observations performed in the individual of interest \mathbf{W} . We proceed in three steps: first, we build a region $\mathcal{R}_\xi^\alpha(\mathbf{W})$

by assuming that $\boldsymbol{\xi}$ is known, secondly, we plug-in the estimate $\widehat{\boldsymbol{\xi}}$ of $\boldsymbol{\xi}$ obtained using the sample $(\mathbf{Y}_i)_{i \in \{1, \dots, N\}}$ into $\mathcal{R}_{\boldsymbol{\xi}}^\alpha(\mathbf{W})$ to get $\mathcal{R}_{\widehat{\boldsymbol{\xi}}}^\alpha(\mathbf{W})$. Of course, because the estimate $\widehat{\boldsymbol{\xi}}$ is a random variable, this plug-in estimator does not guarantee a coverage of $1 - \alpha$. This is the reason why we propose and compare three different corrections for the estimated prediction region using this plug-in estimator in the third step.

3 Estimation of parameters

Let us denote $\boldsymbol{\Lambda}_i = \boldsymbol{\Lambda}_i(\boldsymbol{\rho}, \boldsymbol{\Sigma}) = \text{var}(\boldsymbol{\varepsilon}_i)$. Assume for a while that the parameter $\boldsymbol{\xi}$ is known and equal to $\boldsymbol{\xi}_0 = (\boldsymbol{\theta}_0, \boldsymbol{\Omega}_0, \boldsymbol{\Sigma}_0, \boldsymbol{\rho}_0)$. Then we have

$$\begin{pmatrix} \mathbf{Y}_i - \mathbf{A}_i \boldsymbol{\theta}_0 \\ \boldsymbol{\Psi}_i \end{pmatrix} \sim N \left(0, \begin{pmatrix} \mathbf{Z}_i \boldsymbol{\Omega}_0 \mathbf{Z}_i' + \boldsymbol{\Lambda}_i & \mathbf{Z}_i \boldsymbol{\Omega}_0 \\ \boldsymbol{\Omega}_0 \mathbf{Z}_i' & \boldsymbol{\Omega}_0 \end{pmatrix} \right)$$

so, by Shur lemma (Zhang, 2010), we obtain

$$\begin{aligned} \mathbf{m}_{i,0} &\triangleq \mathbb{E}_{\boldsymbol{\xi}_0}(\boldsymbol{\Psi}_i | \mathbf{Y}_i - \mathbf{A}_i \boldsymbol{\theta}_0) = \mathbb{E}_{\boldsymbol{\xi}_0}(\boldsymbol{\Psi}_i | \mathbf{Y}_i) = \boldsymbol{\Omega}_0 \mathbf{Z}_i' [\mathbf{Z}_i \boldsymbol{\Omega}_0 \mathbf{Z}_i' + \boldsymbol{\Lambda}_i]^{-1} [\mathbf{Y}_i - \mathbf{A}_i \boldsymbol{\theta}_0] \\ \mathbf{V}_{i,0} &\triangleq V_{\boldsymbol{\xi}_0}(\boldsymbol{\Psi}_i | \mathbf{Y}_i) = \boldsymbol{\Omega}_0 - \boldsymbol{\Omega}_0 \mathbf{Z}_i' [\mathbf{Z}_i \boldsymbol{\Omega}_0 \mathbf{Z}_i' + \boldsymbol{\Lambda}_i]^{-1} \mathbf{Z}_i \boldsymbol{\Omega}_0. \end{aligned}$$

To estimate the parameter $\boldsymbol{\xi}$ we propose to use the EM algorithm (Dempster et al., 1977). At iteration k , $\boldsymbol{\xi}_{k-1}$ is available. The EM algorithm alternates between two steps: the E step computes $Q(\boldsymbol{\xi}, \boldsymbol{\xi}_{k-1}) = \mathbb{E}[\log L(\mathbf{Y}_1, \dots, \mathbf{Y}_n; \boldsymbol{\xi}) | \mathbf{Y}_1, \dots, \mathbf{Y}_n; \boldsymbol{\xi}_{k-1}]$ and the M step solves $\boldsymbol{\xi}_k = \text{argsup}_{\boldsymbol{\xi}} Q(\boldsymbol{\xi}, \boldsymbol{\xi}_{k-1})$. As the \mathbf{Y}_i are independent, $Q(\boldsymbol{\xi}, \boldsymbol{\xi}_{k-1}) = \sum_i \mathbb{E}_{\boldsymbol{\xi}_{k-1}}[\log \mathcal{L}(\mathbf{Y}_i, \boldsymbol{\Psi}_i) | \mathbf{Y}_i]$.

Reminding that

$$-2 \log \mathcal{L}(\mathbf{Y}_i, \boldsymbol{\Psi}_i) = (\mathbf{Y}_i - \mathbf{A}_i \boldsymbol{\theta} - \mathbf{Z}_i \boldsymbol{\Psi}_i)' \boldsymbol{\Lambda}_i^{-1} (\mathbf{Y}_i - \mathbf{A}_i \boldsymbol{\theta} - \mathbf{Z}_i \boldsymbol{\Psi}_i) + \log |\boldsymbol{\Lambda}_i| + \boldsymbol{\Psi}_i' \boldsymbol{\Omega}^{-1} \boldsymbol{\Psi}_i + \log |\boldsymbol{\Omega}|,$$

we thus obtain

$$\begin{aligned} -2Q(\boldsymbol{\xi}, \boldsymbol{\xi}_{k-1}) &= \sum_{i=1}^N \text{tr}(\boldsymbol{\Lambda}_i^{-1} \mathbf{Z}_i \mathbf{V}_{i,k-1} \mathbf{Z}_i') + (\mathbf{Y}_i - \mathbf{A}_i \boldsymbol{\theta} - \mathbf{Z}_i \mathbf{m}_{i,k-1})' \boldsymbol{\Lambda}_i^{-1} (\mathbf{Y}_i - \mathbf{A}_i \boldsymbol{\theta} - \mathbf{Z}_i \mathbf{m}_{i,k-1}) \\ &\quad + \text{tr}(\boldsymbol{\Omega}^{-1} \mathbf{V}_{i,k-1}) + \mathbf{m}_{i,k-1}' \boldsymbol{\Omega}^{-1} \mathbf{m}_{i,k-1} + \log |\boldsymbol{\Lambda}_i| + \log |\boldsymbol{\Omega}|. \end{aligned}$$

The minimization of $-2Q(\boldsymbol{\xi}, \boldsymbol{\xi}_{k-1})$ with respect to $\boldsymbol{\Omega}$ and $\boldsymbol{\theta}$ leads respectively to

$$\begin{aligned} \boldsymbol{\Omega}_k &= \frac{1}{N} \sum_{i=1}^N [\mathbf{V}_{i,k-1} + \mathbf{m}_{i,k-1} \mathbf{m}_{i,k-1}'], \\ \boldsymbol{\theta}_k &= \left(\sum_{i=1}^N \mathbf{A}_i' \boldsymbol{\Lambda}_i^{-1} \mathbf{A}_i \right)^{-1} \sum_{i=1}^N (\mathbf{A}_i' \boldsymbol{\Lambda}_i^{-1} (\mathbf{Y}_i - \mathbf{Z}_i \mathbf{m}_{i,k-1})). \end{aligned}$$

Let us denote $\mathbf{U}_i = \mathbf{Y}_i - \mathbf{A}_i\boldsymbol{\theta} - \mathbf{Z}_i\mathbf{m}_{i,0}$ and $\mathbf{P}_i = \mathbf{Z}_i\mathbf{V}_{i,0}\mathbf{Z}'_i + \mathbf{U}_i\mathbf{U}'_i$. We want to compute

$$\arg \min_{\boldsymbol{\Sigma}, \boldsymbol{\rho}} -2Q(\boldsymbol{\Lambda}, \boldsymbol{\Lambda}_0) = \arg \min_{\boldsymbol{\Sigma}, \boldsymbol{\rho}} \text{tr} \left(\sum_{i=1}^N \boldsymbol{\Lambda}_i^{-1} \mathbf{P}_i \right) + \sum_{i=1}^N \log |\boldsymbol{\Lambda}_i|$$

that is

$$\arg \min_{\boldsymbol{\Sigma}, \boldsymbol{\rho}} -2Q(\boldsymbol{\Lambda}, \boldsymbol{\Lambda}_0) = \arg \min_{\boldsymbol{\Sigma}, \boldsymbol{\rho}} \text{tr} \left(\sum_{i=1}^N \mathbf{D}_i \mathbf{R}_i \mathbf{D}_i \mathbf{P}_i \right) - 2 \sum_{i=1}^N \log |\mathbf{D}_i|.$$

Let us denote $\mathbf{J}_{j,i} = \frac{\delta \mathbf{D}_i^{-1}}{\delta \sigma_j^{-1}}$. We have

$$-2 \frac{\partial Q}{\partial \sigma_j^{-1}} = 2 \left[\sum_{i=1}^N \text{tr}(\mathbf{D}_i \mathbf{F}_i^j) - \frac{n_i}{\sigma_j^{-1}} \right]$$

where $\mathbf{F}_i^j = \mathbf{P}_i \mathbf{J}_{j,i} \mathbf{R}_i$ is a $n_i r \times n_i r$ block-partitioned matrix with $n_i \times n_i$ matrices $(\mathbf{f}_i^j)_{k,l \in \{1, \dots, r\}}$. Thus,

$$-2 \frac{\delta Q}{\delta \sigma_j^{-1}} = 2 \sum_{k=1}^r \sum_{i=1}^N \left(\sigma_k^{-1} \text{tr} [(\mathbf{f}_i^j)_{kk}] - \frac{n_i}{\sigma_j^{-1}} \right)$$

and it remains to solve the system of equations $\frac{\delta Q}{\delta \sigma_j^{-1}} = 0, j = 1, \dots, r$ to get the maximum of Q with respect to $\sigma_1^{-1}, \dots, \sigma_r^{-1}$. Obviously, this system is not linear. However, we can note that it writes as $\mathbf{x}' \mathbf{B} = 1/\mathbf{x}$ where

$$\mathbf{x}' = (\sigma_1^{-1}, \sigma_2^{-1}, \dots, \sigma_r^{-1}), 1/\mathbf{x} = (\sigma_1, \sigma_2, \dots, \sigma_r) \text{ and } \mathbf{B} = \left(\frac{\sum_{i=1}^N \text{tr} [(\mathbf{f}_i^j)_{kk}]}{\sum_{i=1}^N n_i} \right)_{j,k \in \{1, \dots, r\}},$$

which can be solved iteratively using the recursion formula $\mathbf{x}^m \mathbf{B} = 1/\mathbf{x}^{m+1}$.

As explained in Section 2, α_{jk} and ρ_{jk} lies in $[-1, 1] \forall j, k \in \{1, \dots, r\}$. Thus, we can easily optimize numerically Q on this interval to obtain the estimates of α_{jk} and ρ_{jk} .

A good choice of starting values speeds up the convergence of this algorithm. These starting values can be easily derived from univariate analyses. The analysis of the j^{th} variable using model (1) gives $\hat{\beta}_j$, and $\hat{\boldsymbol{\Phi}}_{ij} \hat{\boldsymbol{\zeta}}_{ij}$ for all $i \in \{1, \dots, N\}$. The empirical variance of $\hat{\boldsymbol{\Phi}}_{ij}$ can be used as starting value of $\boldsymbol{\Omega}$, $\hat{\beta}_j$ can be used for β_j . Starting values for intra-individual variance components can be obtained as follows

$$\Sigma_{jj'} = \sum_{i=1}^N \sum_{k=1}^{n_i} \hat{\zeta}_{ijk} \hat{\zeta}_{ij'k} / (n_i N), \sigma_j = \sqrt{\Sigma_{jj}}, \alpha_{jj'} = \Sigma_{jj'} / (\sigma_j \sigma_{j'}).$$

The starting values for $\rho_{jj'}$ were computed by maximizing the likelihood obtained by assuming that $\hat{\boldsymbol{\zeta}}_i$ were distributed according to a $\mathcal{N}(0, \boldsymbol{\Lambda}_i(\boldsymbol{\rho}, \boldsymbol{\Sigma}))$. During this maximization, $\boldsymbol{\Sigma}$ was set to its initial value. The computer time needed for parameter estimation is less than one second using an ordinary laptop.

4 Building prediction regions

Remind that we assume that observations \mathbf{W} for the r variables are available in a new individual. We are going to build a prediction region for the next observation \mathbf{U} for this new individual. From the model defined in (2), we have

$$\mathbf{U} = \mathbf{A}_u \boldsymbol{\theta} + \mathbf{Z}_u \boldsymbol{\psi}_u + \boldsymbol{\varepsilon}_u. \quad (3)$$

According to Section 3, we assume in this section that all the model parameters are known. We denote

$$\mathbf{E} = \text{vec}(\boldsymbol{\varepsilon}_w, \boldsymbol{\varepsilon}_u) \sim N \left(0; \begin{pmatrix} \boldsymbol{\Lambda}_w(\boldsymbol{\rho}, \boldsymbol{\Sigma}) & \mathbf{M}_{wu}(\boldsymbol{\rho}, \boldsymbol{\Sigma})' \\ \mathbf{M}_{wu}(\boldsymbol{\rho}, \boldsymbol{\Sigma}) & \boldsymbol{\Lambda}_u(\boldsymbol{\rho}, \boldsymbol{\Sigma}) \end{pmatrix} \right)$$

where $\boldsymbol{\Lambda}_w(\boldsymbol{\rho}, \boldsymbol{\Sigma})$ and $\boldsymbol{\Lambda}_u(\boldsymbol{\rho}, \boldsymbol{\Sigma})$ are defined in the first section and $\mathbf{M}_{wu}(\boldsymbol{\rho}, \boldsymbol{\Sigma})$ is a $r \times (rn_w)$ matrix with

$$\mathbf{M}_{wu}(\boldsymbol{\rho}, \boldsymbol{\Sigma}) = \text{cov}(\boldsymbol{\varepsilon}_w; \boldsymbol{\varepsilon}_u) = \begin{pmatrix} \text{cov}(\varepsilon_u^1; \boldsymbol{\varepsilon}_{k=1, \dots, n_w}^1) & \dots & \text{cov}(\varepsilon_u^1; \boldsymbol{\varepsilon}_{k=1, \dots, n_w}^r) \\ \vdots & \dots & \vdots \\ \text{cov}(\varepsilon_u^r; \boldsymbol{\varepsilon}_{k=1, \dots, n_w}^1) & \dots & \text{cov}(\varepsilon_u^r; \boldsymbol{\varepsilon}_{k=1, \dots, n_w}^r) \end{pmatrix}$$

where ε_u^i is the i^{th} term of $\boldsymbol{\varepsilon}_u$ and $\boldsymbol{\varepsilon}_{k=1, \dots, n_w}^j$ is a n_w dimensional vector for variable j and individual \mathbf{W} and

$$\text{cov}(\varepsilon_u^i; \boldsymbol{\varepsilon}_{k=1, \dots, n_w}^j) = \left(\sum_{ij} \rho_{ij}^{|t_u - t_1|}, \dots, \sum_{ij} \rho_{ij}^{|t_u - t_{n_w}|} \right).$$

Using these notations and Schur lemma, we obtain the following proposition.

Proposition 4.1 *Let α be any real number in $[0; 1]$ and $\chi_{r, 1-\alpha}^2$ be the $1 - \alpha$ quantile of a chi-square distribution with r degrees of freedom. Let us consider the vector $\mathbf{m}(\boldsymbol{\xi}, \mathbf{W})$ and the matrix $\mathbf{V}(\boldsymbol{\xi})$ defined by*

$$\begin{aligned} \mathbf{m}(\boldsymbol{\xi}, \mathbf{W}) &= \mathbf{A}_u \boldsymbol{\theta} + (\mathbf{Z}_u \boldsymbol{\Omega} \mathbf{Z}_w' + \mathbf{M}_{wu}(\boldsymbol{\rho}, \boldsymbol{\Sigma})) (\mathbf{Z}_w \boldsymbol{\Omega} \mathbf{Z}_w' + \boldsymbol{\Lambda}_w(\boldsymbol{\rho}, \boldsymbol{\Sigma}))^{-1} (\mathbf{W} - \mathbf{A}_w \boldsymbol{\theta}), \\ \mathbf{V}(\boldsymbol{\xi}) &= (\mathbf{Z}_u \boldsymbol{\Omega} \mathbf{Z}_u' + \boldsymbol{\Lambda}_u(\boldsymbol{\rho}, \boldsymbol{\Sigma})) \\ &\quad - (\mathbf{Z}_u \boldsymbol{\Omega} \mathbf{Z}_w' + \mathbf{M}_{wu}(\boldsymbol{\rho}, \boldsymbol{\Sigma})) (\mathbf{Z}_w \boldsymbol{\Omega} \mathbf{Z}_w' + \boldsymbol{\Lambda}_w(\boldsymbol{\rho}, \boldsymbol{\Sigma}))^{-1} (\mathbf{Z}_u \boldsymbol{\Omega} \mathbf{Z}_w' + \mathbf{M}_{wu}(\boldsymbol{\rho}, \boldsymbol{\Sigma}))'. \end{aligned}$$

A $(1 - \alpha)$ prediction region of \mathbf{U} , conditionally to \mathbf{W} , is the set

$$S = \{ \mathbf{u} \in \mathbb{R}^r; \|\mathbf{V}(\boldsymbol{\xi})^{-1/2} (\mathbf{u} - \mathbf{m}(\boldsymbol{\xi}, \mathbf{W}))\|^2 \leq \chi_{r, 1-\alpha}^2 \} \quad (4)$$

where $\mathbf{V}(\boldsymbol{\xi})^{-1/2}$ is the inverse of the Cholesky transformation of $\mathbf{V}(\boldsymbol{\xi})$.

When $r > 1$, the prediction region for \mathbf{U} is thus an ellipsoid centered on $\mathbf{m}(\boldsymbol{\xi}, \mathbf{W})$. This ellipsoid degenerates to the interval $\left[m(\boldsymbol{\xi}, \mathbf{W}) - \tau_{(1-\alpha/2)} \sqrt{V(\boldsymbol{\xi})}; m(\boldsymbol{\xi}, \mathbf{W}) + \tau_{(1-\alpha/2)} \sqrt{V(\boldsymbol{\xi})} \right]$, where $\tau_{(1-\alpha/2)}$ is the $(1 - \alpha/2)$ quantile of the standard gaussian distribution, when one wants

to predict the next value U of a single variable (i.e. $r = 1$).

In this case, if $\rho = 0$ and assuming that there is no covariable, the j^{th} observation in the i^{th} individual writes

$$X_{ij} = Y_{ij} = \theta + \psi_i + \varepsilon_{ij}$$

with $\psi_i \sim N(0, \omega^2)$ and $\varepsilon_{ij} \sim N(0, \sigma^2)$. Using Schur complement, the $100(1-\alpha)\%$ prediction interval for the future value when $k-1$ observations are already available in an individual has the following expression :

$$\left[\frac{\theta}{1 + \gamma^2(k-1)} + \frac{\gamma^2(k-1)}{1 + \gamma^2(k-1)} \overline{W}_{k-1} - \tau_{(1-\alpha/2)} \sqrt{\frac{1 + \gamma^2 k}{1 + \gamma^2(k-1)} \sigma^2}, \right. \\ \left. \frac{\theta}{1 + \gamma^2(k-1)} + \frac{\gamma^2(k-1)}{1 + \gamma^2(k-1)} \overline{W}_{k-1} + \tau_{(1-\alpha/2)} \sqrt{\frac{1 + \gamma^2 k}{1 + \gamma^2(k-1)} \sigma^2} \right], \quad (5)$$

where $\gamma = \omega/\sigma$ and \overline{W}_{k-1} is the average of the $k-1$ available observations. Note that γ measures the benefit of the individualization compared to the usual reference interval built with a single value per individual (CLSI, 2008). When γ is high, the prediction interval is close to $[\overline{W} \pm \tau_{(1-\alpha/2)}\sigma]$ and the individualization is beneficial. When $k = 1$, i.e. when no observation is available for the new individual, the interval degenerates to the usual $[\theta \pm \tau_{(1-\alpha/2)}\sqrt{\sigma^2 + \omega^2}]$. As this model does not include any correlation between observations within each individual, the time interval between an observation and a future value does not modify the width of the prediction interval.

The plug-in estimator of S is \widehat{S} obtained by replacing in Equation (4) the parameter $\boldsymbol{\xi}$ by its estimate. The obtained prediction region fails to guarantee an exact coverage of $(1-\alpha)$. This occurs when the sample size n is low leading to some imprecise estimate for $\boldsymbol{\xi}$. To prevent this problem, we propose different strategies to correct the coverage of our prediction region.

5 Plug-in corrections

Let us define $K_{\mathbf{W}}(\boldsymbol{\xi}_0, \boldsymbol{\xi}) = \|V^{-1/2}(\boldsymbol{\xi})(\mathbf{U} - \mathbf{m}(\boldsymbol{\xi}, \mathbf{W}))\|^2$ where $(\mathbf{U}|\mathbf{W}) \sim N(\mathbf{m}(\boldsymbol{\xi}_0, \mathbf{W}), \mathbf{V}(\boldsymbol{\xi}_0))$. It is obvious that for all $\boldsymbol{\xi}$, $K_{\mathbf{W}}(\boldsymbol{\xi}, \boldsymbol{\xi})$ is a random variable distributed conditionally to \mathbf{W} according to a χ^2 distribution. This is no longer the case for $K_{\mathbf{W}}(\boldsymbol{\xi}_0, \widehat{\boldsymbol{\xi}})$. Let $\widehat{\boldsymbol{\xi}}$ be a n consistent estimate of $\boldsymbol{\xi}$ with a known distribution \mathcal{L} . As $(\mathbf{W}' \mathbf{U})'$ is assumed to be independent of the observations \mathbf{Y}_i , (\mathbf{W}, \mathbf{U}) and $\widehat{\boldsymbol{\xi}}$ are independent. We want to estimate $\mathbf{x}_{\alpha, \boldsymbol{\xi}_0, \widehat{\boldsymbol{\xi}}, \mathbf{W}}$ such that $P\left(K_{\mathbf{W}}\left(\boldsymbol{\xi}_0, \widehat{\boldsymbol{\xi}}\right) \leq \mathbf{x}_{\alpha, \boldsymbol{\xi}_0, \widehat{\boldsymbol{\xi}}, \mathbf{W}} \mid \mathbf{W}\right) = 1 - \alpha$. We define in the following sub-sections three different corrections which will be denoted by $\mathbf{x}_{\alpha, \boldsymbol{\xi}_0, \widehat{\boldsymbol{\xi}}, \mathbf{W}}^1$, $\mathbf{x}_{\alpha, \boldsymbol{\xi}_0, \widehat{\boldsymbol{\xi}}, \mathbf{W}}^2$ and $\mathbf{x}_{\alpha, \boldsymbol{\xi}_0, \widehat{\boldsymbol{\xi}}, \mathbf{W}}^3$. Theoretical properties of these corrections are not discussed here and we refer the interested reader to the references (Beran, 1990; Hall et al., 1999; Ueki and Fueda, 2007; Vidoni, 2009; Fonseca et al., 2012, 2014).

5.1 First method

A rough estimation $\mathbf{x}_{\alpha, \hat{\xi}, \hat{\xi}}^4$ of $\mathbf{x}_{\alpha, \xi_0, \hat{\xi}, \mathbf{W}}$ can be obtained by assuming that the data are drawn from a distribution with parameters $\hat{\xi}$ instead of ξ_0 . As $K_{\mathbf{W}}(\hat{\xi}, \hat{\xi})$ is distributed conditionally to \mathbf{W} according to a χ_r^2 distribution, $\mathbf{x}_{\alpha, \hat{\xi}, \hat{\xi}}^4 = \chi_{r, 1-\alpha}^2$ the $(1-\alpha)$ -quantile of this distribution. Because $\hat{\xi}$ is the maximum-likelihood estimate of ξ_0 , it achieves first order accuracy (Ueki and Fueda, 2007; Vidoni, 2009; Fonseca et al., 2012) and we have

$$1 - \alpha^+(\xi_0, \mathbf{W}) = P_{\xi_0} \left(K_{\mathbf{W}}(\xi_0, \hat{\xi}) \leq \mathbf{x}_{\alpha, \hat{\xi}, \hat{\xi}}^4 | \mathbf{W} \right) = 1 - \alpha + \delta(\xi_0, \mathbf{W})/N + O(N^{-3/2}). \quad (6)$$

It follows that

$$P_{\xi_0} \left(K_{\mathbf{W}}(\xi_0, \hat{\xi}) \leq \mathbf{x}_{\alpha - \delta(\xi_0, \mathbf{W})/N, \hat{\xi}, \hat{\xi}}^4 | \mathbf{W} \right) = 1 - \alpha + O(N^{-3/2}). \quad (7)$$

We thus obtain third order accuracy by correcting α by $\delta(\xi_0, \mathbf{W})/N$. But we have

$$\delta(\xi_0, \mathbf{W})/N = (1 - \alpha^+(\xi_0, \mathbf{W})) - (1 - \alpha) + O(N^{-3/2}).$$

As ξ_0 is unknown, this correcting term is also unknown. However, we can estimate it by using a parametric bootstrap method leading to $(1 - \alpha^+(\hat{\xi}, \mathbf{W}))$. So we have

$$\delta(\hat{\xi}, \mathbf{W})/N = (1 - \alpha^+(\hat{\xi}, \mathbf{W})) - (1 - \alpha) + O(N^{-3/2}) \quad (8)$$

and as $\delta(\hat{\xi}, \mathbf{W}) - \delta(\xi, \mathbf{W}) = O(1/\sqrt{N})$, we obtain $\delta(\hat{\xi}, \mathbf{W})/N = \delta(\xi, \mathbf{W})/N + O(N^{-3/2})$.

So it remains to evaluate $1 - \alpha^+(\hat{\xi}, \mathbf{W})$. Remind that if ξ^* is a bootstrap sample of ξ , we have asymptotically $\mathcal{L}(\hat{\xi} | \xi_0) \approx \mathcal{L}(\xi^* | \hat{\xi})$. Here we have 2 options. First, we know that $\hat{\xi} - \xi_0 \sim N(0, \mathbf{I}^{-1/2}(\xi_0)/N)$ (usual central limit theorem) and because $(\xi^* - \hat{\xi}) | \hat{\xi} \sim N(0, \mathbf{I}^{-1/2}(\hat{\xi})/N)$, we draw ξ^* in a $N(\hat{\xi}, \mathbf{I}^{-1/2}(\hat{\xi})/N)$ distribution. In the second option, we simulate $(\mathbf{Y}_i^*)_{1 \leq i \leq N}$ using the distribution of $(\mathbf{Y}_i)_{1 \leq i \leq N}$ with parameter $\hat{\xi}$ and we get ξ^* by estimating the parameter of the model (1) using $(\mathbf{Y}_i^*)_{1 \leq i \leq N}$ instead of $(\mathbf{Y}_i)_{1 \leq i \leq N}$.

Using Equations (7) and (8) and as $K_{\mathbf{W}}(\hat{\xi}, \xi^*) = \|(\mathbf{V}(\xi^*))^{-1/2}(\mathbf{U} - \mathbf{m}(\xi^*, \mathbf{W}))\|^2$ where $\mathbf{U} \sim N(\mathbf{m}(\hat{\xi}, \mathbf{W}); \mathbf{V}(\hat{\xi}))$, if we have a set of $(\xi_h^*)_{1 \leq h \leq H}$ we can deduce that

$$1 - \alpha^+(\hat{\xi}, \mathbf{W}) = \frac{1}{H} \sum_{h=1}^H \frac{1}{L} \sum_{l=1}^L \mathbf{1} \left[\|(\mathbf{V}(\xi_h^*))^{-1/2}(\mathbf{U}_l - \mathbf{m}(\xi_h^*, \mathbf{W}))\|^2 \leq \mathbf{x}_{\alpha, \hat{\xi}, \hat{\xi}}^4 \right]$$

where H and L are respectively the bootstrap sample sizes of (ξ^*) and \mathbf{U} . We conclude that

$$\mathbf{x}_{\alpha, \xi_0, \hat{\xi}, \mathbf{W}}^4 = \mathbf{x}_{\alpha - \delta(\hat{\xi}, \mathbf{W})/N, \hat{\xi}, \hat{\xi}}^4 + O(N^{-3/2}).$$

5.2 Second method

Let $F_{\xi_0, \mathbf{W}}(\mathbf{x}) = P\left(K_{\mathbf{W}}(\xi_0, \hat{\xi}) \leq \mathbf{x} | \mathbf{W}\right)$. We want $\mathbf{x}_{\alpha, \xi_0, \hat{\xi}, \mathbf{W}}^2$ such that $F_{\xi_0, \mathbf{W}}\left(\mathbf{x}_{\alpha, \xi_0, \hat{\xi}, \mathbf{W}}^2\right) = 1 - \alpha$ but $F_{\xi_0, \mathbf{W}}$ is unknown. As shown in (6), we have

$$1 - \alpha^+(\xi_0, \mathbf{W}) = F_{\xi_0, \mathbf{W}}\left(\mathbf{x}_{\alpha, \hat{\xi}, \hat{\xi}}^4\right) = 1 - \alpha + \frac{\delta(\xi_0, \mathbf{W})}{N} + O(N^{-3/2}).$$

Consequently, we obtain

$$\mathbf{x}_{\alpha, \xi_0, \hat{\xi}, \mathbf{W}}^2 = F_{\xi_0, \mathbf{W}}^{-1}(1 - \alpha) = F_{\xi_0, \mathbf{W}}^{-1}\left\{1 - \alpha^+(\xi_0, \mathbf{W}) - \delta(\xi_0, \mathbf{W})/N + O(N^{-3/2})\right\}$$

whose asymptotic expansion gives

$$\mathbf{x}_{\alpha, \xi_0, \hat{\xi}, \mathbf{W}}^2 = F_{\xi_0, \mathbf{W}}^{-1}\left\{1 - \alpha^+(\xi_0, \mathbf{W})\right\} - \frac{\delta(\xi_0, \mathbf{W})}{N} [F_{\xi_0, \mathbf{W}}^{-1}]' \left\{1 - \alpha^+(\xi_0, \mathbf{W})\right\} + O(N^{-3/2}).$$

It leads us to

$$\mathbf{x}_{\alpha, \xi_0, \hat{\xi}, \mathbf{W}}^2 = \mathbf{x}_{\alpha, \hat{\xi}, \hat{\xi}}^4 - \frac{\delta(\xi_0, \mathbf{W})}{N f_{\xi_0, \mathbf{W}}(\mathbf{x}_{\alpha, \hat{\xi}, \hat{\xi}}^4)} + O(N^{-3/2}) \quad (9)$$

where $f_{\xi_0, \mathbf{W}}$ is the probability density function of $K_{\mathbf{W}}(\xi_0, \hat{\xi})$. Of course $f_{\xi_0, \mathbf{W}}$ is unknown but because the model is regular we have for all \mathbf{x} , $(f_{\xi_0, \mathbf{W}}(\mathbf{x}) - f_{\hat{\xi}, \mathbf{W}}(\mathbf{x})) = O(1/\sqrt{n})$. We have seen in the previous section that $K_{\mathbf{W}}(\hat{\xi}, \hat{\xi})$ is a random variable distributed according to a χ^2 distribution. Consequently, the probability density function of the χ^2 distribution can approximate $f_{\xi_0, \mathbf{W}}$ and be plugged into (9) to give

$$\mathbf{x}_{\alpha, \xi_0, \hat{\xi}, \mathbf{W}}^2 = \mathbf{x}_{\alpha, \hat{\xi}, \hat{\xi}}^4 - \frac{\delta(\xi_0, \mathbf{W})}{N f_{\hat{\xi}, \mathbf{W}}(\mathbf{x}_{\alpha, \hat{\xi}, \hat{\xi}}^4)} + O(N^{-3/2}).$$

It remains to estimate $\delta(\xi_0, \mathbf{W})$ that can be obtained as with the previous method.

5.3 Third method

The first two methods can be read as delta-methods. The third method is an application of a simple parametric bootstrap method. As explained in the first method, we have for all \mathbf{x}

$$F_{\xi_0, \mathbf{W}}(\mathbf{x}) = P\left(K_{\mathbf{W}}(\xi_0, \hat{\xi}) \leq \mathbf{x} | \mathbf{W}\right) \approx P\left(K_{\mathbf{W}}(\hat{\xi}, \xi^*) \leq \mathbf{x} | \mathbf{W}\right).$$

Therefore, using notations of previous subsections, $F_{\xi_0, \mathbf{W}}(\mathbf{x})$ can be estimated by

$$\hat{F}_{\xi_0, \mathbf{W}}(\mathbf{x}) = \frac{1}{H} \sum_{h=1}^H \frac{1}{L} \sum_{l=1}^L \mathbb{1}\left[\|(\mathbf{V}(\xi_h^*))^{-1/2}(\mathbf{U}_{l-m}(\xi_h^*, \mathbf{W}))\|^2 \leq \mathbf{x}\right].$$

Denoting $\widehat{F}^{-1}(y) = \inf\{\mathbf{x} : F(\mathbf{x}) \geq y\}$, we have

$$\mathbf{x}_{\alpha, \xi_0, \widehat{\xi}, \mathbf{W}}^3 \approx \widehat{F}_{\xi_0, \mathbf{W}}^{-1}(1 - \alpha).$$

As discussed in Fonseca et al. (2014), this building is in line with Proposition 3B of Beran (1990) that shows that

$$P\left(K_{\mathbf{W}}(\xi_0, \widehat{\xi}) \leq \mathbf{x}_{\alpha, \xi_0, \widehat{\xi}, \mathbf{W}}^3 \mid \mathbf{W}\right) = 1 - \alpha + O(N^{-2}).$$

6 Illustrations

6.1 The dataset

The data come from a prospective study aimed at evaluating variations over time of several biochemical variables in healthy cats. This study was carried in the clinics of the Veterinary College, which usually received sick animals or healthy animals for sterilization (obviously only once). This is the reason why only $N = 20$ healthy cats could have been included in this study. The main variables for renal follow-up are the urea X_1 , the creatinine X_2 and the protein X_3 which are plotted in Figure 1 for the 20 healthy cats. There is no reason to think that these variables are not stable over time in healthy cats (Reynolds et al., 2010; Lefebvre, 2011). Univariate analyses were performed and the effect of time was found not significant. Every cat but three were measured five times: 0, 3, 6, 12 and 24 months after inclusion. The remaining three were sampled only for the first four times. Note that the entire study is performed on the log transformation of the variables as usual.

So, according to Section 2, we propose the following model

$$\mathbf{X}_i = \mathbf{B}_i \boldsymbol{\beta} + \mathbf{T}_i \boldsymbol{\Phi}_i + \boldsymbol{\zeta}_i \quad (10)$$

where \mathbf{X}_i is a $n_i \times 3$ matrix with n_i the number of observation for the i^{th} cat (4 or 5), \mathbf{B}_i and \mathbf{T}_i are vectors of length n_i such that $\mathbf{B}_i = \mathbf{T}_i = (1, \dots, 1)$, $\boldsymbol{\beta} = (\beta_1, \beta_2, \beta_3)$ and $\boldsymbol{\phi}_i = (\phi_{i1}, \phi_{i2}, \phi_{i3})$ and $\boldsymbol{\zeta}_i$ is a $n_i \times 3$ matrix. According to Section 2, in this application we have $n_i = 4$ or 5 , $r = 3$, $p = 1$, $q = 1$ and $N = 20$. As we have no available covariable (age, sex), the matrix \mathbf{B}_i does not incorporate any information but this kind of information can easily be inserted in our model as in Sottas et al. (2007).

6.2 Estimation of the parameters

Using the method proposed in Section 3, we estimated the parameters and found the following values : $\widehat{\sigma}_1 = 0.083$, $\widehat{\sigma}_2 = 0.062$, $\widehat{\sigma}_3 = 0.032$, $\widehat{\theta}_1 = 2.054$, $\widehat{\theta}_2 = 4.834$, $\widehat{\theta}_3 = 4.368$,

$$\widehat{\boldsymbol{\alpha}} = \begin{pmatrix} 1 & & \\ 0.306 & 1 & \\ 0.074 & 0.211 & 1 \end{pmatrix}, \widehat{\boldsymbol{\rho}} = \begin{pmatrix} 0.411 & 0.001 & 0.010 \\ 0.013 & 0.005 & 0.010 \\ 0.009 & 0.896 & 0.005 \end{pmatrix} \text{ and } \widehat{\boldsymbol{\Omega}} = \begin{pmatrix} 0.022 & & \\ 0.011 & 0.015 & \\ -0.001 & -0.001 & 0.003 \end{pmatrix}.$$

To illustrate the use of this estimates, here are estimates of correlations between variable 2 and 3 at times t (in months) and $t' > t$ carried out in the same individual:

$$\hat{\omega}_{23} = \begin{pmatrix} 1 & & & & \\ \hat{\rho}_{22}^{t'-t} & & & & \\ \hat{\alpha}_{23} & 1 & & & \\ \hat{\alpha}_{23}\hat{\rho}_{23}^{t'-t} & \hat{\alpha}_{23}\hat{\rho}_{32}^{t'-t} & 1 & & \\ \hat{\alpha}_{23}\hat{\rho}_{23}^{t'-t} & \hat{\alpha}_{23} & \hat{\rho}_{33}^{t'-t} & 1 & \end{pmatrix} = \begin{pmatrix} 1 & & & & \\ 0.005^{t'-t} & & & & \\ 0.211 & & 1 & & \\ 0.211 * 0.010^{t'-t} & & 0.211 * 0.896^{t'-t} & & \\ 0.211 * 0.010^{t'-t} & & 0.211 & & 0.005^{t'-t} & 1 \end{pmatrix}.$$

In this example, the correlation between two successive measurements carried out in the same individual is rather low for practical use with $t' - t > 1$ month. More surprisingly, it appears that no variable is an earlier marker than the others to detect kidney insufficiency. In other words, there is no major correlation between two different variables at two different times. This result could not be anticipated. With this result, the benefit of the individualization can be roughly measured by the ratio $\gamma = \omega/\sigma$ (see (5)) which is equal to 1.8, 2.0 and 1.7 for urea, creatinine and protein respectively. As these ratios are greater than one, one can expect the individualized region (4) to be narrower than the population counterpart.

6.3 Comparison of the three corrections

Because our first motivation is to propose a reference region for variables measuring kidney insufficiency in cats, we compared the three proposal corrections in this context. To compare the three different corrections obtained with the three proposed methods and the standard χ^2 threshold $\mathbf{x}_{\alpha, \hat{\xi}, \hat{\xi}}^4$ without correction, we carried out a simulation study.

The goal of the study was to compare $P\left(K_{\mathbf{W}}\left(\xi_0, \hat{\xi}\right) \leq \mathbf{x} \mid \mathbf{W}\right)$ for each \mathbf{x} in the set $\left\{\mathbf{x}_{\alpha, \xi_0, \hat{\xi}, \mathbf{W}}^1, \mathbf{x}_{\alpha, \xi_0, \hat{\xi}, \mathbf{W}}^2, \mathbf{x}_{\alpha, \xi_0, \hat{\xi}, \mathbf{W}}^3, \mathbf{x}_{\alpha, \hat{\xi}, \hat{\xi}}^4\right\}$. So we first need to know $\xi_0 = (\theta_0, \Omega_0, \Sigma_0, \rho_0)$ and we take ξ_0 as the maximum likelihood estimate for ξ in the cat's dataset. Using ξ_0 , we simulate 500 new datasets $\left(\left(\mathbf{Y}_i^k\right)_{i=1, \dots, 20}\right)_{k=1, \dots, 500}$, where (\mathbf{Y}_i^k) is the vector of size $rn_i = 15$ containing the values of the three variables at the five sampling times for the i^{th} cat of the k^{th} dataset. The three corrections we want to compare depends on the past values observed in the individual of interest \mathbf{W} . Because we want to compare the average performances of these corrections, we chose to explore the possible \mathbf{W} values that could be observed in an individual. This is the reason why, for each dataset, we chose to predict for a single cat the values \mathbf{U}_k from \mathbf{W}_k which contains the $n_w = 4$ measurements at 0, 3, 6 and 12 months for the three variables. We simulated 5000 pairs $(\mathbf{W}_k, \mathbf{U}_k)$ drawn from a $\mathcal{N}(\mathbf{A}_w \theta_0, \mathbf{Z}_w \Omega_0 \mathbf{Z}_w' + \Lambda(\rho_0, \Sigma_0))$ where \mathbf{U}_k contains the corresponding measurements at time $t_u = 24$ months. For each dataset $\left(\left(\mathbf{Y}_i^k\right)_{i=1, \dots, 20}\right)_k$, we first computed the maximum likelihood estimate $\hat{\xi}_k$. We took $H = 200$ and $L = 1$ to compute the three corrections and, for each dataset k , we compute each correction $\left(\mathbf{x}_{\alpha, \xi_0, \hat{\xi}_k, \mathbf{W}_k}^l\right)_{l=1, 2, 3}$ and the classical value $\mathbf{x}_{\alpha, \hat{\xi}_k, \hat{\xi}_k}^4$ that does not depend on \mathbf{W}_k . Note that we only performed one estimate of the corrections by bootstrap sample. Conditionally to \mathbf{W}_k , the prediction regions were then calculated and we checked whether or not \mathbf{U}_k belonged to these regions.

Finally, we estimated $P\left(K_{\mathbf{W}}(\boldsymbol{\xi}_0, \widehat{\boldsymbol{\xi}}_k) \leq \mathbf{x}\right)$ by its empirical average version

$$\frac{1}{5000} \sum_{k=1}^{5000} \mathbb{1}_{[K_{\mathbf{W}}(\boldsymbol{\xi}_0, \widehat{\boldsymbol{\xi}}_k) \leq \mathbf{x}]}$$

for each \mathbf{x} in the set $\left\{\mathbf{x}_{\alpha, \boldsymbol{\xi}_0, \widehat{\boldsymbol{\xi}}_k, \mathbf{W}_k}^1, \mathbf{x}_{\alpha, \boldsymbol{\xi}_0, \widehat{\boldsymbol{\xi}}_k, \mathbf{W}_k}^2, \mathbf{x}_{\alpha, \boldsymbol{\xi}_0, \widehat{\boldsymbol{\xi}}_k, \mathbf{W}_k}^3, \mathbf{x}_{\alpha, \widehat{\boldsymbol{\xi}}_k, \widehat{\boldsymbol{\xi}}_k}^4\right\}$. The results of these corrections are shown in Table 1.

Table 1: Comparison of average coverage probabilities (standard deviations) for the three corrections and the standard χ^2 threshold for $\alpha = 0.95$.

Method	1 st correction	2 nd correction	3 rd correction	Standard χ^2
Coverage probability	0.955 (0.003)	0.956 (0.003)	0.948 (0.003)	0.933 (0.004)

We can note that the third correction gives the narrower prediction region. In the case of $\mathbf{x}_{\alpha, \widehat{\boldsymbol{\xi}}, \widehat{\boldsymbol{\xi}}}^4$, the imprecision of the estimation is not taken into account in the computation of the prediction region leading to a small prediction region. This is not the case for the first and the second correction that give close results. Contrary to the results obtained by Vidoni (2009), they both overestimate the coverage and lead to wide prediction regions. Note that similar results were obtained for other confidence levels. Comparing the average coverage probabilities is certainly desirable but it does not allow the size of the prediction regions to be compared, which is why we represent the distributions of the corrections in Figure 2. For a given value of $\widehat{\boldsymbol{\xi}}_k$, the size of the prediction region is governed by the different \mathbf{x}^l . Indeed, a correction would not be necessary if $\boldsymbol{\xi}_0$ were known and not estimated. Here, the knowledge of $\boldsymbol{\xi}_0$ enables the targeted 95% quantile to be calculated for each $(\mathbf{W}_k, \widehat{\boldsymbol{\xi}}_k)$, and, consequently, the optimal size for the prediction region. Deconditioning on $(\mathbf{W}, \widehat{\boldsymbol{\xi}})$ gives the distribution of the actual 95% quantile, in red in Figure 2. This distribution is estimated empirically with 5000 values of $(\mathbf{W}_k, \widehat{\boldsymbol{\xi}}_k)$. As we can see, the distribution of \mathbf{x}^3 has a lower variance than the others that insures a small variation of the prediction region size from one sample to another. Furthermore, its distribution is the closest to that of the actual quantile. According to this simulation study, we will provide results in the next section using the third correction $\mathbf{x}_{\alpha, \boldsymbol{\xi}_0, \widehat{\boldsymbol{\xi}}, \mathbf{W}}^3$ and it will be calculated as the average of 5000 bootstrap samples. Note that the best performances of the third method are only established here for the used dataset and could not be true in general. A possible explanation could be that the first correction is based on a χ^2 approximation for the distribution of $K_{\mathbf{W}}(\boldsymbol{\xi}_0, \widehat{\boldsymbol{\xi}})$ and a bootstrap adjustment for the targeted coverage. Furthermore the second one is a truncated asymptotic expansion of the first one and consequently is less accurate. By its very construction, the third correction was expected to be the best one because it does not *a priori* assume a χ^2 distribution for $K_{\mathbf{W}}(\boldsymbol{\xi}_0, \widehat{\boldsymbol{\xi}})$. Its only approximation is to substitute the real distribution of $\widehat{\boldsymbol{\xi}}$ by its bootstrap counterpart. As expected it achieves a coverage probability very close to the targeted one.

6.4 Prediction region

We have a new cat for which we possess four measurements (at 0, 3, 6 and 12 months) for each variable. Using the proposed method, we can build an individual reference region (an ellipsoid) for future values for these variables. If its future measures lie outside this region, this cat has a low probability of being healthy. The results on this new cat are plotted in Figure 1. Because clinicians are not accustomed to matrix calculus, it is not easy to check whether or not a new point on the given cat belongs to its prediction region. This is the reason why we proposed to represent the projection of the ellipsoid $K_{\mathbf{W}}(\boldsymbol{\xi}_0, \widehat{\boldsymbol{\xi}}) \leq \mathbf{x}_{\alpha, \boldsymbol{\xi}_0, \widehat{\boldsymbol{\xi}}, \mathbf{W}}^3$ for each variable. This gives an interval of prediction for each variable and each future time of measurement. Note that these intervals are presented to give a graphical representation. As they were obtained by projection they do not guarantee the right coverage contrarily to the ellipsoids defined by Proposition 4.1. So, they can not be used separately to diagnose a cat as the three variables are strongly related. As soon as a value of a variable is outside the prediction region, the cat can be considered as probably not healthy.

We can remark that our prediction intervals are very different and narrower than the so-called "reference intervals" and therefore lead to different clinical decisions. As an example, a log(Creatinine) of 5.15 at fifteen months would be detected as suspicious for the new cat using the standard reference intervals while the individualization does not trigger such a false alarm. On the other hand, a log(Urea) value of 1.8 would be detected as abnormal by our method but not by the usual reference intervals. The reduction of width for the prediction region decreases the probability for each individual of being detected as a false-negative. Despite the considerable difference between the χ^2 threshold $\mathbf{x}_{\alpha, \widehat{\boldsymbol{\xi}}, \widehat{\boldsymbol{\xi}}}^4$ and $\mathbf{x}_{\alpha, \boldsymbol{\xi}_0, \widehat{\boldsymbol{\xi}}, \mathbf{W}}^3$, the corresponding prediction regions are very close. In this case, this can be explained by a small variance in a future value conditional on the observations. This cannot easily be anticipated by a simple glance on the parameter estimations because this conditional variance depends on a complicated function of all the variance parameters (see Proposition 4.1).

7 Discussion

In this paper, we introduce a new model to build prediction regions. This approach is new because it is individualized and multidimensional. Indeed, every individual gets its own prediction region which takes into account the possible correlations between all the variables at all the different times. These advantages enable us to build narrower prediction regions than the usual "reference intervals" method. Using our methodology, clinicians will be alerted with more precision to a potential unhealthy animal or person.

Nevertheless, our model is based on a strong Gaussian assumption which can be false. An alternative could be the use of a nonparametric framework but it would need more individuals and, by consequence, it can not be applied to our practical problem. This assumption is also a classical one at least up to a Box-Cox transformation (CLSI, 2008).

On the other hand, an assumption was also made on the exponential decrease of the correlation over time that can appear restrictive. To the best of our knowledge, this kind of

problem has already been modeled by an $AR(p)$ -process (Wang and Fan, 2010): an assumption difficult to check. In this respect, the model we propose can be seen in continuous time as a first order approximation of such chains.

Our parametric framework allows us to use a plug-in approach for the model estimation, but the plug-in approach adds errors in the estimation of the prediction region and fails to guarantee an exact coverage rate. To circumvent this problem, we use three different corrections and the method 3 seems to give better results on our dataset.

Recently, some robust extensions concerning the multivariate linear mixed model have been proposed. These extensions are based on the multivariate t and skew normal distributions and provide approaches with a more general family of distributions (Wang and Fan, 2011, 2012; Wang, 2013; Lin and Wang, 2013). An interesting future work would be an extension of our results for these models.

References

- Barndorff-Nielsen, O. E. and Cox, D. R. (1996). Prediction and asymptotics. *Bernoulli*, 2(4):319–340.
- Beran, R. (1990). Calibrating prediction regions. *Journal of the American Statistical Association*, 85(411):715–723.
- Bhatia, R. (2009). *Positive Definite Matrices*. Princeton Series in Applied Mathematics. Princeton University Press.
- Chi, E. M. and Reinsel, G. C. (1989). Models for longitudinal data with random effects and AR(1) errors. *Journal of the American Statistical Association*, 84(406):452–459.
- CLSI (2008). *Defining, Establishing, and Verifying Reference Intervals in the Clinical Laboratory: Approved Guideline*. CLSI document C28-A3. Wayne, PA: CLSI.
- Davidian, M. and Giltinan, D. M. (1995). *Nonlinear models for repeated measurement data*. Monographs on statistics and applied probability. Chapman & Hall, London, New York.
- Dempster, A. P., Laird, M. N., and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society: Series B*, 39:1–22.
- Fonseca, G., Giummolè, F., and Vidoni, P. (2012). A note about calibrated prediction regions and distributions. *Journal of Statistical Planning and Inference*, 142(9):2726–2734.
- Fonseca, G., Giummolè, F., and Vidoni, P. (2014). Calibrating predictive distributions. *Journal of Statistical Computation and Simulation*, 84:1–11.
- Hall, P., Peng, L., and Tajvidi, N. (1999). On prediction intervals based on predictive likelihood or bootstrap methods. *Biometrika*, 86(4):871–880.

- Lefebvre, H. P. (2011). Renal function testing. In J., B. and J., P. D., editors, *Nephrology and urology of small animals*, pages 91–96. Blackwell Publishing, Ames, IA, USA.
- Lin, T. I. and Wang, W.-L. (2013). Multivariate skew-normal at linear mixed models for multi-outcome longitudinal data. *Statistical Modelling*, 13:199–221.
- Reynolds, B., Concordet, D., Germain, C., Daste, T., Boudet, K., and Lefebvre, H. (2010). Breed dependency of reference intervals for plasma biochemical values in cats. *Journal of Veterinary Internal Medicine*, 24(4):809–818.
- Sottas, P.-E., Baume, N., Saudan, C., Schweizer, C., Kamber, M., and Saugy, M. (2007). Bayesian detection of abnormal values in longitudinal biomarkers with an application to t/e ratio. *Biostatistics*, 8(2):285–296.
- Ueki, M. and Fueda, K. (2007). Adjusting estimative prediction limits. *Biometrika*, 94(2):509–511.
- Verbeke, G. and Molenberghs, G. (2000). *Linear mixed models for longitudinal data*. Springer Series in Statistics. Springer-Verlag, New-York.
- Vidoni, P. (2009). Improved prediction intervals and distribution functions. *Scandinavian Journal of Statistics*, 36(4):735–748.
- Wang, W. L. (2013). Multivariate t linear mixed models for irregularly observed multiple repeated measures with missing outcomes. *Biometrical Journal*, 55:554–571.
- Wang, W. L. and Fan, T. H. (2010). ECM-based maximum likelihood inference for multivariate linear mixed models with autoregressive errors. *Computational Statistics and Data Analysis*, 54(5):1328–1341.
- Wang, W. L. and Fan, T. H. (2011). Estimation in multivariate t linear mixed models for multiple longitudinal data. *Statistica Sinica*, 21:1857–1880.
- Wang, W. L. and Fan, T. H. (2012). Bayesian analysis of multivariate t linear mixed models using a combination of ibf and gibbs samplers. *Journal of Multivariate Analysis*, 105(1):300–310.
- Zhang, F. (2010). *The Schur Complement and Its Applications*. Numerical Methods and Algorithms. Springer.
- Zorzoli, M. and Rossi, F. (2010). Implementation of the biological passport: The experience of the international cycling union. *Drug Testing and Analysis*, 2(11-12):542–547.

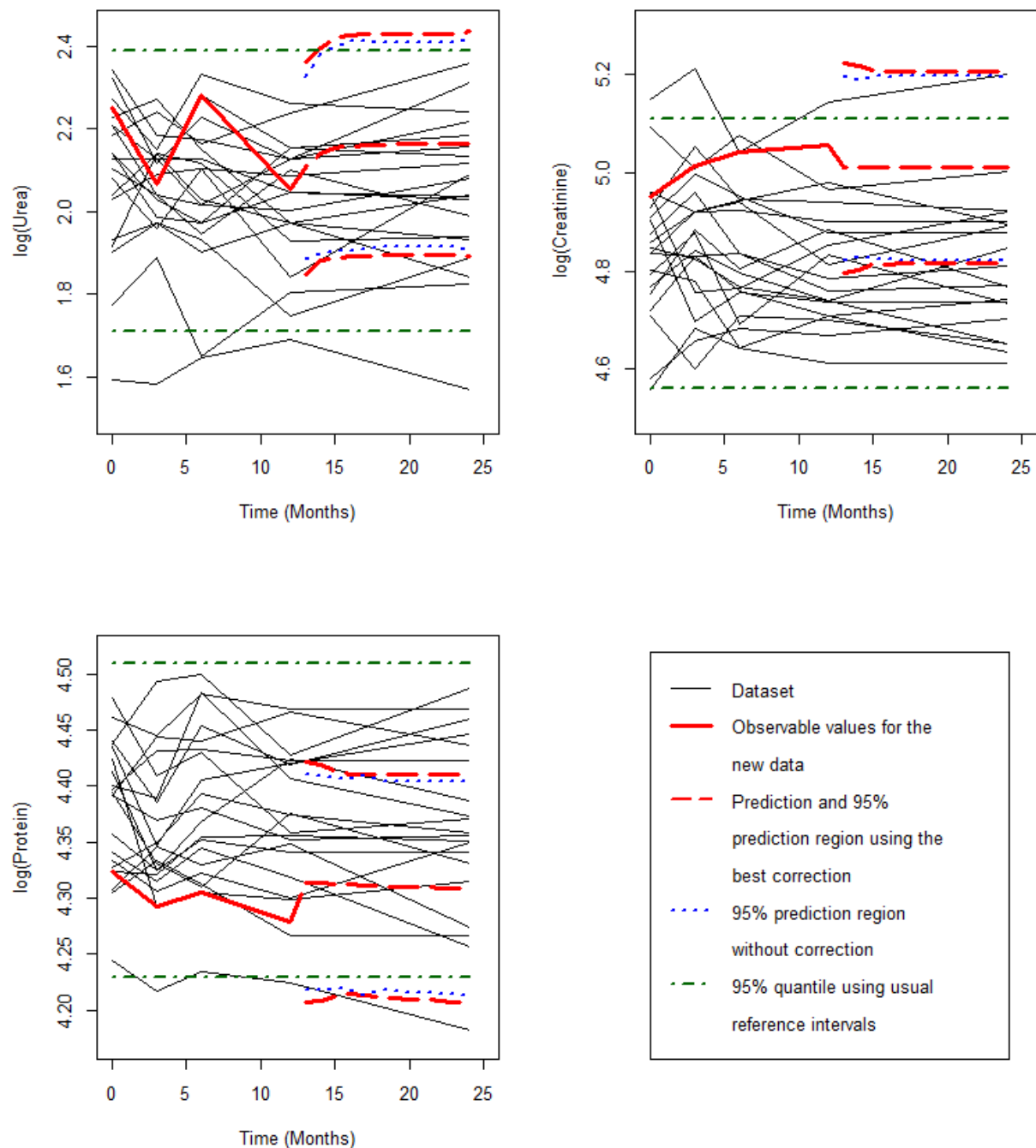


Figure 1: The three variables of interest are plotted for the whole dataset and the new cat in bold. The usual reference intervals (in dotted-dashed lines) are wider than the individual ones (in dashed lines). The prediction region using the standard χ^2 threshold $\mathbf{x}_{0.95, \hat{\xi}, \hat{\xi}}^4$ is close to those obtained using $\mathbf{x}_{0.95, \xi_0, \hat{\xi}, \mathbf{w}}^3$. This figure appears in color in the electronic version of this article.

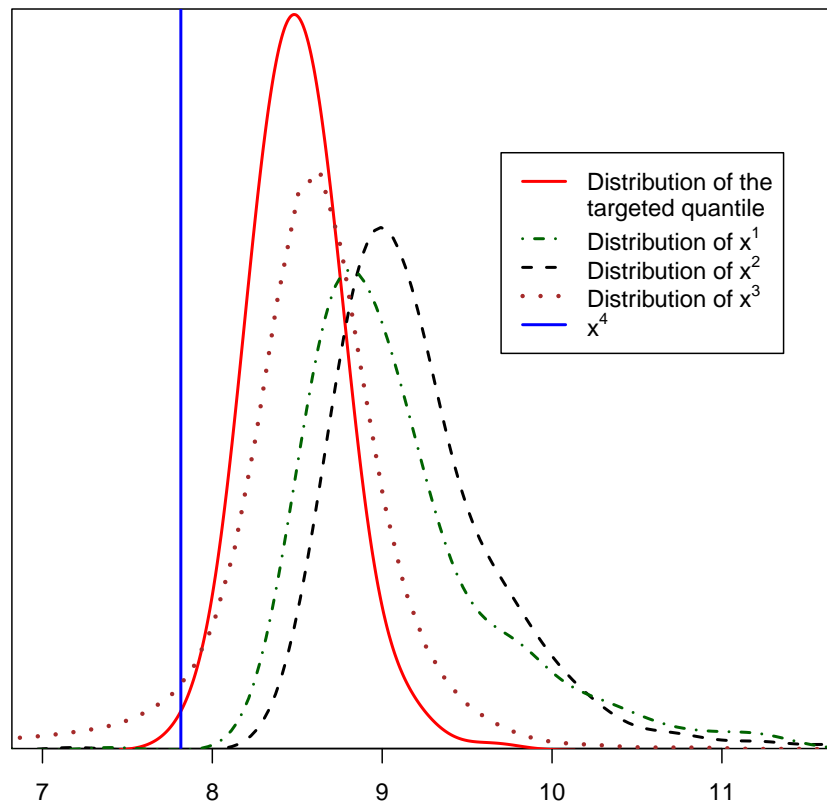


Figure 2: Distributions of the targeted quantile and of the different corrections are represented using a kernel estimate, \mathbf{x}^l standing for $\mathbf{x}_{0.95, \xi_0, \hat{\xi}, \mathbf{W}}^l$ for $l = 1, 2, 3$ and \mathbf{x}^4 for $\mathbf{x}_{0.95, \hat{\xi}, \hat{\xi}}$. The solid line distribution shows how the actual quantile changes with \mathbf{W} . The distribution of $\mathbf{x}_{0.95, \xi_0, \hat{\xi}, \mathbf{W}}^3$ is the closest to the actual distribution. Overall, methods 1 and 2 overestimate the actual quantiles while method 4 underestimates them. This figure appears in color in the electronic version of this article.