Journal of Pharmacokinetics and Pharmacodynamics, Vol. 32, No. 1, February 2005 (© 2005)
DOI: 10.1007/s10928-005-2104-x

Non-Linear Mixed-Effects Models with Stochastic Differential Equations: Implementation of an Estimation Algorithm

Rune V. Overgaard,[1,2,*] Niclas Jonsson,[3] Christoffer W. Tornøe,[1,3,4] and Henrik Madsen[1]

Received May 26, 2004—Final March 22, 2005

Pharmacokinetic/pharmacodynamic modelling is most often performed using non-linear mixed-effects models based on ordinary differential equations with uncorrelated intra-individual residuals. More sophisticated residual error models as e.g. stochastic differential equations (SDEs) with measurement noise can in many cases provide a better description of the variations, which could be useful in various aspects of modelling. This general approach enables a decomposition of the intra-individual residual variation $\boldsymbol{\epsilon}$ into system noise $\boldsymbol{w}$ and measurement noise $\boldsymbol{e}$. The present work describes implementation of SDEs in a non-linear mixed-effects model, where parameter estimation was performed by a novel approximation of the likelihood function. This approximation is constructed by combining the First-Order Conditional Estimation (FOCE) method used in non-linear mixed-effects modelling with the Extended Kalman Filter used in models with SDEs. Fundamental issues concerning the proposed model and estimation algorithm are addressed by simulation studies, concluding that system noise can successfully be separated from measurement noise and inter-individual variability.

KEY WORDS: SDE; PK/PD; Kalman filter; population modelling; system noise; correlated residuals.

INTRODUCTION

Non-linear mixed effects modelling has proven to be a useful tool in the characterization of pharmacokinetic (PK) and pharmacodynamic (PD) properties of drugs (see (1) and (2)). The models used in this mode of

[1]Informatics and Mathematical Modelling, Technical University of Denmark, Lyngby, Denmark.
[2]Experimental Medicine, Novo Nordisk A/S, Bagsvaerd, Denmark.
[3]Department of Pharmaceutical Biosciences, Uppsala University, Sweden.
[4]Experimental Medicine, Ferring Pharmaceuticals A/S, Copenhagen, Denmark.
*To whom correspondence should be addressed. Telephone: +45-30-24-7876; fax: +45-45-88-2673; e-mail: rvo@imm.dtu.dk

analysis are most frequently based on ordinary differential equations (ODEs), or the solutions thereof, supplemented by a model for the inter-individual variations in the structural model parameters and a model for the variation of the residuals that assumes independence, so that the residuals are uncorrelated. However, correlations between residuals are not uncommon, and it is well known that a violation of this basic statistical assumption may lead to erroneous estimates, for example of the inter-individual variations as demonstrated in (3).

In NONMEM (4), which is the most commonly used software for PK/PD analysis using nonlinear mixed-effects models, it is possible to handle correlated residual errors using an AR(1) model (3). Simulations suggest that the introduction of a model for correlated residuals may lead to (1) better estimates of the inter-individual variation, (2) better estimates of the structural parameters, (3) and a diagnostic tool giving a measure of the model improvement.

An alternative approach to model correlated residuals is to use stochastic differential equations (SDEs). This model structure includes the statistical functionality of the continuous AR(1) model, but is more flexible with respect to specifying models for different residual error correlation patterns.

With SDEs, the differences between individual predictions and observations are explained by two fundamentally different types of noise: (1) the dynamic noise, which enters through the dynamics of the system and may originate from model deficiencies or true random fluctuations within the system, and (2) the measurement noise, which represents the uncorrelated part of the residual variability, may be due to assay error or if the sample concentration is not representative for the true concentration in plasma. This could e.g. occur for samples during the distribution phase of an intravenous (IV) bolus administrated drug. The difference between two measurements at the same time point will therefore only be due to measurement noise. In a recent book on PK/PD modelling, it is suggested that the intra-individual variations may be more appropriately modelled by using SDEs rather than ODEs (5).

In addition to separating the residual error into dynamic noise and measurement noise, SDEs also allow the dynamic noise to be attributed to different model components. For example, if the absorption process of an orally administered drug cannot be well described by the model, this may lead to correlated residuals. However, if an AR(1) model was used to account for this, the auto-correlation pattern would be assumed effective along the whole concentration time profile. With SDEs, the dynamic noise component could be put directly on the state equation for the absorption meaning that the auto-correlation pattern is only assumed effective as long

as absorption occurs. Furthermore, SDEs could facilitate the estimation of the actual absorption profile in a way similar to deconvolution, and thus reveal the misspecifications of the model (6). The same approach can be used for other model components as well, and SDEs therefore has the potential to be a useful model building tool, as well as a diagnostic tool (7).

Besides the increased functionality, SDEs may indeed offer practical benefits in terms of easier PK/PD modelling, particularly when more complicated mechanistic models are used. In this case, the number and complexity of the mechanisms involved may be too great for inclusion in a model used for estimation. Here SDEs may be included to describe some of these mechanisms, while only the major mechanisms are treated by the parametric model.

SDEs have been used for individual non-linear analysis of PK/PD data and have proven to be useful, both with respect to parameter estimation, model building, and simulation, e.g. in (8) and (9).

Both non-linear mixed-effects models and SDEs are highly non-trivial statistical problems where an analytical likelihood function can rarely be found. The combination of the two should therefore be treated with care. The problem has previously been addressed by the Markov Chain Monte Carlo (MCMC) method (10) for a combined minimal model of glucose disposal and insulin secretion. In the present paper, we combine the Gaussian approximation of the non-linear mixed-effects models with the Gaussian approximation of SDEs with measurement noise. The approximations are facilitated by the Extended Kalman Filter (EKF) to approximate the intra-individual likelihood function (11) and the First-Order Conditional Estimation (FOCE) method to approximate the population likelihood function (12).

The focus of the present study is on two fundamental issues concerning the implementation of SDEs in non-linear mixed effects models. The first is how the likelihood function of nonlinear mixed-effects models with SDEs can be approximated to facilitate estimation in these models. The second focus concerns identifiability: Can the inter-individual variability, the measurement- and the system noise be separated? Or in other words, will significant system noise be predicted by the algorithm when none is used in the simulations (Type I error), and will the algorithm fail to detect significant system noise when it is truly present in the data (Type II error).

THEORY

Non-linear mixed-effects models can be thought of as a hierarchical model structure where the variability in concentration/effect is split into intra-individual variability described by the first-stage distribution and

inter-individual variability described by the second-stage distribution (13, 14). While the introduction of SDEs do not change this fundamental hierarchical structure, they do change the entities in the first stage density and the construction thereof. This section describes the notation for non-linear mixed-effects models used in the present paper, and proceeds with an explanation of the extensions needed to include SDEs.

Notation of Non-Linear Mixed-Effects Models

Non-linear mixed-effects models are used to describe, understand, and simulate data structured as

$$\boldsymbol{y}_{ij}, \quad i = 1, \ldots, N \quad j = 1, \ldots, n_i \tag{1}$$

where the observation $\boldsymbol{y}_{ij}$ in general is a vector of responses for the ith individual at the jth time point, N is the number of individuals and n_i is the number of measurements for the ith individual.

The structural model used to describe the intra-individual data typically consists of a set of ODEs or the solution thereof. These ODEs are supplemented by a model of the residual variation that describes the differences between the structural model and the observations. This gives rise to the following equations for the first stage model:

$$\frac{d\boldsymbol{x}_i}{dt} = \boldsymbol{g}(\boldsymbol{x}_i, \boldsymbol{d}_i, t, \boldsymbol{\phi}_i) \tag{2}$$

$$\boldsymbol{y}_{ij} = \boldsymbol{f}(\boldsymbol{x}_i(t_{ij}), \boldsymbol{d}_i(t_{ij}), \boldsymbol{\phi}_i) + \boldsymbol{\epsilon}_{ij} \tag{3}$$

where for the ith individual, $\boldsymbol{x}_i(t)$ is a vector of state variables, e.g. the amount of drug in a PK model, $\boldsymbol{d}_i$ is a vector of inputs, e.g. dose administration, t is time, t_{ij} is the jth measurement time, $\boldsymbol{\phi}_i$ is a vector of the individual parameters, and $\boldsymbol{\epsilon}_{ij}$ is the jth residual. The residuals are typically modelled as serial independent with covariance matrix Σ, which may depend on the states, input, time, and/or individual parameters. Σ is most often a scalar, except in the general case of multidimensional measurements. $\boldsymbol{g}(\cdot)$ and $\boldsymbol{f}(\cdot)$ are nonlinear vector functions describing the dynamics of the states and the relationship between the states and the observations, respectively.

The second-stage model describes the inter-individual variations, which in the present work is accomplished through the following model for the individual parameters

$$\boldsymbol{\phi}_i = \boldsymbol{h}(\boldsymbol{\theta}, \boldsymbol{Z}_i) \exp(\boldsymbol{\eta}_i) \tag{4}$$

where $\boldsymbol{h}(\cdot)$ denotes the structural type parameter model, which is a function of the fixed-effects parameters $\boldsymbol{\theta}$ and typically also some

covariates $\boldsymbol{Z}_i$. The random-effects $\boldsymbol{\eta}_i$ are independent and multivariate normally distributed with zero mean and covariance matrix $\boldsymbol{\Omega}$, resulting in a multivariate log-normal distribution for the individual parameters.

The total set of population parameters to be estimated in the non-linear mixed effects model can now be summarized as intra-individual variability, inter-individual variability, and fixed-effects parameters, given by the set ($\boldsymbol{\Sigma}, \boldsymbol{\Omega}$, and $\boldsymbol{\theta}$).

Non-Linear Mixed-Effects Models with Stochastic Differential Equations

Non-linear mixed-effects models based on SDEs extend the usual non-linear mixed-effects models by including system noise as an additional source of variation in the first-stage model. This extended model describes the intra-individual variation in data through two sources of noise, which in the present work will be described as two different types of population parameters: (1) $\boldsymbol{\Sigma}$ describing the covariance matrix of the measurement noise rather than that of the residuals, (2) the new parameter matrix $\boldsymbol{\sigma}_w$ describing the magnitude of the system noise. By this extension, the complete set of population parameters in non-linear mixed-effects models based on SDEs becomes ($\boldsymbol{\Sigma}, \boldsymbol{\sigma}_w, \boldsymbol{\Omega}$, and $\boldsymbol{\theta}$).

The formulation of the first-stage model can be separated into the design of a structural model and the addition of variations. Designing the structural model is equivalent to selecting the model structure for conventional PK/PD modelling, which means that exploratory data analysis and physiological knowledge may be applied in the usual manner. This process consists of formulating a set of ODEs as those given in Eq. 2. Once the structural part of the first-stage model has been formulated, we can add variations to account for the differences between the measured and the predicted values. When SDEs are involved, the usual residuals will no longer be independent, and we need to describe the underlying variations of $\boldsymbol{\epsilon}$ by system noise $\boldsymbol{w}$ as well as measurement noise $\boldsymbol{e}$. The system noise can be added directly to the differential equations allowing for some random variations in the evolution of the states, which may be appropriate whenever the modeler fears that the true evolution of the states does not comply strictly with the structural model. Measurement noise is added to the model in the usual measurement equation, such that a first-stage model with SDEs can be written as

$$d\boldsymbol{x}_i = \boldsymbol{g}(\boldsymbol{x}_i, \boldsymbol{d}_i, t, \boldsymbol{\phi}_i)dt + \boldsymbol{\sigma}_w d\boldsymbol{w} \tag{5}$$

$$\boldsymbol{y}_{ij} = \boldsymbol{f}(\boldsymbol{x}_i(t_{ij}), \boldsymbol{d}_i(t_{ij}), \boldsymbol{\phi}_i) + \boldsymbol{e}_{ij} \tag{6}$$

where $\boldsymbol{x}_i(t), \boldsymbol{d}_i, t, t_{ij}, \boldsymbol{\phi}_i, \boldsymbol{g}(\cdot)$ and $\boldsymbol{f}(\cdot)$ are identical to what has previously been defined for ODEs in Eqs. 2 and 3. $\boldsymbol{e}_{ij}$ are the independent identically

distributed Gaussian measurement errors with covariance matrix $\boldsymbol{\Sigma}$, and $\boldsymbol{\sigma}_w d\boldsymbol{w}$ gives the system noise, where both $\boldsymbol{\Sigma}$ and $\boldsymbol{\sigma}_w$ may depend on the states, input, time, and/or individual parameters. If the magnitude of the system noise $\boldsymbol{\sigma}_w$ is zero, then the entire system noise term will vanish and the remaining part of the SDE will simply be the differential form of the ODE given in Eq. 2. SDEs are usually written on the differential form given above, because the term $d\boldsymbol{w}$ has a mathematical interpretation as the infinitesimal increments in the noise process ($\boldsymbol{w}$), whereas the corresponding derivative $d\boldsymbol{w}/dt$ cannot be treated mathematically. The individual specific system noise $\boldsymbol{w}$ is a standard vector Wiener process, i.e. a continuous time Gaussian process where the mean and variance of the differences between two time points are

$$E[\boldsymbol{w}_{t_2} - \boldsymbol{w}_{t_1}] = 0 \tag{7}$$

$$V[\boldsymbol{w}_{t_2} - \boldsymbol{w}_{t_1}] = |t_2 - t_1| I \tag{8}$$

where I is the identity matrix. The Gaussian process can be understood to originate from the sum of many identically distributed stochastic events giving rise to the difference between the true evolution of the state and the evolution described by the structural term $\boldsymbol{g}(\cdot)$. Furthermore, the variance of the Wiener process increases linearly in time, which can be interpreted as a linear increase in the number of stochastic events contributing to the dynamic noise. Refs. (11) and (15) provide an introduction to applied stochastic differential equations, and (16) gives a more thorough mathematical introduction.

THE LIKELIHOOD FUNCTION FOR THE NON-LINEAR MIXED-EFFECTS MODEL WITH SDES

When the first-stage model is extended to include SDEs rather than ODEs, the first-stage probability density function can no longer be computed analytically. In the present section, we shall describe how the likelihood function is now formulated, how it can be approximated, and how it is combined with the second-stage density to form the population likelihood function. This is all summarized in Table I, presented at the end of the section.

When the intra-individual model contains correlations in the residuals, the first-stage distribution must be factorized as a product of densities that are conditioned not only upon ($\boldsymbol{\phi}_i, \boldsymbol{\Sigma}, \boldsymbol{\sigma}_w$, and $\boldsymbol{d}_i$), but also on all previous measurements. Conditioning on previous measurements is central for the present text, giving rise to the term conditional densities used in the following. To see how factorization of the first stage density comes

Table I. Summary of the Likelihood Evaluation Algorithm

Function: Approximate Individual a posteriori Log-Likelihood ($\boldsymbol{l}_i$)
1: Function of $\boldsymbol{\eta}_i$ given ($\boldsymbol{\Sigma}$,$\boldsymbol{\sigma}_w$, $\boldsymbol{\Omega}$, $\boldsymbol{\theta}$, $\boldsymbol{Z}_i$, $\boldsymbol{d}_i$, and $\mathcal{Y}_{in_i}Y$)
2: Use (4) to compute $\boldsymbol{\phi}_i$
3: Initialize the state prediction and state covariance.
See the numerical implementation below.
4: Use EKF in Appendix A to compute $\boldsymbol{\epsilon}_{ij}$ and $\boldsymbol{R}_{i(j|j^{-1}})$
5: **Return** l_i as computed in (l5)
Function: Approximate Population Likelihood (L)

1: Function of ($\boldsymbol{\Sigma}$, $\boldsymbol{\sigma}_w$, $\boldsymbol{\Omega}$, and $\boldsymbol{\theta}$), given $\boldsymbol{Z}_i$, $\boldsymbol{d}_i$, and all observations
2: **For** $i = 1$ to N **do**
3: $\hat{\boldsymbol{\eta}}_i = \arg\min(\boldsymbol{l}_i)$
4: Use EKF in Appendix A to compute $\boldsymbol{\epsilon}_{ij}|\hat{\boldsymbol{\eta}}_i$, $\nabla\boldsymbol{\epsilon}_{ij}|\hat{\boldsymbol{\eta}}_i$, and $\boldsymbol{R}_{i(j\backslash j^{-1}}|\hat{\boldsymbol{\eta}}_i$
5: Use (l8) and (l9) to compute $\boldsymbol{l}_i|\hat{\boldsymbol{\eta}}_i$ and $\triangle\boldsymbol{l}_i|\hat{\boldsymbol{\eta}}_i$
6: **end for**
7: **Return** $L(\boldsymbol{\theta}, \boldsymbol{\Sigma}, \boldsymbol{\sigma}w, \boldsymbol{\Omega})$ as computed in (l7)

about, start with the distribution for the initial observation and successively add one observation at a time by the use of a reformulation of Bayes rule $P(A \cap B) = P(B|A)P(A)$. This gives us the following first stage density for the ith individual

$$p_1(\mathcal{Y}_{in_i}|\boldsymbol{\phi}_i, \boldsymbol{\Sigma}, \boldsymbol{\sigma}_w, \boldsymbol{d}_i) = \left(\prod_{j=2}^{n_i} p(\boldsymbol{y}_{ij}|\mathcal{Y}_{i(j-1)}, \cdot)\right) p(\boldsymbol{y}_{i1}|\cdot) \tag{9}$$

The so called conditional densities are given on the right hand side, $\mathcal{Y}_{ij} = [\boldsymbol{y}_{i1}, \ldots, \boldsymbol{y}_{ij}]$ represents all observations of the ith individual up to time t_{ij}, and conditioning on $\boldsymbol{\phi}_i$, $\boldsymbol{\Sigma}$, $\boldsymbol{\sigma}_w$, and $\boldsymbol{d}_i$ is represented by "$\cdot$".

If the SDEs are reduced to ODEs, the residuals will be uncorrelated, and the conditional densities will be identical to the unconditional densities, such that the likelihood function will reduce to the product of unconditional densities, as known for ODEs.

An analytical determination of the conditional densities requires a solution of a so called general non-linear filtering problem. This entails to start with the initial distribution and then successively solving Kolmogorov's forward equation for the SDE and applying Bayes' rule (11). In practice, this approach involves the numerical solution of a partial differential equation for each time increment, which is too time consuming, making it computational infeasible, and an alternative is needed. Various methods have been proposed and are still investigated for parameter estimation in the general setup, and a consensus of a preferred method has not yet been reached (17).

In the present work we shall use a quasi likelihood method, i.e. a method that uses the Gaussian approximation, so we assume that the conditional densities are well approximated by Gaussian densities. This is a particularly useful choice since filtering techniques used for SDEs with measurement noise rely on separability of the first and second order moment, and since Gaussian densities are easily combined with the methods usually used in non-linear mixed-effects modelling. The calculation of the conditional densities for the intra-individual model is facilitated by the Extended Kalman Filter (EKF) (11). In the case where the differential equations are linear and both system noise and measurement noise are state independent, the EKF reduces to the ordinary Kalman Filter, which in this case gives the exact likelihood function. When proportional system noise or proportional measurement noise is needed, one can typically log-transform the states or the measurements to approach the Gaussian distribution. These transformations will most often give rise to more severe non-linearities in the functions for the structural model $g(\cdot)$ and the measurement equation $f(\cdot)$ in Eqs. 5 and 6. However, it is our experience that the conditional densities are well described by Gaussian distributions, also for highly non-linear systems. The assumption of Gaussian conditional densities is easily tested by the distribution of the standardized residuals, while also more advanced methods have been developed, see (18).

The EKF approximates the conditional densities with Gaussian distributions, which is described in detail in Appendix A. The conditional densities describe the distribution of the following measurement conditioned on all the previous measurements, so that the mean of the distribution is identical to the prediction of the following measurement, i.e. the one-step prediction $\hat{\boldsymbol{y}}_{i(j|j-1)}$. Likewise, the covariance of the conditional density will be the one-step prediction covariance $\boldsymbol{R}_{i(j|j-1)}$. We have thus completely described the approximate Gaussian conditional densities by the conditional mean and covariance, which are

$$\hat{\boldsymbol{y}}_{i(j|j-1)} = E(\boldsymbol{y}_{ij}|\mathcal{Y}_{i(j-1)}, \cdot) \tag{10}$$

$$\boldsymbol{R}_{i(j|j-1)} = V(\boldsymbol{y}_{ij}|\mathcal{Y}_{i(j-1)}, \cdot) \tag{11}$$

The notation above is also used for the mean and variance of the first prediction, such that $\hat{\boldsymbol{y}}_{i(1|0)}$ is the unconditioned model prediction of the first observation, and $\boldsymbol{R}_{i(1|0)}$ is the covariance of the first prediction error.

The one-step prediction error $\boldsymbol{\epsilon}_{ij}$ is given by

$$\boldsymbol{\epsilon}_{ij} = \boldsymbol{y}_{ij} - \hat{\boldsymbol{y}}_{i(j|j-1)} \in N(0, \boldsymbol{R}_{i(j|j-1)}) \tag{12}$$

Using the notation above, the Gaussian approximation of the first-stage distribution density function in Eq. 9 can be written as

$$p_1(\mathcal{Y}_{in_i}|\cdot) \approx \prod_{j=1}^{n_i} \frac{\exp\left(-\frac{1}{2}\boldsymbol{\epsilon}_{ij}^T \boldsymbol{R}_{i(j|j-1)}^{-1}\boldsymbol{\epsilon}_{ij}\right)}{\sqrt{|2\pi \boldsymbol{R}_{i(j|j-1)}|}} \tag{13}$$

The second-stage density (4) can be written as $p_2(\boldsymbol{\eta}_i|\boldsymbol{\Omega})$, which is included in the same way as for ordinary differential equations. This gives us the full non-linear mixed-effects likelihood function

$$L(\boldsymbol{\theta}, \boldsymbol{\Sigma}, \boldsymbol{\sigma}_w, \boldsymbol{\Omega}) \propto \prod_{i=1}^{N} \int p_1(\mathcal{Y}_{in_i}|\boldsymbol{\eta}_i, \boldsymbol{\theta}, \boldsymbol{\Sigma}, \boldsymbol{\sigma}_w, \boldsymbol{d}) p_2(\boldsymbol{\eta}_i|\boldsymbol{\Omega}) d\boldsymbol{\eta}_i = \prod_{i=1}^{N} \int \exp(\boldsymbol{l}_i) d\boldsymbol{\eta}_i \tag{14}$$

where

$$\boldsymbol{l}_i = -\frac{1}{2}\sum_{j=1}^{n_i}\left(\boldsymbol{\epsilon}_{ij}^T \boldsymbol{R}_{i(j|j-1)}^{-1}\boldsymbol{\epsilon}_{ij} + \log|2\pi \boldsymbol{R}_{i(j|j-1)}|\right) - \frac{1}{2}\boldsymbol{\eta}_i^T \boldsymbol{\Omega}^{-1}\boldsymbol{\eta}_i - \frac{1}{2}\log|2\pi\boldsymbol{\Omega}| \tag{15}$$

is the approximate a posteriori log-likelihood function for the random effects of the ith individual. It is observed that the likelihood function is based on the one-step prediction error.

In the case of no system noise ($\boldsymbol{\sigma}_w = 0$) the SDE in the model reduces to the more familiar ODE, and the one-step prediction covariances will reduce to the residual covariance $\boldsymbol{R}_{i(j|j-1)} = \boldsymbol{\Sigma}$. The one-step prediction errors will be identical to the usual unconditioned prediction errors, such that the likelihood function above will reduce to the one known for non-linear mixed-effects models based on ODEs.

As usual for non-linear mixed-effects models, the likelihood function cannot be solved analytically. Approximations, therefore, have to be made in order to estimate the parameters, which will be considered in the following.

Approximations of the Population Likelihood Function

The Gaussian structure of the individual likelihood functions allows us to use the well known Laplacian approximation, as well as the other approximation schemes frequently used in non-linear mixed-effects modelling to obtain the population likelihood function. The entities within the likelihood function structure ($\boldsymbol{\epsilon}$ and $\boldsymbol{R}$) are however extended compared to the standard problem. In the following, we shall discuss the approximation when applied to these extended entities.

For non-linear mixed-effects models, the likelihood function is usually approximated by performing a second-order Taylor series expansion of the

a posteriori individual log-likelihood function $\boldsymbol{l}_i$ around some value of the random effects, e.g. zero or the value of $\boldsymbol{\eta}_i$ that minimizes $\boldsymbol{l}_i$. The same approach is taken for mixed-effects model with SDEs, so the Laplacian approximation of (14) becomes

$$L(\boldsymbol{\theta}, \boldsymbol{\Sigma}, \boldsymbol{\sigma}_w, \Omega) \propto \prod_{i=1}^{N} \int \exp(\boldsymbol{l}_i) d\boldsymbol{\eta}_i \approx \prod_{i=1}^{N} |\triangle \boldsymbol{l}_i|^{-1/2} \exp\left[\boldsymbol{l}_i - \frac{1}{2} \nabla \boldsymbol{l}_i^T \triangle \boldsymbol{l}_i^{-1} \nabla \boldsymbol{l}_i\right] \tag{16}$$

The gradient $\nabla \boldsymbol{l}_i$ of the *a posteriori* individual log-likelihood with respect to the random effects will vanish when the expansion is made around the true minimum, but it has been included here to account for the more general case. The Hessian $\triangle \boldsymbol{l}_i$ is thus a key element in the evaluation of the population likelihood function.

The numerical evaluation of double derivatives to form the Hessian is usually quite sensitive leading to uncertainty in the objective function and optimization problems. Several approximations have been developed specifically to avoid numerical calculations of double derivatives related to this Hessian in non-linear mixed-effects models. In the present work, we have used the First-Order Conditional Estimation (FOCE) method (4).

First-Order Conditional Estimation method

The FOCE method uses only first-order derivatives in the evaluation of the population likelihood function where the derivatives are evaluated at the conditional estimates of the random effects $\hat{\boldsymbol{\eta}}_i$. The likelihood function can thus be written as

$$L(\boldsymbol{\theta}, \boldsymbol{\Sigma}, \boldsymbol{\sigma}_w, \boldsymbol{\Omega}) \approx \prod_{i=1}^{N} |\Delta \boldsymbol{l}_i|^{-1/2} \exp(\boldsymbol{l}_i)|_{\hat{\boldsymbol{\eta}}_i} \tag{17}$$

where the second-order derivatives are disregarded such that the individual *a posteriori* log-likelihood function and its Hessian can be written as

$$\boldsymbol{l}_i = -\frac{1}{2} \sum_{j=1}^{n_i} \left(\boldsymbol{\epsilon}_{ij}^T \boldsymbol{R}_{i(j|j-1)}^{-1} \boldsymbol{\epsilon}_{ij} + \log |2\pi \boldsymbol{R}_{i(j|j-1)}| \right) - \frac{1}{2} \boldsymbol{\eta}_i^T \boldsymbol{\Omega}^{-1} \boldsymbol{\eta}_i - \frac{1}{2} \log |2\pi \boldsymbol{\Omega}| \tag{18}$$

$$\triangle \boldsymbol{l}_i \approx -\sum_{j=1}^{n_i} \left(\nabla \boldsymbol{\epsilon}_{ij}^T \boldsymbol{R}_{i(j|j-1)}^{-1} \nabla \boldsymbol{\epsilon}_{ij} \right) - \boldsymbol{\Omega}^{-1} \tag{19}$$

We note that the variance of the one-step predictions $\boldsymbol{R}_{i(j|j-1)}$ is generated through the dynamics of the system and will thus depend inherently

upon the parameters of the model. This dependence leads to interactions between the intra-individual residuals and the random effects even in models that are usually homoscedastic. In the FOCE approximation given above, we are disregarding first-order derivatives of $\boldsymbol{R}_{i(j|j-1)}$, which is an approximation in the case of interactions, and thus a more crude approximation when modelling with SDEs as compared to ODEs. This approximation would be worse if directly including interactions between the random effects and system noise, which has been avoided in the present study, but could be included if one feels that it is necessary.

Interactions were included in the FOCE method presented here, so both the predictions and the covariances are evaluated at the conditional estimate. One could also, as it is implemented in NONMEM version V, use the conditionally estimated predictions together with covariances computed by the population predictions, corresponding to having no interactions, or even expand the likelihood function around the population average, corresponding to the first-order (FO) method.

METHODS

Several simulation studies are used to test whether the likelihood function formulated above can be advocated for estimating parameters in a non-linear mixed-effects model based on SDEs. These simulation studies are based on the PK model and the numerical implementation described below.

Model Used for Simulation

We shall simulate experiments using a one-compartment PK model with IV bolus dose with a constant coefficient of variation for the uncorrelated measurement error and additive Wiener noise such that the ith individual is modelled by

$$dA_i = -\frac{CL_i}{V_i} A_i dt + \sigma_w dw \tag{20}$$

$$y_{ij} = \log \frac{A(t_{ij})}{V_i} + e_{ij}, \qquad e_{ij} \in N(0, \sigma_e^2) \tag{21}$$

where A_i is the amount of drug in the central compartment, y_{ij} is the measurement at time t_{ij}, e_{ij} is the measurement error and σ_e is the coefficient of variation for the measurement error, w is a standard Wiener process, and σ_w is the magnitude of the system noise. V_i and CL_i are the individual parameters for volume of distribution and clearance, respectively. These are composed of a fixed effect and a random effect, i.e.

$$V_i = V\exp(\eta_i^V), \qquad \eta_i^V \in N(0, \omega_V^2) \tag{22}$$

$$CL_i = CL\exp(\eta_i^{CL}), \qquad \eta_i^{CL} \in N(0, \omega_{CL}^2) \tag{23}$$

Hence, the six population parameters to be estimated are (V, CL, ω_V, ω_{CL}, σ_w, and σ_e). The model is used to simulate two different experimental setups. First, the type I and type II errors are investigated using a somewhat data rich situation with 25 individuals, where the plasma concentration is sampled 12 times at times 0, 0.5, 1, 2, 4, 6, 8, 10, 12, 16, 20, and 24 h. In the second experimental setup, we simulate 100 individuals, where the plasma concentration is sampled three times at times 0, 6, and 24 h. This second analysis is included to illustrate that parameters in models based on SDEs may also be successfully estimated when only a few samples per individual are taken.

All simulations are performed using the same structural parameter values ($V = 10$ and $CL = 0.5$), while the simulated values of the noise parameters (ω_V, ω_{CL}, σ_w, and σ_e) all vary between 0.01 and 0.4. Examples of simulated individual profiles for various levels and types of intra-individual noise are presented in Fig. 1.

Numerical Implementation

Evaluation of the approximate likelihood function, the optimization, and the simulations of the mixed-effects models based on stochastic differential equations were performed in MATLAB.

As previously described, the population log-likelihood can be approximated using the individual *a posteriori* log-likelihood functions and the Hessian of each of these. The individual likelihood functions can be calculated from the one-step predictions and covariances given by the Extended Kalman Filter as described in Appendix A depending on the initialization of the covariance of the states. The covariance of the states at the first measurement was set to the amount of system noise accumulating in a time span equal to that between the first and second measurement. This particular choice has proven successful in other software implementations (19).

The Hessian of the individual *a posteriori* log-likelihood function can be approximated using the gradients of the one-step predictions as previously described. These gradients were calculated numerically by a central differencing algorithm given in (20). We note that other methods such as automatic differentiation may be useful, especially in a more general software implementation.

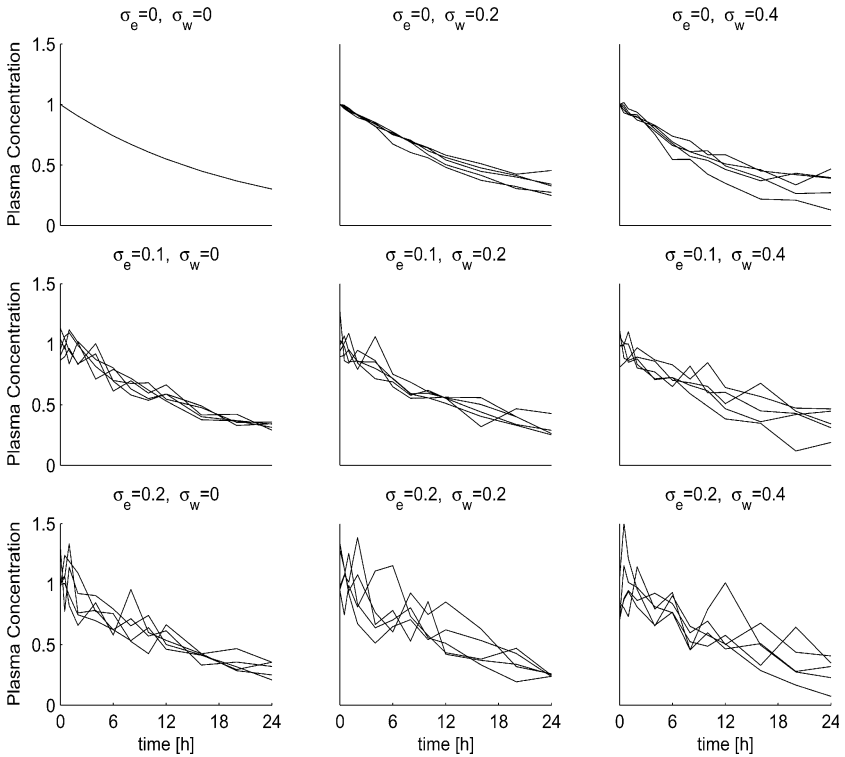

Fig. 1. Simulated individual plasma concentration profiles for various levels and types of intra-individual noise. Simulations are made with no inter-individual variation, using $V = 10$ and $CL = 0.5$. Each individual was sampled 12 times, as in the somewhat frequently sampled experiment. The simulated measurements are connected with lines.

In order to ensure stability in the calculation of the objective function, simple constraints were introduced on the parameters, i.e.

$$\theta_{\min} < \theta < \theta_{\max} \tag{24}$$

These constraints were satisfied by solving the optimization problem with respect to a transformation of the original parameters, i.e.

$$\tilde{\theta} = \ln\left(\frac{\theta - \theta_{\min}}{\theta_{\max} - \theta}\right) \tag{25}$$

Regardless of the true parameter value, minimum and maximum allowed values were chosen to be 10^{-5} and 50, respectively.

Two sets of initial parameter values were used in the optimization, one at 0.8 times the simulated values and one at 1.2 times the simulated

values. Experience with the present implementation of the optimization problem tells us that the initial values should not be too low, so the smallest allowable set was chosen to be 0.04 and 0.06. In the figures showing the results, the estimate is plotted for both of these initial values, but any difference can rarely be seen. This gives some assurance of the size of the region governed by the local- if not global-minima reached by the estimation procedure.

The asymptotic standard errors of the estimates were computed through the Hessian of the population log-likelihood, which was approximated by a finite difference algorithm given in (20). The algorithm depends on the accuracy in the numerical evaluation of the likelihood function. By graphical means, the number of significant digits in the likelihood function was found to be around 10, such that the number of significant digits in the Hessian should be around $10/3 \approx 3$, see (20).

The computation time on a standard laptop PC (1400 MHz Pentium IV processor) for one evaluation of the population likelihood function was between 1 and 3 s depending on the simulated example and a few minutes for a complete optimization. The relative swiftness compared to e.g. MCMC methods is a serious advantage of the presented algorithm, and this may enable estimation of SDEs in more general models for practical purposes.

RESULTS

Successive simulation and estimation of many different experiments have been used to investigate the type I and type II error of the presented algorithm, and whether SDEs could be used in studies with only a few measurements per individual. In the present section we describe the results of these simulation studies.

System Noise is Separable From Other Sources of Variation

Any use of the presented algorithm is naturally dependent upon its ability to separate the three levels of noise proposed in the present paper. Separability should thus be among the fundamental issues to be addressed prior to investigation of real data or implementation in more general software. Since system noise is the central addition to the model setup, the central concern is towards separability of system noise e.g. whether system noise is detected when it is truly present in data (type II error).

Separability of system noise and type II errors were investigated in the first simulation study consisting of 40 simulated experiments with 25 individuals each sampled 12 times. Each simulated experiment was

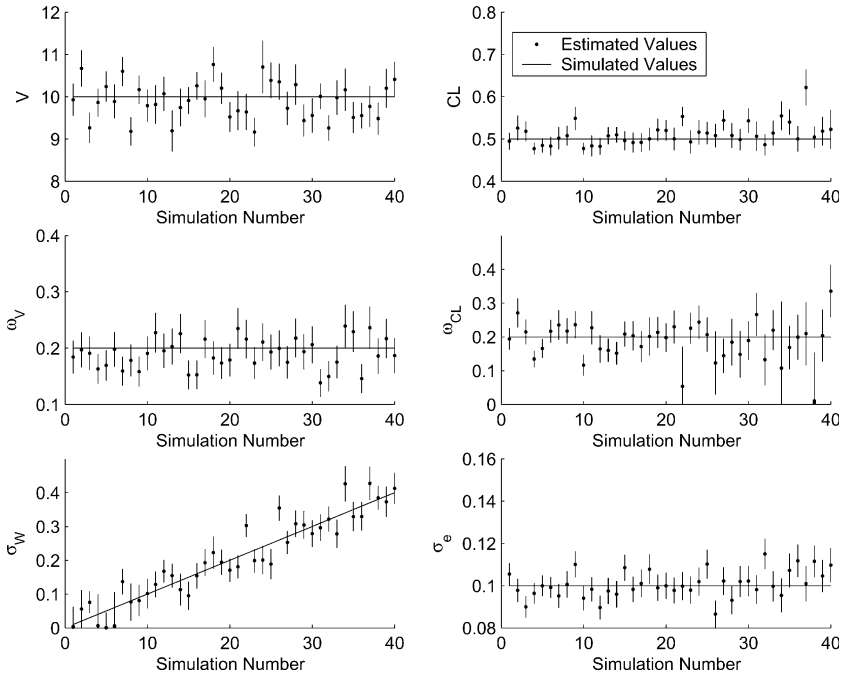

Fig. 2. Simulated parameter values are given by the line and estimated parameters values are given by a dot supplemented by error bars representing plus/minus one standard error. The value of the system noise is satisfactorily inferred at increasing levels of system noise, and the bias introduced to the remaining parameters seems to remain small.

performed using a fixed level of inter-individual variation and measurement noise and with an increasing amount of system noise. The results presented in Fig. 2 provide a visual confirmation that higher levels of system noise do not produce neither additional measurement noise nor inter-individual variability, illustrating that system noise is in fact satisfactorily separable from the remaining noise parameters.

Furthermore, the variability and the standard error of the estimated values of inter-individual variation of the clearance is seen to increase when system noise increases. This indicates that the system noise does make estimation of this particular noise parameter more difficult, while the remaining noise parameters are not influenced to the same extent.

Insignificant Bias in the Estimated System Noise

A fundamental feature in maximum likelihood estimation of mixed-effects models is that the individual predictions are regressed towards the

mean (the population prediction). This introduces manifest correlations in the intra-individual residuals, which in the present model could be estimated as system noise introducing bias to the estimates. Bias in the estimated system noise could also originate in more general separability problems where inter-individual variation or intra-individual measurement noise is estimated as system noise. If a significant level of system noise is estimated when none is used in the simulation, then a type I error has occurred.

Bias in the system noise and type I errors were investigated in the second simulation study consisting of 40 simulated experiments with 25 individuals, each sampled 12 times. Each simulated experiment was performed using no system noise but with an increasing amount of inter-individual variation and measurement noise. This investigation might reveal a potential relationship between bias in the system noise and the level of the remaining noise parameters. The results presented in Fig. 3 demonstrate that this relationship is small and that the existing bias is insignificant such that only few type I errors occur. The largest estimates of system noise was around 0.2, which in Fig. 1 is seen to be a modest level of noise compared to the corresponding measurement noise.

Estimation Based on Sparsely Sampled Individuals

A data set consisting of many sparsely sampled individuals is not uncommon within PK/PD modelling. However, parameter estimation in SDEs based on sparsely sampled individuals has previously been complicated by the need of rich sampling to separate system noise from measurement noise and the limitations of single subject estimation algorithms. Mixed-effects modelling has in this situation enabled parameter estimation of ODEs, making it interesting whether also SDEs can be successfully treated.

Fifty experiments with 100 individuals each sampled three times has been simulated and subsequently estimated. The results presented in Fig. 4 demonstrate that estimation can be performed successfully with sparsely sampled individuals in the chosen model. We note that estimation based on two samples per individual failed to yield the same level of success, so three samples per individual may indeed be the lower limit when system noise is included simultaneously with measurement noise and inter-individual variability.

Statistics of the estimation results from Fig. 4 are given in Table II, which demonstrates that the mean values of the estimated parameters are close to the true values used in the simulations and that the standard deviations of the estimates are relatively close to the mean of the standard

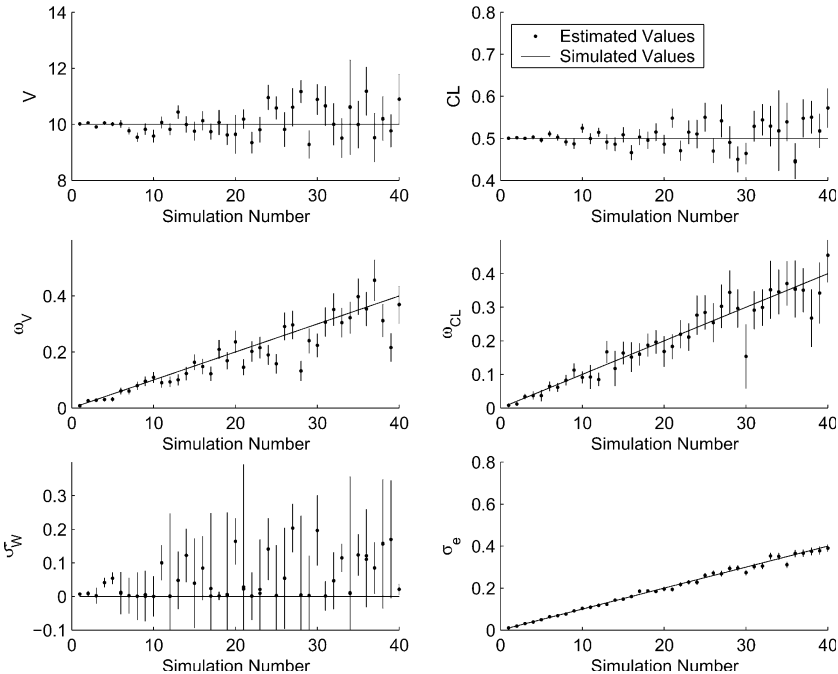

Fig. 3. Simulated parameter values are given by the solid line and estimated parameters values are given by a dot supplemented by error bars representing plus minus one standard error. The estimated system noise is insignificant in most of the simulated experiments, also when the other noise levels increase dramatically. Each simulation was estimated with two sets of initial conditions, which in a few cases gave two slightly different parameter estimates, such that two dots are seen for the same simulation number.

error estimates. It should be mentioned that due to the relatively high number of simulated experiments, one is able to demonstrate bias in the estimated clearance values on a 95% confidence level. Note that bias exists in practically any algorithm for parameter estimation in complicated systems. This entails that bias in all parameter estimates will be found if the number of simulated experiments is increased sufficiently.

DISCUSSION

SDEs offer a general intra-individual error structure where the evolution of the states is allowed to deviate from the structural model. This work presented a novel approximation of the likelihood function for non-linear mixed-effects models based on SDEs and addresses some fundamental issues regarding parameter estimation in this new type of model.

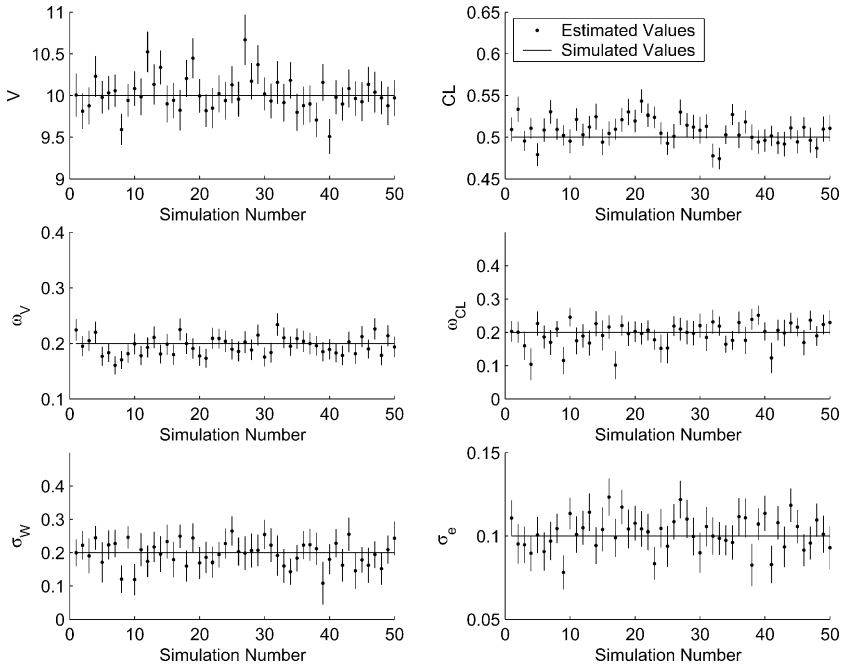

Fig. 4. Simulated parameter values are given by the solid line and estimated parameters values are given by a dot supplemented by error bars representing plus/minus one standard error. Each simulated experiment consists of 100 individuals sampled three times.

Table II. Statistics of the Estimation Performance When Simulating Experiments with 100 Individuals each Sampled Three Times.

	V	CL	ω_v	ω_{CL}	σ_w	σ_e
Simulated value	10	0.5	0.2	0.2	0.2	0.1
Mean of estimates	10.02	0.508	0.195	0.195	0.198	0.102
Standard deviation of estimates	0.212	0.015	0.016	0.035	0.037	0.010
Mean of standard error estimates	0.220	0.014	0.018	0.035	0.046	0.011

When SDEs are implemented, one-step predictions and their variances were seen to take up the role of the usual unconditioned predictions and intra-individual variance in many ways. In the theory section, it was emphasized that these objects reduce to the usual predictions and variances when the system noise vanishes, and that they can be used to construct the likelihood function. The fundamental model assumption of uncorrelated prediction errors is converted into the assumption of uncorrelated one-step prediction errors, such that diagnostic plots should

be slightly changed. Predicted vs. previous prediction should be replaced by one-step prediction vs. previous one-step prediction. Furthermore, weighted residuals vs. time or covariates or similar diagnostic plots should be replaced by corresponding plots of the one-step prediction error weighted by the one-step prediction variance.

Stochastic differential equations are complicated mathematical entities with peculiarities such as dependence in the solution upon the chosen interpretation of the infinitesimal correlation structure, or equivalently on the selected interpretation of the integral part of the SDE. The system noise is called multiplicative or additive when the diffusion term is dependent or independent of the state variable ($\boldsymbol{x}$), respectively. In the multiplicative case, the SDE proposed in Eq. 5 will be dependent upon the chosen interpretation of the infinitesimal correlation structure, and special attention is needed. In the present work, we restrict ourselves to the additive case where the solutions are independent of the interpretation, since for this case the EKF is recognized to work best. However, situations may arise where multiplicative noise is necessary, as already seen for PK/PD modelling in e.g. (10). Luckily, a large class of SDEs with multiplicative diffusion can be transformed into SDEs with additive diffusion (21). One may choose to do this transformation before implementing the diffusion, such that only additive diffusion is needed. This line of approach, as undertaken e.g. in (10), is consistent with the Stratonovich interpretation of the infinitesimal correlations, which is recommended for physical modelling (11).

The presented combination of SDEs and measurement noise constitute a general framework to describe the intra-individual variations, which includes many previous implementations of statistically sophisticated intra-individual models. Motivated by these efforts, we try to identify some of the potential benefits that the literature indicates SDEs can offer nonlinear mixed-effects models of PK/PD: (1) improve estimates of inter-individual variability, (2) improve structural parameter estimates, (3) give a diagnostic test of the model, (4) pinpoint model defects to be used in successive model improvement, (5) improve simulation properties of the model, (6) provide a more realistic description of the observed variations, (7) allow fluctuations in physiological parameters, (8) enable deconvolution, (9) extend deconvolution to nonlinear PK/PD models, (10) improve estimation of the states in the system.

Points 1–3 were argued based on a simulation study including the AR(1) in mixed-effects modelling (3). Points 3 and 4 were also demonstrated by SDEs with measurement noise for single subject PK/PD data in a work in progress using ready available software (19). Points 5–7 were demonstrated by various case-by-case implementations of SDEs to

PK/PD data, see e.g. (8–10). Point 8 was demonstrated for PK/PD models in (22), where deconvolution was facilitated by a random walk, which is a discrete version of a SDE. Since the SDE setup includes non-linear models, point 9 gives itself. Point 10 was demonstrated using a mixed-effects setup of random walks to improve estimation of the area under the curve (23).

The focus of the present investigation was implementation of SDEs in non-linear mixed-effects models, which could potentially boost the use of SDEs within PK/PD. Fundamental issues concerning the implementation was addressed by simulation and successive estimation of a non-linear mixed-effects model based on SDEs corresponding to a one-compartment model. Three specific concerns were addressed: (1) Will bias in the estimated parameter values lead to significant estimates of system noise when none is used in the simulations (type I errors)? (2) will the system noise be separable from the remaining noise parameters? and (3) can the parameters in the proposed model also be determined with sparsely sampled individuals? The relationship between bias in the system noise and the level of the remaining noise parameters was found to be small and only few type I errors occurred. It was demonstrated that a significant level of system noise can be detected when it is truly present in data (no type II errors). Successful estimation was performed with population data including only three samples per individual, which may indeed be the lower limit, when three different types of noise are used.

In conclusion, it is confirmed that inter-individual variability, measurement- and system noise can be separated for the chosen model, which is necessary for non-linear mixed-effects models based on SDEs to be treated meaningfully. However, the presented model and study setup were quite simple, and the routines used in the numerical implementation were not state of the art. So the present investigation should be seen as a pilot study preceding a more general implementation, which more easily allows SDEs in non-linear mixed-effects models. A work in progress demonstrates how the EKF can be implemented in the control stream of NONMEM Version VI beta.

APPENDIX A: THE EXTENDED KALMAN FILTER

The Extended Kalman filter can be used to calculate the one-step predictions and the one-step prediction variances for a stochastic differential equation with additive diffusion and measurement noise. The algebra presented in the following appendix is all performed on the individual level, so to ease the notation, the i index referring to the individual has

been dropped. The general intra-individual model treated here can be written as

$$d\boldsymbol{x} = \boldsymbol{g}(\boldsymbol{x}, \boldsymbol{d}, t, \boldsymbol{\phi})dt + \boldsymbol{\sigma}_w d\boldsymbol{w} \tag{A.1}$$

$$\boldsymbol{y}_j = \boldsymbol{f}(\boldsymbol{x}(t_j), \boldsymbol{d}(t_j), \boldsymbol{\phi}) + \boldsymbol{e}_j \tag{A.2}$$

where $\boldsymbol{x}$ is the vector of state variables, $\boldsymbol{y}_j$ is the vector of measurements at time t_j, $\boldsymbol{e}_j$ are the associated normally distributed measurement errors with covariance matrix $\boldsymbol{\Sigma}$, and $\boldsymbol{\sigma}_w d\boldsymbol{w}$ is the system noise, where both $\boldsymbol{\Sigma}$ and $\boldsymbol{\sigma}_w$ may depend on input $\boldsymbol{d}$, time t and/or individual parameters $\boldsymbol{\phi}$.

The following notation for the derivatives is applied

$$\boldsymbol{A}_t = \left.\frac{\partial \boldsymbol{g}}{\partial \boldsymbol{x}}\right|_{\boldsymbol{x}=\hat{\boldsymbol{x}}t|j-1}, \quad \boldsymbol{C}_j = \left.\frac{\partial \boldsymbol{f}}{\partial \boldsymbol{x}}\right|_{\boldsymbol{x}=\hat{\boldsymbol{x}}t|j-1} \tag{A.3}$$

One needs to initiate the EKF with a prediction of the initial state $\hat{\boldsymbol{x}}_{1|0}$ and a prediction of the covariance of the initial state $\boldsymbol{P}_{1|0}$. From this point, the EKF is a recursive algorithm starting with the calculation of the one step prediction of the measurement and its associated covariance matrix. This is achieved by

$$\hat{\boldsymbol{y}}_{j|j-1} = \boldsymbol{f}(\boldsymbol{x}_{j|j-1}, \boldsymbol{d}_j, \boldsymbol{\phi}) \tag{A.4}$$

$$\boldsymbol{R}_{j|j-1} = \boldsymbol{C}_j \boldsymbol{P}_{j|j-1} \boldsymbol{C}_j^T + \boldsymbol{\Sigma}_{j|j^{-1}} \tag{A.5}$$

Given the actual measurement, we can update our state prediction and variance to be predictions conditioned also on the jth measurement. This is performed by the update equations, i.e.

$$\hat{\boldsymbol{x}}_{j|j} = \hat{\boldsymbol{x}}_{j|j_{-1}} + \boldsymbol{K}_j(\boldsymbol{y}_j - \hat{\boldsymbol{y}}_{j|j_{-1}}) \tag{A.6}$$

$$\boldsymbol{P}_{j|j} = \boldsymbol{P}_{j|j_{-1}} - \boldsymbol{K}_j \boldsymbol{R}_{j|j_{-1}} \boldsymbol{K}_j^T \tag{A.7}$$

$$\boldsymbol{K}_j = \boldsymbol{P}_{j|j_{-1}} \boldsymbol{C}^T \boldsymbol{R}_{j|j_{-1}}^{-1} \tag{A.8}$$

where $\boldsymbol{K}_j$ is the Kalman gain.

The final step in the recursive algorithm is to predict the state and the state variance at the time of the following measurement. This is performed by solving the prediction equations, i.e.

$$d\hat{\boldsymbol{x}}_{t|j}/dt = \boldsymbol{g}(\boldsymbol{x}_{t|j}, \boldsymbol{d}, t, \boldsymbol{\phi}) \tag{A.9}$$

$$d\boldsymbol{P}_{t|j}/dt = \boldsymbol{A}_t \boldsymbol{P}_{t|j} + \boldsymbol{P}_{t|j} \boldsymbol{A}_t^T + \boldsymbol{\sigma}_w \boldsymbol{\sigma}_w^T \tag{A.10}$$

Table A1. Algorithm for the Extended Kalman Filter

Algorithm: Kalman Filtering

1: Given parameters and initial prediction $\boldsymbol{\phi}$, $\hat{\boldsymbol{x}}_{1|0}$ and $\boldsymbol{P}_{1|0}$
2: **For** $j = 1$ to n_i **do**
3: Use Eq. A.4 and A.5 to compute, $\hat{\boldsymbol{y}}_{j|j-1}$ and $\hat{\boldsymbol{R}}_{j|j-1}$
4: Use Eq. A.8 to compute the Kalman Gain, $\boldsymbol{K}_j$
5: Use Eq. A.6 and A.7 to compute updates, $\hat{\boldsymbol{x}}_{j|j}$ and $\boldsymbol{P}_{j|j}$
6: Use Eq. A.9 and A.10 to compute $\hat{\boldsymbol{x}}_{(j+1)|j}$ and $\hat{\boldsymbol{P}}_{(j+1)|j}$
7: **end for**
8: **Return** (for all j) $\boldsymbol{\epsilon}_j = \hat{\boldsymbol{y}}_j - \hat{\boldsymbol{y}}_{j|j-1}$ and $\boldsymbol{R}_{j|j-1}$

After the prediction of the state value at the following measurement, we start again with predictions of the actual measurements until all the one-step predictions $\boldsymbol{y}_{j|j-1}$ and all the one-step prediction variances $\boldsymbol{R}_{j|j-1}$ have been calculated. The algorithm is summarized in Table A1.

REFERENCES

1. L. Aarons. Pharmacokinetic and pharmacodynamic modelling in drug development. *Stat. Methods Med. Res.* **8**(3):181–182 (1999).
2. L. B. Sheiner and J. L. Steimer. Pharmacokinetic/pharmacodynamic modeling in drug development. *Annu. Rev. Pharmacol. Toxicol.* **40**:67–95 (2000).
3. M. O. Karlsson, S. L. Beal, and L. B. Sheiner. Three new residual error models for population PK/PD analyses. *J. Pharmacokinet. Biopharm.* **23**(6):651–672 (1995).
4. S. L. Beal and L. B. Sheiner. *NONMEM User's Guides*, NONMEM Project Group, University of California, San Francisco, 1994.
5. Rajesh Krishna. *Applications of Pharmacokinetic Principles in Drug Development*, Kluwer Academic/Plenum Publishers, New York, 2004.
6. N. R. Kristensen. A method for systematic improvemenr of stochastic Grey-Box models. *Comput. Chem. Eng.* **28**(8):1431–1439 (2004).
7. C. W. Tornøe, J. L. Jacobsen, O. Pedersen, T. Hansen, and H. Madsen. Grey-box modelling of pharmacokinetic/pharmacodynamic systems. *J. Pharmacokinet. Pharmacodyn.* **31**(5):401–417 (2004).
8. C. W. Tornøe, J. L. Jacobsen, and H. Madsen. Grey-box pharmacokinetic/pharmacodynamic modelling of a euglycaemic clamp study. *J. Math. Biol.* **48**(6):591–604 (2004).
9. P. Sen, D. Bell, and D. Mohr. A calcium model with random absorption: a stochastic approach. *J. Theor. Biol.* **154**(4):485–493 (1992).
10. K. E. Andersen and M. Højbjerre. A population-based Bayesian approach to the minimal model of glucose and insulin homeostasis. Technical Report R-2003–25. Aalborg University (2003).
11. A. H. Jazwinski. *Stochastic Processes and Filtering Theory*, Academic Press, New York, 1970.
12. M. L. Lindstrom and D. M. Bates. Nonlinear mixed effects models for repeated measures data. *Biometrics* **46**(3):673–687 (1990).
13. A. Racine-Poon and J. Wakefield. Statistical methods for population pharmacokinetic modelling. *Stat. Methods Med. Res.* **7**(1):63–84 (1998).
14. L. Sheiner and J. Wakefield. Population modelling in drug development. *Stat. Methods Med. Res.* **8**(3):183–193 (1999).

15. P. E. Kloeden and E. Platen. *Numerical Solutions of Stochastic Differential Equations*, 2nd edition. Springer-Verlag, Heidelberg, 1995.
16. B. Øksendal. *Stochastic Differential Equations*, 4th edition. Springer-Verlag, Heidelberg, 1995.
17. J. N. Nielsen, H. Madsen, and P. C. Young. Parameter Estimation in Stochastic Differential Equations: An Overview. *Ann. Rev. Control* **24**:83–94 (2000).
18. J. Bak, H. A. Nielsen, and H. Madsen. Goodness of fit of stochastic differential equations. 21. Symp. Applied Statistics, ISBN 87–90117–24–7, 1999.
19. N. R. Kristensen, H. Madsen, and S. B. Jørgensen. Parameter estimation in stochastic Grey-Box Models. *Automatica* **40**:225–237 (2004).
20. J. E. Dennis and R. B. Schnabel. *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*, Prentice-Hall, NJ, 1983.
21. J. N. Nielsen and H. Madsen. Applying the EKF to stochastic differential equations with level efects. *Automatica* **37**:107–112 (2001).
22. A. Capderou, D. Douguet, T. Similowski, A. Aurengo, and M Zelter. Non-invasive assessment of technetium-99 m albumin transit time distribution in the pulmonary circulation by first-pass angiocardiography. *Eur. J. Nucl. Med.* **24**(7):745–753 (1997).
23. P. Magni, R. Bellazzi, G. De Nicolao, I. Poggesi, and M. Rocchetti. Nonparametric AUC estimation in population studies with incomplete sampling: a Bayesian approach. *J. Pharmacokinet. Pharmacodyn.* **29**(5–6):445–471 (2002).