Journal of Pharmacokinetics and Pharmacodynamics, Vol. 32, No. 1, February 2005 (© 2005)
DOI: 10.1007/s10928-005-2105-9

Using Stochastic Differential Equations for PK/PD Model Development

Niels Rode Kristensen,[1,*] Henrik Madsen,[2] and Steen Hvass Ingwersen[1]

Received July 7, 2004—Final March 22, 2005

A method for PK/PD model development based on stochastic differential equation models is proposed. The new method has a number of advantages compared to conventional methods. In particular, the new method avoids the exhaustive trial-and-error based search often conducted to determine the most appropriate model structure, because it allows information about the appropriate model structure to be extracted directly from data. This is accomplished through quantification of the uncertainty of the individual parts of an initial model, by means of which tools for performing model diagnostics can be constructed and guidelines for model improvement provided. Furthermore, the new method allows time-variations in key parameters to be tracked and visualized graphically, which allows important functional relationships to be revealed. Using simulated data, the performance of the new method is demonstrated by means of two examples. The first example shows how, starting from a simple assumption of linear PK, the method can be used to determine the correct nonlinear model for describing the PK of a drug following an oral dose. The second example shows how, starting from a simple assumption of no drug effect, the method can be used to determine the correct model for the nonlinear effect of a drug with known PK in an indirect response model.

KEY WORDS: stochastic differential equations; parameter estimation; extended Kalman filtering; smoothing; random walk; nonparametric modelling.

INTRODUCTION

Selecting an appropriate model structure is of critical importance for the final result when analysing individual pharmacokinetic (PK) and pharmacodynamic (PD) data using compartmental analysis. On one hand,

[1]Pharmacokinetics and Biomodelling, Novo Nordisk A/S, Novo Nordisk Park, DK-2760 Målov, Denmark.
[2]Informatics and Mathematical Modelling, Technical University of Denmark, Building 321, DK-2800 Lyngby, Denmark.
*To whom correspondence should be addressed. Tel: +45-44424028; E-mail: nikr@novonordisk.com.

using an overly simplified model structure may give biased estimates of key PK and PD parameters due to serial correlation between the residuals [1]. On the other hand, using an overly complex model structure may compromise parameter identifiability [2]. A suitable trade-off must therefore be found, taking into account also the aim of the specific analysis, but few systematic methods are available for this purpose. Exploratory data analysis may provide valuable insights and, in combination with knowledge of the physiology involved, such analysis sometimes allows an appropriate model structure to be determined. This is not always the case, however, and a more exhaustive trial-and-error based search is often conducted. In the present paper a new and more systematic method is proposed, the aim of which is to avoid this search by extracting from the data as much information as possible about the appropriate model structure.

The new method is based on stochastic differential equation (SDE) models instead of ordinary differential equation (ODE) models, and this facilitates extraction of otherwise inaccessible information. More specifically, the SDE models used are stochastic state space models consisting of a set of SDEs describing the dynamics of the states of the system in continuous time and a set of discrete-time measurement equations describing the relationship between the states and the observations obtained. Such models have a number of advantages compared to ODE models. Most importantly, the inclusion of a diffusion term, which constitutes the primary difference between an SDE model and the corresponding ODE model, allows the SDE model to explain a larger portion of the variation in a given data set and facilitates decomposition of prediction error. This means that effects of model uncertainty, e.g. due to model structure misspecifications, can be decoupled from effects of measurement noise. Thus, for parameter estimation, using SDE models instead of ODE models is a powerful approach to dealing with serially correlated residuals.

Since SDE models can account for model uncertainty, it can also be quantified, and this feature can be used to construct tools for performing model diagnostics. Similar diagnostic tools in the form of models that specifically allow serial correlation to be estimated have previously been proposed [1]. However, the information obtained by using tools based on SDE models may even be used to pinpoint model structure misspecifications. For example, it may be determined that what was initially modelled as a first order linear process with a fixed rate constant is in fact a nonlinear process, where the rate constant changes as a function of time due to variations in other variables.

Using SDE models not only makes it possible to pinpoint model structure misspecifications. By utilizing the prediction properties of SDE

models, time variations in key parameters, e.g. the rate constant mentioned above, can be tracked, and through graphical visualization of these variations, possibly in combination with nonparametric methods for feature extraction, it may be possible to determine a more appropriate model structure.

In the remainder of the paper the new method is described in detail and two examples using simulated data are given for illustration.

METHODS

This section contains a detailed description of the proposed method for determining appropriate model structures. Descriptions of the examples used for illustration and of the software used in the examples are also given.

The method proposed here is based on a recently proposed general method for iterative model improvement and some of the material is therefore similar to material presented in the original references [3, 4].

Methodology

The proposed method for determining appropriate model structures is illustrated in Fig. 1. Starting from the initial ODE model, which can be set up using conventional methods, the idea is to use SDE models to iteratively develop the final ODE model by means of a systematic procedure for iterative model building. The procedure is described in more detail in the following.

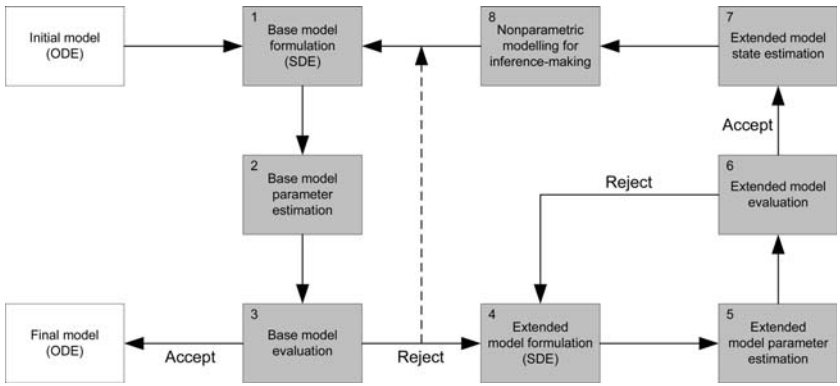

Fig. 1. Schematic illustration of the proposed procedure for iterative model building.

Base Model Formulation

The first step of the proposed procedure deals with the formulation of a basic SDE model. This involves deriving an initial set of ODEs to describe the dynamics of the states of the system, translating these ODEs into SDEs and adding an appropriate set of measurement equations to describe the relationship between the states and the observations obtained.

Deriving the initial set of ODEs is equivalent to selecting an initial model structure for conventional PK/PD modelling, which means that exploratory data analysis and knowledge of the physiology involved may be applied in the usual manner. In general, this gives rise to equations of the following type:

$$\frac{d\boldsymbol{x}_t}{dt} = \boldsymbol{f}(\boldsymbol{x}_t, \boldsymbol{u}_t, t, \boldsymbol{\theta}) \tag{1}$$

where $t \in \mathbb{R}$ is time, $\boldsymbol{x}_t \in \mathbb{R}^n$ is a vector of state variables (e.g. concentrations), $\boldsymbol{u}_t \in \mathbb{R}^m$ is a vector of input variables (e.g. infusion rates) and $\boldsymbol{\theta} \in \mathbb{R}^p$ is a vector of parameters, and where $\boldsymbol{f}(\cdot) \in \mathbb{R}^n$ is a vector function.

Translating the ODEs into SDEs is also straightforward, which is easily appreciated by considering the general form of the resulting equations:

$$d\boldsymbol{x}_t = \boldsymbol{f}(\boldsymbol{x}_t, \boldsymbol{u}_t, t, \boldsymbol{\theta})dt + \boldsymbol{\sigma}(\boldsymbol{u}_t, t, \boldsymbol{\theta})d\boldsymbol{\omega}_t \tag{2}$$

where the two terms on the right-hand side are called the *drift* term and the *diffusion* term, respectively. In the diffusion term $\boldsymbol{\sigma}(\cdot) \in \mathbb{R}^{n \times n}$ is a matrix function and $d\boldsymbol{\omega}_t$ is an infinitesimal increment in the stochastic process $\{\boldsymbol{\omega}_t\}$, which is an n-dimensional so-called *Wiener process*. This is a continuous stochastic process, which has increments (between points in time) that are independent and zero mean Gaussian with a covariance equal to the size of the time increment [5]. The Wiener process can be understood to originate from the sum of many identically distributed stochastic events giving rise to a difference (which increases with time) between the true evolution of $\boldsymbol{x}_t$ and the evolution described by $\boldsymbol{f}(\cdot)$. In other words, the diffusion term can be said to account for effects of model uncertainty, e.g. due to model structure misspecifications. Indeed, if the diffusion term is zero, (2) reduces to the differential form of (1). In the general case, however, (2) cannot be written in derivative form, because the derivative $\frac{d\boldsymbol{\omega}_t}{dt}$ cannot be treated mathematically.

Strictly speaking, even the commonly used differential form given in (2) is not a correct way of specifying a set of SDEs, because it does not say how $d\boldsymbol{\omega}_t$ should be integrated to obtain the solution. In the general

case, this integration may be performed either in the sense of Stratonovich or in the sense of Itô [5]. In the present case, though, the two approaches are identical, because $\boldsymbol{\sigma}(\cdot)$ is independent of $\boldsymbol{x}_t$. For a more detailed account of SDE theory, see [6].

The solution to an SDE is a stochastic process and not a well-defined function of time like the solution to an ODE. Solving an SDE is therefore often complicated and only very few SDEs can be solved explicitly. Numerical methods are available [7], but these are not as powerful as those available for ODEs, and rather than the unique curve obtained from an ODE simulation, only single sample paths or random realizations can be obtained for the evolution of the state variables when performing an SDE simulation. Consequently, proper characterization of the solution to an SDE generally requires Monte Carlo simulation and subsequent statistical analysis. Fortunately, only the first- and second-order moments of the solution are needed in the context of the present paper, and these can be obtained in a simpler manner, which involves only ODE solution and matrix algebra. This is discussed in more detail below.

Translating the ODEs of (1) into the SDEs of (2) does not affect the physical interpretability of the parameters of the drift term, which means that (2) may be used instead of (1) when estimating the unknown parameters of the model. However, by simultaneously estimating the parameters of the diffusion term, a quantitative measure of model uncertainty can also be obtained.

The formulation used in (2) corresponds to the formulation used in the original references [3, 4], but in order to construct a measure of model uncertainty that allows efficient quantification of the uncertainty in the individual parts of the model separately, a simpler formulation is used in the following, i.e.:

$$d\boldsymbol{x}_t = \boldsymbol{f}(\boldsymbol{x}_t, \boldsymbol{u}_t, t, \boldsymbol{\theta})dt + \boldsymbol{\sigma}(\boldsymbol{\theta})d\boldsymbol{\omega}_t \tag{3}$$

where $\boldsymbol{\sigma}(\cdot)$ is an $n \times n$-matrix with elements $\boldsymbol{\sigma_1}, \ldots, \boldsymbol{\sigma_n}$ along the diagonal and possibly some off-diagonal elements in order to satisfy fundamental balance equations (more details on the design of this matrix are given in the examples).

The only part that remains in formulating the basic SDE model is to add an appropriate set of measurement equations. This is done in accordance with the formulation used in the original references [3, 4], i.e.:

$$\boldsymbol{y}_k = \boldsymbol{h}(\boldsymbol{x}_k, \boldsymbol{u}_k, t_k, \boldsymbol{\theta}) + \boldsymbol{e}_k \tag{4}$$

where t_k, $k = 0, \ldots, N$, are sampling instants, $\boldsymbol{y}_k \in \mathbb{R}^l$ is a vector of measured output variables (e.g. plasma concentrations), $\boldsymbol{h}(\cdot) \in \mathbb{R}^l$ is a vector function and $\{\boldsymbol{e}_k\}$ is an l-dimensional white noise process with

$\boldsymbol{e}_k \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{S}(\boldsymbol{\theta}))$, where $\boldsymbol{S}(\boldsymbol{\theta})$ is a covariance matrix. The formulation in (4) corresponds to an additive error model, but by introducing a logarithmic transformation of the observations, it may also be used to approximate a proportional error model.

Base Model Parameter Estimation

The second step of the proposed procedure deals with estimation of the unknown parameters of the base model. A brief outline of an approximate maximum likelihood scheme that can be used for this purpose is given in the following. A more detailed account is given in the original references [3, 4].

Given a sequence of observations $\boldsymbol{y}_0, \boldsymbol{y}_1, \ldots, \boldsymbol{y}_k, \ldots, \boldsymbol{y}_{N-1}, \boldsymbol{y}_N$, estimates of the unknown parameters in (3)–(4) can be determined by finding the parameters $\boldsymbol{\theta}$ that maximize the likelihood function, i.e. the joint probability density:

$$L(\boldsymbol{\theta}; \mathcal{Y}_N) = p(\mathcal{Y}_N|\boldsymbol{\theta}) = p(\boldsymbol{y}_N, \boldsymbol{y}_{N-1}, \ldots, \boldsymbol{y}_1, \boldsymbol{y}_0|\boldsymbol{\theta}) \tag{5}$$

or equivalently in terms of conditional probability densities:

$$L(\boldsymbol{\theta}; \mathcal{Y}_N) = \left(\prod_{k=1}^{N} p(\boldsymbol{y}_k|\mathcal{Y}_{k-1}, \boldsymbol{\theta}) \right) p(\boldsymbol{y}_0|\boldsymbol{\theta}) \tag{6}$$

In general, performing an exact evaluation of the likelihood function is computationally infeasible, because it implies solving a complicated, so-called *general nonlinear filtering problem* [5]. However, since the increments of the Wiener process are Gaussian, it is reasonable to assume that the conditional probability densities in (6) can be well approximated by Gaussian densities. The Gaussian density is completely characterized by its first and second order moments, and this means that a method based on the so-called *extended Kalman filter* (EKF) [5] can be applied instead. The EKF describes the evolution of the first and second order moments of the conditional probability densities in terms of ODEs and algebraic equations (see Appendix A.1), and this makes it a computationally feasible approach. In order to formally introduce the assumption of Gaussianity, the following notation is defined:

$$\hat{\boldsymbol{y}}_{k|k-1} = E\{\boldsymbol{y}_k|\mathcal{Y}_{k-1}, \boldsymbol{\theta}\} \tag{7}$$

$$\boldsymbol{R}_{k|k-1} = V\{\boldsymbol{y}_k|\mathcal{Y}_{k-1}, \boldsymbol{\theta}\} \tag{8}$$

$$\boldsymbol{\epsilon}_k = \boldsymbol{y}_k - \hat{\boldsymbol{y}}_{k|k-1} \tag{9}$$

and the likelihood function is rewritten in the following way:

$$L(\boldsymbol{\theta};\mathcal{Y}_N) = \left(\prod_{k=1}^{N} \frac{\exp\left(-\frac{1}{2}\boldsymbol{\epsilon}_k^T \boldsymbol{R}_{k|k-1}^{-1}\boldsymbol{\epsilon}_k\right)}{\sqrt{\det\left(\boldsymbol{R}_{k|k-1}\right)}\left(\sqrt{2\pi}\right)^l}\right) p(\boldsymbol{y}_0|\boldsymbol{\theta}) \tag{10}$$

The parameter estimates can now be determined by further conditioning on $\boldsymbol{y}_0$ and solving the following nonlinear optimisation problem:

$$\hat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\{-\ln(L(\boldsymbol{\theta};\mathcal{Y}_N|\boldsymbol{y}_0))\} \tag{11}$$

where, for each value of $\boldsymbol{\theta}$ in the optimisation, $\boldsymbol{\epsilon}_k$ and $\boldsymbol{R}_{k|k-1}$, $k = 1,\ldots,N$, can be computed recursively by means of the EKF (see Appendix A.1).

The estimation scheme described here has been implemented in a computer program called CTSM [8, 9], within which the nonlinear optimisation problem (11) is solved by means of a quasi-Newton method. More details about this method and about how robustness towards outliers and missing observations has been incorporated into the estimation scheme are given in the original references [3,4], where the general efficiency and consistency of the scheme is also demonstrated, especially with respect to the parameters of the diffusion term, which is important for the procedure proposed in the present paper.

Base Model Evaluation

The third step of the proposed procedure deals with evaluation of the base model, when the unknown parameters have been estimated. This involves assessing the uncertainty of the parameter estimates, applying statistical tests and, most importantly, analysing the predictive capabilities of the model.

The uncertainty of the parameter estimates is easily assessed, because the estimator in (11) is asymptotically Gaussian with mean $\boldsymbol{\theta}$ and covariance:

$$\Sigma_{\hat{\boldsymbol{\theta}}} = \boldsymbol{H}^{-1} \tag{12}$$

where the matrix $\boldsymbol{H}$ can be approximated with the Hessian of the objective function in (11) at its minimum. To obtain a measure of the uncertainty of the individual parameter estimates, the covariance matrix can be decomposed:

$$\Sigma_{\hat{\boldsymbol{\theta}}} = \boldsymbol{\sigma}_{\hat{\boldsymbol{\theta}}}\boldsymbol{R}\boldsymbol{\sigma}_{\hat{\boldsymbol{\theta}}} \tag{13}$$

into $\boldsymbol{\sigma}_{\hat{\boldsymbol{\theta}}} = \mathrm{diag}(\sigma_{\hat{\theta}_1},\ldots,\sigma_{\hat{\theta}_p})$, which is a diagonal matrix of the standard errors of the parameter estimates, and $\boldsymbol{R}$, which is the corresponding correlation matrix. The asymptotic Gaussianity of the estimator in (11) also allows marginal t-tests to be performed to test the simple hypothesis:

$$H_0 : \theta_j = 0 \tag{14}$$

against the corresponding alternative:

$$H_1 : \theta_j \neq 0 \tag{15}$$

i.e. to test whether a given parameter θ_j is insignificant or not. The test quantity is the value of the parameter estimate $\hat{\theta}_j$ divided by the standard error of the estimate $\sigma_{\hat{\theta}_j}$ and under H_0 this quantity is asymptotically t-distributed with a number of degrees of freedom that equals the number of data points minus the number of estimated parameters, i.e.:

$$z(\hat{\theta}_j) = \frac{\hat{\theta}_j}{\sigma_{\hat{\theta}_j}} \sim t((N+1)l - p) \tag{16}$$

Due to correlations between the individual parameter estimates, a series of such marginal tests cannot be used to test the more complex hypothesis that a subset of the parameters, $\boldsymbol{\theta}_* \subset \boldsymbol{\theta}$, are simultaneously insignificant:

$$H_0 : \boldsymbol{\theta}_* = \mathbf{0} \tag{17}$$

against the alternative that they are not:

$$H_1 : \boldsymbol{\theta}_* \neq \mathbf{0} \tag{18}$$

and a test that takes correlations into account must be used, e.g. a test based on Wald's W-statistic [see e.g. [10]]. Under H_0 the test quantity for this test has an asymptotic χ^2-distribution with a number of degrees of freedom that equals the number of parameters tested. It can be computed as follows:

$$W(\hat{\boldsymbol{\theta}}_*) = \hat{\boldsymbol{\theta}}_*^T \Sigma_{\hat{\boldsymbol{\theta}}_*}^{-1} \hat{\boldsymbol{\theta}}_* \sim \chi^2(\dim(\hat{\boldsymbol{\theta}}_*)) \tag{19}$$

where $\hat{\boldsymbol{\theta}}_* \subset \hat{\boldsymbol{\theta}}$ is the subset of the parameter estimates subjected to the test and $\Sigma_{\hat{\theta}_*}$ is the corresponding covariance matrix. Tests for insignificance of one or more parameters can be applied to determine if the base model is overparameterized, and this may be done in all iterations of the proposed procedure, including the first, because no re-estimation is required. In all iterations but the first, the current base model may also be compared to the previous base model, which can be done by means of a likelihood ratio test, which is a test that measures whether or not expanding the base model has given a significant improvement of fit. More specifically, the following hypothesis is tested:

$$H_0 : \boldsymbol{\theta} \in \mathcal{M}_0 \tag{20}$$

against the corresponding alternative:

$$H_1 : \boldsymbol{\theta} \in \mathcal{M}_1 \tag{21}$$

where $\mathcal{M}_0 \subset \mathcal{M}_1$, and the test quantity, which under H_0 has an asymptotic χ^2-distribution with a number of degrees of freedom that equals the difference in dimension between $\mathcal{M}_0$ and $\mathcal{M}_1$, can be computed as follows:

$$LR = -2 \ln \frac{L(\hat{\boldsymbol{\theta}}_0; \mathcal{Y}_N | \boldsymbol{y}_0)}{L(\hat{\boldsymbol{\theta}}_1; \mathcal{Y}_N | \boldsymbol{y}_0)} \sim \chi^2(\dim(\mathcal{M}_1) - \dim(\mathcal{M}_0)) \tag{22}$$

where $L(\hat{\boldsymbol{\theta}}_0; \mathcal{Y}_N | \boldsymbol{y}_0)$ and $L(\hat{\boldsymbol{\theta}}_1; \mathcal{Y}_N | \boldsymbol{y}_0)$ are the maximized likelihood functions corresponding to the two hypotheses H_0 and H_1, respectively.

The above tests all provide valuable information about the quality of the base model, but analysing the predictive capabilities of the model is even more important. Given the fact that the proposed procedure aims at developing a final model consisting of ODEs rather than SDEs, the predictive capabilities of interest are the pure simulation capabilities, i.e. the ability of the drift term to explain the variation in the data set, when the diffusion term is set to zero. This ability can be investigated by plotting simulated output values against observed values and by calculating and analysing the corresponding residuals, e.g. by plotting them against time or against the simulated output values.

Thus far, the procedure proposed in the present paper is similar to conventional PK/PD modelling, but when using the results of the above methods for model evaluation as a basis for deciding whether to accept or reject the current base model, there is an important difference. In principle, parameters deemed insignificant by the above tests are parameters that may be eliminated, and the presence of such parameters is therefore an indication that the current base model is overparameterized. However, because of the particular nature of the SDE model in (3)–(4), where the diffusion term is included to account for effects of model uncertainty, e.g. due to model structure misspecifications, the presence of significant parameters in the diffusion term is an indication of an inability of the corresponding drift term to explain significant parts of the variation in the data set. Evaluating the significance of the parameters of the diffusion term therefore provides a quantitative measure of model uncertainty, and a such measure is not available in conventional PK/PD modelling.

The individual methods for model evaluation mentioned here can be summarized by stating a set of criteria for accepting the current base model:

(1) The pure simulation capabilities of the model must be acceptable.
(2) All of the parameters of the diffusion term should be insignificant.
(3) None of the parameters of the drift term should be insignificant.
(4) The current model should give a better fit than the previous model.

If these criteria, listed in order of decreasing importance, are all fulfilled, the current base model is accepted and its drift term can be used to formulate the final ODE model. If not, the current base model is rejected and an improved base model must be developed. If a given model fails on the first criterion, it usually fails on the second criterion as well, and if this is the case, the fourth step of the proposed procedure must be used. If it only fails on the third or fourth criterion, one can return directly to the first step of the proposed procedure and reformulate the model, e.g. by removing insignificant parameters.

Extended Model Formulation

The fourth step of the proposed procedure is the first step towards determining how to improve the base model, if it fails on one of the first two criteria listed above. It deals with determining the specific cause of failure and is facilitated by the diffusion term formulation used in (3), which allows model structure misspecifications to be pinpointed. More specifically, a significant diagonal element of the diffusion term can be used as an indicator of misspecifications in the corresponding element of the drift term, and if a specific part of such an element is selected for further analysis, the proposed procedure also provides a technique to determine whether or not that part is in fact misspecified. Model parts that are often misspecified are rate expressions, i.e. absorption rates, distribution rates or elimination rates in PK models or rates of buildup or loss of response in PD models. This is due to the fact that simple expressions are often used to characterize these potentially complex processes. If e.g. a process is in fact nonlinear, the rate constant associated with a linear approximation of that process will not be constant but may vary significantly with time. By specifically investigating if this is the case, it can therefore be determined whether or not the linear approximation is correct, and this is the idea behind the technique used to determine whether or not a specific model part is misspecified. More specifically, the assumption that any given parameter φ in the base model is constant can be checked by including this parameter as an additional state variable φ_t in a new SDE model, i.e.:

$$d\boldsymbol{x}_t^* = \boldsymbol{f}^*(\boldsymbol{x}_t^*, \boldsymbol{u}_t, t, \boldsymbol{\theta})dt + \boldsymbol{\sigma}^*(\boldsymbol{\theta})d\boldsymbol{\omega}_t^* \tag{23}$$

$$\boldsymbol{y}_k = \boldsymbol{h}(\boldsymbol{x}_k^*, \boldsymbol{u}_k, t_k, \boldsymbol{\theta}) + \mathbf{e}_k,\ \mathbf{e}_k \sim \mathcal{N}(\mathbf{0}, \boldsymbol{S}(\boldsymbol{\theta})) \tag{24}$$

and estimating the parameters of this extended model, where $\boldsymbol{x}_t^* = [\boldsymbol{x}_t^T \varphi_t]^T$ is an extended state vector, $\boldsymbol{f}^*(\boldsymbol{x}_t^*, \boldsymbol{u}_t, t, \boldsymbol{\theta}) = [\boldsymbol{f}(\boldsymbol{x}_t, \boldsymbol{u}_t, t, \boldsymbol{\theta})^T\ 0]^T$ is an extended drift term, $\boldsymbol{\sigma}^*(\boldsymbol{\theta}) = [\boldsymbol{\sigma}(\boldsymbol{\theta})\,\mathbf{0};\ \mathbf{0}\,\sigma_{n+1}]$ is an extended diffusion term and $\{\boldsymbol{\omega}_t^*\}$ is an $(n+1)$-dimensional Wiener process. Using this formulation, φ_t is constant if σ_{n+1} is zero, but varies with time (continuous random walk) if σ_{n+1} is nonzero. This means that estimating the parameters of the extended model and evaluating the significance of σ_{n+1} provides a method for checking whether or not φ is constant. Details are given in the following.

Extended Model Parameter Estimation

The fifth step of the proposed procedure deals with estimation of the unknown parameters of the extended model using the same data set as was used for estimation of the unknown parameters of the base model. The approximate maximum likelihood scheme outlined for the base model estimation can also be used here, e.g. as implemented in the computer program CTSM [8, 9].

Extended Model Evaluation

The sixth step of the proposed procedure deals with evaluation of the extended model, when the unknown parameters have been estimated, and this involves applying two types of statistical tests. First of all, it must be assessed whether or not the extended model has given a significant improvement of fit compared to the base model. This can be done by means of a likelihood ratio test, because the two models fulfil the condition of the base model being a real subset of the extended model. The test statistic, which can be computed from the maximized likelihood functions for the two models, has an asymptotic χ^2-distribution with one degree of freedom, corresponding to the diffusion term parameter σ_{n+1}. If the result of the test indicates that the extended model does not give a significant improvement of fit, the current extended model is rejected and one must return to the fourth step of the proposed procedure and formulate an alternative extended model, e.g. by including a different parameter as an additional state variable. If, on the other hand, the result of the test indicates that the extended model gives a significantly better fit, the current extended model is accepted. Marginal t-tests should subsequently be performed to assess the significance of the individual parameters, and if σ_{n+1}

is deemed significant, which is likely to be the case when the extended model gives a better fit than the base model, it indicates that φ varies with time and hence that the original assumption of constant φ is invalid and should be replaced by an assumption of φ being a function of other variables.

Extended Model State Estimation

The seventh step of the proposed procedure deals with obtaining information about the variations in φ that will allow an appropriate assumption about the functional dependence of φ on other variables to be formulated. The approximate maximum likelihood scheme used for parameter estimation in both the base model and the extended model uses the EKF to compute the innovations $\boldsymbol{\epsilon}_k = \boldsymbol{y}_k - \hat{\boldsymbol{y}}_{k|k-1}$ and their covariances $\boldsymbol{R}_{k|k-1}$ for $k = 1, \ldots, N$ (see Appendix A.1). Within the EKF, however, state estimates $\hat{\boldsymbol{x}}^*_{k|k-1}$ and $\hat{\boldsymbol{x}}^*_{k|k}$, $k = 1, \ldots, N$, are also computed, and because φ is a state variable in the extended model, both of these sets of estimates can be used to obtain information about the variations in φ from the data set used for estimation. An even better set of estimates can be constructed, however. By using a nonlinear smoothing algorithm based on the EKF [11] (see Appendix A.2), smoothed state estimates $\hat{\boldsymbol{x}}^*_{k|N}$, $k = 1, \ldots, N$ can be computed, including estimates of φ, i.e. $\hat{\varphi}_{k|N}$, $k = 1, \ldots, N$. Using smoothed state estimates to obtain quantitative information about the variations in φ bears great resemblance with deconvolution, i.e. the reconstruction of an unknown input profile to a dynamic system [12, 13], and the SDE model based technique used here can in fact be used as an alternative to standard deconvolution techniques.

Nonparametric Modelling for Inference-making

The eighth and, in principle, last step of the proposed procedure deals with analysing the variations in φ in order to formulate an appropriate assumption about the functional dependence of φ on other variables. The data used for this purpose are the smoothed state estimates obtained from the extended model, in particular the estimates of φ, i.e. $\hat{\varphi}_{k|N}$, $k = 1, \ldots, N$, which first and foremost should be plotted as a function of time to reveal the time-course of the variations, but the estimates of φ may also be plotted as a function of estimates of other state or input variables. If such direct visual inspection does not lead to any conclusions, a nonparametric regression method may be applied, which will allow significant features to be extracted to facilitate inference-making. Several nonparametric regression techniques are available [14],

but for the procedure proposed in the present paper, generalized additive models [15] are preferred, because fitting such models circumvents the curse of dimensionality, which tends to render nonparametric regression infeasible in higher dimensions, and because results obtained with such models are particularly easy to visualize. Using additive models, the variations in φ can be decomposed into variations that can be attributed to each of a given set of state or input variables, and the result can be visualized by means of partial dependence plots with associated confidence intervals. In this manner, it may be possible to reveal the true structure of the function describing φ, i.e. $\varphi(\boldsymbol{x}_t,\boldsymbol{u}_t,t,\boldsymbol{\theta})$, and obtain valuable information about how to re-formulate the base model, when returning to the first step of the proposed procedure for the next iteration. Knowledge of the physiology involved should of course be applied at this stage, both when selecting variables for nonparametric regression, but also when re-formulating the base model for the next iteration.

Summary of Methodology

Summarizing the procedure proposed in the present paper, the idea is to start from the initial ODE model and use SDE models to iteratively develop the final ODE model, using the steps described above. Within the proposed procedure two different types of SDE models are used. Base models, the first one of which is derived from the initial ODE model and from the last one of which the final ODE model is derived, and extended models, which are merely used as an aid in the iterative development of the base models. The primary criterion for finally accepting a given base model is that the pure simulation capabilities of the model are acceptable, and this usually coincides with all parameters of the diffusion term becoming simultaneously insignificant, which is the most important secondary criterion, because it indicates that there are no serious model structure misspecifications. Using the proposed procedure therefore helps to assure that the structure of the final ODE model is appropriate, and the parameters of the model may subsequently be re-estimated using conventional PK/PD modelling software such as WinNonlin [16,17].

Example Studies Design

The following briefly describes the design of the two examples reported in the next section. Both examples use simulated data for illustration.

Example 1: Absorption Kinetics

The first simulation example was designed to demonstrate that the proposed method can be used to determine what type of absorption

kinetics is appropriate for describing the PK of a drug following an oral dose. The simulated data set used for this purpose was generated with the following one-compartment model with nonlinear absorption and first order elimination:

$$\frac{dQ}{dt} = -\frac{V_{\max} Q}{K_m + Q} \tag{25}$$

$$\frac{dC}{dt} = \frac{V_{\max} Q}{(K_m + Q)V} - \frac{CL}{V} C \tag{26}$$

where $t \in [0,300]$ (min) is time, Q (mg) is amount of drug in the GI tract, and C (mg/l) is concentration of drug in plasma. The initial conditions and parameter values used were $Q_0 = 5$ mg, $C_0 = 0$ mg/l, $CL = 0.05$ l/min, $V = 5$ l, $V_{\max} = 0.1$ mg/min and $K_m = 0.5$ mg. A proportional error model was used, corresponding to the following measurement equation:

$$y_k = C_k(1 + e_k), \quad e_k \sim \mathcal{N}(0, S) \tag{27}$$

where $S = 0.01$, and where $k = 1,\ldots,N$ is an index corresponding to sampling times 1, 2, 4, 6, 8, 10, 15, 20, 30, 40, 50, 60, 70, 80, 90, 100, 150, 200, 250, 300 (min.), which gives $N = 20$ data points. The data set is shown in Fig. 2.

Example 2: Pharmacodynamics

The second simulation example was designed to demonstrate that the proposed method can also be used to determine what type of model is appropriate for describing the PD of a drug. The simulated data set used for this purpose was generated with the following one-compartment, indirect response model:

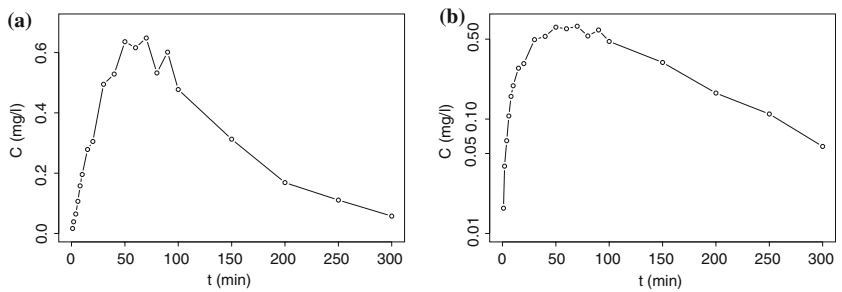

Fig. 2. Simulated concentration data for Example 1: (a) absolute scale, (b) semi-log scale.

$$\frac{dC}{dt} = -\frac{CL}{V}C \tag{28}$$

$$\frac{dR}{dt} = k_{\text{in}} - k_{\text{out}}\left(1 - \frac{I_{\max}C}{\text{IC}_{50} + C}\right)R \tag{29}$$

where $t \in [0,300]$ (min) is time, C (mg/l) is concentration of drug in plasma, and R is response. The initial conditions and parameter values used were $C_0 = 1$ mg/l, $R_0 = 1$ (pharmacodynamic steady state), $CL = 0.1$ l/min, $V = 5$ l, $k_{\text{in}} = 1$, $k_{\text{out}} = 1$, $I_{\max} = 1$ and $IC_{50} = 0.1$ mg/l. A proportional error model was used, corresponding to the following measurement equation:

$$y_k = R_k(1 + e_k), \quad e_k \sim \mathcal{N}(0, S) \tag{30}$$

where $S = 0.0025$, and $k = 1,\ldots,N$ is an index corresponding to sampling times 0, 2, 4, 6, 8, 10, 15, 20, 30, 40, 50, 60, 70, 80, 90, 100, 150, 200, 250, 300 (min), which gives $N = 20$ data points. The data set is shown in Fig. 3.

Software

The above described, simulated data sets for the two examples reported in the next section were generated using S-PLUS, version 6.1, from Insightful Corporation, Seattle, WA, USA. S-PLUS was also used to perform basic data manipulation, analysis and visualization. Parameter estimation in SDE models was performed using CTSM [8,9], version 2.3, publicly available from http://www.imm.dtu.dk/ctsm. CTSM was also used to generate simulated and smoothed state estimates for evaluation of the SDE models.

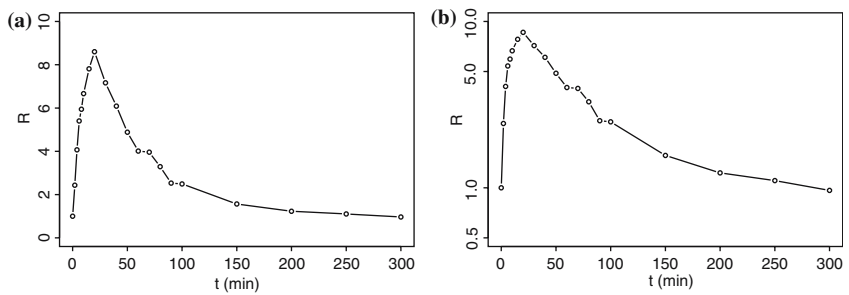

Fig. 3. Simulated response data for Example 2: (a) absolute scale, (b) semi-log scale.

RESULTS

In this section the performance of the proposed method for determining appropriate model structures is illustrated by means of two examples, using, for illustration, the simulated data sets briefly described in the previous section.

For both examples results are reported according to the flowchart shown in Fig. 1, the steps of which constitute the basis of the proposed method.

Example 1: Absorption Kinetics

The first example demonstrates how the proposed method can be used to determine what type of absorption kinetics is appropriate for describing the PK of a drug following an oral dose. Data from Fig. 2 is used.

Base Model Formulation

It is assumed that nothing is known in advance about the PK of the drug, and since the concentration profiles in Fig. 2 show no immediate need for including any nonlinearities or more than one compartment to describe the disposition of the drug, a reasonable initial model is a one-compartment model with first-order absorption and first order elimination, i.e.:

$$\frac{dQ}{dt} = -k_A Q \tag{31}$$

$$\frac{dC}{dt} = \frac{k_A Q}{V} - \frac{CL}{V} C \tag{32}$$

Translating this model into an SDE model, by including a diffusion term, gives rise to a base model with the following set of system equations:

$$dQ = -k_A Q dt + \sigma_1 d\omega_t^1 \tag{33}$$

$$dC = \left(\frac{k_A Q}{V} - \frac{CL}{V} C\right) dt - \frac{\sigma_1}{V} d\omega_t^1 + \sigma_2 d\omega_t^2 \tag{34}$$

The diagonal elements $\sigma_1 d\omega_t^1$ and $\sigma_2 d\omega_t^2$ have been included to allow separate quantification of the uncertainty in the two equations, and the off-diagonal element $-\frac{\sigma_1}{V} d\omega_t^1$ has been included to satisfy the mass balance for the drug, i.e. to reflect that all that is absorbed from the dosing compartment (GI tract) must appear in the central compartment (plasma). In other words, everything on the right-hand side of (33) must

appear negated on the right-hand-side of (34). The measurement equation of the base model is equivalent to (27).

Base Model Parameter Estimation

Estimating the unknown parameters of the base model with CTSM [8, 9] using the data in Fig. 2 gives the results shown in Table I.

Base Model Evaluation

Evaluating the quality of the base model, the four criteria for accepting a base model that were listed in the methodology section should be considered. First of all, the pure simulation capabilities of the model should be tested. Figure 4 shows a comparison between the observed data and a simulated concentration profile based on the base model with the diffusion term set to zero. From the plot of the residuals, it is evident that the base model is unable to capture important parts of the dynamics of the system, giving rise to significant serial correlation. The model thus fails on the first of the four criteria, and, as it turns out, also on the second, because one of the diffusion term parameters (σ_1) is significantly different from zero, but this is a natural consequence of the inability of

Table I. Estimation Results for the Base Model in (33)–(34) Using the Data From Fig. 2

Parameter	Estimate	Standard error	*t*-score	*p*-value
k_A	1.6186E−02	3.5558E−03	4.5521	0.0007
CL	5.2110E−02	1.6075E−03	32.4178	0.0000
V	3.2194E+00	7.2219E−01	4.4578	0.0008
σ_1	1.8910E−02	7.6761E−03	2.4635	0.0283
σ_2	1.1146E−07	8.9960E−05	0.0012	0.9990
S	5.7424E−03	3.0170E−03	1.9034	0.0786
	Neg. log. like.	−14.8020		

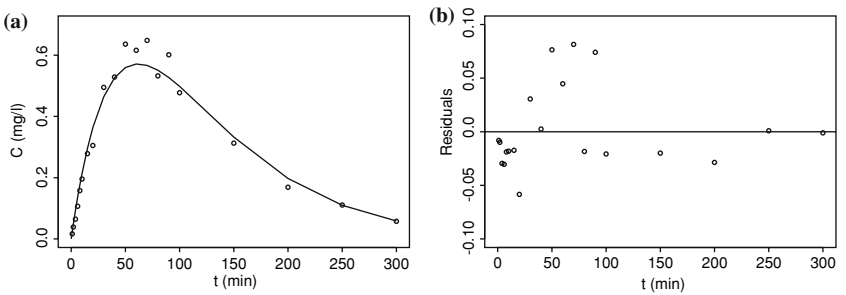

Fig. 4. Pure simulation comparison and residuals for the model in (33)–(34).

the drift term to capture the dynamics of the system. Seeing failure on the first two, the remaining criteria are not relevant.

Extended Model Formulation

Failing to fulfil the criteria for accepting a base model, the base model should be reformulated, and in order to determine how to do this, the *t*-scores and *p*-values shown in Table I are examined. The fact that σ_1 is significant, whereas σ_2 is not, is an indication that the inability of the model to fully capture the dynamics of the system may be due to incorrectly modelled absorption. Thus it may be worthwhile to form an extended model with k_A as an additional state variable, i.e. a model with the following system equations:

$$dQ = -k_A Q\, dt + \sigma_1\, d\omega_t^1 \tag{35}$$

$$dC = \left(\frac{k_A Q}{V} - \frac{CL}{V} C\right) dt - \frac{\sigma_1}{V} d\omega_t^1 + \sigma_2 d\omega_t^2 \tag{36}$$

$$dk_A = \sigma_3 d\omega_t^3 \tag{37}$$

The measurement equation of the extended model remains equivalent to (27).

Extended Model Parameter Estimation

Estimating the unknown parameters of the extended model with CTSM using the same data as before gives the results shown in Table II.

Extended Model Evaluation

Evaluating the quality of the extended model, the two types of statistical tests mentioned in the methodology section are applied. Because the

Table II. Estimation Results for the Extended Model in (35)–(37) Using the Data from Fig. 2

Parameter	Estimate	Standard error	*t*-score	*p*-value
$k_{A,0}$	1.7136E−02	2.1356E−03	8.0241	0.0000
CL	5.2187E−02	9.0733E−04	57.5175	0.0000
V	4.6358E+00	2.5781E−01	17.9816	0.0000
σ_1	3.0032E−20	5.4598E−16	0.0001	1.0000
σ_2	3.3896E−09	1.5970E−05	0.0002	0.9998
σ_3	1.6303E−03	6.7444E−04	2.4172	0.0322
S	5.1587E−03	1.7177E−03	3.0032	0.0111
	Neg. log. like.	−19.2063		

extended model reduces to the base model for $\sigma_3 = 0$, the two models are nested and a likelihood ratio test (using a χ^2-distribution with one degree of freedom) can be applied to determine if the extended model gives a better fit, which turns out to be the case (compare the values of the negative log. likelihoods in Tables II). Examining the *t*-scores and *p*-values in Table II, both σ_1 and σ_2 are now insignificant, whereas σ_3 is significant, and this indicates that the improvement in fit has been obtained by allowing k_A to vary with time.

Extended Model State Estimation

Attempting to determine a proper functional expression to account for the time-variations in k_A, the variations should first be examined. This is accomplished by generating smoothed state estimates from the extended model, as shown in Fig. 5. The state estimates reveal that k_A seems to increase from its initial value until a higher value is reached, whereafter it remains constant.

Nonparametric Modelling for Inference-making

As the next step in attempting to determine a functional expression to account for the time-variations in k_A, the smoothed state estimates are

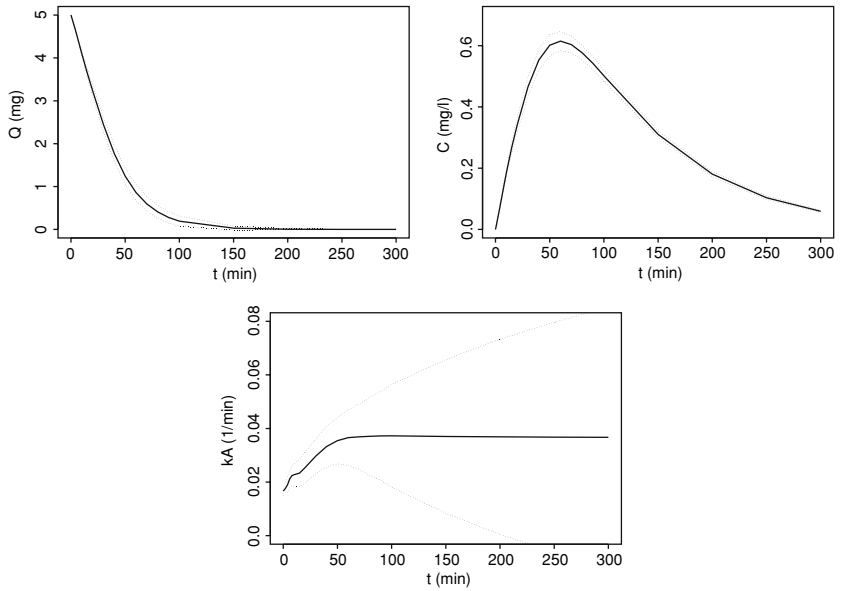

Fig. 5. Smoothed state estimates $\hat{Q}_{k|N}, \hat{C}_{k|N}$ and $\hat{k}_{A,k|N}$, $k = 1, \ldots, N$, for the model in (35)–(37). Solid lines: mean values, dotted lines: 95% confidence limits.

used to fit an additive model with k_A as the response variable and Q and C as predictor variables, using the S-PLUS function `gam`. The results, which are shown in Fig. 6, indicate that k_A is a decreasing function of Q and largely independent of C. The particular dependence on Q (lower rate constant at higher amounts in the GI tract) is indicative of saturable absorption kinetics, in turn suggesting that it may be a good idea to use Michaelis–Menten absorption kinetics.

Completing Model Development

Returning to the base model formulation step, the base model is therefore reformulated by introducing Michaelis–Menten absorption kinetics, which gives rise to a new base model with the following system equations:

$$dQ = -\frac{V_{\max} Q}{K_{\mathrm{m}} + Q}\, dt + \sigma_1 d\omega_t^1 \tag{38}$$

$$dC = \left(\frac{V_{\max} Q}{(K_{\mathrm{m}} + Q)V} - \frac{CL}{V} C\right) dt - \frac{\sigma_1}{V} d\omega_t^1 + \sigma_2 d\omega_t^2 \tag{39}$$

Estimating the unknown parameters of this model with CTSM using the same data as before gives the results shown in Table III, and evaluating the quality of the model, it turns out that it fulfils all four criteria for accepting a base model. More specifically, Fig. 7 shows that the pure simulation capabilities of the model are acceptable (no serial correlation of residuals) and Table III shows that both of the diffusion term parameters (σ_1 and σ_2) are insignificant, whereas all of the drift term parameters are significant. Finally, comparing the negative log. likelihood of the

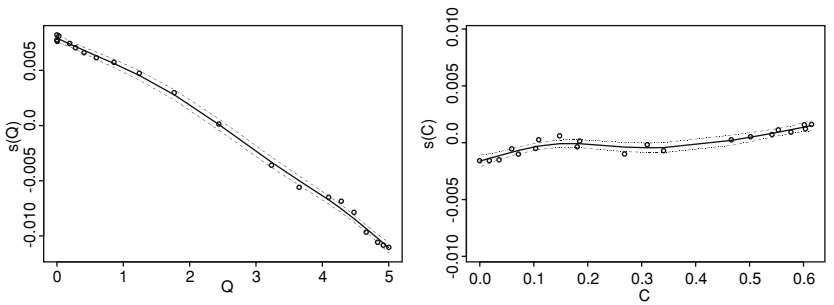

Fig. 6. Additive model fits of $\hat{k}_{A,k|N}$ vs. $\hat{Q}_{k|N}$ and $\hat{C}_{k|N}$, $k = 1, \ldots, N$, for the model in (35)–(37) (created using the S-PLUS functiong `gam`). Solid lines: Smoothing spline fits, dotted lines: 95% confidence limits, dots: partial residuals.

Table III. Estimation Results for the New Base Model in (38)–(39) Using the Data From Fig. 2

Parameter	Estimate	Standard error	*t*-score	*p*-value
V_{max}	1.1451E−01	6.4349E−03	17.7958	0.0000
K_m	9.8934E−01	7.2427E−02	13.6598	0.0000
CL	5.2199E−02	1.1369E−03	45.9134	0.0000
V	4.9924E + 00	2.1079E−01	23.6847	0.0000
σ_1	4.2244E−03	6.0448E−03	0.6989	0.4971
σ_2	1.2767E−08	2.2552E−06	0.0057	0.9956
S	5.3721E−03	1.3791E−03	3.8955	0.0024
	Neg. log. like.	−23.2659		

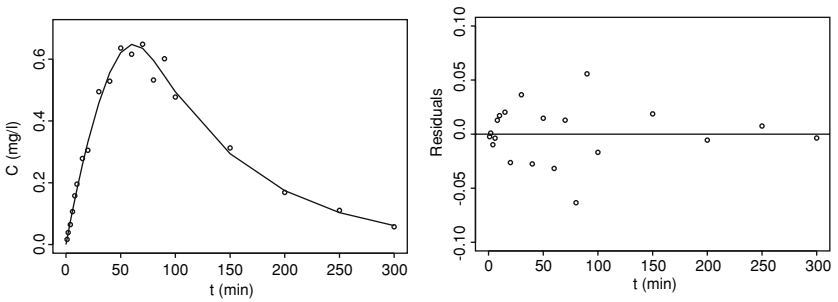

Fig. 7. Pure simulation comparison and residuals for the model in (38)–(39).

new base model to the negative log. likelihood of the original base model, it is clear that the new base model gives a significantly better fit. The new base model is therefore accepted and model development is complete. To summarize, this example has demonstrated how, starting from a simple assumption of linear kinetics, the proposed method could be used to determine a more appropriate model for describing the kinetics of a drug following an oral dose. In particular, it has been demonstrated that a not immediately visible nonlinearity could be effectively uncovered.

Example 2: Pharmacodynamics

The second example demonstrates how the proposed method can be used to determine what type of model is appropriate for describing the PD of a drug with known PK following an intravenous bolus. Data from Fig. 3 is used.

Base Model Formulation

It is assumed that the PK of the drug is known but that nothing is known in advance about the PD of the drug, except that it inhibits the loss of response in an indirect response model. The response profiles in Fig. 3 do not reveal the nature of the inhibition and since no other information is available, a very simple model with no drug effect is taken to be the starting point, i.e.:

$$\frac{dR}{dt} = k_{\text{in}} - k_{\text{out}} R \tag{40}$$

Translating this model into an SDE model, by including a diffusion term, gives rise to a base model with the following system equation:

$$dR = (k_{\text{in}} - k_{\text{out}} R)dt + \sigma_1 d\omega_t^1 \tag{41}$$

The measurement equation of the base model is equivalent to (30).

Base Model Parameter Estimation

Estimating the unknown parameters of the base model with CTSM [8,9] using the data in Fig. 3 gives the results shown in Table IV (note that the initial condition was also estimated, i.e. pharmacodynamic steady state was ignored).

Base Model Evaluation

Including no drug effect, the base model is bound to be inappropriate, and the comparison in Fig. 8 between the observed data and a simulated response profile based on the base model clearly shows this. The model is unable to capture the dynamics of the system and thus fails on the first of the four criteria for accepting a base model. It also fails on the second, because the diffusion term parameter (σ_1) is significantly different from zero as a consequence of the inability of the drift term to capture the dynamics of the system.

Table IV. Estimation Results for the Base Model in (41) Using the Data From Fig. 3

Parameter	Estimate	Standard error	*t*-score	*p*-value
R_0	1.1518E+00	4.0240E−01	2.8624	0.0126
k_{in}	6.8589E−02	3.3926E−02	2.0217	0.0622
k_{out}	1.7344E−02	1.3181E−02	1.3158	0.2084
σ_1	5.1687E−01	9.0102E−02	5.7365	0.0001
S	4.8131E−14	3.1145E−09	0.0000	1.0000
	Neg. log. like.	7.9792		

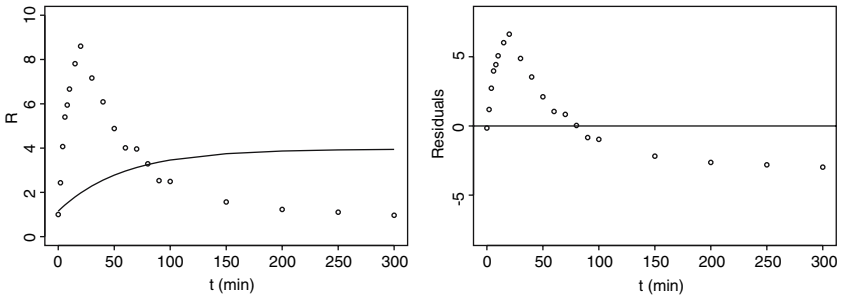

Fig. 8. Pure simulation comparison and residuals for the model in (41).

Extended Model Formulation

The base model should obviously be reformulated by including an inhibitory drug effect on the loss of response. In order to determine how to do this, an extended model, where k_{out} is included as an additional state variable, is therefore formed, i.e. a model with the following set of system equations:

$$dR = (k_{\text{in}} - k_{\text{out}}R)\,dt + \sigma_1 d\omega_t^1 \tag{42}$$

$$dk_{\text{out}} = \sigma_2 d\omega_t^2 \tag{43}$$

The measurement equation of the extended model remains equivalent to (30).

Extended Model Parameter Estimation

Estimating the unknown parameters of the extended model with CTSM using the same data as before gives the results shown in Table V.

Table V. Estimation Results for the Extended Model in (42)–(43) Using the Data From Fig. 3

Parameter	Estimate	Standard error	*t*-score	*p*-value
R_0	9.9500E−01	6.0443E−02	16.4618	0.0000
$k_{\text{out},0}$	2.9364E−02	5.0768E−02	0.5784	0.5723
k_{in}	8.0920E−01	9.8048E−02	8.2531	0.0000
σ_1	2.8991E−07	1.7698E−04	0.0016	0.9987
σ_2	1.4295E−02	2.7874E−03	5.1284	0.0003
S	8.5402E−04	9.1065E−04	0.9378	0.3644
	Neg. log. like.	−13.7552		

Extended Model Evaluation

The extended model reduces to the base model for $\sigma_2 = 0$, which means that the two models are nested, and a likelihood ratio test (using a χ^2-distribution with one degree of freedom) can therefore be applied to determine if the extended model gives a better fit. As expected, this turns out to be the case (compare the values of the negative log-likelihoods in Tables IV and V), and from the *t*-scores and *p*-values in Table V, which show that σ_1 is now insignificant, whereas σ_2 is significant, it is clear that the improvement in fit has been obtained by allowing k_{out} to vary significantly with time.

Extended Model State Estimation

Attempting to determine a proper functional expression to account for the time-variations in k_{out}, the variations are first examined by generating smoothed state estimates from the extended model. These are shown along with simulated concentrations C based on the known PK of the drug in Fig. 9, which reveals that k_{out} seems to increase almost linearly with time.

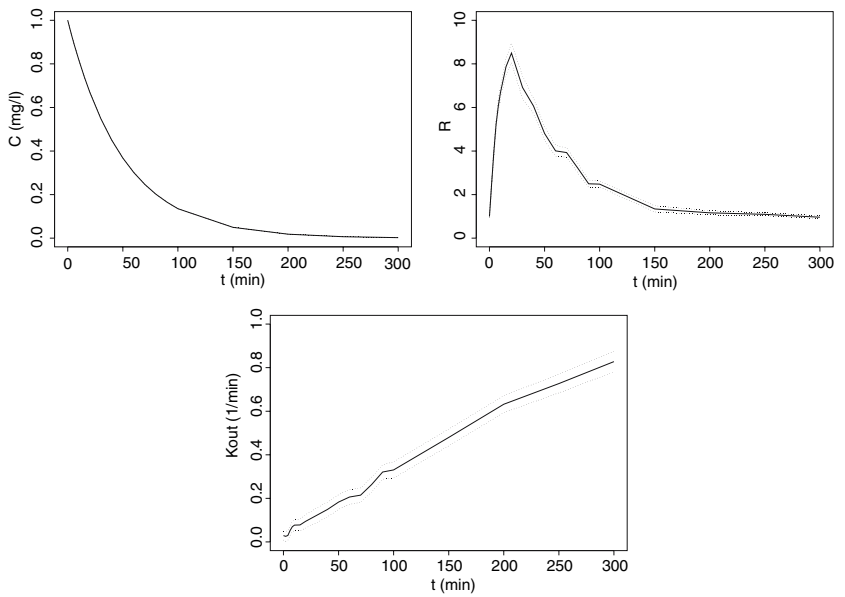

Fig. 9. Simulated concentrations $C_{k|0}$, $k = 1, \ldots, N$, based on the known PK of the drug, and smoothed state estimates $\hat{R}_{k|N}$ and $\hat{k}_{\text{out},k}$, $k = 1, \ldots, N$, for the model in (42)–(43). Solid lines: mean values, dotted lines: 95% confidence limits.

Nonparametric Modelling for Inference-making

As the next step, the simulated concentrations and smoothed state estimates are used to fit an additive model with k_{out} as the response variable and C and R as predictor variables, using the S-PLUS function `gam`. The results, which are shown in Fig. 10, indicate that k_{out} is a decreasing function of C and independent of R. The particular dependence on C (asymptotic decrease towards a lower limit for higher concentrations) in turn suggests that it may be a good idea to introduce an assumption of an $I_{\max}$-type drug effect.

Completing Model Development

Returning to the base model formulation step, the base model is reformulated by introducing the assumption mentioned above, which gives rise to a new base model with the following system equation:

$$dR = \left(k_{\text{in}} - k_{\text{out}}\left(1 - \frac{I_{\max}C}{IC_{50} + C}\right)R\right)dt + \sigma_1 d\omega_t^1 \tag{44}$$

Estimating the unknown parameters of this model with CTSM using the same data as before gives the results shown in Table VI, and evaluating the quality of the model, it turns out that it fulfils all four criteria for accepting a base model. More specifically, Fig. 11 shows that the pure simulation capabilities of the model are acceptable (no serial correlation of residuals) and Table VI shows that the diffusion term parameter (σ_1) is insignificant, whereas all of the drift term parameters are significant. Finally, comparing the negative log-likelihood of the new base model to the negative log-likelihood of the original base model, it is clear that the new base model gives a significantly better fit. It is therefore accepted and

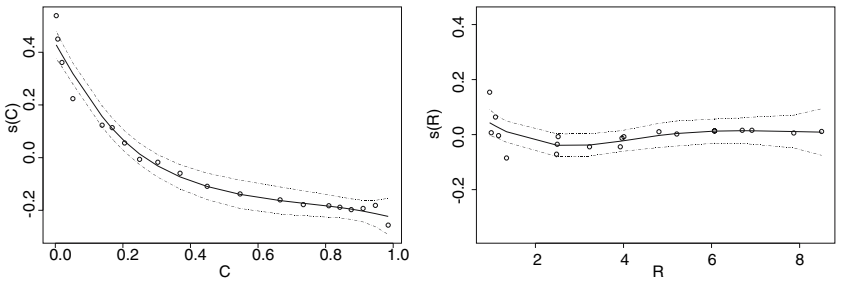

Fig. 10. Additive model fits of $k_{\text{out},k|N}$ vs. $\hat{R}_{k|N}$ and $C_{k|0}$, $k = 1,\ldots,N$, for the model in (42)–(43) (created using the S-PLUS function `gam`). Solid lines: Smoothing spline fits, dotted lines: 95% confidence limits, dots: partial residuals.

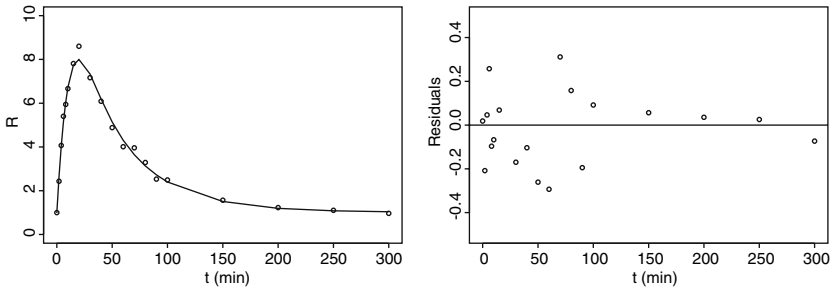

Fig. 11. Pure simulation comparison and residuals for the model in (44).

Table VI. Estimation Results for the New Base Model in (44) Using the Data From Fig. 3

Parameter	Estimate	Standard error	*t*-score	*p*-value
R_0	9.8133E−01	4.6141E−02	21.2682	0.0000
k_{in}	9.8782E−01	4.7552E−02	20.7735	0.0000
k_{out}	9.7464E−01	6.0339E−02	16.1527	0.0000
I_{max}	1.0122E+00	1.2520E−02	80.8437	0.0000
IC_{50}	1.0701E−01	9.6240E−03	11.1187	0.0000
σ_1	6.3329E−05	1.5672E−02	0.0040	0.9968
S	2.4014E−03	7.5878E−04	3.1648	0.0084
	Neg. log. like.	−31.9382		

model development is complete. To summarize, this example has demonstrated how, starting from a simple assumption of no drug effect, the proposed method could be used to determine an appropriate model for describing the nature of the inhibitory effect of a drug with known PK on the loss of response in an indirect response model.

DISCUSSION

The SDE-based method for developing PK/PD models that has been proposed in the present paper has a number of advantages compared to conventional methods, some of which have been demonstrated through the examples given in the previous section. In an overall perspective, the key advantage of the proposed method is that it avoids the exhaustive trial-and-error based search often conducted to determine the most appropriate model structure, which is accomplished by allowing information about the appropriate model structure to be extracted directly from data. This was illustrated, using simulated data, by showing how, starting from

a simple assumption of linear PK, the proposed method could be used to determine the correct nonlinear model for describing the PK of a drug following an oral dose. Similar results were obtained for PD, again using simulated data, by showing how, starting from a simple assumption of no drug effect, the proposed method could be used to determine the correct model for describing the nonlinear nature of the inhibitory effect of a drug with known PK on the loss of response in an indirect response model.

What allows the SDE-based method to facilitate extraction of information about the appropriate model structure from data is the inclusion of a diffusion term, which allows the SDE model to explain a larger portion of the variation in a given data set than the corresponding ODE model. Estimation of the parameters of the diffusion term therefore facilitates quantification of model uncertainty, and this not only allows tools for performing model diagnostics to be constructed, but it also provides guidelines for model improvement. Another very strong feature is that the inclusion of the diffusion term allows state filtering and smoothing techniques to be applied, by means of which time-variations in key parameters can be tracked and visualized graphically. This in turn allows nonparametric methods for feature extraction to be applied in order to determine if there are functional relationships between the key parameters and other important variables. In this manner, not only the structure of such relationships, but also reasonable initial estimates of the parameters needed to characterize them, may be determined graphically.

Using the proposed method, it may be possible to reveal nonlinear PK and PD phenomena using fewer dose levels than what is common today. Indeed, a single dose level should be enough, provided it is high enough to excite the nonlinearity. However, as the proposed method, in its current form, is a method for analysing individual PK and PD data, rich data is a requirement, especially in view of the fact that by including a diffusion term the number of parameters to be estimated increases. As sparse population data is often more readily available, initiatives have also been taken to extend the proposed method to a more general population form [18, 19]. In particular, it has been demonstrated that parameters in nonlinear mixed-effects models based on SDEs can be correctly estimated using sparse population data [18], but further developments, especially with respect to smoothing techniques and with respect to nonparametric methods for feature extraction, are needed. Given such developments, a population form of the proposed method may prove even more useful than the individual form. Indeed, it has already been demonstrated that in terms of the ability to detect nonlinearities, using nonlinear mixed-effects modelling is superior to using a standard two-stage population analysis approach, where parameters are estimated separately for each individual [20].

APPENDIX A

In the following, algorithms for performing extended Kalman filtering and nonlinear smoothing based on the extended Kalman filter (EKF) are described (for formal derivations of the two algorithms, see [5,11]). Both algorithms have been implemented in the computer program CTSM [8,9]. The EKF is an integral part of the approximate maximum likelihood scheme for parameter estimation, which constitutes the core of the program, but both algorithms may be used independently for generating different types of state estimates.

Extended Kalman Filtering

Consider an SDE model with the following system and measurement equation:

$$d\boldsymbol{x}_t = \boldsymbol{f}(\boldsymbol{x}_t, \boldsymbol{u}_t, t, \boldsymbol{\theta})dt + \boldsymbol{\sigma}(\boldsymbol{u}_t, t, \boldsymbol{\theta})d\boldsymbol{\omega}_t \tag{A.1}$$

$$\boldsymbol{y}_k = \boldsymbol{h}(\boldsymbol{x}_k, \boldsymbol{u}_k, t_k, \boldsymbol{\theta}) + \boldsymbol{e}_k,\ \boldsymbol{e}_k \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{S}(\boldsymbol{\theta})) \tag{A.2}$$

Given a sequence of observations $\boldsymbol{y}_0, \boldsymbol{y}_1, \ldots, \boldsymbol{y}_k, \ldots, \boldsymbol{y}_{N-1}, \boldsymbol{y}_N$, the EKF can be used to recursively compute the values of $\boldsymbol{\epsilon}_k$ and $\boldsymbol{R}_{k|k-1}$, $k = 1, \ldots, N$, needed to evaluate the objective function in (11) for a given value of $\boldsymbol{\theta}$.

Initial conditions for the EKF are $\hat{\boldsymbol{x}}_{1|0} = E\{\boldsymbol{x}_{t_0}\}$ and $\boldsymbol{P}_{1|0} = V\{\boldsymbol{x}_{t0}\}$, i.e. the mean and covariance of the initial states in (A.1). Given the initial conditions, the first step of the EKF recursions involves the *output prediction* equations:

$$\hat{\boldsymbol{y}}_{k|k-1} = \boldsymbol{h}(\hat{\boldsymbol{x}}_{k|k-1}, \boldsymbol{u}_k, t_k, \boldsymbol{\theta}) \tag{A.3}$$

$$\boldsymbol{R}_{k|k-1} = \boldsymbol{C}_k \boldsymbol{P}_{k|k-1} \boldsymbol{C}_k^T + \boldsymbol{S}_k \tag{A.4}$$

which predict the mean and covariance of the outputs at time t_k given the information available at time t_{k-1} (e.g., for $k = 1$, the mean and covariance of the outputs at time t_1 given the mean and covariance of the initial states). $\boldsymbol{C}_k$ is the Jacobian of $\boldsymbol{h}(\cdot)$ with respect to $\boldsymbol{x}_k$ (evaluated using $\hat{\boldsymbol{x}}_{k|k-1}$) and $\boldsymbol{S}_k = \boldsymbol{S}(\boldsymbol{\theta})$. The second step of the recursions involves the *innovation* equation:

$$\boldsymbol{\epsilon}_k = \boldsymbol{y}_k - \hat{\boldsymbol{y}}_{k|k-1} \tag{A.5}$$

which determines the one-step-ahead prediction residual or innovation at time t_k, i.e. the difference between the observed outputs and the predicted

outputs. The third step of the recursions involves the *Kalman gain* equation:

$$\boldsymbol{K}_k = \boldsymbol{P}_{k|k-1}\boldsymbol{C}_k^T\boldsymbol{R}_{k|k-1}^{-1} \tag{A.6}$$

which determines the so-called Kalman gain at time t_k. This is essentially a matrix of weights determining the degree of updating that is going to take place in the fourth recursion step. This step involves the *updating* equations:

$$\hat{\boldsymbol{x}}_{k|k} = \hat{\boldsymbol{x}}_{k|k-1} + \boldsymbol{K}_k\boldsymbol{\epsilon}_k \tag{A.7}$$

$$\boldsymbol{P}_{k|k} = \boldsymbol{P}_{k|k-1} - \boldsymbol{K}_k\boldsymbol{R}_{k|k-1}\boldsymbol{K}_k^T \tag{A.8}$$

where the predicted mean and covariance of the states at time t_k are updated based on the observed outputs via the innovation and the Kalman gain. The fifth and final recursion step involves the state prediction equations:

$$\frac{d\hat{\boldsymbol{x}}_{t|k}}{dt} = \boldsymbol{f}(\hat{\boldsymbol{x}}_{t|k}, \boldsymbol{u}_t, t, \boldsymbol{\theta}) \tag{A.9}$$

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

which are solved for $t \in [t_k,\ t_{k+1}]$ to predict the mean and covariance of the states at time t_{k+1}. $\boldsymbol{A}_t$ is the Jacobian of $\boldsymbol{f}(\cdot)$ with respect to $\boldsymbol{x}_t$ (evaluated using $\hat{\boldsymbol{x}}_{t|k}$) and $\boldsymbol{\sigma}_t = \boldsymbol{\sigma}(\boldsymbol{u}_t, t, \boldsymbol{\theta})$. Using the new predicted values of the mean and covariance of the states, the recursions are repeated from (A.3) to (A.4).

In terms of the initial conditions for the EKF, $\hat{\boldsymbol{x}}_{1|0} = E\{\boldsymbol{x}_{t_0}\}$ can either be pre-specified or estimated as a part of the overall parameter estimation problem in CTSM, and it is in general reasonable to compute $\boldsymbol{P}_{1|0} = V\{\boldsymbol{x}_{t_0}\}$ as follows:

$$\boldsymbol{P}_{1|0} = P_s \int_{t_0}^{t_1} e^{\boldsymbol{A}_s s}\boldsymbol{\sigma}_s\boldsymbol{\sigma}_s^T(e^{\boldsymbol{A}_s s})^T ds \tag{A.11}$$

i.e. as the integral of the Wiener process and the system dynamics over the first sample, scaled by a pre-specified scaling factor $P_s \geq 1$, which in CTSM is set to 1. The EKF algorithm is formally summarized in Table VII.

The above formulation of the EKF is specifically tailored to allow the innovations $\boldsymbol{\epsilon}_k$ and their covariances $\boldsymbol{R}_{k|k-1}$ to be computed for $k = 1, \ldots, N$, but this is not the most common formulation. The EKF is

Table VII. The Extended Kalman Filter (EKF) Algorithm 1

1: Given the SDE model in (A.1)–(A.2), a parameter vector $\boldsymbol{\theta}$, and initial states $\boldsymbol{x}_{t_0}$.
2: Set the initial mean $\hat{\boldsymbol{x}}_{1|0} = \boldsymbol{x}_{t_0}$ and use (A.11) for the initial covariance $\boldsymbol{P}_{1|0}$.
3: **FOR** $k = 1$ to N **DO**:
4: Use (A.3)–(A.4) to compute $\hat{\boldsymbol{y}}_{k|k-1}$ and $\boldsymbol{R}_{k|k-1}$.
5: Use (A.5) and (A.6) to compute $\boldsymbol{\epsilon}_k$ and $\boldsymbol{K}_k$.
6: Use (A.7)–(A.8) to compute $\hat{\boldsymbol{x}}_{k|k}$ and $\overline{\boldsymbol{P}}_{k|k}$.
7: Use (A.9)–(A.10) to compute $\hat{\boldsymbol{x}}_{k+1|k}$ and $\overline{\boldsymbol{P}}_{k+1|k}$.
8: **END FOR**
9: Return $\boldsymbol{\epsilon}_k$ and $\boldsymbol{R}_{k|k-1}$ for $k = 1, \ldots, N$ (to construct the likelihood function).

more frequently used to compute state estimates, e.g. $\hat{\boldsymbol{x}}_{k|k-1}$ or $\hat{\boldsymbol{x}}_{k|k}$ and the corresponding covariance matrices $\boldsymbol{P}_{k|k-1}$ or $\boldsymbol{P}_{k|k}$ for $k = 1, \ldots, N$. The above formulation still allows this, however, and in CTSM both types of state estimates and the corresponding covariance matrices can also be computed (see Table VII).

Nonlinear Smoothing based on the EKF

Consider again the SDE model in (A.1)–(A.2). Given a sequence of observations $\boldsymbol{y}_0, \boldsymbol{y}_1, \ldots, \boldsymbol{y}_k, \ldots, \boldsymbol{y}_N - 1, \boldsymbol{y}_N$, a nonlinear smoothing algorithm based on the EKF can be used to compute state estimates that have smaller variance than the ones provided by the EKF formulation given above. While the estimates provided by the EKF are conditioned on either past information $(\hat{\boldsymbol{x}}_{k|k-1},\ k = 1, \ldots, N)$ or past and present information $(\hat{\boldsymbol{x}}_{k|k},\ k = 1, \ldots, N)$ smoothed state estimates are conditioned on all available information $(\hat{\boldsymbol{x}}_{k|N},\ k = 1, \ldots, N)$ i.e. both past, present and "future" information.

The nonlinear smoothing algorithm is based on the following formulas:

$$\hat{\boldsymbol{x}}_{k|N} = \boldsymbol{P}_{k|N}(\boldsymbol{P}_{k|k-1}^{-1}\hat{\boldsymbol{x}}_{k|k-1} + \overline{\boldsymbol{P}}_{k|k}^{-1}\hat{\overline{\boldsymbol{x}}}_{k|k}) \tag{A.12}$$

$$\boldsymbol{P}_{k|N} = (\boldsymbol{P}_{k|k-1}^{-1} + \overline{\boldsymbol{P}}_{k|k}^{-1})^{-1} \tag{A.13}$$

which state that the smoothed estimate can be computed by combining a forward filter estimate based only on past information with a backward filter estimate based only on present and "future" information.

The forward filter estimates $\hat{\boldsymbol{x}}_{k|k-1}$ and their covariance matrices $\boldsymbol{P}_{k|k-1}$ can be computed for $k = 1, \ldots, N$ by means of the EKF (see Table VII). The backward filter estimates $\hat{\overline{\boldsymbol{x}}}_{k|k}$ and their covariance matrices $\overline{\boldsymbol{P}}_{k|k}$, on the other hand, must be computed for $k = N, \ldots, 1$ using a different set of formulas.

Table VIII. The Algorithm for Nonlinear Smoothing Based on the EKF

1. Given the SDE model in (A.1)–(A.2), a parameter vector $\boldsymbol{\theta}$, and initial states $\boldsymbol{x}_{t_0}$.
2. Use the EKF algorithm in Table VII to compute and $\boldsymbol{P}_{k|k-1}$ for $k = 1, \ldots, N$.
3. Set the initial conditions $\boldsymbol{s}_{N|N+1} = 0$ and $\overline{\boldsymbol{P}}_{N|N+1} = \boldsymbol{0}$.
4. **FOR** $k = N$ to 1 **DO**:
5. Use (A.18)–(A.19) to compute $\boldsymbol{s}_{k|k}$ and $\overline{\boldsymbol{P}}^{-1}_{k|k}$
6. Use (A.20)–(A.21) to compute $\boldsymbol{s}_{k\text{-}1|k}$ and $\overline{\boldsymbol{P}}^{-1}_{k-1|k}$.
7. **END FOR**
8. Use (A.16)–(A.17) to compute $\hat{\boldsymbol{x}}_{k|N}$ and $\boldsymbol{P}_{k|N}$ for $k = 1, \ldots, N$.

In this set of formulas a transformation of the time variable is used, i.e. $\tau = t_N - t$, and the backward filter is based on a rewritten SDE model:

$$d\boldsymbol{x}_{t_N-\tau} = -\boldsymbol{f}(\boldsymbol{x}_{t_N-\tau}, \boldsymbol{u}_{t_n-\tau}, t_N - \tau, \boldsymbol{\theta})d\tau - \boldsymbol{\sigma}(\boldsymbol{u}_{t_N-\tau}, t_N - \tau, \boldsymbol{\theta})d\boldsymbol{\omega}_\tau \tag{A.14}$$

$$\boldsymbol{y}_k = \boldsymbol{h}(\boldsymbol{x}_k, \boldsymbol{u}_k, t_k, \boldsymbol{\theta}) + \boldsymbol{e}_k, \;\; \boldsymbol{e}_k \sim N(\boldsymbol{\theta}, \boldsymbol{S}(\boldsymbol{\theta})) \tag{A.15}$$

where $\tau \in [0, t_N]$. For ease of implementation a coordinate transformation is also introduced, i.e. $\mathbf{s}_t = \overline{\boldsymbol{P}}_t^{-1}\hat{\boldsymbol{x}}_t$, and (A.12)–(A.13) are rewritten as follows:

$$\hat{\boldsymbol{x}}_{k|N} = \boldsymbol{P}_{k|N}(\boldsymbol{P}^{-1}_{k-1}\hat{\boldsymbol{x}}_{k|k-1} + \boldsymbol{s}_{k|k}) \tag{A.16}$$

$$\boldsymbol{P}_{k|N} = (\boldsymbol{P}^{-1}_{k|k-1} + \overline{\boldsymbol{P}}^{-1}_{k|k})^{-1} \tag{A.17}$$

Starting from the initial conditions $\mathbf{s}_{N|N+1} = 0$, and $\overline{\boldsymbol{P}}^{-1}_{N|N+1} = 0, \boldsymbol{s}_{k|k}$ and $\overline{\boldsymbol{P}}^{-1}_{k|k}, k = N, \ldots, 1$, are computed recursively in a two-step manner. The first step of the recursions involves the following *updating* equations:

$$\boldsymbol{s}_{k|k} = \boldsymbol{s}_{k|k+1} + \boldsymbol{C}_k^T\boldsymbol{S}_k^{-1}(\boldsymbol{y}_k - \boldsymbol{h}(\hat{\boldsymbol{x}}_{k|k-1}, \boldsymbol{u}_k, t_k, \boldsymbol{\theta}) + \boldsymbol{C}_k\hat{\boldsymbol{x}}_{k|k-1}) \tag{A.18}$$

$$\overline{\boldsymbol{P}}^{-1}_{k|k} = \overline{\boldsymbol{P}}^{-1}_{k|k+1} + \boldsymbol{C}_k^T\boldsymbol{S}_k^{-1}\boldsymbol{C}_k \tag{A.19}$$

where $\boldsymbol{C}_k$ is the Jacobian of $\boldsymbol{h}(\cdot)$ with respect to $\boldsymbol{x}_k$ (evaluated using $\hat{\boldsymbol{x}}_{k|k-1}$) and $\boldsymbol{S}_k = \boldsymbol{S}(\boldsymbol{\theta})$. The second recursion step involves the *prediction* equations:

$$\begin{aligned}\frac{d\boldsymbol{s}_{t_N-\tau|k}}{d\tau} = {} & \boldsymbol{A}_\tau^T\boldsymbol{s}_{t_N-\tau|k} - \overline{\boldsymbol{P}}^{-1}_{t_N-\tau|k}\boldsymbol{\sigma}_\tau\boldsymbol{\sigma}_\tau^T\boldsymbol{s}_{t_N-\tau|k} - \overline{\boldsymbol{P}}^{-1}_{t_N-\tau|k} \\ & \times (\boldsymbol{f}(\hat{\boldsymbol{x}}_{t_N-\tau|k}, \boldsymbol{u}_{t_N-\tau}, t_N - \tau, \boldsymbol{\theta}) - \boldsymbol{A}_\tau\hat{\boldsymbol{x}}_{t_N-\tau|k})\end{aligned} \tag{A.20}$$

$$\frac{d\overline{\boldsymbol{P}}^{-1}_{t_N-\tau|k}}{d\tau} = \overline{\boldsymbol{P}}^{-1}_{t_N-\tau|k}\boldsymbol{A}_\tau + \boldsymbol{A}_\tau^T\overline{\boldsymbol{P}}^{-1}_{t_N-\tau|k} - \overline{\boldsymbol{P}}^{-1}_{t_N-\tau|k}\boldsymbol{\sigma}_\tau\boldsymbol{\sigma}_\tau^T\overline{\boldsymbol{P}}^{-1}_{t_N-\tau|k} \tag{A.21}$$

which are solved for $\tau \in [\tau_k, \tau_{k+1}]$. $\boldsymbol{A}_\tau$ is the Jacobian of $\boldsymbol{f}(\cdot)$ with respect to $\boldsymbol{x}_{t_N-\tau}$ (evaluated using $\hat{\boldsymbol{x}}_{t_N-\tau|k}$) and $\boldsymbol{\sigma}_\tau = \boldsymbol{\sigma}(\boldsymbol{u}_{t_N-\tau}, t_N - \tau, \boldsymbol{\theta})$. Using the new predicted values of $\boldsymbol{s}_{k-1|\,k}$ and $\overline{\boldsymbol{P}}^{-1}_{k-1|k}$, the recursions are repeated.

The initial conditions for the backward filter i.e. $\boldsymbol{s}_{N|N+1} = \boldsymbol{0}$ and $\overline{\boldsymbol{P}}^{-1}_{N|N+1} = \boldsymbol{0}$, can be derived from an alternative formulation of (A.16)–(A.17):

$$\hat{\boldsymbol{x}}_{k|N} = \boldsymbol{P}_{k|N}(\overline{\boldsymbol{P}}^{-1}_{k|N}\hat{\boldsymbol{x}}_{k|k} + \boldsymbol{s}_{k|k+1}) \tag{A.22}$$

$$\boldsymbol{P}^{-1}_{k|N} = (\boldsymbol{P}^{-1}_{k|k} + \overline{\boldsymbol{P}}^{-1}_{k|k+1})^{-1} \tag{A.23}$$

by realizing that the smoothed estimate must coincide with the forward filter estimate for $k = N$. The nonlinear smoothing algorithm is formally summarized in Table VIII. Smoothed state estimates and the corresponding covariance matrices can also be straightforwardly computed in CTSM.

REFERENCES

1. 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).
2. C. Cobelli and J. J. DiStefano III. Parameter and structural identifiability concepts and ambiguities: A critical review and analysis. *Am. J. Physiol.* **239**:R7–R24 (1980).
3. N. R. Kristensen, H. Madsen, and S. B. Jørgensen. Parameter estimation in stochastic grey-box models. *Automatica* **40**(2):225–237 (2004).
4. N. R. Kristensen, H. Madsen, and S. B. Jørgensen. A method for systematic improvement of stochastic grey-box models. *Comput. Chem. Eng.* **28**(8):1431–1449 (2004).
5. A. H. Jazwinski. *Stochastic Processes and Filtering Theory*, Academic Press, New York, USA, 1970.
6. B. Øksendal. *Stochastic Differential Equations—An Introduction with Applications*, 5 edn., Springer-Verlag, Berlin, Germany, 1998.
7. P. E. Kloeden and E. Platen. *Numerical Solution of Stochastic Differential Equations*, Springer–Verlag, Berlin, Germany, 1992.
8. N. R. Kristensen and H. Madsen. *Continuous Time Stochastic Modelling-CTSM 2.3–Mathematics Guide*, Technical University of Denmark, Lyngby, Denmark, 2003.
9. N. R. Kristensen and H. Madsen. *Continuous Time Stochastic Modelling-CTSM 2.3–User's Guide*, Technical University of Denmark, Lyngby, Denmark, 2003.
10. S. Kotz and N. L. Johnson. *Encyclopedia of Statistical Sciences*, 5 vol. Wiley, New York, USA, 1985.
11. A Gelb. *Applied Optimal Estimation*, The MIT Press, Cambridge, USA, 1974.
12. G. De Nicolao, G. Sparacino, and C. Cobelli. Nonparametric input estimation in physiological systems: Problems, methods, and case studies. *Automatica* **33**(5):851–870 (1997).
13. G. Sparacino, G. Pillonetto, M. Capello, G. De Nicolao, and C. Cobelli. WINSTODEC: A stochastic deconvolution interactive program for physiological and pharmacokinetic systems. *Comput. Meth. Programs Biomed.* **67**:67–77 (2001).
14. T. J. Hastie, R. J. Tibshirani, and J. Friedman. *The Elements of Statistical Learning–Data Mining, Inference and Prediction*, Springer-Verlag, New York, USA, 2001.
15. T. J. Hastie and R. J. Tibshirani. *Generalized Additive Models*, Chapman & Hall, London, England, 1990.

16. Pharsight. *WinNonlin 4.0–Examples Guide*, Pharsight Corporation, Mountain View, USA, 2002.
17. Pharsight. *WinNonlin 4.0–Reference Guide*, Pharsight Corporation, Mountain View USA, 2002.
18. R. V. Overgaard, N. Jonsson, C. W. Tornøe, and H. Madsen. Nonlinear mixed-effects models with stochastic differential equations: Implementation of an estimation algorithm. *J. Pharmacokinet. Pharmacodyn.*, In Press, 2005.
19. C. W. Tornøe, R. V. Overgaard, H. Agersø, H. A. Nielsen, H. Madsen, and E. N. Jonsson. Stochastic differential equations in NONMEM: Implementation, application, and comparison with ordinary differential equations. *Pharm. Res.* 22: 1247–1258 (2005).
20. E. N. Jonsson and M. O. Karlsson. Nonlinearity detection: Advantages of nonlinear mixed-effects modelling. *AAPS Pharm. Sci.* **2**(3): (2000).