

THESIS FOR THE DEGREE OF LICENTIATE OF ENGINEERING

Modelling of drug-effect on time-varying biomarkers

Felix Held

Department of Mathematical Sciences
Division of Applied Mathematics and Statistics
Chalmers University of Technology and University of Gothenburg
Göteborg, Sweden 2018

Modelling of drug-effect on time-varying biomarkers

Felix Held

© Felix Held, 2018

Department of Mathematical Sciences
Division of Applied Mathematics and Statistics
Chalmers University of Technology and University of Gothenburg
SE-412 96 Göteborg
Sweden
Telephone: +46 (0)31-772 1000

Chalmers Reproservice
Göteborg, Sweden 2018

Modelling of drug-effect on time-varying biomarkers

Felix Held

Department of Mathematical Sciences
Division of Applied Mathematics and Statistics
Chalmers University of Technology and University of Gothenburg

Abstract

Model-based quantification of drug effect is an efficient tool during pre-clinical and clinical phases of drug trials. Mathematical modelling can lead to improved understanding of the underlying biological mechanisms, help in finding short-comings of experimental design and suggest improvements, or be an effective tool in simulation-based analyses. This thesis addresses the modelling of time-varying biomarkers both with and without drug-treatment. Pharmacokinetic/pharmacodynamic models were used to describe observed drug concentrations and biomarkers. These are modelled in the framework of compartmental modelling described by ordinary differential equations.

This thesis contains two papers in manuscript-form. In the first paper, a meta-analysis was performed of an existing model and previously published data for the stress-hormone cortisol and the drug dexamethasone. Cortisol exhibits a circadian rhythm, resembling oscillations, and is therefore a time-varying target for treatment. The aim was to utilize the model for prediction of the outcome of a medical test used in veterinary treatments on horses. In addition to model parameters, inter-individual variability was modelled and estimated in a Bayesian framework. This allowed simulation of test outcomes for the whole population, which in turn were used to evaluate available test protocols and suggest improvements.

In the second paper, an improved model was constructed for the cytokine TNF_α after challenge with LPS in addition to intervention treatment. TNF_α is not measurable in healthy subjects but release into blood plasma can be provoked by challenge with LPS. The result is a short-lived turnover of TNF_α . A test compound targeting intervention of TNF_α release was included in the study. Comprehensive experimental data from two studies was available and allowed to model features of TNF_α release, that were not addressed in previously published models. The final model was then used to analyse the current experimental design and correlations between LPS challenge and test compound effectiveness. The paper provides suggestions for future experimental designs.

Keywords: pharmacokinetic/pharmacodynamic modelling, turnover model, TNF_α , cortisol, hierarchical modelling

List of publications

This thesis consists of an extended summary and the following appended manuscripts

Paper 1: **Held F**, Ekstrand C, Cvijovic M, Gabrielsson J, Jirstrand M (2018) Modelling of oscillatory cortisol response using a Bayesian population approach for evaluation of dexamethasone suppression test protocols, *Submitted to Journal of Pharmacokinetics and Pharmacodynamics*

Paper 2: **Held F**, Hoppe E, Cvijovic M, Jirstrand M, Gabrielsson J (2018) Challenge model of TNF_α turnover at varying LPS and drug provocations, *Manuscript in preparation*

Additional publications not included in this thesis:

Loryan I, Hoppe E, Hansen K, **Held F**, Kless A, Linz K, Marossek V, Nolte B, Ratcliffe P, Saunders D, Terlinden R, Wegert A, Welbers A, Will O, and Hammarlund-Udenaes M (2017), Quantitative Assessment of Drug Delivery to Tissues and Association with Phospholipidosis: A Case Study with Two Structurally Related Diamines in Development. *Mol Pharmaceutics* 14(12):4362–4373

Author contributions

Paper 1: Refined and analysed the existing model, performed all computational work—including the implementation of the parameter estimation in Stan, deployment to the cluster and simulations—, derived approximate expressions for the simulation of sensitivity and specificity, created all figures, participated in the interpretation of the results, and drafted and edited the manuscript.

Paper 2: Participated in model creation, tested the model, performed all computational work—including the implementation of the parameter estimation in Monolix and simulations—, derived theoretical results about a connection between the integral of TNF_α time courses and administered challenge amount as well as drug dose, created all the figures, participated in the interpretation of the results, and drafted and edited the manuscript.

Acknowledgements

I want to thank my supervisor Mats Jirstrand for guidance and stimulating discussions throughout the course of this project, my co-supervisor Marija Cvijovic for warmly welcoming me to her research group, support in challenging periods and many interesting discussions about academia's inner workings, Johan Gabrielsson for sharing his domain knowledge, creating the opportunity for this project and his active participation, Edmund Hoppe from Grünenthal GmbH for his pharmacological domain knowledge, genuine interest in modelling and never-ending optimistic support, Carl Ekstrand for his veterinary domain knowledge and a successful collaboration. Especially, I want to thank Grünenthal GmbH for enabling this project financially.

Other people have been a big support during the last two years. I want to thank Joachim Almquist, Johannes Borgqvist, Barbara Schnitzer and Linnea Österberg for stimulating scientific discussions and Niek Welkenhuysen for pushing me through phases of procrastination.

I want to thank all other current or former colleagues at Fraunhofer-Chalmers Centre and Mathematical Sciences. A special shout-out to (in no particular order) Jacob Leander, Annikka Polster, Lucas Brynte, Emilio Jorge, Anna Lokrantz, Helga Kristín Ólafsdóttir, Mariana Buongiorno Pereira, Tim Cardilin, Patrick Reith, Sviatlana Shashkova, Anders Sjöberg, Aurélien Hontabat.

This last one goes out to friends and family, some close some far away. Distance may part us, but I am always glad to have you.

Göteborg, 2018

Contents

Abstract	i
List of publications	iii
Author contributions	iv
Acknowledgements	v
1 Introduction	1
1.1 Background	1
1.2 Problem formulation	2
1.3 Contributions	2
1.4 Thesis structure	2
2 Modelling	3
2.1 Pharmacokinetic/pharmacodynamic modelling	3
2.2 Compartmental modelling	3
2.3 Turnover models	4
2.4 Turnover models with a time-varying baseline	5
2.5 Challenge models	6
2.6 Inter-individual variability and hierarchical models	6
3 Parameter estimation	11
3.1 Preliminaries	11
3.2 Estimation in hierarchical models	13
3.3 Note on ODE models	13
3.4 Bayesian posterior distribution	14
3.5 Maximum likelihood estimates	14
4 Summary of papers	17
Paper 1: Modelling of oscillatory cortisol response using a Bayesian population approach for evaluation of dexamethasone suppression test protocols	17
Paper 2: Challenge model of TNF_α turnover at varying LPS and drug provocations	18
5 Outlook	19
References	21
Attached papers	25

1 | Introduction

1.1 Background

Mathematical modelling has become a standard tool in drug development, being routinely used during pre-clinical and clinical phases of drug discovery and development. In addition, the created models often accompany requests of approval to government agencies like the European Medicines Agency or the U.S. Food and Drug Administration. A model typically encompasses not just the drug itself but also one or multiple targeted biological markers, used to quantify drug effect. Some examples of these so-called biomarkers are protein-, cytokine-, or hormone concentration in blood plasma, activity of gene expression or blood pressure. While mathematical modelling has long been recognized to be of value in drug development, many challenges still arise due to limited availability of data and limited knowledge or high complexity of the underlying biology. In the current scientific landscape, especially in the field of machine learning, ever more complicated black-box models exist, which are capable of describing data in surprising detail. However, simpler models describing the key mechanisms are often preferred, favouring interpretability over accuracy.

Two different biological systems were addressed in this work. In Paper 1, the hormone cortisol that shows a circadian rhythm, i.e. an oscillating pattern with a 24 h period, and its response to the drug dexamethasone, aimed at reducing cortisol concentration, was explored. The oscillating pattern as well as sensitivity to stress and pulsatile production make cortisol a complex biological system. In Paper 2, the cytokine tumor necrosis factor alpha (TNF_α), a small protein important in cell-to-cell signalling in case of inflammation, was investigated. The complexity of TNF_α arises in that it is not measurable in blood plasma in the healthy body. An external challenge, leading to TNF_α release is therefore necessary to be able to observe TNF_α time courses and its response to test compounds, aimed at reducing TNF_α release during challenge. A common and important property of both of these biological systems is that the biomarker under study is time-varying, both with and without treatment.

Observed data can be used to calibrate a model and therefore parameter estimation stands in close relation to modelling. The data in this work came from pre-clinical studies collected on groups of test subjects, exposed to one or multiple dosing regimen. Since test subjects are typically selected to be similar, they can be considered to be from a common population. A hierarchical parameter framework was used in both papers, which adds an additional layer of parameters to describe the common population in relation to individual subjects. This allows to explicitly quantify variability between subjects, which is of importance when using the model for prediction purposes (Paper 1). Additionally, if the number of parameters is large compared to the amount of data available per individual, the parameter hierarchy is beneficial in regularising the parameter estimation process (Paper 1 and 2).

1.2 Problem formulation

A thematic question throughout both papers was how mathematical modelling can help to understand complex biological systems that are time-varying both with and without treatment. Therefore, the goal of the modelling phase was to focus on key features of the respective time-varying biomarker dynamics. Apart from this, a primary goal in Paper 1 was to use model prediction to answer questions concerning the design of a medical test protocol used during veterinary treatment of horses. In Paper 2, the main goal was the construction of an improved model to describe experimental data and to quantify response to a test compound. Secondary goals in both papers were to answer questions about drug-effect on the biomarker as well as to give suggestions for improvements of experimental design in future studies.

1.3 Contributions

The main contributions of Paper 1 are (1) the derivation of an explicit solution to the model equations in a special case, resulting in initial values and diagnostic plots; (2) successful application of a Bayesian parameter estimation workflow to implement the model in a Bayesian hierarchical framework, with the goal to capture uncertainty stemming from varying data quality, manifesting itself in parameter uncertainty; (3) illustration of a workflow to analyse the weaknesses (e.g. false positive rate and sensitivity to the oscillating baseline) of a medical test protocol through model predictions as well as suggestions for its improvement.

The main contributions of Paper 2 are (1) development of a new model for TNF_α after an external challenge including test compound intervention based on biological and engineering principles; (2) determination of key parameters from a drug development perspective and characterisation of test compound effect; (3) discovery of short-comings in experimental design through model predictions and suggestions for improvement, e.g. pointing out sparsely sampled periods or the benefits of crossover designs.

1.4 Thesis structure

The rest of this introductory part of the thesis (Chapters 1–3) introduces relevant topics concerning modelling and parameter estimation, which are necessary to fully appreciate the content of the two appended papers. This is followed by short summaries of each paper and a short outlook illustrating ideas for future research.

2 | Modelling

2.1 Pharmacokinetic/pharmacodynamic modelling

Data was modelled in the framework of pharmacokinetic/pharmacodynamic (PK/PD) modelling [1, 2]. PK/PD models are typically formulated as a set of time-dependent ordinary differential equations (ODEs) [3] and anchored in ideas from compartmental modelling [4] and chemical reaction equations [5]. Typically, PK/PD models are split into two parts. The PK model describes how drug concentration changes over time, therefore describing the effect of the body's physiology on the drug. The PD model on the other hand describes the dynamics of one or multiple biomarkers and how their response to drug treatment. These models are constructed to describe observed time-courses and to determine key quantities relating to the drug or drug response. An example of a quantity inherent to the drug is the clearance, the rate at which one unit of volume of drug is eliminated per unit of time. Potency, the concentration at which the drug achieves 50% of its effect, is an example of an important quantity when quantifying drug response. In both papers, the focus was on the PD model. The role of the PK model was to resemble the data empirically and drive the biological response through its effect on the PD model.

2.2 Compartmental modelling

Compartments can be thought of as containers/buckets containing either an amount or a concentration of a substance. A compartment is said to be well-mixed in the sense that incoming or outgoing substance does not lead to a heterogeneous distribution of substance in the compartment. Instead, contained substance instantaneously uniformly distributes itself in the compartment. A consequence of this property is that substance in a compartment is well-described by either its amount or volume and concentration. Compartments are often given a loose biological interpretation, e.g. a central compartment for concentration of drug in blood plasma and another compartment for drug in other tissues. This reflects that drug or any other substance of interest is often not only present in the blood stream, but also in other parts of the body.

The models considered in this work were all open systems, in the sense that there was an inflow, bringing in substance from outside the system, as well as outflow, permanently eliminating substance. Apart from external inflow and outflow, there are also two types of flow between compartments. The first is mass transfer of a substance from one compartment to another during which it may react with another substance. This type of flow respects mass balance. A typical example is the transport of drug from a gut compartment to a central compartment, representing concentration in blood plasma (Fig. 2.1). The second type is

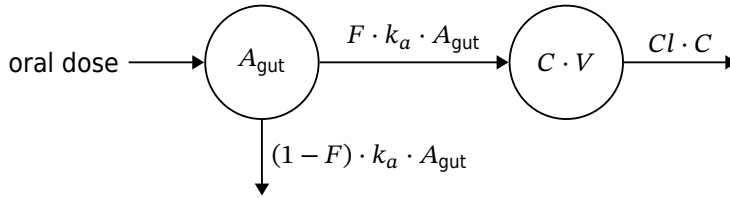


Fig. 2.1 A schematic representation of a compartment model describing the fate of a drug after administration of an oral dose. The drug dose arrives first in the gut compartment, where the latter contains the amount A_{gut} of drug. It is then transported with rate of absorption k_a and bioavailability F to a central blood plasma compartment. This compartment is described by volume V and drug concentration C , resulting in contained drug amount $V \cdot C$. Drug amount is finally eliminated from the body at a rate proportional to drug concentration C . The proportionality constant is called the clearance Cl .

control flows, leading to stimulation or inhibition of the production or elimination of one substance through another. Substance in the controlling compartment is not directly affected by its effect on the controlled substance.

To demonstrate how a simple compartmental PK model translates to ODEs, consider the model shown in Fig. 2.1, which is described by the following equations:

$$\begin{aligned} \frac{dA_{\text{gut}}}{dt} &= -k_a \cdot A_{\text{gut}}, & A_{\text{gut}}(0) &= \text{Dose} \\ V \cdot \frac{dC}{dt} &= F \cdot k_a \cdot A_{\text{gut}} - Cl \cdot C, & C(0) &= 0 \end{aligned} \quad (2.1)$$

where A_{gut} is the amount of drug in the gut compartment, C is the concentration of drug in the blood plasma compartment with volume V , k_a the rate of absorption, F the bioavailability of the drug, a number between 0 and 1 determining how much of drug in the gut actually arrives in the blood stream, and Cl the clearance of drug from the blood plasma compartment.

A more detailed introduction to compartmental modelling in the context of PK/PD modelling can be found in [1].

2.3 Turnover models

The turnover model, also called the indirect response model, is a commonly used pharmacodynamic model and is used to model physiological responses whose dynamics is governed by production as well as elimination processes [1, 6, 7]. The biomarker which is described by the model, in the following called R , is typically called the response, even if describing a situation without drug treatment. The physical dimensions of R , i.e. whether it is an amount, a concentration or something else, depend on the biomarker under investigation. In its simplest form, a turnover model can be thought of as a one-compartment model with an inflow described by a constant rate of production and an outflow proportional to the amount or concentration contained in the main compartment (Fig. 2.2). The model is described by the ODE

$$\frac{dR}{dt} = k_{\text{in}} - k_{\text{out}} \cdot R, \quad R(0) = R_0 \quad (2.2)$$

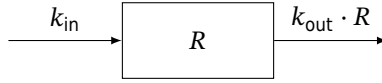


Fig. 2.2 Example of a turnover model with one compartment. The pharmacological response is represented by its concentration R . Additionally, k_{in} is the turnover rate and k_{out} the fractional turnover rate.

where R is the response, k_{in} is the turnover rate, describing the rate of production of response, and k_{out} is the fractional turnover rate, describing the rate of elimination proportional to R .

A useful property of this simple model is the existence of a steady state. After an initial transient phase, the response R will converge towards k_{in}/k_{out} , independently of its initial state. When modelling drug effect, it is often assumed that the test subject's response is at steady state before treatment. This is equivalent to choosing the initial state as

$$R_0 = \frac{k_{in}}{k_{out}}. \quad (2.3)$$

Drug effect on a turnover model is described as stimulation or inhibition of the production or elimination rate. In both papers presented in this work, the pharmacokinetics of the drug were assumed to affect the turnover rate. Stimulation and inhibition functions can be linear or non-linear functions of drug concentration and possibly show saturated behaviour for increasing drug concentration. Saturation-limited and therefore non-linear interactions have been used exclusively in this work. These can be modelled by Hill functions which have their origin in Michaelis-Menten enzyme kinetics and receptor binding [5]. Stimulation $S(C)$ and inhibition $I(C)$ dependent on drug concentration C are then described by the following equations:

$$S(C) = \frac{S_{max} \left(\frac{C}{SC_{50}} \right)^n}{1 + \left(\frac{C}{SC_{50}} \right)^n} \quad \text{and} \quad I(C) = 1 - \frac{I_{max} \left(\frac{C}{IC_{50}} \right)^n}{1 + \left(\frac{C}{IC_{50}} \right)^n} \quad (2.4)$$

where $0 < S_{max}$ is the maximal stimulatory rate, $0 < I_{max} \leq 1$ is the maximal inhibitory capacity, SC_{50} and IC_{50} are respectively the potency of drug for stimulatory and inhibitory interaction and n is a Hill exponent. Note that both S_{max} and I_{max} are typically unitless entities. As an example, in a model with stimulatory drug effect on response production Eq. 2.2 becomes

$$\frac{dR}{dt} = k_{in} \cdot S(C) - k_{out} \cdot R. \quad (2.5)$$

2.4 Turnover models with a time-varying baseline

It is possible to replace turnover rate and/or fractional turnover rate by time-dependent functions, as was done in Paper 1. There, the turnover rate was replaced by a cosine function with a vertical shift and a period of 24 h. This was done to model oscillating production relative to an average production rate, governed by a circadian rhythm. This was described by

the function

$$k_{\text{in}}(t) = k_{\text{avg}} \cdot (1 + \alpha \cdot \cos(\omega(t - t_0))), \quad \omega = \frac{2\pi}{24} \text{ h}^{-1} \quad (2.6)$$

where $0 < k_{\text{avg}}$ is the average turnover rate, $0 < \alpha < 1$ is the proportional size of the amplitude of the oscillations in relation to k_{avg} and t_0 is a phase shift. In this case, the turnover model loses its constant baseline but gains a time-dependent baseline. In this particular case, the baseline is oscillating, following a scaled version of the turnover rate. One challenge in connection with time-varying baselines is that drug effect becomes a function of administration time. This was exemplified in Paper 1 during the analysis of a medical test protocol. Additionally, determination of ODE initial values accommodating baseline oscillations and quantification of the time-varying baseline at different levels of drug exposure were addressed in Paper 1.

2.5 Challenge models

Another variant of the turnover model is considered in Paper 2. There, the considered physiological response did not show a baseline in a healthy subject. Rather, it was necessary to physically provoke a disease state by means of an external challenge. In terms of a turnover model this means that there is no endogenous production of the response, i.e. the turnover rate is zero without administration of the challenge. In the model considered in Paper 2, this was reflected by formulating the model as

$$\frac{dR}{dt} = S(t) - k_{\text{out}} \cdot R, \quad R(0) = 0 \quad (2.7)$$

where $S(t)$ is a time-dependent stimulation term, dependent on the external challenge and described by a separate part of the model. Response will start at zero, increase due to stimulation at challenge and then tend to zero again as stimulation disappears. This short-lived transient nature of the response makes these models different from those showing also a time-varying but cyclic baseline. As before, administration time, now in relation to time of challenge, is important during compound testing and can have a large influence on the compounds effectiveness. Another difficulty can be to separate the effect of challenge from the compound effect, as both are transient effects on the response. Both of these topics have been addresses in Paper 2.

2.6 Inter-individual variability and hierarchical models

A PK/PD model describes drug concentrations and as well as response in one test subject. Multiple subjects can be described by the same model, using a different set of parameters for each subject. In most cases, some or all parameters will show variability between subjects, so-called inter-individual variability (IIV). Variability is the result of differences between subjects in drug absorption, distribution and elimination as well as differences in baseline response and varying response to drug effect. In experimental studies, subjects are chosen carefully to be as similar as possible. However, differences in weight, metabolism and genetic variability [8, 9] can lead to noticeable differences in the above-mentioned physiological functions and therefore to IIV. In addition to this, other factors can have an influence, such as

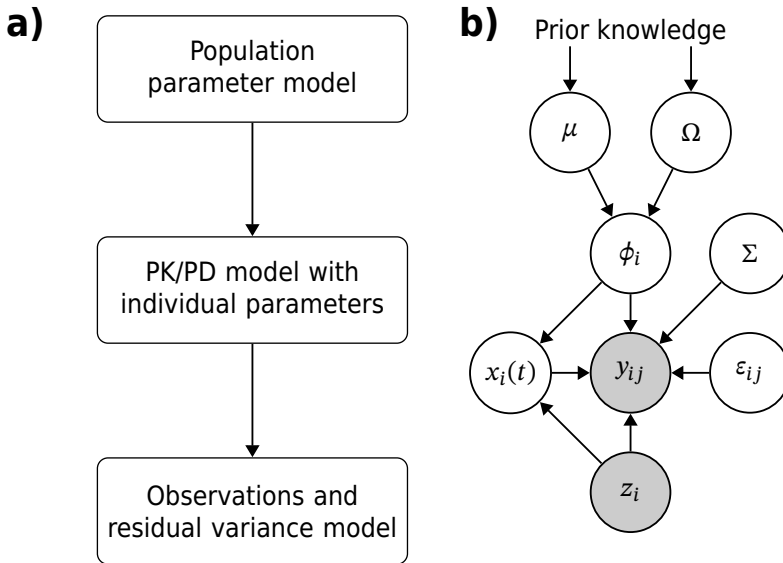


Fig. 2.3 a) Schematic representation of the hierarchy in a population PK/PD model. **b)** Graphical representation of all involved variables. This includes prior knowledge informing population parameters μ and Ω , and thereby the individual parameters ϕ_i . Model predicted time courses $x_i(t)$ are related to the individual parameters ϕ_i and covariates z_i through the ODE right hand side f (Eq. 2.8). Measurements are related to $x_i(t)$, z_i and ϕ_i through model function g (Eq. 2.9). Additionally, measurements are related to residual errors with parameters Σ through model function h (Eq. 2.9). Quantities in greyed out nodes denote observed quantities (z_i and y_{ij}). All other quantities need to be estimated.

environmental circumstances and subject handling during the experiment. In the present work IIV as a modelling tool was crucial in both papers. In Paper 1, populations of unobserved subjects could be simulated by assuming that variability in the observed subjects captured variability in the general population of subjects. This allowed the prediction of potentially observable ranges of concentrations and response, which in turn could be used for further analysis. This would not have been possible otherwise. In Paper 2, modelling with IIV was used as a tool to reduce model complexity during parameter estimation and for subsequent analysis of the estimated model. Only some parameters were estimated for all subjects, while others were assumed to be the same for all subjects. This reduced the amount of parameters to be estimated and led to improved accuracy.

IIV can be included into a PK/PD model by the use of hierarchical modelling [10]. In the field of drug modelling this is historically called non-linear mixed effects modelling [11]. The idea is to augment a PK/PD model by an additional layer capturing parameter variability. This is shown conceptually in Fig. 2.3 a. Experimental observations are described by a PK/PD model on a per subject basis, each with their own set of individual parameters. A residual variance model allows for measurement noise and other random influences which are not captured by the model mechanisms. The statistical distribution of individual parameters is then described

by an additional population parameter model.

To be able to give an unambiguous description of the parameter model and parameter estimation problem in the following chapter, the description above will now be formalised. For notational convenience, the time-derivative of a quantity will be indicated by a dot above the quantity. Assume in the following that there are N subjects, indicated by $i = 1, \dots, N$. For subject i there are M_i observations, indicated by $j = 1, \dots, M_i$. These observations are called y_{ij} and were collected at time t_{ij} . A graphical representation of all involved quantities and their dependencies is shown in Fig. 2.3 b.

Let $x_i(t)$ be a time-dependent vector-valued function describing observed and unobserved concentration- and response-time courses for subject i according to a PK/PD model. In this work, only PK/PD models described by ODEs were considered and it can therefore be assumed that

$$\dot{x}_i(t) = f(t, x_i(t); \phi_i, z_i), \quad x_i(0) = x_0(\phi_i, z_i), \quad (2.8)$$

where $f(t, x_i(t); \phi_i, z_i)$ describes the rate of change for each model component, ϕ_i is a d -dimensional parameter vector, z_i is a vector of covariates—such as dose, weight, age—and x_0 is the initial state of model components at the starting time of the experiment, which itself can depend on parameters or covariates.

The goal is to describe vector valued observations y_{ij} of drug concentration(s) and/or response(s) at time t_{ij} for subject i . Observations y_{ij} are described by model predictions and a residual error. This can be written as

$$y_{ij} = g(t_{ij}, x_i(t_{ij}); \phi_i, z_i) + h(t_{ij}, x_i(t_{ij}); \phi_i, z_i, \Sigma) \cdot \varepsilon_{ij}. \quad (2.9)$$

Here, $g(t_{ij}, x_i(t_{ij}); \phi_i, z_i)$ is a function describing the relation between all model components and the observed values. This could be as simple as picking one or multiple components of $x_i(t_{ij})$ or it could be more involved like a non-linear function of multiple components of $x_i(t_{ij})$ that cannot be observed separately. The residual error is described by ε_{ij} , a standard normally distributed vector, i.e. with zero mean and identity covariance matrix, scaled by $h(t_{ij}, x_i(t_{ij}); \phi_i, z_i, \Sigma)$ where Σ is an additional vector of parameters. This makes it possible to use, e.g. additive errors by setting $h \equiv \sigma$ or proportional errors with

$$h(t_{ij}, x_i(t_{ij}); \phi_i, z_i, \Sigma) = 1 + \sigma \cdot g(t_{ij}, x_i(t_{ij}); \phi_i, z_i) \quad (2.10)$$

where $\sigma > 0$ is a parameter in Σ . Different error models can be used for the different components of the vector y_{ij} .

Individual parameter vectors ϕ_i are described by an additional layer in the model hierarchy, the population parameter model. While restrictive, the following parameter model based on a normal distribution is in widespread use and sufficient for most use cases. Individual parameters can be bounded from one or both sides, e.g. most rate parameters are positive. Assume therefore that $\theta_i = F(\phi_i)$ transforms all individual parameters to the unbounded real axis. It is then assumed that

$$\theta_i \sim N(\mu, \Omega) \quad \text{for all } i, \quad (2.11)$$

where μ is the mean vector and Ω a covariance matrix, both of which are common to all subjects. Often, Ω is assumed to be diagonal, but correlations between individual parameters

can be included. Note that some components of the individual parameter vector ϕ_i might not be modelled to include IIV and are then excluded from the distributional assumption in Eq. 2.11. This leads to these particular components being common to all subjects.

As an example, a parameter ϕ bounded between 0 and 1 can be transformed to an unbounded variable via a logit transform

$$\theta = \log\left(\frac{\phi}{1-\phi}\right). \quad (2.12)$$

Conversely, θ can be transformed back to a bounded variable with the logistic transform

$$\phi = \frac{1}{1 + \exp(-\theta)}. \quad (2.13)$$

Assuming a normal distribution on θ is the same as assuming a logit-normal distribution on ϕ .

Parameters in hierarchical PK/PD models are typically unknown and need to be estimated. Methods for parameter estimation will be addressed in the next chapter.

3 | Parameter estimation

This chapter starts with a short recapitulation of the main ideas behind maximum likelihood estimation with missing data and sampling of the Bayesian posterior distribution. Thereafter, the general parameter estimation problem and reviews of the estimation approaches used in the papers are presented. In Paper 1, a Bayesian posterior distribution of the parameters was estimated, whereas in Paper 2, maximum likelihood estimates were estimated. In the following, the probability density of a variable x will be denoted by $p(x)$. The conditional density of one variable x dependent on another y is denoted by $p(x | y)$. The terms *distribution* and *density* will be used inter-changeably, since there is no confusion in this context. Additionally, a Bayesian mindset [10] is adopted and all parameters as well as data are initially assumed to be random. When quantities like the vector of individual parameters ϕ_i appear without an index, then ϕ is the collection of all vectors ϕ_i .

3.1 Preliminaries

In the following, assume that y is some observed quantity and θ a parameter used in a model describing y . Using Bayes rule [10], the joint distribution of y and θ can then be written as

$$p(y, \theta) = p(y | \theta)p(\theta). \quad (3.1)$$

Here, $p(y | \theta)$ is the likelihood and captures the modelled dependency between θ and y . The distribution $p(\theta)$ is called the prior distribution of θ , capturing prior beliefs about the parameter.

3.1.1 Maximum likelihood estimation

In the context of maximum likelihood estimation, the prior distribution is disregarded or, equivalently, all possible values of θ are considered to be *a priori* equally likely. After observation of the data y , the likelihood is considered as a function of θ , i.e. $L(\theta) = p(y | \theta)$. The maximum likelihood estimate (MLE) is the parameter value that maximizes the likelihood [12].

3.1.2 Expectation-Maximization

Assume now that there is a second parameter ϕ , which is not directly of interest. As an example, in case of a hierarchical model, θ could be a population parameter and ϕ a vector of individual parameters. The individual parameters are necessary to describe the model, but their estimates might not be of interest. One possible algorithm to calculate the MLE of θ in presence of such missing data or such nuisance parameters is Expectation-Maximization (EM)

[12]. The idea behind the algorithm is to alternate between two different types of likelihoods. $p(y, \phi | \theta)$ is called the complete data likelihood and

$$p(y | \theta) = \int p(y, \phi | \theta) d\phi \quad (3.2)$$

the marginal likelihood. EM alternates between the following steps:

1. Expectation step: Calculation of

$$Q_n(\theta) = \mathbb{E}_\phi [\log p(y, \phi | \theta) | y, \theta_{k-1}] \quad (3.3)$$

2. Maximization step: Calculation of

$$\theta_n = \underset{\theta}{\operatorname{argmax}} Q_n(\theta) \quad (3.4)$$

The feasibility of the algorithm is based on whether or not the expectation in Step 1 can be calculated. Step 2 is typically solved by a numerical optimization routine and therefore usually straight-forward. Convergence of this algorithm to a local maximum of the likelihood has been proved [13].

3.1.3 Sampling from the Bayesian posterior distribution

In a Bayesian context, the goal is the estimation of the posterior distribution of θ after the data y has been observed. Using Bayes rule, the posterior distribution is proportional to

$$p(\theta | y) \propto p(y | \theta)p(\theta). \quad (3.5)$$

In many cases, it is not possible to calculate the posterior distribution in closed form. It is therefore necessary to find a numerical approximation, for which knowledge of the proportional terms, as in Eq. 3.5, is sufficient. Popular estimation algorithms (Metropolis-Hastings [14], Metropolis-adjusted Langevin algorithm [15] or Hamiltonian Monte Carlo [16]) all follow the same general scheme [17].

A parameter θ_0 is given as a starting value.

1. A new parameter value θ' is proposed based on the previous parameter θ_{n-1} , following a proposal distribution $q(\theta' | \theta_{n-1})$.
2. The acceptance probability is calculated as

$$\alpha = \min \left(1, \frac{p(\theta' | y)}{p(\theta_{n-1} | y)} \cdot \frac{q(\theta_{n-1} | \theta')}{q(\theta' | \theta_{n-1})} \right) \quad (3.6)$$

3. Based on a random sample u from a uniform distribution on $[0, 1]$, the new value θ' is

- rejected if $u > \alpha$, i.e. $\theta_n = \theta_{n-1}$, or
- accepted if $u \leq \alpha$, i.e. $\theta_n = \theta'$.

The algorithms mentioned above differ in how they choose new proposals. Running this procedure for many iterations returns a set of parameters that can be used to approximate the posterior distribution of θ . As an example, the expected value of θ can be approximated with N samples by

$$\mathbb{E}[\theta] = \frac{1}{N} \sum_{n=1}^N \theta_n. \quad (3.7)$$

3.2 Estimation in hierarchical models

As a short re-capitulation, summarising Eqns. 2.8, 2.9, 2.11, shortening notation and omitting covariates, the full model for hierarchical PK/PD models is formulated as

$$\begin{aligned} \dot{x}_i &= f(t, x_i; \phi_i), & x_i(0) &= x_0(\phi_i), \\ y_{ij} &= g(t_{ij}, x_i(t_{ij})) + h(x_i(t_{ij}), \Sigma) \cdot \varepsilon_{ij}, \\ \theta_i &= F(\phi_i) \sim N(\mu, \Omega). \end{aligned} \quad (3.8)$$

The goal of the parameter estimation is to get estimates of all unobserved quantities μ , Ω , ϕ and Σ after the data and covariates have been observed.

Using Bayes rule, the joint distribution for all involved quantities is

$$p(y, \mu, \Omega, \Sigma, \phi) = \prod_i \left[\prod_j (p(y_{ij} | \phi_i, \Sigma)) p(\phi_i | \mu, \Omega) \right] p(\mu, \Omega, \Sigma). \quad (3.9)$$

Here, $p(y_{ij} | \phi_i, \Sigma)$ is called the likelihood of y_{ij} , $p(\phi_i | \mu, \Omega)$ is the distribution of the individual parameters, dependent on population parameters μ and Ω and $p(\mu, \Omega, \Sigma)$ is the prior distribution for μ , Ω and Σ , holding prior beliefs about these parameters. The data y will be observed during experimentation and by using Eq. 3.9, it is possible to gain information about unobserved parameters. However, not all unobserved parameters are of equal importance. Often, population parameters in μ and Ω are of more interest than individual parameters ϕ . Individual parameters can be marginalised out to focus on data and population parameters

$$p(y, \mu, \Omega, \Sigma) = \prod_i \left[\int \left(\prod_j p(y_{ij} | \phi_i, \Sigma) \right) p(\phi_i | \mu, \Omega) d\phi_i \right] p(\mu, \Omega, \Sigma). \quad (3.10)$$

A number of well-known parameter estimation algorithms for hierarchical PK/PD models ([18, 19, 20]) take advantage of this marginalisation, estimating individual parameters in hindsight or through computational tricks. One major reason for using Eq. 3.10 instead of Eq. 3.9 directly, is that the number of parameters can be reduced tremendously, if there is a large number of subjects involved. In this work, only few subjects per study were available, and parameter reduction was of lesser importance.

3.3 Note on ODE models

Most PK/PD models that are formulated as a system of ODEs cannot be solved analytically. It is therefore necessary to discretise the time variable t_k , and use numerical solvers to approximate

model predicted concentrations $x(t_k)$ by \hat{x}_k . The PK/PD models analysed in this work do not exhibit any chaotic behaviour and fulfil the assumptions of most ODE solvers for reasonable choices of the parameter vector ϕ , guaranteeing high accuracy of the approximate solution. It is therefore assumed that numerical approximation error, stemming from the ODE solver, is negligible for the purposes of parameter estimation. Note that this is not always the case for ODE models [21].

3.4 Bayesian posterior distribution

In Paper 1, a Bayesian posterior distribution was estimated for the unknown parameters. A major goal in that study was to use model predictions to gain an understanding of the variability of possible response-time courses. After initial model testing and based on results from a previous study, it was clear that the data showed fluctuations that were not captured by the model. Considerable uncertainty in parameter estimates was therefore to be expected. A Bayesian approach was chosen to incorporate parameter uncertainty into model predictions and to avoid bias or overconfidence in predictions. Additional advantages of Bayesian methods are regularisation and incorporation of prior knowledge through the prior distribution. In this study this was used to set soft boundaries on parameter estimates and use knowledge from a prior study. A disadvantage of Bayesian posterior sampling, compared to available maximum likelihood methods, was the runtime of sampling algorithms.

The posterior distribution for hierarchical PK/PD models can be determined from the joint parameter distribution in Eq. 3.9. Using Bayes rule, the posterior distribution is proportional to

$$p(\mu, \Omega, \phi, \Sigma | y) \propto \prod_i \left[\prod_j (p(y_{ij} | \phi_i, \Sigma)) p(\phi_i | \mu, \Omega) \right] p(\mu, \Omega, \Sigma) \quad (3.11)$$

Here, $p(y_{ij} | \phi_i, \Sigma)$ is the likelihood of the individual observations given individual parameters and residual error variance parameters. The likelihood is a function of ϕ_i and Σ and no longer a distribution, since y_{ij} is an observed value.

A numerical approximation of the posterior distribution was calculated using the No-U-Turn sampler (NUTS) [22], an extension to the Hamiltonian Monte Carlo (HMC) algorithm [16], implemented in the software Stan [23]. HMC uses gradient information of the likelihood with respect to the unobserved parameters as well as ideas from Hamiltonian mechanics to choose improved proposals during sampling. The NUTS extension automates adaptation of HMC's tuning parameters and potentially improves choice of proposals even further. HMC as implemented by Stan has been shown to be efficient for parameter estimation in hierarchical models [24].

3.5 Maximum likelihood estimates

For Paper 2, maximum likelihood estimates of the involved parameters were obtained. Here, the focus was on model building and quick iterations, for which maximum likelihood methods are well suited. Compared to the dataset in Paper 1, the available dataset was much cleaner

and substantially less parameter uncertainty was expected. When the MLEs are used for simulation from the model, then parameter uncertainty is dropped. Therefore, parameter uncertainty was considered more important than in the Bayesian case and efforts were made to keep it low. As a consequence, the number of variables was intentionally kept as small as possible and not all model parameters were considered to vary between individuals.

Parameters were estimated using the Stochastic Approximation Expectation-Maximization (SAEM) [20] algorithm, implemented in Monolix [25]. This algorithm uses the marginalised form of the estimation problem (Eq. 3.10), to obtain a likelihood

$$L(\mu, \Omega, \Sigma | y) = \prod_i \left[\int \left(\prod_j p(y_{ij} | \phi_i, \Sigma) \right) p(\phi_i | \mu, \Omega) d\phi_i \right]$$

for the population parameters. The expectation step cannot be solved analytically in this case, since $p(y_{ij} | \phi_i, \Sigma)$ depends on the solution of an ODE that depends non-linearly on its parameters. To find the MLEs for the population parameters the expectation step of the EM algorithm is replaced by the following two steps:

1. Simulation: Draw a small number of samples $\phi_n^{(k)}$, $k = 1, \dots, K$ from $p(\phi | y, \mu_{n-1}, \Omega_{n-1}, \Sigma_{n-1})$, using for example the Metropolis-Hastings algorithm
2. Stochastic approximation: Set

$$Q_n(\mu, \Omega, \Sigma) = Q_{n-1}(\mu, \Omega, \Sigma) + \gamma_n \left(\frac{1}{K} \sum_k \log p(y, \phi_n^{(k)} | \mu, \Omega, \Sigma) - Q_{n-1}(\mu, \Omega, \Sigma) \right)$$

This concludes the presentation of background material introduced in Chapters 1–3. The papers, which are summarised in the following chapter, build on the presented modelling and parameter estimation techniques, but go further by demonstrating the construction of a new model and by using a finished model for model predictions.

4 | Summary of papers

Paper 1: Modelling of oscillatory cortisol response using a Bayesian population approach for evaluation of dexamethasone suppression test protocols

In this paper, we present a meta-study of a previously published model and dataset [26]. The model describes time courses of the hormone cortisol as well as the drug dexamethasone and its effect on cortisol. The previous model was simplified as well as modified to ensure positivity of predicted cortisol concentrations. In addition, inter-individual variability (IIV) was included in the model. The main goal was to apply the improved model in the analysis of overnight dexamethasone suppression test (DST) protocols. The overnight DST is a medical test used as a tool in the diagnose of a degenerative disease in horses, based on a single measurement of cortisol concentration.

Challenges were posed by cortisol's oscillating circadian rhythm, a repeating 24 h pattern, its sensitivity to stress, which is particularly relevant during experimental handling, and cortisol's pulsatile production. Earlier investigations of the data made it clear that uncertainty in estimated model parameters was not insignificant and would potentially bias the resulting model prediction. To avoid this bias or overconfidence in predictions, it was necessary to propagate the parameter uncertainty through the model. To this end the parameter estimation problem was formulated in a Bayesian framework. The posterior distribution was numerically estimated using Hamiltonian Monte Carlo (HMC) in the software Stan and then used during model prediction.

Using ideas from Fourier analysis [27], an analytical solution to the PD model was obtained. It was used for the proper initialisation of the parameter estimation problem as well as to show the influence of variability on the interaction of steady-state dexamethasone exposure and the cortisol oscillation. Model simulations were used to investigate two DST protocols. In particular, how their false positive rates are influenced by protocol design in the context of IIV and cortisol oscillations. Additionally, the distributions of sensitivity and specificity were estimated for both DST protocols and compared to previously published experimental values.

Paper 2: Challenge model of TNF $_{\alpha}$ turnover at varying LPS and drug provocations

In this paper, we present a new model for tumor necrosis factor alpha (TNF $_{\alpha}$) concentration after challenge with lipopolysaccharides (LPS), additionally incorporating intervention of TNF $_{\alpha}$ release with a test compound. A difficulty in this study was that the test compound had to be applied to a time-varying disease model, only showing a short-lived effect after LPS administration. Data from two pre-clinical studies, that we received from Grünenthal GmbH, was used during the development and testing of the model. The data included LPS challenge at multiple levels as well as multiple levels of test compound intervention at a fixed challenge. Most previously published work on TNF $_{\alpha}$ after LPS challenge only considers one fixed dose of LPS challenge.

The model was constructed using mechanistic ideas about how LPS triggers TNF $_{\alpha}$ release and known drug mechanisms. Since data was sparse and the biology behind TNF $_{\alpha}$ is complex and not fully understood, engineering principles were used to fill in the gaps and build an empirical model based on observations made during an exploratory data analysis. An additional difficulty was that LPS time courses were not available. It was therefore necessary to make assumptions about LPS time course behaviour and connect these to observed TNF $_{\alpha}$ and drug concentration measurements.

IIV was modelled on key parameters. To determine these, a sensitivity analysis was done to show which parameters had the largest impact on the individual model. Additional knowledge from test-runs was used in cases where results from sensitivity analysis were ambiguous. Parameters were then estimated using the Stochastic Approximation Expectation-Maximization (SAEM) algorithm, implemented in the software Monolix.

The final model was used to scrutinize the current experimental design and to determine correlations between LPS challenge and the effectiveness of the test compound. Suggestions for future experimental designs were given. They include (1) increased sampling after LPS challenge, to improve capturing the peak location of TNF $_{\alpha}$ response; (2) using a crossover design including TNF $_{\alpha}$ challenge with and without drug intervention on the same subject, to be able to determine LPS challenge parameters separately from drug intervention parameters; and (3) sampling of systemic LPS exposure.

5 | Outlook

In the work presented here, it was shown that mathematical modelling can be fruitful when analysing already collected experimental data. Models can be built guided by observed features in the data combined with knowledge of the underlying biological mechanisms, as was done in Paper 2. However, some knowledge about the biomarker of interest is typically available, even in pre-clinical studies. This allows for the construction of a first version of a model before experiments are conducted. Analysing this model can lead to the discovery of e.g. time period during which large changes in biomarker time-courses are to be expected or which model parameters the model output is sensitive to. Using model-based experimental design can avoid some short-comings as discovered in Paper 2, e.g. sampling sparsely around the expected location of peak concentration, which itself was of interest in the study. This inverted approach is not new and used routinely in many studies, especially once a test compound goes forward into more advanced trial stages. However, it seems to be rather seldom used in a pre-clinical context.

Using Bayesian methods for parameter estimation in hierarchical PK/PD models is not new (e.g. [28]). However, even though there are clear benefits with using Bayesian methods, their use is still not widespread. Reasons for this might be that the theory behind Bayesian estimation is not as well-known as for maximum likelihood, but also due to long computational run-times. Even when available optimization techniques, such as within-chain parallelisation (the likelihood calculation is parallelised over subjects during the computation of a new sample from the posterior distribution), proper parameter tuning and the best possible model parametrisation were used, running times for the model presented in Paper 1 were still considerably longer than comparable runs during maximum likelihood estimation. The number of subjects in this study was small and an even greater run-time is to be expected for studies based on a larger population of test-subjects. This is of lesser impact when parameters are estimated for the final version of a model is run and reasonable outcomes can be expected. During model development however, estimation time is a severe limiting factor. Variational Bayesian (VB) methods [29] could help to improve on this issue. Similarly to the Expectation-Maximization algorithm, VB aims to approximate intractable integrals, such as those arising during marginalisation of the individual parameters (Eq. 3.10).

Fluctuations in cortisol baseline are not accounted for in the model in Paper 1. Stochastic differential equations (SDEs) could be used as a tool to include unmodelled physiological processes as system noise. This has been shown to regularise the parameter estimation [30] and might be able to give hints to the modeller, which parts of the model could be extended. Using SDE models as an extension to hierarchical PK/PD models is still relatively new in the field, posing challenges due to scarce availability of estimation algorithms and communication of results to others who are unfamiliar with the concept of SDEs. Some work in maximum likelihood estimation in this context has been done (e.g. [31]), whereas Bayesian

posterior distribution estimation has not been addressed so far. From a technical point of view, probability distributions need to be estimated, which describe the possible range(s) of SDE state variables. These estimation procedures often rely on Gaussian approximations, such as in the Extended Kalman filter (as in [31]). While these algorithms are fast, they might oversimplify the actual distribution and do not allow for multiple modes. To alleviate these issues it can be beneficial to use particle filters [32] in the context of particle MCMC methods [33] for state estimation. This has also previously been done in maximum likelihood estimation [34] but has not been attempted in a Bayesian setting (to our knowledge).

The main reason for the slow sampling in fully Bayesian methods seems to be two-fold. The complexity of hierarchical PK/PD models requires the usage of a sophisticated algorithm, such as Hamiltonian Monte Carlo. However, advanced algorithms require more evaluations of the likelihood, its gradient and potentially even higher derivatives to efficiently explore the parameter space. The necessity to solve ODEs during each of these steps creates a substantial computational load at each step. Some ideas have been proposed [35, 36, 37] suggesting to replace the numerical solution of the ODEs. One suggestion is to model the data with a Gaussian process [38] and then estimate ODE parameters by comparing the ODE right hand side with the derivative of the estimated Gaussian process [35]. However, so far no published method (to our knowledge) has outperformed Bayesian sampling with explicit numerical integration.

When the biological response is in focus, as was the case in both papers in this thesis, then PK concentration-time data is typically only modelled empirically by resembling the shape of the data, without much thought of the underlying physiological mechanisms—as would be the case in a physiologically based PK (PBPK) model [39]. This step could be replaced by an empirical, data-driven model based on a Gaussian process model. This would make it unnecessary to choose between different compartmental models. A choice which is often based on a purely statistical measure (e.g. Akaike’s information criterion) without relating to the underlying biology. One unsolved problem would be how inter-individual variability can be included when PK concentration time-courses are described by a Gaussian process.

In conclusion, this work provides contributions to hierarchical PK/PD modelling and model-based prediction as well as some smaller contributions regarding implementation of hierarchical PK/PD models in a Bayesian framework and mathematical model analysis.

References

1. Gabrielsson J, Weiner D (2016) *Pharmacokinetic & Pharmacodynamic Data Analysis: Concepts and Applications*, 5th edn. Swedish Pharmaceutical Press, Stockholm
2. Ette EI, Williams PJ (2007) *Pharmacometrics: The Science of Quantitative Pharmacology*. Wiley-Interscience, New York
3. Teschl G (2012) *Ordinary Differential Equations and Dynamical Systems*. American Mathematical Society
4. Anderson DH (1983) *Compartmental Modeling and Tracer Kinetics*. Springer-Verlag, Berlin Heidelberg
5. Atkins P, de Paula J, Keeler J (2017) *Atkins' Physical Chemistry*, 11th edn. Oxford University Press
6. Dayneka NL, Garg V, Jusko WJ (1993) Comparison of Four Basic Models of Indirect Pharmacodynamic Responses. *J Pharmacokinet Biopharm* 21:457–478
7. Sharma A, Jusko WJ (1998) Characteristics of indirect pharmacodynamic models and applications to clinical drug responses. *Br J Clin Pharmacol* 45:229–239
8. Wilkinson GR (2005) Drug Metabolism and Variability among Patients in Drug Response. *N Engl J Med* 352(21):2211–2221, DOI 10.1056/NEJMra032424
9. Roden DM, George Jr AL (2002) The genetic basis of variability in drug responses. *Nat Rev Drug Discov* 1:37–44, DOI 10.1038/nrd705
10. Gelman A, Carlin JB, Stern HS, Dunson DB, Vehtari A, Rubin DB (2013) *Bayesian Data Analysis*, 3rd edn. CRC Press, Boca Raton, Florida
11. Davidian M, Giltinan DM (1995) *Nonlinear Models for Repeated Measurement Data*. Chapman & Hall/CRC, Boca Raton, Florida
12. Casella G, Berger RL (2002) *Statistical Inference*, 2nd edn. Duxbury, Pacific Grove, CA
13. Wu CFJ (1983) On the Convergence Properties of the EM Algorithm. *Ann Stat* 11:95–103
14. Hastings WK (1970) Monte Carlo Sampling Methods Using Markov Chains and Their Applications. *Biometrika* 57:97–109, DOI 10.1093/biomet/57.1.97
15. Roberts GO, Rosenthal JS (1998) Optimal scaling of discrete approximations to Langevin diffusions. *J R Stat Soc Series B* 60:255–268, DOI 10.1111/1467-9868.00123
16. Neal RM (2011) MCMC Using Hamiltonian Dynamics. In: Brooks S, Gelman A, Jones GL, Meng XL (eds) *Handbook of Markov Chain Monte Carlo*, CRC Press, New York
17. Robert CP, Casella G (2004) *Monte Carlo Statistical Methods*, 2nd edn. Springer Science+Business Media, New York

18. Wang Y (2007) Derivation of various NONMEM estimation methods. *J Pharmacokinet Pharmacodyn* 34:575–593, DOI 10.1007/s10928-007-9060-6
19. Almquist J, Leander J, Jirstrand M (2015) Using sensitivity equations for computing gradients of the FOCE and FOCEI approximations to the population likelihood. *J Pharmacokinet Pharmacodyn* 42:191–209, DOI 10.1007/s10928-015-9409-1
20. Delyon B, Lavielle M, Moulines E (1999) Convergence of a stochastic approximation version of the EM algorithm. *Ann Stat* 27:94–128, DOI 10.1214/aos/1018031103
21. Chkrebtii OA, Campbell DA, Calderhead B, Girolami MA (2016) Bayesian Solution Uncertainty Quantification for Differential Equations. *Bayesian Anal* 11:1239–1267
22. Hoffman MD, Gelman A (2014) The No-U-Turn Sampler: Adaptively Setting Path Lengths in Hamiltonian Monte Carlo. *J Mach Learn Res* 15:1593–1623, URL <http://jmlr.org/papers/v15/hoffman14a.html>
23. Carpenter B, Gelman A, Hoffman MD, Lee D, Goodrich B, Betancourt M, Brubaker M, Guo J, Li P, Riddell A (2017) Stan: A Probabilistic Programming Language. *J Stat Softw* 76, DOI 10.18637/jss.v076.i01, URL <http://www.jstatsoft.org/v76/i01/>
24. Betancourt M, Girolami M (2013) Hamiltonian Monte Carlo for Hierarchical Models. URL <http://arxiv.org/abs/1312.0906>
25. Lixoft SAS (2018) Monolix version 2018R2. URL <http://lixoft.com/products/monolix/>
26. Ekstrand C, Ingvast-Larsson C, Olsen L, Hedeland M, Bondesson U, Gabrielsson J (2016) A quantitative approach to analysing cortisol response in the horse. *J Vet Pharmacol Ther* 39:255–263, DOI 10.1111/jvp.12276
27. Folland GB (2009) *Fourier Analysis and its Applications*. American Mathematical Society, Belmont, California
28. Lunn DJ, Best N, Thomas A, Wakefield J, Spiegelhalter D (2002) Bayesian analysis of population PK/PD models: general concepts and software. *J Pharmacokinet Pharmacodyn* 29:271–307
29. Beal MJ (2003) *Variational algorithms for approximate Bayesian inference*. PhD thesis, University College London
30. Leander J, Lundh T, Jirstrand M (2014) Stochastic differential equations as a tool to regularize the parameter estimation problem for continuous time dynamical systems given discrete time measurements. *Math Biosci* 251:54–62, DOI 10.1016/j.mbs.2014.03.001
31. Olafsdottir HK, Leander J, Almquist J, Jirstrand M (2018) Exact gradients improve parameter estimation in nonlinear mixed effects models with stochastic dynamics. *AAPS J* 20, DOI 10.1208/s12248-018-0232-7
32. Del Moral P (2004) *Feynman-Kac Formulae: Genealogical and Interacting Particle Systems with Applications*. Probability and its Applications, Springer, New York, NY
33. Schön TB, Svensson A, Murray L, Lindsten F (2018) Probabilistic learning of nonlinear dynamical systems using sequential monte carlo. *Mech Syst Signal Process* 104:866–883

34. Donnet S, Samson A (2011) EM algorithm coupled with particle filter for maximum likelihood parameter estimation of stochastic differential mixed-effects models. MAP5 2010-24, URL <https://hal.archives-ouvertes.fr/hal-00519576v2>
35. Calderhead B, Girolami M, Lawrence ND (2008) Accelerating Bayesian inference over nonlinear differential equations with Gaussian processes. *Advances in Neural Information Processing Systems* 21:217–224
36. Dondelinger F, Filippone M, Rogers S, Husmeier D (2013) ODE parameter inference using adaptive gradient matching with Gaussian processes. In: *Sixteenth International Conference on Artificial Intelligence and Statistics*
37. Wang Y, Barber D (2014) Gaussian Processes for Bayesian Estimation in Ordinary Differential Equations. In: *Proceedings of the 31st International Conference on Machine Learning*, vol 32
38. Rasmussen CE, Williams CKI (2006) *Gaussian Processes for Machine Learning*. MIT Press
39. Rowland M, Peck C, Tucker G (2011) Physiologically-based pharmacokinetics in drug development and regulatory science. *Annu Rev Pharmacol Toxicol*

