Approximate Repeated Administration Models for Pharmacometrics

Balazs Nemeth[1], Tom Haber[1,2], Jori Liesenborgs[1], and Wim Lamotte[1]

[1] Hasselt University - tUL - Expertise Center for Digital Media,
Wetenschapspark 2, 3590 Diepenbeek, Belgium
{balazs.nemeth,tom.haber,jori.liesenborgs,wim.lamotte}@uhasselt.be
[2] Exascience Lab, Imec, Kapeldreef 75, B-3001 Leuven, Belgium

Abstract. Improving performance through parallelization, while a common approach to reduce running-times in high-performance computing applications, is only part of the story. At some point, all available parallelism is exploited and performance improvements need to be sought elsewhere. As part of drug development trials, a compound is periodically administered, and the interactions between it and the human body are modeled through pharmacokinetics and pharmacodynamics by a set of ordinary differential equations. Numerical integration of these equations is the most computationally intensive part of the fitting process. For this task, parallelism brings little benefit. This paper describes how to exploit the nearly periodic nature of repeated administration models by numerical application of the method of averaging on the one hand and reusing previous computational effort on the other hand. The presented method can be applied on top of any existing integrator while requiring only a single tunable threshold parameter. Performance improvements and approximation error are studied on two pharmacometrics models. In addition, automated tuning of the threshold parameter is demonstrated in two scenarios. Up to 1.7-fold and 70-fold improvements are measured with the presented method for the two models respectively.

Keywords: Pharmacometrics · Monte Carlo Sampling · Hamiltonian Monte Carlo · High-Performance Computing · Hierarchical Models · Approximation · Importance Sampling

1 Introduction

One of the key questions of drug development, which pharmacometrics is concerned with, is what dosage regimen is safe and effective for individuals within a population. In this field, models from pharmacokinetics (PK) and pharmacodynamics (PD) characterize the interactions between a drug and an organism. Here, PK describes how a drug is affected by the organism, and PD describes the effect of the compound on the organism. The use of tools in this field requires both theoretical knowledge of biological systems and statistical expertise [14]. Therefore, methods that are easy to use, like the one described in this paper, are of great interest.

ICCS Camera Ready Version 2019
To cite this paper please use the final published version:
DOI: 10.1007/978-3-030-22734-0_46

Due to the complexity of these models, sufficient data is required to derive meaningful conclusions, but clinical data is typically sparse. Therefore, the common approach is to pool data from multiple drug trails and subjects within those trials. In this context, it is imprecise to merely consider the data as an unstructured collection of observations. Rather, with each observation, additional valuable information is available. This includes from which subject an observation is taken, his or her weight and height.

To incorporate this information, mixed effect models are used. Since PK and PD models typically rely on ordinary differential equations (ODEs), simulation requires computationally intensive numerical methods. An integrator is configured to ensure some level of accuracy in the result. Depending on the ODEs, the size of the steps that are taken is limited. More importantly, models with repeated administration hamper performance further. In these models, the simulation of dosing events causes the integrator to invalidate any gathered knowledge about the ODEs and take small steps. In addition, after a dosing event, computational time is spent on determining what step size to use.

Estimating parameters for these models in a reasonable amount of time requires not only the right mathematical tools, but also techniques from computer science. For example, within a drug trial, a compound is tested on multiple subjects and to determine the model parameter quality, each subject can be simulated in parallel. After parallelization, the most computationally intensive part is the numerical integration. Although parallel numerical integration has been studied [11], only limited improvements are possible [13].

Instead, the method outlined in this paper exploits the periodic behavior of models in pharmacometrics by reusing previous computations and employing the method of averaging to form an approximation of the model. It is applicable on top of any numerical integrator and besides a single parameter, no additional input from the user is required. To de-emphasize the existence of the parameter, it is important to note that it can be tuned automatically in a use-case dependent manner. Two examples are discussed to demonstrate this.

The remainder of this paper is structured as follows. Section 2 lists related work. Two examples of repeated administration models are discussed in Section 3. Section 4 discusses how these are used when data is sparse. The approximation method is presented in Section 5. Next, experimental results are shown in Section 6, and the paper is concluded and directions for future work are provided in Section 7.

2 Related Work

Dunne et al. [5] studied the application of the method of averaging in pharmacometrics, but their approach consisted of transforming the model by hand followed by solving it symbolically. The automated method presented in Section 5 partially relies on the same observations but differs in two ways. First, it does not require the user to manually alter the model. Second, for models that combine both PK and PD, all portions of the model are handled while the ap-

DOI: 10.1007/978-3-030-22734-0_46

proach outlined by Dunne et al. focuses mainly on dealing with the PD portion where no periodicity is observed.

Conrad et al. [3] tackle computationally expensive models by constructing and gradually refining approximations of the posterior for Bayesian inference during Markov Chain Monte Carlo (MCMC) sampling. Their approximation method uses previous evaluations in a shrinking region to interpolate the posterior function. Similarly, Gong et al. [6] propose an adaptive refinement strategy that builds a surrogate model to explore a target distribution. Compared to these approaches where no knowledge of the underlying model is used, the approximation described in this paper works at the level of the model itself. As such, the two approaches are complementary.

Rasmussen [15] considers Hybrid Monte Carlo (HMC) on Bayesian integrals. In his work, gradients of the posterior are approximated using a Gaussian Process. He notes that to guarantee that the samples generated by HMC are unbiased, accurate posterior evaluations are only required at the end of a set of leapfrog iterations. Similarly, in Section 6, gradients are computed from the approximation and the final accept-reject step relies on the real model.

3 Repeated Administration Models

This paper considers two models to exemplify what is seen in drug development when patients are administered a compound periodically. While the details of the models are less important for the work presented in this paper, they are listed here to describe their structure. Each model in this paper, denoted by f, is built using a set of ODEs parametrized by a vector ϕ. The set of q equations in f is denoted by $S = \{S_i(t)\}_1^q$.

Data to which these models are fit consists of a dosage regimen D and a sequence of observations (y_j, x_j). Each dosing event (a, c, t) in D adds some amount a of a compound to any state identified by c in model f at time t. Without loss of generality, the first dose is administered at $t = 0$, and all observations and dosing events are sorted by increasing time t. To fit ϕ, prediction $\hat{y}_j$ need only be made at x_j and Algorithm 1 outlines how to obtain predictions. It relies on a subroutine that implements an integrator of which the state is stored in $\mathcal{I}$.

The execution time of the integrator is mainly determined by the range spanned by x_j and the number of dosing events falling in that range since. Repeatedly stopping the integrator to simulate dosing events is the main cause for slowdown; as noted in Section 1, the integrator cannot take large steps when the internal state is changed. The method presented in Section 5 avoids this.

3.1 Nimotuzumab Model

The first model characterizes PK behavior of Nimotuzumab, a humanized monoclonal antibody mAb, in patients with advanced breast cancer [16]. The system of coupled differential equations in Equation (1) describes the dynamics of this model.

ICCS Camera Ready Version 2019
To cite this paper please use the final published version:
DOI: 10.1007/978-3-030-22734-0_46

Algorithm 1: Using an integrator to collect predictions $\hat{y}_j$.

```
Input: $x_1, \ldots, x_n$, $D$, and $S$
Result: $\hat{y}_1, \ldots, \hat{y}_n$
$k = 1$; $\mathcal{I} = $ InitializeIntegrator($S$)
$(a, c, t) = $ GetDose($D$, $k$)
for $j = 1, \ldots, n$ do
    while $t \leq x_j$ do
        IntegrateTo($\mathcal{I}$, $t$)
        AddToState($\mathcal{I}$, $c$, $a$)
        $k = k + 1$; $(a, c, t) = $ GetDose($D$, $k$)
    end
    IntegrateTo($\mathcal{I}$, $x_j$)
    $\hat{y}_j = $ GetState($\mathcal{I}$)
end
```

$$\begin{cases} \frac{\mathrm{d}C_{\mathrm{tot}}(t)}{\mathrm{d}t} & = -(k_{\mathrm{e}} + k_{\mathrm{pt}}) \cdot C(t) + k_{\mathrm{tp}} \cdot A_{\mathrm{t}}(t) - \left(\frac{k_{\mathrm{int}} \cdot R_{\mathrm{tot}} \cdot C(t)}{k_{\mathrm{ss}} + C(t)}\right) \\ \frac{\mathrm{d}A_{\mathrm{t}}(t)}{\mathrm{d}t} & = k_{\mathrm{pt}} \cdot C(t) \cdot v_1 - k_{\mathrm{tp}} \cdot A_{\mathrm{t}}(t) \\ \frac{\mathrm{d}R_{\mathrm{tot}}(t)}{\mathrm{d}t} & = k_{\mathrm{syn}} - k_{\mathrm{deg}} \cdot R_{\mathrm{tot}}(t) - \left(\frac{(k_{\mathrm{int}} - k_{\mathrm{deg}}) \cdot C(t) \cdot R_{\mathrm{tot}}(t)}{k_{\mathrm{ss}} + C(t)}\right) \\ C(t) & = 0.5 \cdot \Big[C_{\mathrm{tot}}(t) - R_{\mathrm{tot}}(t) - k_{\mathrm{ss}} \\ & \quad + \sqrt{(C_{\mathrm{tot}}(t) - R_{\mathrm{tot}}(t) - k_{\mathrm{ss}})^2 + 4 \cdot k_{\mathrm{ss}} \cdot C_{\mathrm{tot}}}(t)\Big] \end{cases} \tag{1}$$

Observations to which this model is fit consist of measured free concentrations of the mAb compound $C(t)$, at a particular time t, determined by the total mAb concentrations $C_{\mathrm{tot}}(t)$, the total target concentration $R_{\mathrm{tot}}(t)$ and the steady state rate constant k_{ss}. The change in the amount of free mAb in tissue compartments $A(t)$ depends on $C(t)$ and k_{pt} and k_{tp} which denote tissue-serum and serum-tissue rate constants respectively. The other constants that need to be estimated are the elimination rate k_{el}, the degradation rate k_{deg}, zero-order kinetic synthesis k_{syn} and irreversible internalization rate k_{int}. Note that there is a bidirectional influence between the compartments and $C(t)$ since it also appears on the right hand side. The model parameter vector ϕ is $[cl, v_1, Q, v_2, k_{\mathrm{ss}}, k_{\mathrm{int}}, k_{\mathrm{syn}}, k_{\mathrm{deg}}]$, where $k_{\mathrm{e}} = cl/v_1$, $k_{\mathrm{pt}} = Q/v_1$ and $k_{\mathrm{tp}} = Q/v_2$.

Figure 1 shows an example of the evolution of ODE states in time for the Nimotuzumab model from Equation (1) with parameters $cl = 9.93 \times 10^{-4}$, $v_1 = 1.38$, $Q = 4.00 \times 10^{-3}$, $v_2 = 44$, $k_{\mathrm{ss}} = 12.71$, $k_{\mathrm{int}} = 3$, $k_{\mathrm{syn}} = 1$ and $k_{\mathrm{deg}} = 7$. There are ten dosing events, each adding 50 milliliters intravenously. Programmatically, this is done by adding the same amount to $C_{\mathrm{tot}}(t)$ at each dosing event. During the first few dosing intervals, the concentration of the compound increases until the rate at which it is eliminated balances the rate at which the compound is added to the system. While $A_{\mathrm{t}}(t)$ increases perpetually due to the bidirectional interplay between it and the compartments, *nearly* periodic behavior is observed in $C_{\mathrm{tot}}(t)$ and $R_{\mathrm{tot}}(t)$. Note that measurements are also taken after the final dosing event as $C(t)$ drops.

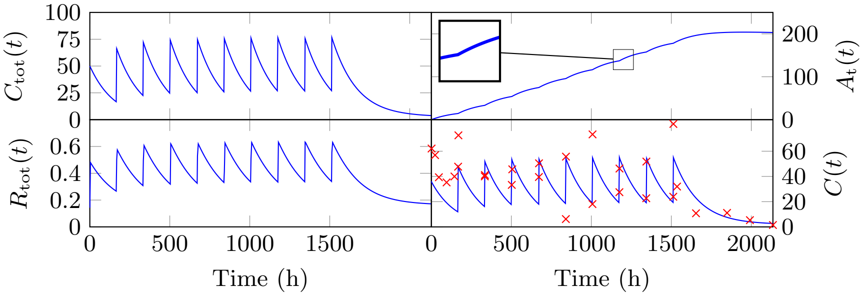

Fig. 1. The ODE states from the Nimotuzumab three-compartment model with ten dosing events. The state for $C_{\mathrm{tot}}(t)$, $A_{\mathrm{t}}(t)$ and $R_{\mathrm{tot}}(t)$ in function of time is shown on the left and top right, and the projected value $C(t)$ with observations shown as red crosses on the bottom right. After the first few dosing events, $C_{\mathrm{tot}}(t)$ and $R_{\mathrm{tot}}(t)$ exhibit close to periodic behavior. The plots were created by supplying a dense sequence of time points for x_j to Algorithm 1. The inset on $A_{\mathrm{t}}(t)$ is discussed in Section 5.

3.2 Canagliflozin Model

Canagliflozin is a drug for type-2 diabetes treatment. The model in Equation (2) for this drug consists of both a PK and a PD portion. The former is modelled by a two-compartment model [9] denoted by the gut compartment $A_G(t)$, the central compartment $A_C(t)$ and the peripheral $A_P(t)$. Following Dunne et al. [4], the latter is captured by glycated haemoglobin (HbA1c) denoted by $H(t)$.

$$\begin{cases} \frac{\mathrm{d}A_{\mathrm{G}}(t)}{\mathrm{d}t} &= -k_{\mathrm{a}} \cdot A_{\mathrm{G}}(t) \\ \frac{\mathrm{d}A_{\mathrm{C}}(t)}{\mathrm{d}t} &= k_{\mathrm{a}} \cdot A_{\mathrm{G}}(t) - k_{23} \cdot A_{\mathrm{C}}(t) + k_{32} \cdot A_{\mathrm{P}}(t) - k_{\mathrm{e}} \cdot A_{\mathrm{C}}(t) \\ \frac{\mathrm{d}A_{\mathrm{P}}(t)}{\mathrm{d}t} &= k_{23} \cdot A_{\mathrm{C}}(t) - k_{32} \cdot A_{\mathrm{P}}(t) \\ \frac{\mathrm{d}H(t)}{\mathrm{d}t} &= k_{\mathrm{in}} + Ef - k_{\mathrm{out}} \cdot H(t) \\ C(t) &= A_{\mathrm{C}}(t)/v \\ Ef &= (Ef_{\mathrm{c}} + Ef_{\mathrm{p}})\frac{H(0)-5}{8-5} \\ Ef_{\mathrm{c}}(t) &= E_{\mathrm{max}}\frac{C(t)}{EC_{50}+C(t)} \end{cases} \tag{2}$$

For this model, $\phi = [k_{\mathrm{out}}, H(0), Ef_{\mathrm{p}}, EC_{50}, E_{\mathrm{max}}]$, where Ef_{p} represents the placebo effect, $k_{\mathrm{in}} = H(0) \cdot k_{\mathrm{out}}$, EC_{50} is the exposure that gives half-maximal effect and E_{max} is the maximal effect of the drug. The remaining parameters

ICCS Camera Ready Version 2019
To cite this paper please use the final published version:
DOI: 10.1007/978-3-030-22734-0_46

are fixed. A simulation with $k_{\text{out}} = 10.24 \times 10^{-4}$, $H(0) = 7.72$, $Ef_{\text{p}} = -0.482$, $EC_{50} = 60.34$ and $E_{\max} = -0.736$ is shown in Figure 2. The remaining parameters are $k_{\text{a}} = 3.86$, $k_{23} = 0.101$, $k_{32} = 0.0928$, $k_{\text{e}} = 0.174$ and $v = 92.2260$. Similarly to Nimotuzumab, periodic behavior is observed for the PK portion.

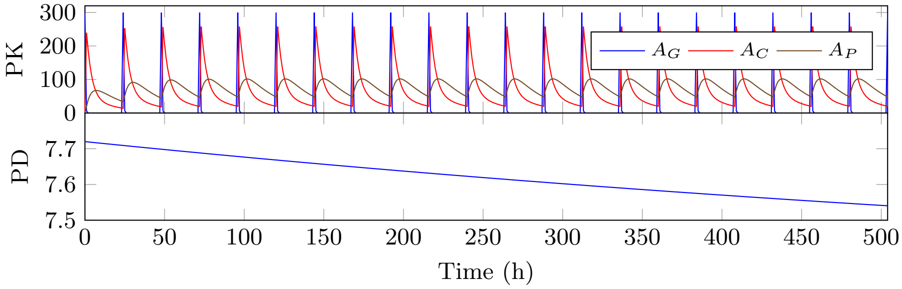

Fig. 2. Canagliflozin PK/PD model for the first 21 dosing events. Periodic behavior is observed after a few dosing events for the PK portion of the model shown at the top. The PD portion, shown at the bottom, does not stabilize.

4 Hierarchical Models

Pharmacometrics deals with models where the amount of available data is limited. Therefore, mixed effects models are used where data is grouped and structured into a hierarchy according to some classification [2]. The data considered in this paper is structured as shown in Equation (3).

$$y_{ij} = f(x_{ij}, \phi_i) + \epsilon_{ij}, \quad i = 1, \ldots, M, \quad j = 1, \ldots, n_i \tag{3}$$

A one-way classification is used resulting in a hierarchy with two layers. The first layer represents the population as a whole, and the second layer consists of individuals. The number of individuals is denoted by M, each of which has n_i observations. The function f, parameterized by ϕ, describes the structural model exemplified by those from Section 3. As these models capture PK or PD behavior or both, x_{ij} will be the j^{th} time point at which an observation was taken for the i^{th} individual. The residuals $\epsilon_{ij} \sim \mathcal{N}(0, \sigma)$ account for the intra-individual variance. With a slight abuse of notation, the individual parameters ϕ_i are given by $\mu + \eta_i$ where $\eta_i \sim \mathcal{N}(0, \Omega)$ and Ω captures inter-individual variance. Here, η_i and μ are called the random and fixed effect respectively. While Equation (3) only allows for additive error, its purpose is to be illustrative. It is important to note that the framework is general enough for other likelihood models as well. The goal is to estimate μ, Ω, and σ.

DOI: 10.1007/978-3-030-22734-0_46

5 Approximating Models

In a model, states are classified either as periodic or non-periodic. Typically, the PK portion is periodic and the PD portion is non-periodic, but this need not be the case. In the integrated states, three phases are distinguished. The first phase spans over all dosing events for which the system has not yet entered periodicity. The second phase is the periodic phase typically taking up the majority of time in repeated dosing models as noted in Section 3. The start of this phase is detected based on a threshold τ that defines when a state is classified as periodic. The final phase starts at the last dosing event and ends at the last observation. In Figure 1, depending on τ, the second phase could start at 500 hours.

The goal is to avoid stopping and altering the state of the integrator to simulate dosing events since this increases execution time substantially. During the first interval of the second phase, all periodic states for the remaining observations are collected. The value of all non-periodic states is collected during the full length of the second phase by applying the method of averaging numerically.

In clinical trials, it is common to have dosage regimens where all dosing events add the same amount of a compound in the same way, i.e. $a_i = a_j$ and $c_i = c_j$ for any pair of dosing events i and j in Algorithm 1. However, it is possible to generalize the presented method where multiple runs of periodic behavior are observed. Since the models targeted in this paper only use dosage regimens with a fixed dosing amount, such extensions are left as future work. As will be shown in Section 6, the efficacy of the presented method depends on the time spent in periodic phases.

In reality, doses will never be spaced *exactly* uniformly throughout time. For example, one of the individuals in the Nimotuzumab data set with 10 dosing events, has the last dose administered at 1512.2 hours after the start of the trial. The average dosing interval is thus approximately 168.02, but the dosing intervals for this individual are between 167.33 and 170.07. In case varying intervals are captured by the model, noise is added complicating periodicity detection. Therefore, a preprocessing step ensures that the events are spaced equally at the cost of potentially introducing some error in the final approximation.

If the mean time between doses is $\Delta t = t_{|D|}/(|D| - 1)$, then the time for dosing event k is set to $t'_k = (k-1) \cdot \Delta t$. Next, each observation j is shifted according to the offset to the dosing event before it. Concisely, x_j is shifted to $x'_j = t'_k + z_j$, where z_j is computed as follows. If t_k denotes the time of the dosing event before it, then $z_j = \min(x_j - t_k, \Delta t - \epsilon)$. Here, capping the offset at $\Delta t - \epsilon$ ensures that the observation is not shifted to the next interval when it is close to the end since doing so introduces a large error due to the rapid rise in compound concentration after a dose. Figure 3 illustrates this process for an exaggerated example; as for the Nimotuzumab example shown above, the variance in dosing intervals for real use cases is typically much smaller. For models in which the dosing intervals are fixed, like for the Canagliflozin model, data need not be preprocessed.

After preprocessing, integration can start. For any model f, three different sets of equations S, S' and $\tilde{S}$ are used. Here, S is the original unaltered set of

DOI: 10.1007/978-3-030-22734-0_46

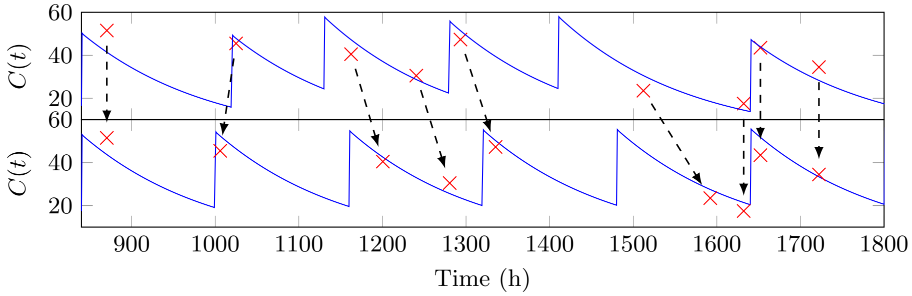

Fig. 3. Dosing events are shifted to ensure that each dosing interval is the same. All observations, shown as red crosses, associated with each dosing interval are shifted accordingly. The 7th observation is an example of an observation that, without capping, would be shifted to the next interval.

equations used during the first and third phase. During the first interval of the second phase, S' is used and the method of averaging is applied numerically during the remaining intervals in the second phase using $\tilde{S}$. The details of these sets of equations will be introduced next.

Integration commences on the set of equations $S = \{S_i(t)\}_1^q$ in f. At each dosing event k, all states in S are partitioned into r periodic states $P = \{P_i(t)\}_1^r$ and $q-r$ non-periodic states $N = \{N_i(t)\}_{r+1}^q$ by using some threshold τ and the criteria $|(S_i(t'_k) - S_i(t'_{k-1}))/S_i(t'_k)| < \tau$. If $|P| > 0$, the state of the integrator $\mathcal{I}_{\text{real}}$ is copied to $\mathcal{I}_{\text{approx}}$. At this time, denoted by t_α below, the second phase is entered and integration continues using $\mathcal{I}_{\text{approx}}$.

During the first interval of the second phase, integration continues with S', a set of equations constructed by adding the equations $\mathrm{d}P'_i(t)/\mathrm{d}t = P_i(t)$ to those in S for a total of $2|P| + |N|$ equations. The value of $P'_i(t_\alpha)$ is set to 0. These additional equations will be used to compute the average for use in the remaining intervals of the second phase. After one dosing interval, integration continues using $\tilde{S}$, constructed by taking the equations $\tilde{P} = \{\mathrm{d}\tilde{P}_i(t)/\mathrm{d}t = 0\}_1^r$ together with the states in N. The initial value for the states in $\tilde{P}$ is $P'_i(t_\alpha + \Delta t)/\Delta t$. In other words, the states in P are replaced by a constant equal to the mean value during a dosing interval. This is how the method of averaging is applied numerically. The values of the states in N are then collected during the second phase at each x'_j. Finally, at the last dose, integration continues using S restoring the state of the states in P to those saved in $\mathcal{I}_{\text{real}}$. The top left of Figure 4 demonstrates when each of these sets is used.

The states of P during the second phase are collected at times $t_\alpha + z_j$ for all observations j for which $x'_j > t_\alpha$. Note that if integration can only continue forward in time, all z_j need to be sorted. This can be seen as moving observations to the first interval of the second phase. Figure 4 shows the output for the Nimotuzumab model from Figure 1. Note that except for a different value of the integrated states, preprocessing and shifting of observations and events is not reflected in the output.

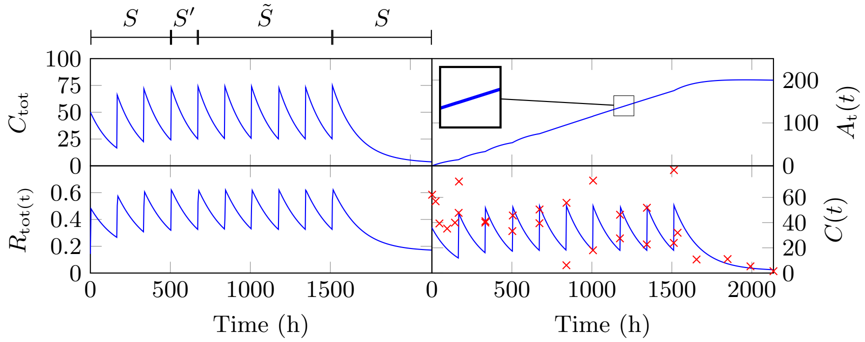

Fig. 4. Approximation of the Nimotuzumab three-compartment model with ten dosing events. Different sets of equations are used at different times. The sets are $S = P \cup N$, $S' = P' \cup N$ and $\tilde{S} = \tilde{P} \cup N$. These are only shown in the top left, but the change in equations effects all states. The choice for τ defines the phases. Here, the first phase spans $[0, 504]$, the second phase spans $[504, 1512.2]$ of which the first interval is $[504, 672]$ and the third phase starts at 1512.2. Compare all results with Figure 1 and note how $A_t(t)$ is smoothed out due to applying the method of averaging numerically. However, preprocessing and event shifting happens transparently. The effect of approximation on the other states is barely visible.

Let $c(t_0, t_1, S)$ denote the computational cost of using an integrator between time t_0 and t_1 on a set of equations S. The total cost of integration can be broken down into $c(0, t_\alpha, S)$, $c(t_\alpha, t_{\alpha+\Delta t}, S')$, $c(t_{\alpha+\Delta t}, t_{|D|}, \tilde{S})$ and $c(t_{|D|}, t_{n_i}, S)$. Since doses need not be simulated in $\tilde{S}$, $c(t_{\alpha+\Delta t}, t_{|D|}, \tilde{S}) \ll c(t_{\alpha+\Delta t}, t_{|D|}, S)$. Some overhead is introduced by preprocessing the data and using S' for one interval, but this is typically much smaller than the reduction in execution time obtained by avoiding simulation of doses between $t_{\alpha+\Delta t}$ and $t_{|D|}$.

Note that states in P are distinguished from those in N by τ. If τ is set too low, all states remain non-periodic and there is no second and third phase. In this case, no cost reduction will be made while some error will still be introduced by the preprocessing step. On the other hand, if all states are marked as periodic, then $c(t_\alpha + \Delta t, t_{|D|}, \tilde{S}) = 0$ since it can be skipped completely and larger cost reductions are expected. Note also that if all measurements after the last dose fall within a span of Δt, integration does not need to switch back to S from $\tilde{S}$.

A useful aspect of the outlined approach is that S' and $\tilde{S}$ can be constructed from S without symbolic manipulation. Integrator implementations require the user to provide a function that, given $S_i(t)$, returns a vector of which the i^{th} component represents $\mathrm{d}S_i(t)/\mathrm{d}t$. Multiplying this vector with the bit vector where all the components corresponding to states in N are set to 1 is a straightforward way to transform S into $\tilde{S}$.

DOI: 10.1007/978-3-030-22734-0_46

6 Performance Evaluation

Test data is taken from an online resource [17] for the Nimotuzumab model and is generated synthetically for the Canagliflozin model using the parameter estimates from Dunne et al.[4]. The Stochastic Approximation Expectation Maximization (SAEM) algorithm from Kuhn et al. [10] is used to fit a complete hierarchical model, described in Section 4. It is difficult to obtain a clear understanding of how well the presented approximation performs by comparing SAEM directly. Instead, the SAEM algorithm is run on the real model and the parameters at which the likelihood is evaluated are logged. The CVODE solver from the SUNDIALS software package [8] is used as the integrator implementation.

The evaluation time together with the log-likelihood value of the classical approach from Algorithm 1 is measured for the collected parameters. The same is measured for the approximate model with different choices for τ. Figure 5 illustrate the influence of τ on both the relative error of the log-likelihood and the speedup between the real and the approximate model. For $\tau = 0$, no speedup is expected since no states will be classified as periodic. Since doses are shifted for the Nimotuzumab model, some error is still introduced. This is not the case for the Canagliflozin model as it does not take into account varying dosing intervals. In both models, the slowdown with $\tau = 0$ is due to computing and sorting z_j, and the additional bookkeeping that is needed to compare the value of each state with τ. Note the difference in speedup between the two models. The Canagliflozin data contains individuals with a much larger number of dosing events than those in the data for the Nimotuzumab model.

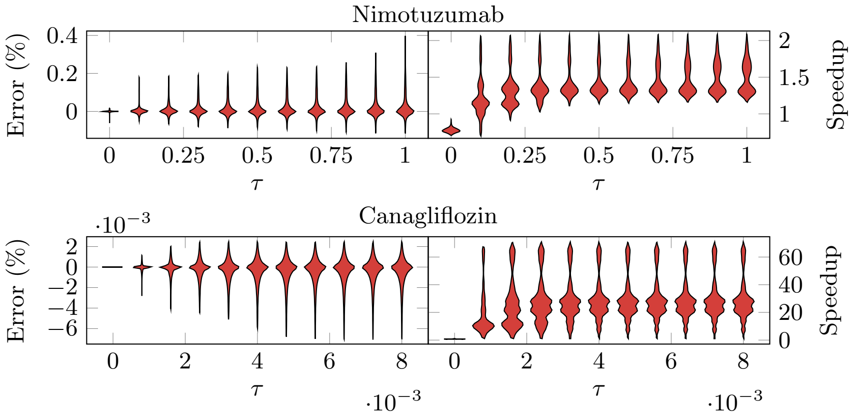

Fig. 5. Violin plots showing relative error and speedup as the threshold τ increases for the Nimotuzumab model at the top and for the Canagliflozin model at the bottom. A larger τ increases the probability of introducing a larger error. At the same time, a higher speedup factor is obtained. While both models show the same behavior as τ increases, there is a difference in scale of the error and τ due to a different number of dosing events in the data and structural differences between the models.

ICCS Camera Ready Version 2019
To cite this paper please use the final published version:
DOI: 10.1007/978-3-030-22734-0_46

Next, data is generated synthetically with an increasing number of doses to show that the total time spent by the integrator in the second phase determines the improvements that can be obtained by using the approximate model. In Figure 6, τ increases from 0 to 0.008, showing that with more dosing events, and hence more periodic behavior, a larger increase in performance is observed.

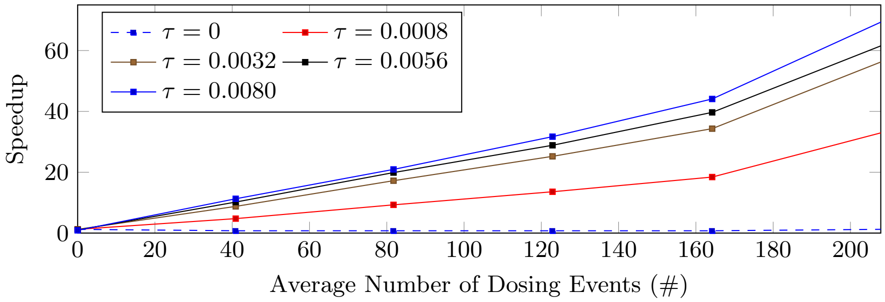

Fig. 6. Speedup for varying τ and varying number of observations for the Canagliflozin model. With more observations, the second phase makes up a larger fraction of the total execution time. Hence, there is a more opportunity to reduce execution time. Although not clearly visible, with $\tau = 0$, a slowdown of up to 25% is seen.

Recall from Section 4 that $\phi_i = \mu + \eta_i$. In algorithms like SAEM, one of the steps involves integrating out random effects η_i for a given individual. Due to the complexity of the models, MCMC samplers are used. Using the approximate model directly in this step results in biased estimates as the introduced errors change the distribution of random effects. As shown above, through the choice of τ, accuracy is sacrificed for performance. Two ways are discussed to use the approximation without introducing bias. A function that weights both the accuracy and the performance aspects is given for each. The same function can then be used to tune τ automatically. While tuning brings with it some computational costs, estimating parameters of hierarchical models takes orders of magnitude longer so it is worth spending some time on the tuning process. The objective is to find a sufficiently good value for τ and not necessarily the optimum. Therefore, tuning can be done on a subset of individuals.

One way to use the approximation is with HMC. Here, new positions are proposed by following the gradient L times and performing an accept-reject step at the final position. If gradients are computed from the approximate model and the accept-reject relies on the real model, the samples obtained remain unbiased [15]. Note that in scenarios where L is large, larger reductions in execution time are possible. Since the gradients are only approximate, proposals will be of lower quality. For example, if the real and the approximate gradients differ too much, the proposed positions will have low mass and many points will be rejected. In turn, this lowers the effective sample size (ESS), a metric used to evaluate the information content of dependent samples. Tuning τ is accomplished

DOI: 10.1007/978-3-030-22734-0_46

by maximizing ESS per unit time. Figure 7 shows this metric for Canagliflozin using $L = 4$ while varying τ. Clearly, the optimal value for τ depends on the choice of L.

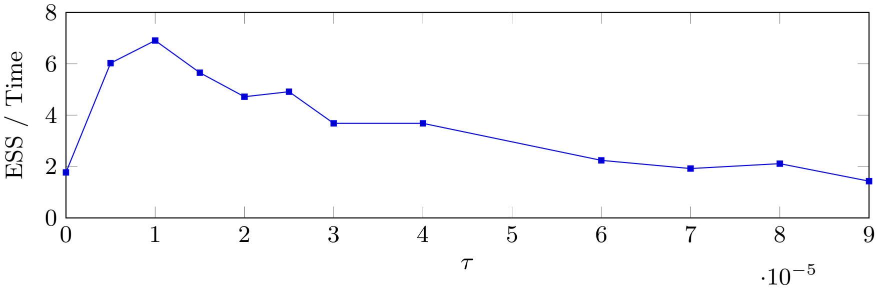

Fig. 7. Effective sample size per unit time while varying τ for the Canagliflozin model. This metric can be used to tune τ automatically.

As noted above, generating samples directly with any MCMC sampler from the random effects distribution built with the approximate model will introduce bias due to the errors. Another way to use the approximation is through importance sampling, where bias is corrected by weighting each sample [12]. These weights, obtained by taking the ratio between the density of the real and the approximate model, can be computed in parallel. If there is too much difference between the importance distribution and the target distribution, expectations computed from samples will exhibit more variance, denoted by σ_τ. An estimator $\hat{\sigma}_\tau$ is built by repeated sampling. A value for τ that trades off between computational efficiency and quality is chosen by minimizing $\hat{\sigma}_\tau$ while keeping time fixed. With multiple random effects, the covariance estimator $\hat{\Sigma}_\tau$ is used instead. Figure 8 shows this for the Nimotuzumab example. In this case, τ is tuned by minimizing $|\hat{\Sigma}_\tau|$, the determinant of the covariance matrix.

7 Conclusion and Future Work

This paper introduces an approximation of repeated administration models that exploits past computation efforts and employs the method of averaging numerically. In case of models with varying dosing intervals, a preprocessing step allows for detection of periodic behavior at the cost of adding some error to the approximation. The actual improvements vary depending on the model and the parameters of the model. On one of the test models, up to 70-fold reductions in run-time were measured while introducing only on the order of 10^{-3} relative error. Since fitting a hierarchical model can take up to hours or even days depending on the configuration parameters of algorithms like SAEM, these improvements have a tremendous impact on the end-users.

ICCS Camera Ready Version 2019
To cite this paper please use the final published version:
DOI: 10.1007/978-3-030-22734-0_46

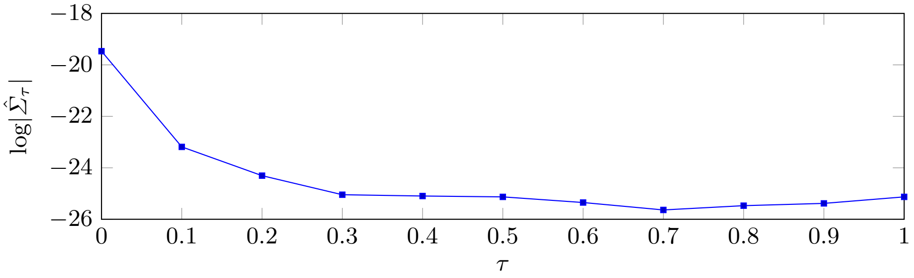

Fig. 8. The value of $\log|\hat{\Sigma}_\tau|$ in function of τ for the importance sampling estimator. By setting τ to 0.7, an appropriate trade-off between approximation accuracy and computational cost is made.

The approximation relies on setting the threshold τ to detect repetitive behavior in ODE states. It determines both the error and speedup of using the approximation instead of the real model. Incorporating a self-adjusting mechanism to automatically set τ for an MCMC sampler was discussed. Different objective functions can be devised depending on the use-case to tune τ, some of which will be studied in future work.

Speculative parallelism is a method to parallelize sequentially dependent tasks [7]. It has previously been applied to the classic Metropolis-Hastings MCMC sampler [1] where the sequence of accept-reject choices are guessed to predict the chain positions. Verification of these predictions then proceeds in parallel. A benefit of the speculative approach is that the collected samples are unaffected. Similarly, the approximation method presented in this paper can be applied to predict the chain, after which verification can occur in parallel. As in Section 6, it is again possible to tune τ. Here, τ trades off between the prediction accuracy and the time spent creating the prediction.

The choice of τ does not bound the error in the approximation. Tolerance bounds are typically already provided as parameters for numerical integration methods. Therefore, a promising direction of future work is to consider the change in integration results by entering the second phase one interval later.

8 Acknowledgments

The work presented in this paper was funded by Johnson & Johnson. The authors would like to thank Pieter Robyns and Nick Michiels for their valuable feedback on this work.

References

1. Elaine Angelino, Eddie Kohler, Amos Waterland, Margo Seltzer, and Ryan P Adams. Accelerating mcmc via parallel predictive prefetching. *arXiv preprint*

arXiv:1403.7265, 2014.

2. VJ Carey and You-Gan Wang. Mixed-effects models in s and s-plus, 2001.
3. Patrick R Conrad, Andrew D Davis, Youssef M Marzouk, Natesh S Pillai, and Aaron Smith. Parallel local approximation mcmc for expensive models. *SIAM/ASA Journal on Uncertainty Quantification*, 6(1):339–373, 2018.
4. Willem de Winter, Adrian Dunne, Xavier Woot de Trixhe, Damayanthi Devineni, Chyi-Hung Hsu, Jose Pinheiro, and David Polidori. Dynamic population pharmacokinetic–pharmacodynamic modelling and simulation supports similar efficacy in glycosylated haemoglobin response with once or twice-daily dosing of canagliflozin. *British journal of clinical pharmacology*, 83(5):1072–1081, 2017.
5. Adrian Dunne, Willem de Winter, Chyi-Hung Hsu, Shiferaw Mariam, Martine Neyens, José Pinheiro, and Xavier Woot de Trixhe. The method of averaging applied to pharmacokinetic/pharmacodynamic indirect response models. *Journal of pharmacokinetics and pharmacodynamics*, 42(4):417–426, 2015.
6. Wei Gong and Qingyun Duan. An adaptive surrogate modeling-based sampling strategy for parameter optimization and distribution estimation (asmo-pode). *Environmental Modelling & Software*, 95:61–75, 2017.
7. Ananth Grama, George Karypis, Vipin Kumar, and Anshul Gupta. *Introduction to Parallel Computing (2nd Edition)*. 01 2003.
8. Alan C Hindmarsh, Peter N Brown, Keith E Grant, Steven L Lee, Radu Serban, Dan E Shumaker, and Carol S Woodward. SUNDIALS: Suite of nonlinear and differential/algebraic equation solvers. *ACM Transactions on Mathematical Software (TOMS)*, 31(3):363–396, 2005.
9. Eef Hoeben, Willem De Winter, Martine Neyens, Damayanthi Devineni, An Vermeulen, and Adrian Dunne. Population pharmacokinetic modeling of canagliflozin in healthy volunteers and patients with type 2 diabetes mellitus. *Clinical pharmacokinetics*, 55(2):209–223, 2016.
10. Estelle Kuhn and Marc Lavielle. Coupling a stochastic approximation version of EM with an MCMC procedure. *ESAIM: Probability and Statistics*, 8:115–131, aug 2004.
11. Jacques-Louis Lions, Yvon Maday, and Gabriel Turinici. A "parareal" in time discretization of pde's. *Comptes Rendus de l'Académie des Sciences. Série I. Mathématique*, 332, 01 2001.
12. David JC MacKay and David JC Mac Kay. *Information theory, inference and learning algorithms*. Cambridge university press, 2003.
13. Michael Minion. A hybrid parareal spectral deferred corrections method. *Communications in Applied Mathematics and Computational Science*, 5(2):265–301, 2011.
14. Joel S Owen and Jill Fiedler-Kelly. *Introduction to population pharmacokinetic/pharmacodynamic analysis with nonlinear mixed effects models*. John Wiley & Sons, 2014.
15. Carl Edward Rasmussen. Gaussian processes to speed up hybrid monte carlo for expensive bayesian integrals. *Bayesian statistics*, 7:651–659, 2003.
16. Leyanis Rodríguez-Vera, Mayra Ramos-Suzarte, Eduardo Fernández-Sánchez, Jorge Luis Soriano, Concepción Peraire Guitart, Gilberto Castañeda Hernández, Carlos O Jacobo-Cabral, Niurys de Castro Suárez, and Helena Colom Codina. Semimechanistic model to characterize nonlinear pharmacokinetics of nimotuzumab in patients with advanced breast cancer. *The Journal of Clinical Pharmacology*, 55(8):888–898, 2015.
17. Mirjam N. Trame. Page 2018 nlmixr workshop materials. https://github.com/nlmixrdevelopment/PAGE-2018, 2018.