

# Supplement:

## Bayesian data assimilation to support informed decision-making in individualised chemotherapy

Corinna Maier<sup>1,2</sup>, Niklas Hartung<sup>1</sup>, Jana de Wiljes<sup>1,3</sup>,  
Charlotte Kloft<sup>4</sup>, and Wilhelm Huisinga<sup>1,\*</sup>

<sup>1</sup>Institute of Mathematics, University of Potsdam, Germany,

<sup>2</sup>Graduate Research Training Program PharMetriX: Pharmacometrics & Computational Disease Modelling, Freie Universität Berlin and University of Potsdam, Germany

<sup>3</sup>Department of Mathematics and Statistics, University of Reading, Whiteknights, UK

<sup>4</sup>Department of Clinical Pharmacy and Biochemistry, Institute of Pharmacy, Freie Universität Berlin, Germany

\*corresponding author (huisinga@uni-potsdam.de)

### Contents

<b>1</b>	<b>Introduction</b>	<b>2</b>
<b>2</b>	<b>Lognormal distribution</b>	<b>2</b>
<b>3</b>	<b>Maximum a-posteriori (MAP) estimation</b>	<b>2</b>
<b>4</b>	<b>Normal Approximation (NAP)</b>	<b>4</b>
4.1	Simulation-based approach . . . . .	5
4.2	Delta Method . . . . .	5
<b>5</b>	<b>Sampling Importance Resampling (SIR)</b>	<b>6</b>
<b>6</b>	<b>Markov chain Monte Carlo (MCMC)</b>	<b>8</b>
<b>7</b>	<b>Particle filter (PF)</b>	<b>8</b>
7.1	State augmentation . . . . .	9
7.2	Resampling strategies . . . . .	9
7.3	Rejuvenation . . . . .	9
7.4	Smoothing . . . . .	9
7.5	Alternative sequential DA algorithms . . . . .	9
<b>8</b>	<b>Simulation study: Single cycle Docetaxel</b>	<b>10</b>
<b>9</b>	<b>Simulation study: Multiple cycles Paclitaxel</b>	<b>13</b>

# 1 Introduction

This document should help to reproduce the simulation studies (along with the supplementary MATLAB code), provide some more explanations and details and should be self-contained as many parameters and models were taken from different publications.

## 2 Lognormal distribution

Since pharmacological parameters are typically positive, e.g., volume of distribution, baseline concentrations or rate constants, often a lognormal distribution is appropriate for modelling the inter-individual variability between patients, i.e.,  $\theta_i = \theta^{\text{TV}} \cdot e^{\eta_i}$  with  $\eta_i \sim \mathcal{N}(0, \Omega)$  for some individual  $i$ . The lognormal distribution results from a transformation of a normal random variable. If  $Y \sim \mathcal{N}(\mu, \Sigma)$ , then  $X = e^Y$  is lognormally distributed, i.e.  $X \sim \mathcal{LN}(\mu, \Sigma)$ . The probability density of the multivariate lognormal distribution is given by

$$p(x|\mu, \Sigma) = \frac{1}{(2\pi)^{d/2} |\Sigma|^{1/2}} \cdot \frac{1}{\prod_k x_k} \cdot e^{-\frac{1}{2}(\log x - \mu)^T \Sigma^{-1} (\log x - \mu)}, \quad (\text{S } 1)$$

where  $\mu \in \mathbb{R}^d$  and  $\Sigma \in \mathbb{R}^{d \times d}$  are the parameters of the associated normal distribution and  $\log$  denotes the natural logarithm. Thus, the prior for the individual parameters is given by

$$\begin{aligned} p_{\Theta}(\theta) &= \mathcal{LN}(\theta | \log(\theta^{\text{TV}}), \Omega) \\ &= \frac{1}{(2\pi)^{d/2} |\Omega|^{1/2}} \cdot \frac{1}{\prod_k \theta_k} \cdot e^{-\frac{1}{2}(\log(\theta) - \log(\theta^{\text{TV}}))^T \Omega^{-1} (\log(\theta) - \log(\theta^{\text{TV}}))}. \end{aligned}$$

## 3 Maximum a-posteriori (MAP) estimation

In MAP estimation, one seeks the parameter values that maximise the posterior probability

$$\hat{\theta}_n^{\text{MAP}} = \arg \max_{\theta} p(\theta | y_{1:n}).$$

It is, however, more convenient and numerically more stable to minimise the negative log-posterior instead

$$\begin{aligned} \hat{\theta}_n^{\text{MAP}} &= \arg \min_{\theta} -\log p(\theta | y_{1:n}) \\ &= \arg \min_{\theta} -\log p(y_{1:n} | \theta) - \log p(\theta). \end{aligned}$$

Choosing an additive normal residual error model ( $Y_j = h_j + \epsilon_j$  with  $\epsilon_j \sim_{\text{iid}} \mathcal{N}(0, \sigma^2)$ ) and a lognormal IIV model for the parameters ( $\theta_k = \theta_k^{\text{TV}} \cdot e^{\eta_k}$  with  $\eta_k \sim_{\text{iid}} \mathcal{N}(0, \omega_k^2)$ ) yields

$$\begin{aligned} \hat{\theta}_n^{\text{MAP}} &= \arg \min_{\theta} \frac{n}{2} \log(2\pi) + \frac{n}{2} \log \sigma^2 + \frac{1}{2} \sum_{j=1}^n \frac{(y_j - h_j(\theta))^2}{\sigma^2} \\ &\quad + \frac{d}{2} \log(2\pi) + \frac{1}{2} \sum_{k=1}^d \log \omega_k^2 + \sum_{k=1}^d \log \theta_k + \frac{1}{2} \sum_{k=1}^d \frac{(\log(\theta_k) - \log(\theta_k^{\text{TV}}))^2}{\omega_k^2} \end{aligned}$$

with data  $y_{1:n} = (y_1, \dots, y_n)^T$  observed up to time point  $t_n$ . MAP estimation was performed in Matlab R2017b using the interior-point algorithm in `fmincon`. The Matlab toolbox AMICI [1] was used for simulation of the system of ordinary differential equations (ODEs) and for computations of sensitivities used in gradient calculations (described below).

Gradient descent algorithms can often be improved by providing the gradient and the hessian of the objective function  $J(\theta) = -\log p(\theta | y_{1:n})$ . The gradient for this specific problem is given by

$$\begin{aligned} \frac{\partial J(\theta)}{\partial \theta_l} &= - \sum_{j=1}^n \frac{(y_j - h_j(\theta))}{\sigma^2} \cdot \frac{\partial h_j(\theta)}{\partial \theta_l} \\ &\quad + \frac{1}{\theta_l} + \frac{(\log(\theta_l) - \log(\theta_l^{\text{TV}}))}{\omega_l^2} \cdot \frac{1}{\theta_l}, \end{aligned}$$

and the hessian for  $l \neq m$

$$\frac{\partial^2 J(\theta)}{\partial \theta_l \partial \theta_m} = - \left( \sum_{j=1}^n \frac{(y_j - h_j(\theta))}{\sigma^2} \cdot \frac{\partial^2 h_j(\theta)}{\partial \theta_l \partial \theta_m} - \frac{1}{\sigma^2} \frac{\partial h_j(\theta)}{\partial \theta_l} \frac{\partial h_j(\theta)}{\partial \theta_m} \right)$$

and

$$\begin{aligned} \frac{\partial^2 J(\theta)}{\partial \theta_l^2} &= - \left( \sum_{j=1}^n \frac{(y_j - h_j(\theta))}{\sigma^2} \cdot \frac{\partial^2 h_j(\theta)}{\partial \theta_l^2} - \frac{1}{\sigma^2} \frac{\partial h_j(\theta)}{\partial \theta_l} \frac{\partial h_j(\theta)}{\partial \theta_l} \right) \\ &\quad + \frac{1}{\theta_l^2} \cdot \left[ \frac{1}{w_l^2} \left( 1 - \log(\theta_l) + \log(\theta_l^{\text{TV}}) \right) - 1 \right]. \end{aligned}$$

Note that  $\frac{\partial h_i(\theta)}{\partial \theta_l} = s_l^h$  are the output sensitivities, which are given by

$$s_l^h = \frac{\partial h(x, \theta)}{\partial x} s_l^x + \frac{\partial h(x, \theta)}{\partial \theta_l},$$

using the sensitivities of the states

$$\frac{\partial s_l^x}{\partial t} = \frac{\partial f(x, \theta)}{\partial x} s_l^x + \frac{\partial f(x, \theta)}{\partial \theta_l}, \quad s_l^x(0) = \frac{\partial x_0(\theta)}{\partial \theta_l}.$$

For the computation of the state sensitivities the extended system of ODEs needs to be solved

$$\begin{aligned} \dot{x} &= f(x, \theta), & x(0) &= x_0(\theta) \\ \dot{s}_l^x &= \frac{\partial f(x, \theta)}{\partial x} s_l^x + \frac{\partial f(x, \theta)}{\partial \theta_l}, & s_l^x(0) &= \frac{\partial x_0(\theta)}{\partial \theta_l}. \end{aligned}$$

Alternatively, the gradient could be computed via adjoint sensitivity analysis which is more efficient for models with large number of states and parameters.

Since the hessian matrix requires the computation of the second-order sensitivities  $\frac{\partial^2 h_i(\theta)}{\partial \theta_l \partial \theta_m}$ , which is computationally expensive, often the (expected) Fisher information matrix FIM is used as approximation

$$\mathcal{I}_{lm}(\theta) = - \sum_{i=1}^n \frac{1}{\sigma^2} \cdot \frac{\partial h_i(\theta)}{\partial \theta_l} \cdot \frac{\partial h_i(\theta)}{\partial \theta_m},$$

and

$$\mathcal{I}_{ll}(\theta) = - \left( \sum_{i=1}^n \frac{1}{\sigma^2} \cdot \frac{\partial h_i(\theta)}{\partial \theta_l} \cdot \frac{\partial h_i(\theta)}{\partial \theta_l} + \frac{1}{\omega_l^2 \theta_l^2} \cdot \left( \log(\theta_l^{\text{TV}}) - \log(\theta_l) + 1 \right) \right).$$

It was found that the MAP estimate does not correctly transform under a nonlinear mapping to the most probable observation/quantity of interest. As pharmacometric PK/PD models are often nonlinear, this is a major drawback for decision support in MIPD. Figure S 1 (A) shows the posterior of the drug effect parameter ‘Slope’ on the x-axis and the a-posteriori probability of the nadir concentration on the y-axis. The blue line shows the MAP estimate for the parameter ‘Slope’ and links to the MAP-predicted nadir concentration, which clearly does not correspond to the mode of the a-posteriori probability distribution of the nadir concentration. In addition it is shown how some randomly chosen samples transform to the nadir concentration. Note, however that the nadir concentration does of course not only depend on the ‘Slope’ but also on the other parameters. Figure S 1 (B) demonstrates the same observation for a different quantity of interest, the time of recovery to grade 2, which shows a bimodal posterior distribution. Further we considered as a statistical measure of accuracy the root mean squared error (RMSE) between the model-predicted outcome  $T_i(\mathbb{S}_n)$  given data  $y_{1:n}$  for individual  $i$  and the reference outcome  $T_i^{\text{ref}}$  (for which the data was simulated)

$$\text{RMSE}(T)_n = \sqrt{\frac{1}{N} \sum_{i=1}^N (T_i(\mathbb{S}_n) - T_i^{\text{ref}})^2}.$$

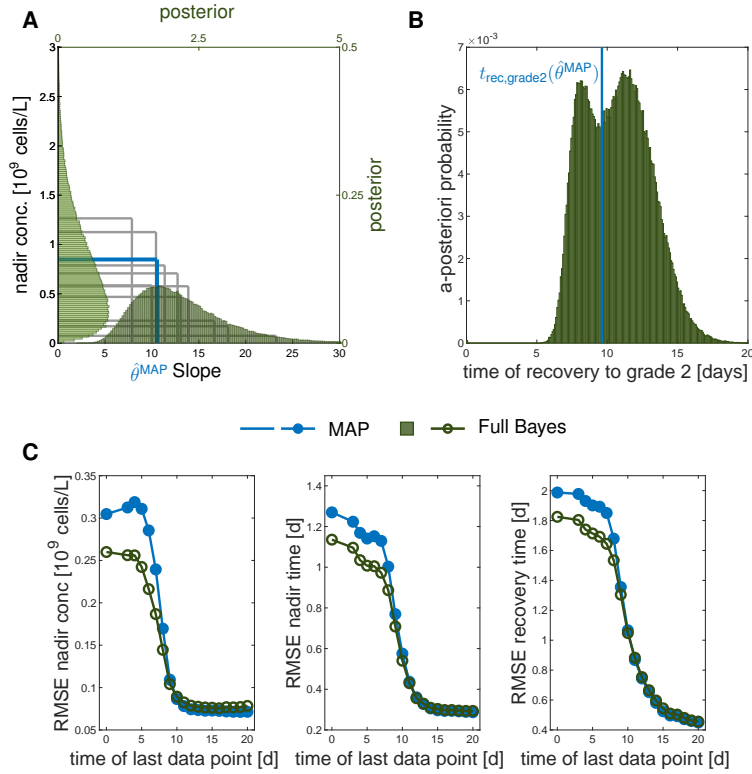


Figure S 1: **Illustration of the unfavourable properties of MAP estimation with regard to reliable decision support.** **A** The MAP estimate (blue line) does not correctly transform under a nonlinear mapping to the most probable nadir concentration (based on [2, Figure 5.2]). The posterior of the parameter 'Slope' is depicted for an exemplary patient after four data points  $y_{1:4}$  were observed on the x-axis and the corresponding a-posteriori probability of the nadir concentration on the y-axis (same scenario as in Figure 3). **B** The mode is not preserved under transformation. The same scenario was considered as for part A. **C** Root mean squared error (RMSE) of selected statistics. Comparison of the accuracy of the computed statistics  $c_{\text{nadir}}$ ,  $t_{\text{nadir}}$  and  $t_{\text{rec}0}$  based on MAP estimation and full Bayesian inference (SIR using  $S = 10^3$  samples). The RMSE was computed across the whole considered virtual population  $N = 100$ .

Figure S 1 (C) shows the prediction accuracy of the point-estimates of MAP-estimation and of the Full Bayesian approach over time. As more data points are taken into account the RMSE decreases for both estimators. In the beginning of the cycle the Full Bayesian approach shows increased accuracy across all considered quantities of interest.

## 4 Normal Approximation (NAP)

To quantify the uncertainty associated to a MAP estimate one may consider a quadratic approximation to the log posterior at its mode (following [3]). A Taylor expansion about the MAP estimate results in

$$\begin{aligned} \log p(\theta|y_{1:n}) &\approx \underbrace{\log p(\hat{\theta}^{\text{MAP}}|y_{1:n})}_{=\text{const.}} + (\theta - \hat{\theta}^{\text{MAP}})^T \underbrace{\left[ \frac{d \log p(\theta|y_{1:n})}{d\theta} \right]_{\theta=\hat{\theta}^{\text{MAP}}}}_{=0} \\ &\quad + \frac{1}{2} (\theta - \hat{\theta}^{\text{MAP}})^T \left[ \frac{d^2 \log p(\theta|y_{1:n})}{d\theta^2} \right]_{\theta=\hat{\theta}^{\text{MAP}}} (\theta - \hat{\theta}^{\text{MAP}}). \end{aligned}$$

If we take the exponential on both sides and normalise, we get as normal approximation to the posterior at the MAP estimate

$$p(\theta|y_{1:n}) \approx \mathcal{N}\left(\hat{\theta}^{\text{MAP}}, \left[ - \frac{d^2 \log p(\theta|y_{1:n})}{d\theta^2} \right]_{\theta=\hat{\theta}^{\text{MAP}}}^{-1} \right).$$

The inverse of the variance can be decomposed using Bayes' formula

$$-\frac{d^2}{d\theta^2} \log p(\theta|y_{1:n}) = -\frac{d^2}{d\theta^2} \log p(y_{1:n}|\theta) - \frac{d^2}{d\theta^2} \log p(\theta), \quad (\text{S } 2)$$

where we retrieve the total observed Fisher information matrix (FIM)

$$\mathcal{I}^{\text{likelihood}}(\theta) = -\frac{d^2}{d\theta^2} \log p(y_{1:n}|\theta).$$

If the definition for the observed FIM of the likelihood is transferred to prior and posterior, the previous decomposition Eq. S 2 can be written as

$$\mathcal{I}^{\text{post}}(\theta) = \mathcal{I}^{\text{likelihood}}(\theta) + \mathcal{I}^{\text{prior}}(\theta),$$

which allows to rewrite the normal approximation of the posterior as

$$p(\theta|y_{1:n}) \approx \mathcal{N}(\hat{\theta}^{\text{MAP}}, [\mathcal{I}^{\text{post}}(\hat{\theta}^{\text{MAP}})]^{-1}). \quad (\text{S } 3)$$

Note, that we defined  $\mathcal{I}(\theta) := \mathcal{I}^{\text{post}}(\theta)$  in the main article. Since the MAP estimator is asymptotically normally distributed in the limit of large sample sizes ( $n \rightarrow \infty$ ) (see [3, appendix B] for a proof), this approximation can be very precise in the case of highly informative data sets.

## 4.1 Simulation-based approach

To propagate the uncertainties from the parameters to the model predictions, we used a simulation-based approach. To this end, we sampled from the normal distribution Eq. S 3 and subsequently propagated each sample.

Step-by-step description of the algorithm

1. Generate posterior samples  $\theta^{(s)}$  from  $\mathcal{N}(\hat{\theta}^{\text{MAP}}, [\mathcal{I}^{\text{post}}(\hat{\theta}^{\text{MAP}})]^{-1})$ .
2. Compute for each sample  $h(t_j, \theta^{(s)})$  for time points of interest  $j = 1, \dots, m$
3. Compute quantiles  $(h_{\alpha/2}, h_{1-\alpha/2})$  for each time point  $t_j$

Similarly, we can compute samples of the quantities of interest  $T^{(s)} = T(\theta^{(s)})$ .

## 4.2 Delta Method

As an alternative to the simulation-based approach, we may use the delta method to determine the limiting distribution of a differentiable function of the parameters  $g(\theta)$  [4, section 5.5]. In our case  $g(\theta) = h(x(t), \theta)$ . Using the Delta method, the uncertainties are propagated from the parameters to the observable via the output sensitivities  $S^h(\hat{\theta}^{\text{MAP}}) = \nabla_{\theta} h_t(\theta)|_{\hat{\theta}^{\text{MAP}}}$  [4, section 5.5]:

$$p(h_t(\hat{\theta}^{\text{MAP}})|y_{1:n}) \approx \mathcal{N}\left(h(x(t), \hat{\theta}^{\text{MAP}}), S^h(\hat{\theta}^{\text{MAP}})^T \Sigma^{\text{MAP}} S^h(\hat{\theta}^{\text{MAP}})\right)$$

with  $\Sigma^{\text{MAP}} = [\mathcal{I}(\hat{\theta}^{\text{MAP}})]^{-1}$ . The credible interval (CI) for the prediction is then given by

$$\text{CI}^{\alpha} = h(x(t), \hat{\theta}^{\text{MAP}}) \pm z_{1-\alpha/2} \sqrt{\sigma_{\text{CI}}^2(t)},$$

where  $z_{1-\alpha/2}$  is the quantile of the standard normal distribution and  $\sigma_{\text{CI}}^2$  is computed using the output sensitivities

$$\sigma_{\text{CI}}^2(t) \approx S^h(t, \hat{\theta}^{\text{MAP}})^T \Sigma^{\text{MAP}} S^h(t, \hat{\theta}^{\text{MAP}}).$$

Alternatively, the quantiles of the Student's t distribution could be used as a more conservative choice, see [5]. For the determination of the prediction interval (PI), the residual variability is additionally taken into account

$$\sigma_{\text{PI}}^2(t) \approx S^h(t, \hat{\theta}^{\text{MAP}})^T \Sigma^{\text{MAP}} S^h(t, \hat{\theta}^{\text{MAP}}) + \sigma_{\text{RUV}}^2.$$

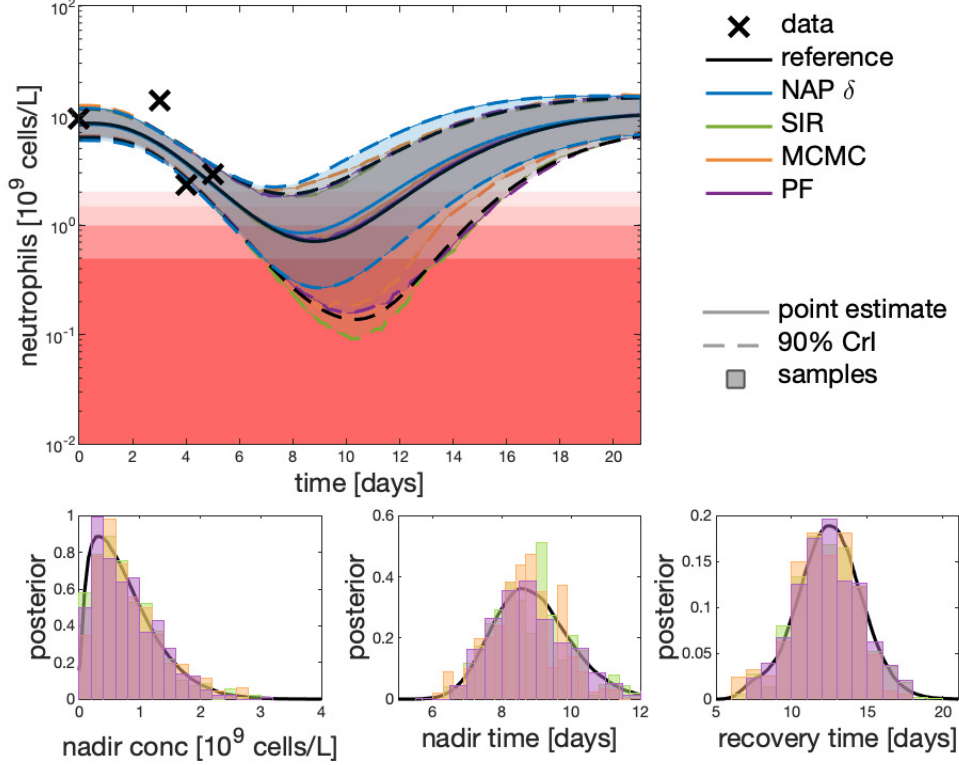


Figure S 2: Propagating uncertainties using the delta method instead of the simulation based approach.

Note that, in the main manuscript  $\sigma_{\text{RUV}}^2 = \sigma^2$ . Consequently, the prediction interval is given by

$$\text{PI}^\alpha = h(t, \hat{\theta}^{\text{MAP}}) \pm z_{1-\alpha/2} \sqrt{\sigma_{\text{PI}}^2(t)},$$

Since the delta method involves differentiation of  $g$ , it is not straightforward to apply the method to any desired quantity of interest, e.g.  $g(\theta) = T(\theta) = C_{\text{nadir}}(\theta)$ .

The Delta method leads to a similar underestimation of the uncertainty as the simulation-based approach (NAP sim), see Figure S 2. In addition it is not straightforward to propagate the uncertainty to quantities of interest (therefore not displayed).

One option to overcome the underestimation of the uncertainty is the use of quantiles of the Student's  $t$  distribution instead of normal quantiles [5]. We have used quantiles of the Student's  $t$  distribution with  $\nu = 4$  degrees of freedom (NAP  $\delta$  t). The credible intervals show an increased width, but now overestimate the uncertainties regarding subtherapeutic areas (grade 0), see Figure S 3. This is also not acceptable as underdosing is in oncology highly undesirable.

## 5 Sampling Importance Resampling (SIR)

Our goal is to generate samples from the posterior distribution at time point  $t_n$  having observed patient-specific data  $y_{1:n}$ . For this importance sampling is used in the SIR algorithm. Let  $\tilde{p}(\theta|y)$  be the unnormalised posterior

$$\tilde{p}(\theta|y_{1:n}) = p(y_{1:n}|\theta) \cdot p(\theta).$$

We use the prior  $p(\theta) = p(\theta|\theta^{\text{TV}}(\text{cov}), \Omega)$  as importance function. This gives for a proposal sample  $\tilde{\theta}_n^{(s)}$  the unnormalised weights

$$\tilde{w}_n^{(s)} = \frac{\tilde{p}(\tilde{\theta}_n^{(s)}|y_{1:n})}{p(\tilde{\theta}_n^{(s)})} = p(y_{1:n}|\tilde{\theta}_n^{(s)}),$$

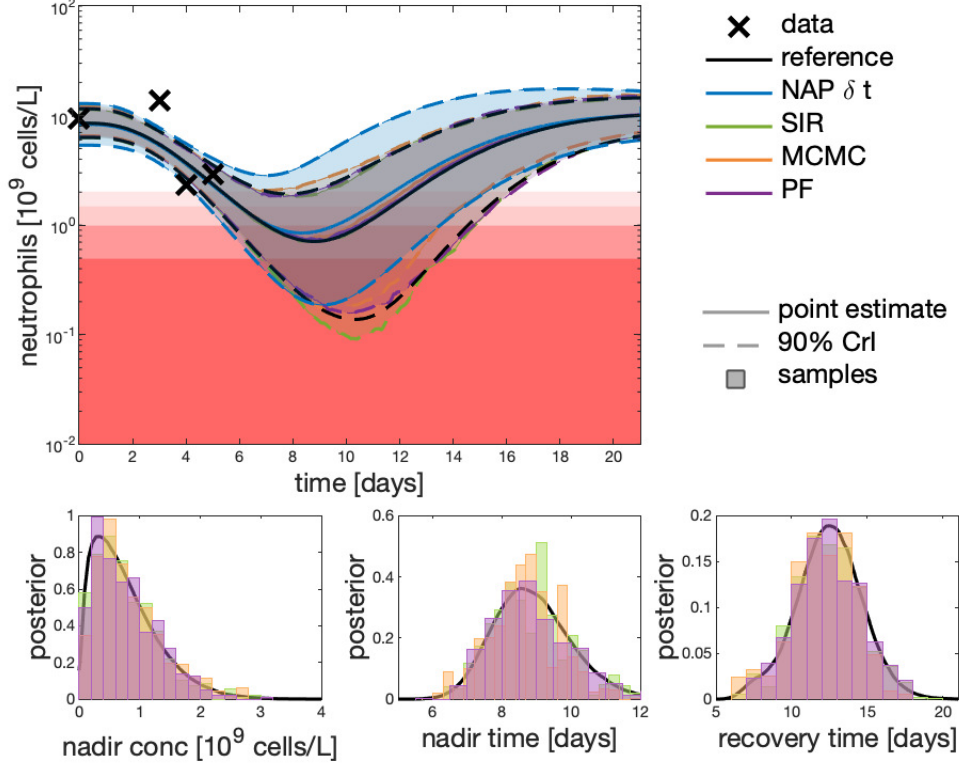


Figure S 3: Propagating uncertainties with the delta method using Student t quantiles.

which need to be normalised for an approximation of the normalised posterior distribution

$$w_n^{(s)} = \frac{p(y_{1:n}|\tilde{\theta}_n^{(s)})}{\sum_{s'} p(y_{1:n}|\tilde{\theta}_n^{(s')})}.$$

Step-by-step description of the algorithm

1. Sample proposals  $\theta_n^{(s)}$  from prior  $p(\theta|\theta^{TV}, \Omega)$
2. Compute unnormalised importance weights  $\tilde{w}_n^{(s)}$
3. Compute normalised importance weights  $w_n^{(s)}$
4. Resample according to normalised importance weights  $w_n^{(s)}$

Generally this algorithm needs a large number of samples, especially if there is a large disagreement between prior and posterior. This is computationally very expensive, but can be run in parallel up to the normalisation of the weights. The SIR algorithm is also used in the population analysis context to improve the estimation of the parameter uncertainty distributions [6]. Thus, the SIR approach might be an option to also take into account uncertainty in the population parameters (hyper parameters).

### Reference posterior

Since the true posterior distribution is analytically intractable, we employ as reference solution the SIR algorithm with a large number of samples ( $S = 10^6$ ), as the algorithm is exact for  $S \rightarrow \infty$ . To validate this reference we have compared it in addition to the posterior derived by the MCMC algorithm using also  $S = 10^6$  samples with a burn-in of 100 samples, see Figure S 7.

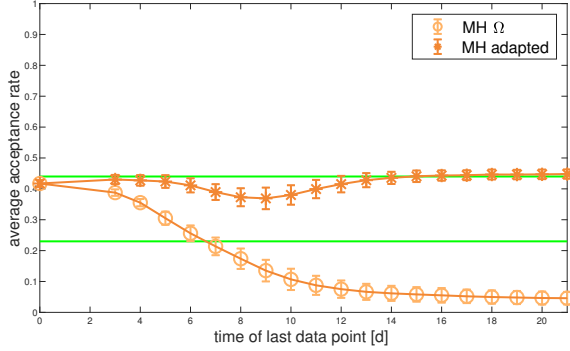


Figure S 4: Acceptance rate of the Metropolis Hastings algorithm with fixed proposal variance (MH $\Omega$ ) and with adapted proposal (MH adapted). The green lines mark the area of a good acceptance rate balancing the trade-off between exploring the space and efficiently moving the chain.

## 6 Markov chain Monte Carlo (MCMC)

We used in the presented study an adaptive version of the well-known Metropolis-Hastings algorithm. In the TDM setting it was previously suggested to use the prior as fixed proposal (independence sampler) [7]. We have, however, found that as the number of data points increases the rejection rate increases (since the posterior is becoming narrower), leading to inefficient sampling of the posterior. However, positioning the proposal ( $\mathcal{LN}(\cdot|\Omega)$ ) at the just accepted proposal also showed high rejection rates. To counteract low acceptance rates, the proposal was therefore not only moved to the just accepted proposal but was also adapted to the previous posterior sampling distribution, e.g.  $\theta_0 = \mathbb{E}[\hat{p}(\log(\theta)|y_{1:n-1})]$  and  $\mathcal{LN}(\cdot|\theta_{s-1}, \text{Cov}[\hat{p}(\log(\theta)|y_{1:n-1})])$ , see Figure S 4.

Since we employed the Metropolis-Hastings algorithm with a lognormal, thus asymmetric distribution as a proposal, we have to use a correction term in the acceptance probability to account for the asymmetry of the proposal distribution. Using Eq. S 1, we get the following acceptance ratio in the case of a lognormal proposal distribution

$$\alpha = \frac{p(\theta^*|y_{1:n})}{p(\theta_{s-1}|y_{1:n})} \cdot \frac{\mathcal{LN}(\theta_{s-1}|\theta^*, \Omega)}{\mathcal{LN}(\theta^*|\theta_{s-1}, \Omega)} = \frac{p(\theta^*|y_{1:n})}{p(\theta_{j-1}|y_{1:n})} \cdot \frac{\prod(\theta^*)_k}{\prod(\theta_{j-1})_k}. \quad (\text{S } 4)$$

The Markov chain was started at the typical value  $\theta^{\text{TV}}(\text{cov})$ . Generally, a certain number of samples in the beginning is discarded, a so called burn-in or warm-up. We chose a burn-in of 100 samples in our analysis.

The steps of the Metropolis-Hastings algorithm at time point  $t_n$  are

1. Start chain at  $\theta_0 = \theta^{\text{TV}}$

For  $s = 1, \dots, S$

2. Generate proposal  $\theta^*$  from the proposal distribution  $\mathcal{LN}(\theta_{s-1}, \Omega)$
3. Accept  $\theta^*$  with probability  $\alpha$  defined in Eq. S 4.

One drawback of MCMC approaches is that standard MCMC methods cannot be used efficiently in sequential inference context, as for every updated posterior distribution  $p(\theta|y_{1:k})$  a new Markov chain has to be generated [8].

## 7 Particle filter (PF)

The particle filter belongs to the class of sequential data assimilation methods. In sequential DA methods, the posterior is iteratively updated via Bayes' formula by combining computer-generated Bayesian forecasts with data in real time. The most well-known sequential data assimilation method is the Kalman filter [9, 10], which relies on the assumptions of linear model dynamics



and Gaussian uncertainty, which reduced the problem to only track the mean and variance of the posterior density over time. However, many problems in application do not satisfy these assumptions. Therefore, particle filters (PF) [11] were developed that allow for non-Gaussian error models and non-linear structural models, which fits the general pharmacometric setting. In this section some more algorithmic details of the particle filter are described.

## 7.1 State augmentation

Filter algorithms were mainly developed for state estimation with fixed parameters. However, the parameters can be added to an augmented state space  $z = (x, \theta)$ ,

$$\begin{aligned}\frac{\partial x}{\partial t}(t) &= f(x(t); \theta, u) \\ \frac{\partial \theta}{\partial t}(t) &= 0.\end{aligned}$$

Since we are considering static parameters (within one cycle), the rate of change of parameters is zero. The filter algorithm was then applied to the augmented state  $z(t) = (x(t), \theta(t))$ .

## 7.2 Resampling strategies

There exist many different strategies on how to resample efficiently and effectively. The most widely used are multinomial, residual and systematic resampling, see e.g. [12]. For this article, we applied residual resampling. Resampling can be performed at every step (bootstrap filter) or, more efficiently, only if the effective sample size is smaller than a threshold, e.g.  $S_{\text{eff}} < S/2$ .

## 7.3 Rejuvenation


Resampling addresses the problem of weight degeneracy in particle filters, however, it introduces the problem of sample impoverishment. Drawing with replacement among the particles results in many identical particles. Since the structural model is deterministic with constant parameters, resampled particles will remain identical over time. Particle rejuvenation can be applied to counteract this sample impoverishment [13]. After each resampling new parameter particles are sampled from a normal distribution centered around the previous parameter values with a small relative variance. Since this only introduces small perturbations in the parameter space, we assumed that  $x(\theta) \approx x(\tilde{\theta})$  for a rejuvenation of  $\theta$  resulting in  $\tilde{\theta}$ .

## 7.4 Smoothing

For smoothing over the past prediction the previously predicted paths are resampled along with the states and parameters according to the current particle weights.

## 7.5 Alternative sequential DA algorithms

There exist many extensions, modifications and add-on techniques for particle filters. Depending on how the analysis ensemble is generated from the forecast ensemble, different filter algorithms are distinguished, see e.g. [14, 15]. Among these the class of ensemble transform filters [15] is very promising as it replaces the stochastic resampling and rejuvenation step of the basic particle filter by a deterministic transformation which allows to ensure certain properties, e.g. 2nd order accuracy [14]. However, in the augmented state space the connection between the parameters and states is lost as larger steps in the parameter steps are undertaken which means that the assumption  $x(\theta) \approx x(\tilde{\theta})$  is no longer valid.

0. Initialisation of particles  $\{x_0^{(s)}, \theta_0^{(s)}, w_0^{(s)}\}_{s=1}^S$    
for all  $s = 1, \dots, S$   
sample  $\theta_0^{(s)}$  from  $p(\theta; \hat{\theta}^{TV}(\text{cov}), \Omega)$ ,  $x_0^s$  from  $p(x_0|\theta)$  and set  $w_0^{(s)} = 1/S$

For  $j = 1, \dots, n$

1. Propagation under model equations  
producing a forecast ensemble

$$\{x_j^{(s)f}, \theta_{j-1}^{(s)}, w_{j-1}^{(s)}\}_{s=1}^S$$

2. Update of importance weights

$$w_j^{(s)} \propto w_{j-1}^{(s)} \cdot p(y_j|x_j^{(s)f}, \Sigma)$$

and normalisation of weights

If  $S_{eff} < S/2$

3. Resampling and rejuvenation  
leads to the analysis ensemble

$$\{x_j^{(s)a}, \theta_j^{(s)}\}_{s=1}^S$$

with uniform weights  $w_j^{(s)}$

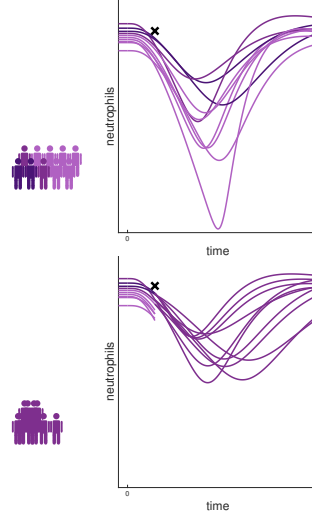


Figure S 5: **Step-by-step description of the particle filter.** The different steps of the particle filter are depicted in the context of forecasting the time-course of neutropenia. For illustration purposes the number of virtual individuals is reduced to  $S = 10$ . The different shades represent the importance weights, the darker the colour the higher the importance weight. From the initial virtual subpopulation only five are resampled (resulting in duplicate samples). Rejuvenation introduces new virtual individuals who are similar to the five resampled ones.

## 8 Simulation study: Single cycle Docetaxel

In the simulation study we use as prior knowledge a population analysis of a clinical study by Kloft et al. [16]. In this section more details about the employed models and parameter estimates is provided. We aimed to be comparable to the setting in [17]. A virtual population ( $N = 100$ ) was generated based on the patient characteristics provided in [16]. The covariates AGE and AAG were sampled from a normal distribution with mean given by the median and an estimated variance from the given observed range. The parameter estimates used for the pharmacokinetic (PK) and pharmacodynamic (PD) model are given in Table S 2. For the MAP estimation we needed to provide parameter bounds to the optimiser. The lower bounds were taken from the code provided by Netterberg [18] and the upper bounds were tested, so that the optimiser did not reach the bound.

As PK model for Docetaxel a three compartment model with first-order elimination was employed [21]. AAG ( $\alpha$ 1-acid glycoprotein), AGE, BSA (body surface area) and ALB (albumin) were found as covariates on clearance. Patient specific parameter values were determined based on covariates. The individual clearance is computed via

$$CL_i = BSA_i \cdot (CL^{TV} + \theta_{CL-AAG} \cdot AAG_i + \theta_{CL-AGE} \cdot AGE_i + \theta_{CL-ALB} \cdot ALB^{TV}) \cdot (1 - \theta_{CL-HEP12} \cdot HEP12),$$

where we used  $ALB^{TV} = 41 \text{ g/L}$  and set  $HEP12=0$  (i.e. no elevated hepatic enzymes). The parameter estimates were taken from [21], see Supplement Table 1 and the system of ODEs for the PK model is given by

$$\begin{aligned} \frac{dCent}{dt} &= u(t) - k_{10}Cent + k_{21}Per1 - k_{12}Cent + k_{31}Per2 - k_{13}Cent, & Cent(0) &= 0 \\ \frac{dPer1}{dt} &= k_{12}Cent - k_{21}Per1, & Per1(0) &= 0 \\ \frac{dPer2}{dt} &= k_{13}Cent - k_{31}Per2, & Per2(0) &= 0. \end{aligned}$$

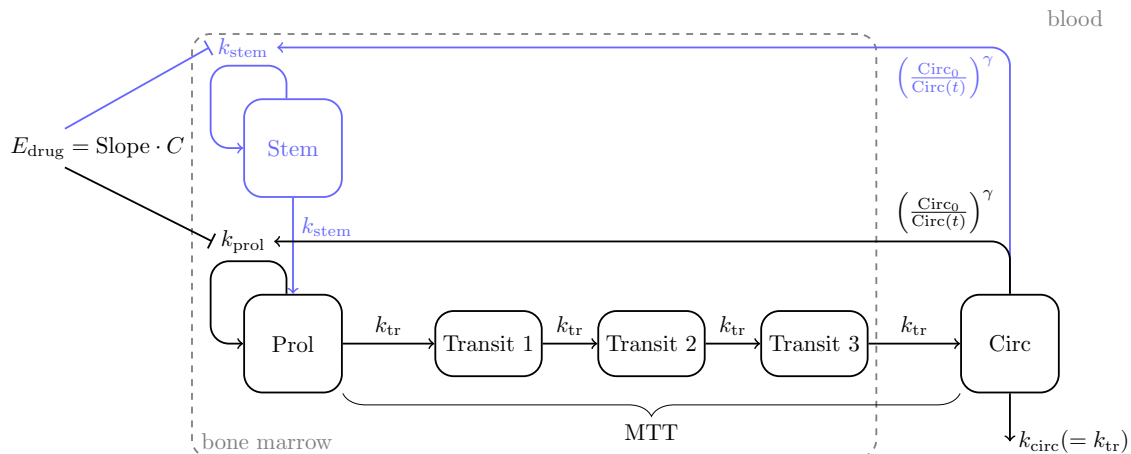


Figure S 6: **Structural models for neutropenia.** The gold-standard model for neutropenia, developed by Friberg et al. [19] (black). Extended bone marrow exhaustion model, which describes cumulative neutropenia over multiple cycles, developed by Henrich et al. [20] (black and blue). The state variables and parameters of the models are described in the text.

Structural submodel		
V	8.31	[L]
CL <sup>TV</sup>	22.1	[L/h]
$k_{10}$	CL/V	[1/h]
$k_{12}$	1.07	[1/h]
$k_{21}$	1.74	[1/h]
$k_{13}$	1.28	[1/h]
$k_{31}$	0.0787	[1/h]
Covariate submodel		
$\theta_{\text{CL-AAG}}$	-3.55	
$\theta_{\text{CL-AGE}}$	-0.095	
$\theta_{\text{CL-ALB}}$	0.225	

Table S 1: Pharmacokinetic parameter estimates for Docetaxel [21].

The semi-mechanistic model by Friberg et al. [19] describes chemotherapy-induced neutropenia and consists of five compartments. The proliferating compartment (Prol) represents rapidly dividing progenitor cells in the bone marrow, which replicate with rate  $k_{\text{prol}}$ . Three transit compartments (Transit 1-3) approximate the maturation chain of progenitor cells in the bone marrow to differentiated neutrophils in the systemic circulation, with transition rate  $k_{\text{tr}}$ . Circulating neutrophils in blood (Circ) are the part of the system that can be observed. Thus, the state vector for the model is given by

$$x(t) = \left( \text{Prol}(t), \text{Transit1}(t), \text{Transit2}(t), \text{Transit3}(t), \text{Circ}(t) \right)^T,$$

with observable

$$h(t) = \text{Circ}(t),$$

and initial conditions

$$\text{Prol}(t_0) = \text{Transit 1}(t_0) = \text{Transit 2}(t_0) = \text{Transit 3}(t_0) = \text{Circ}(t_0) = \text{Circ}_0.$$

The cytotoxic effect of anticancer drugs is implemented linearly on the proliferation rate of susceptible progenitor cells, so the effect is proportional to the drug concentration

$$E_{\text{drug}}(t) = \text{Slope} \cdot C_{\text{drug}}(t),$$

which is the input to the system

$$u(t) = C_{\text{drug}}(t),$$

Structural submodel		
Circ <sub>0</sub>	5.22	[10 <sup>9</sup> cells/L]
MTT	84.2	[h]
Slope	15.6	[L/μmol]
γ	0.145	]
Covariate submodel		
θ <sub>Circ<sub>0</sub>-AAG≤1.34</sub>	0.175	
θ <sub>Circ<sub>0</sub>-AAG&gt;1.34</sub>	0.495	
θ <sub>Circ<sub>0</sub>-SEX</sub>	-0.121	
θ <sub>Circ<sub>0</sub>-PERF</sub>	0.131	
θ <sub>Circ<sub>0</sub>-PC</sub>	-0.147	
θ <sub>Slope-AAG</sub>	-0.351	
Statistical submodel		
σ <sup>2</sup>	0.180	
ω <sub>Circ<sub>0</sub></sub> <sup>2</sup>	0.0606	
ω <sub>MTT</sub> <sup>2</sup>	0.0194	
ω <sub>Slope</sub> <sup>2</sup>	0.122	
ω <sub>γ</sub> <sup>2</sup>	0.0223	

Table S 2: Parameter estimates for the Friberg-model and Docetaxel taken from [16]

parameter	lower bound	upper bound
Circ <sub>0</sub> [10 <sup>9</sup> cells/L]	2	30
MTT [h]	10	250
Slope [L/μmol]	0.01	60

Table S 3: Bounds for MAP estimation. Lower bounds were taken from [18]

where the time-evolution is given by the corresponding PK model for Docetaxel. The effect of the drug on the proliferation rate of the Prol compartement is described by a feedback, which leads to an increase in production of white blood cells if the number of circulating neutrophils is decreasing, and vice versa

$$\text{Feedback}(t) = \left( \frac{\text{Circ}_0}{\text{Circ}(t)} \right)^\gamma .$$

System parameters include the baseline neutrophil concentration (Circ<sub>0</sub>) before the start of treatment, the mean transition time (MTT), which represents the average time for a progenitor cell in the bone marrow to mature to a circulating neutrophil, and the slope parameter (Slope) of the linear inhibitory model for the drug effect. Furthermore, γ is the exponent parameter for the feedback model. Thus, the parameter vector for the model is given by

$$\theta = (\text{Circ}_0, \text{MTT}, \text{Slope}, \gamma)^T .$$

As in [17] we assume γ to be fixed. The system of ordinary differential equations (ODEs) reads

$$\begin{aligned} \frac{d\text{Prol}}{dt} &= k_{\text{prol}}\text{Prol} \cdot (1 - E_{\text{drug}}) \cdot \left( \frac{\text{Circ}_0}{\text{Circ}} \right)^\gamma - k_{\text{tr}}\text{Prol}, & \text{Prol}(0) &= \text{Circ}_0 \\ \frac{d\text{Transit1}}{dt} &= k_{\text{tr}}\text{Prol} - k_{\text{tr}}\text{Transit1}, & \text{Transit1}(0) &= \text{Circ}_0 \\ \frac{d\text{Transit2}}{dt} &= k_{\text{tr}}\text{Transit1} - k_{\text{tr}}\text{Transit2}, & \text{Transit2}(0) &= \text{Circ}_0 \\ \frac{d\text{Transit3}}{dt} &= k_{\text{tr}}\text{Transit2} - k_{\text{tr}}\text{Transit3}, & \text{Transit3}(0) &= \text{Circ}_0 \\ \frac{d\text{Circ}}{dt} &= k_{\text{tr}}\text{Transit3} - k_{\text{circ}}\text{Circ}, & \text{Circ}(0) &= \text{Circ}_0 . \end{aligned}$$

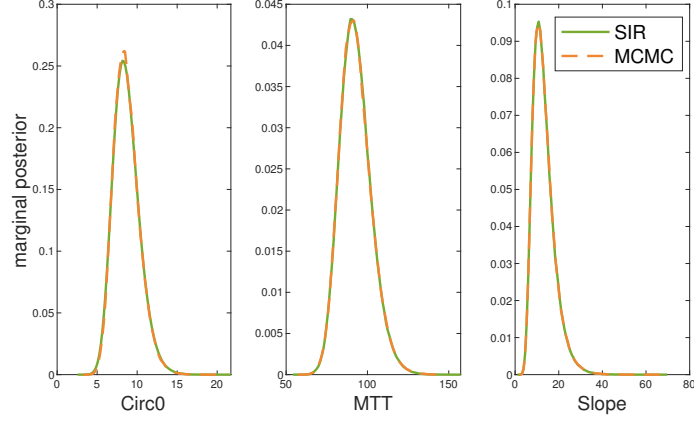


Figure S 7: Comparison of reference posterior. The reference posterior was derived by the SIR algorithm and with the MCMC algorithm using  $S = 10^6$  samples.

## 9 Simulation study: Multiple cycles Paclitaxel

The multiple cycle simulation study is based on a population analysis [20] of the CEPAC-TDM study data. For the data simulation it was assumed that every third day the neutrophil counts were assessed for a patient undergoing chemotherapy. The data was simulated based on a model prediction including interoccasion variability. Paclitaxel pharmacokinetics were described by a three compartment model with nonlinear distribution to the first peripheral compartment and nonlinear elimination [22]. The system of ODEs reads

$$\begin{aligned} \frac{d\text{Cent}}{dt} &= u(t) - \frac{\text{VM}_{\text{EL}} \cdot C_1}{\text{KM}_{\text{EL}} + C_1} + k_{21}\text{Per1} - \frac{\text{VM}_{\text{TR}} \cdot C_1}{\text{KM}_{\text{TR}} + C_1} + k_{31}\text{Per2} - k_{13}\text{Cent}, & \text{Cent}(0) &= 0 \\ \frac{d\text{Per1}}{dt} &= \frac{\text{VM}_{\text{TR}} \cdot C_1}{\text{KM}_{\text{TR}} + C_1} - k_{21}\text{Per1}, & \text{Per1}(0) &= 0 \\ \frac{d\text{Per2}}{dt} &= k_{13}\text{Cent} - k_{31}\text{Per2}, & \text{Per2}(0) &= 0 \end{aligned}$$

where  $C_1(t) = \text{Cent}(t)/V_1$ . We used the parameter values by [20], see Table S 4. The following covariate model for  $\text{VM}_{\text{EL}}$  was given as

$$\text{VM}_{\text{EL},\text{TV},i} = \text{VM}_{\text{EL},\text{pop}} \cdot \left(\frac{\text{BSA}_i}{1.8\text{m}^2}\right)^{\theta_{\text{VM}_{\text{EL}}-\text{BSA}}} \cdot \left(\theta_{\text{VM}_{\text{EL}}-\text{SEX}}\right)^{\text{SEX}_i} \cdot \left(\frac{\text{AGE}_i}{56\text{years}}\right)^{\theta_{\text{VM}_{\text{EL}}-\text{AGE}}} \cdot \left(\frac{\text{BILI}_i}{7\mu\text{mol/L}}\right)^{\theta_{\text{VM}_{\text{EL}}-\text{BILI}}}$$

relating patient characteristics, BSA (body surface area), SEX, AGE and BILI (total bilirubin concentration) to the maximum elimination capacity ( $\text{VM}_{\text{EL}}$ ). In addition to inter-individual variability and residual variability, interoccasion variability was included on the two parameters  $V_1$  and  $\text{VM}_{\text{EL}}$ , where each start of a new cycle is defined as an occasion,

$$\theta_{i,o} = \theta_{\text{TV},i} \cdot e^{\eta_i + \kappa_{i,o}}, \quad \kappa_{i,o} \stackrel{iid}{\sim} \mathcal{N}(0, \Pi^2).$$

Neutropenia has been observed to worsen over several treatment cycles, i.e. the lowest neutrophil concentration (nadir concentration) and the maximum neutrophil concentration decreases over multiple cycles. Bone marrow exhaustion could be one explanation of this cumulative behaviour, which means that the long-term recovery of the bone marrow is also affected. The standard model for neutropenia [19] does not take such a long-term effect into account. Henrich et al. [20] therefore expanded the model by dividing the proliferating compartment into another stem cell compartment, describing pluripotent stem cells with reduced proliferation rate. This model extension made it possible to capture the cumulative long-term effect.

The proliferation rate constants for the proliferating compartment Prol and for the stem cell compartment Stem are given by

$$\begin{aligned} k_{\text{prol}} &= \text{ftr} \cdot k_{\text{tr}} \\ k_{\text{stem}} &= (1 - \text{ftr}) \cdot k_{\text{tr}}, \end{aligned}$$

Structural submodel		
$V_1$	10.8	[L]
$V_3$	301	[L]
$KM_{EL}$	0.667	[ $\mu$ M]
$VM_{EL,POP}$	35.9	[ $\mu$ mol/h]
$KM_{TR}$	1.44	[ $\mu$ M]
$VM_{TR}$	175	[ $\mu$ mol/h]
$k_{21}$	1.12	[1/h]
$Q$	16.8	[1/h]
Covariate submodel		
$\theta_{VM_{EL}-BSA}$	1.14	
$\theta_{VM_{EL}-SEX}$	1.07	
$\theta_{VM_{EL}-AGE}$	-0.447	
$\theta_{VM_{EL}-BILI}$	-0.0942	
Statistical submodel IIV		
$\omega_{V_3}^2$	0.1639	
$\omega_{VM_{EL}}^2$	0.0253	
$\omega_{KM_{TR}}^2$	0.3885	
$\omega_{VM_{TR}}^2$	0.077	
$\omega_{k_{21}}^2$	0.008	
$\omega_Q^2$	0.1660	
Statistical submodel IOV		
$\pi_{V_1}^2$	0.1391	
$\pi_{VM_{EL}}^2$	0.0231	
Statistical submodel RV		
$\sigma^2$	0.0317	

Table S 4: Pharmacokinetic parameter estimates for the anticancer drug Paclitaxel [20].

respectively, where  $f_{tr}$  is the fraction of input in the Prol compartment via proliferation within the compartment. As the drug effect is proportional to the proliferation rate constant, the stem cells are less affected by the treatment than the progenitor cells.

$$\begin{aligned}
\frac{dStem}{dt} &= k_{stem}Stem \cdot (1 - E_{drug}) \cdot \left(\frac{Circ_0}{Circ}\right)^\gamma - k_{tr}StemStem, & Stem(0) &= Circ_0 \\
\frac{dProl}{dt} &= k_{prol}Prol \cdot (1 - E_{drug}) \cdot \left(\frac{Circ_0}{Circ}\right)^\gamma + k_{tr}StemStem - k_{tr}Prol, & Prol(0) &= Circ_0 \\
\frac{dTransit1}{dt} &= k_{tr}Prol - k_{tr}Transit1, & Transit1(0) &= Circ_0 \\
\frac{dTransit2}{dt} &= k_{tr}Transit1 - k_{tr}Transit2, & Transit2(0) &= Circ_0 \\
\frac{dTransit3}{dt} &= k_{tr}Transit2 - k_{tr}Transit3, & Transit3(0) &= Circ_0 \\
\frac{dCirc}{dt} &= k_{tr}Transit3 - k_{circ}Circ, & Circ(0) &= Circ_0
\end{aligned}$$

The baseline parameter  $Circ_0$  was inferred from the baseline data point ( baseline method 2 [23])

$$Circ_{0,i} = y_{Circ_0} \cdot e^{\theta_{RV} \cdot \eta_{Circ_0,i}}, \quad \eta_{Circ_0,i} \sim \mathcal{N}(0, 1).$$

In the bone marrow exhaustion model we have in addition to inter-individual variability (IIV) inter-occasion variability (IOV). The parameter vector, thus, contains parameter values, which are constant across occasions (cycles)  $\theta^{IIV}$  and parameters that are specific for each cycle  $c$ ,  $\theta_c^{IOV}$ .

$$\theta_c = \underbrace{e^{\log(\theta^{TV}) + \eta}}_{=\theta^{IIV}} \cdot \underbrace{e^{\kappa_c}}_{=\theta_c^{IOV}}. \quad (S 5)$$

Structural submodel		
Circ <sub>0</sub>	baseline method [23]	[10 <sup>9</sup> cells/L]
MTT	145	[h]
Slope	13.1	[L/μmol]
γ	0.257	□
ftr	0.787	□
Statistical submodel		
σ <sup>2</sup>	0.2652	
ω <sub>Slope</sub> <sup>2</sup>	0.2007	

Table S 5: Parameter estimates for the bone marrow exhaustion model and the anticancer drug Paclitaxel .

The  $\theta^{\text{IIV}}$  parameters need to be learned across all cycles and the cycle specific parameters  $\theta_c^{\text{IOV}}$  based on the data observed in cycle  $c$ ,  $y_{1:n_c} = (y_1, \dots, y_{n_c})^T$ . The size of the parameter vector that needs to be estimated will, therefore, grow with every occasion if the whole data is processed in a batch. Assuming independence between the inter-individual and interoccasion variability, the objective function for the MAP estimation is then given by

$$\begin{aligned} \hat{\theta}_n^{\text{MAP}} = \arg \min_{\theta^{\text{IIV}}, \theta^{\text{IOV}}} & \frac{1}{2} \left( \sum_{c=1}^C \sum_{j=1}^{n_c} \frac{(y_j - h_j(\theta))^2}{\sigma^2} \right. \\ & + 2 \sum_{k=1}^{n_\theta^{\text{IIV}}} \log(\theta_k^{\text{IIV}}) + \sum_{k=1}^{n_\theta^{\text{IIV}}} \frac{(\log(\theta_k^{\text{IIV}}) - \log(\theta_k^{\text{TV}}(\text{cov})))^2}{\omega_k^2} \\ & \left. + 2 \sum_{c=1}^C \sum_{k=1}^{n_\theta^{\text{IOV}}} \log(\theta_{k,c}^{\text{IOV}}) + \sum_{c=1}^C \sum_{k=1}^{n_\theta^{\text{IOV}}} \frac{(\log(\theta_{k,c}^{\text{IOV}}) - \log(\theta_k^{\text{TV}}(\text{cov})))^2}{\pi_k^2} \right) \end{aligned}$$

for the IIV and IOV model  $\theta_{k,c} = \theta_k^{\text{TV}} \cdot e^{\eta_k + \kappa_{k,c}}$ , with  $\eta_k \sim \mathcal{N}(0, w_k^2)$  and  $\kappa_{k,c} \sim \mathcal{N}(0, \pi_k^2)$ , for an additive normal residual error model  $y_j = h_j + \epsilon_j$ , with  $\epsilon_j \sim \mathcal{N}(0, \sigma^2)$  and for data observed up to time point  $t_n$ , i.e.  $n = \sum_c n_c$ , with  $n_c$  the number of observations made in cycle  $c$ .

The gradient with respect to the IIV parameters is then given by

$$\begin{aligned} \frac{\partial J(\theta)}{\partial \theta_l^{\text{IIV}}} &= - \sum_{j=1}^n \frac{(y_j - h_j(\theta))}{\sigma^2} \cdot \frac{\partial h_j(\theta)}{\partial \theta_l^{\text{IIV}}} \\ &+ \frac{1}{\theta_l^{\text{IIV}}} + \frac{(\log(\theta_l^{\text{IIV}}) - \log(\theta_l^{\text{TV}}))}{\omega_l^2} \cdot \frac{1}{\theta_l^{\text{IIV}}}, \end{aligned}$$

and with respect to the IOV parameters by

$$\begin{aligned} \frac{\partial J(\theta)}{\partial \theta_{l,c}^{\text{IOV}}} &= - \sum_{j=1}^n \frac{(y_j - h_j(\theta))}{\sigma^2} \cdot \frac{\partial h_j(\theta)}{\partial \theta_{l,c}^{\text{IOV}}} \\ &+ \frac{1}{\theta_{l,c}^{\text{IOV}}} + \frac{(\log(\theta_{l,c}^{\text{IOV}}) - \log(\theta_l^{\text{TV}}))}{\omega_l^2} \cdot \frac{1}{\theta_{l,c}^{\text{IOV}}}, \end{aligned}$$

Figure S 8 shows a comparison of the different methods in forecasting the third cycle. The scenario corresponds to the situation presented in Figure 4 in the main manuscript. All full Bayesian methods provide almost overlapping credible intervals as well as point estimates (median). However, the MAP-based forecasted trajectory deviates significantly from the point estimates (median) of the full Bayesian methods.

In Figure S 9 the posterior approximations are compared to the reference on the level of the parameters. In addition, we can observe the deviation of the posterior from the prior for parameters 'Slope' and 'Circ0'. As we do not consider PK samples the knowledge gain about the PK parameters is limited. All approximations show good agreement with the reference and the MAP estimate is located at the mode of the posterior on the level of the parameters.

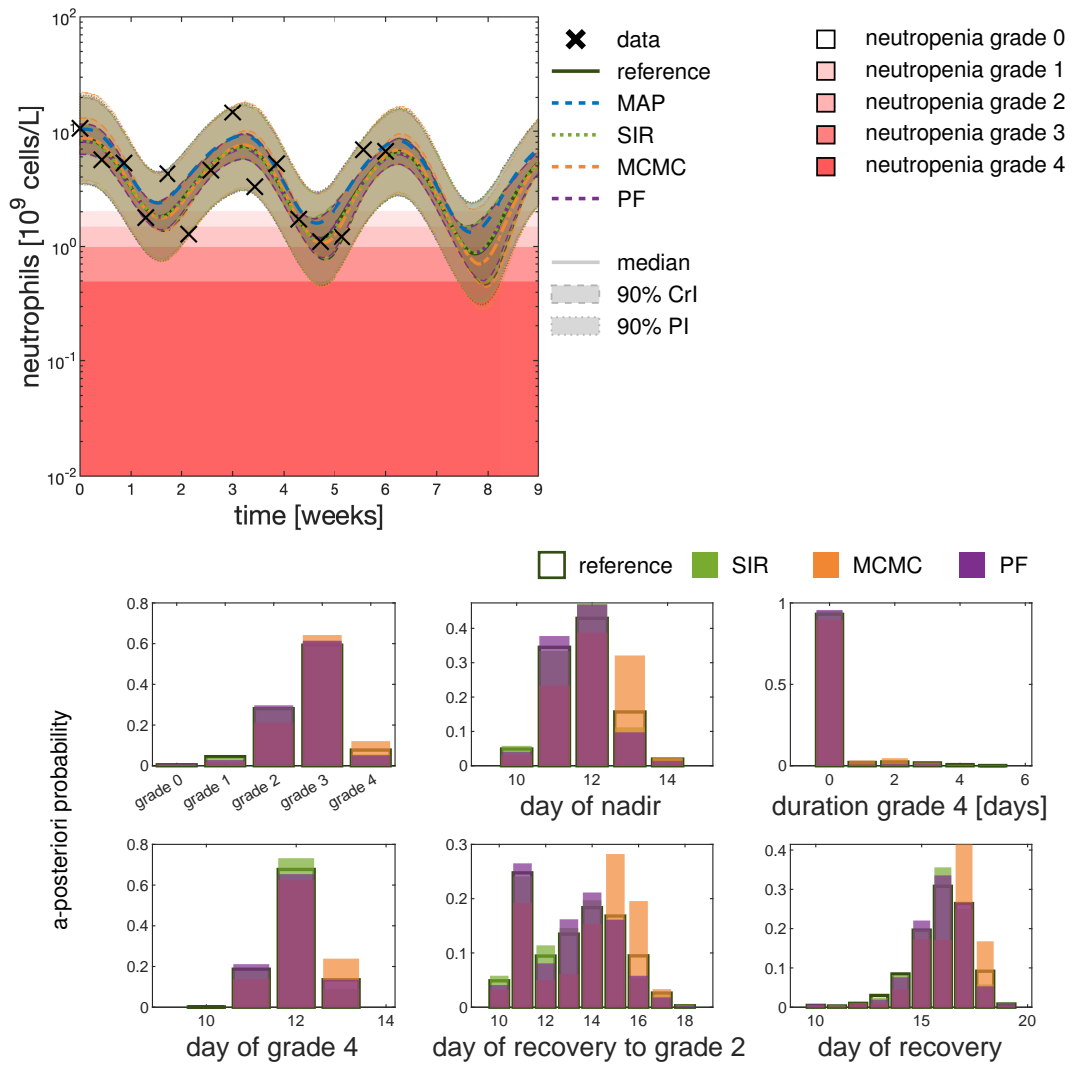


Figure S 8: **Comparison of all methods for the standard dose.** The third cycle is forecasted in case the standard dose is given.



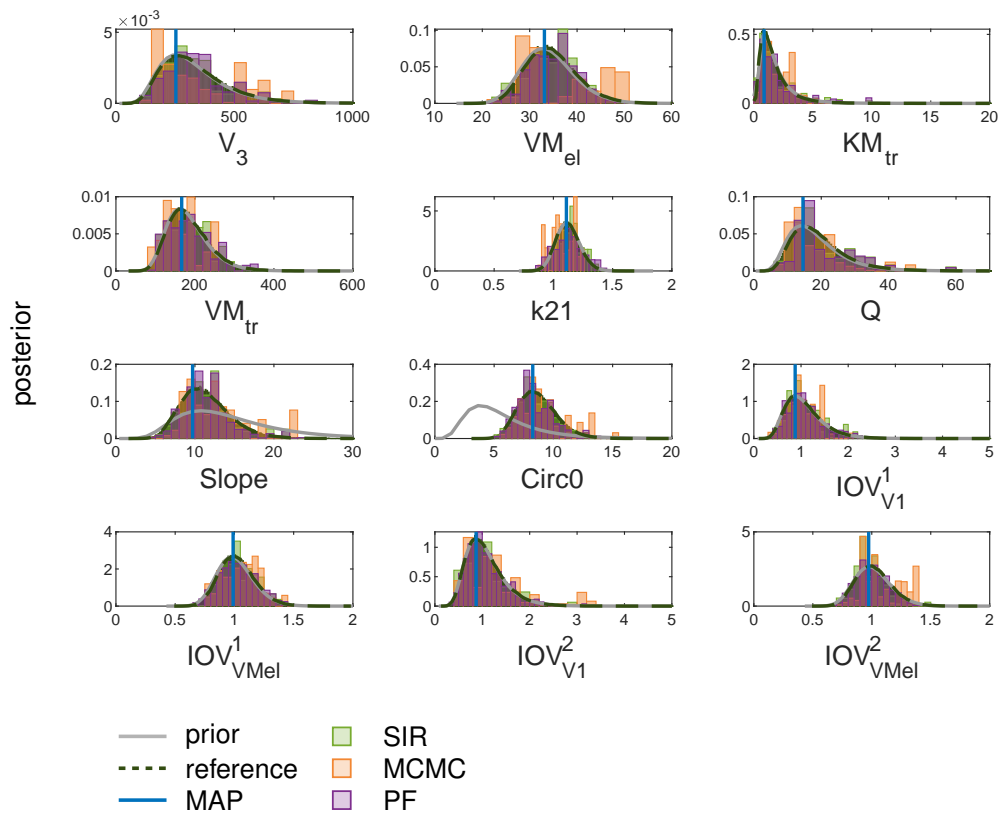


Figure S 9: **Posterior inference of the different methods.** Posterior of the parameters for the scenario considered in Supplement Figure 8

## References

- [1] Fröhlich, F., Theis, F.J., Joachim, O.R., & Hasenauer, J. Parameter estimation for dynamical systems with discrete events and logical operations. *Oxford Bioinforma.* 1–8 (2016). doi:10.1093/bioinformatics/btw764.
- [2] Murphy, K.P. *Machine Learning: A Probabilistic Perspective*, vol. 1, (The MIT Press, Cambridge, MA2012). doi:10.1007/SpringerReference\_18220.
- [3] Gelman, A., Carlin, J.B., Stern, H.S., Dunson, D.B., Vehtari, A., & Rubin, D.B. *Bayesian Data Analysis*. 3rd edn., (Chapman and Hall/CRC, New York2014).
- [4] Wasserman, L. *All of statistics: A concise course in statistical inference*, (Springer Science & Business Media, New York2000).
- [5] Kümmer, A., Bonate, P.L., Dingemans, J., & Krause, A. Confidence and Prediction Intervals for Pharmacometric Models. *CPT Pharmacometrics Syst. Pharmacol.* 7, 360–373 (2018). doi:10.1002/psp4.12286.
- [6] Dosne, A.G., Bergstrand, M., Harling, K., & Karlsson, M.O. Improving the estimation of parameter uncertainty distributions in nonlinear mixed effects models using sampling importance resampling. *J Pharmacokinet Pharmacodyn* 583–596 (2016). doi:10.1007/s10928-016-9487-8.
- [7] Wakefield, J. Bayesian individualization via sampling-based methods. *J. Pharmacokinet. Biopharm.* 24, 103–131 (1996). doi:10.1007/BF02353512.
- [8] Del Moral, P. Sequential Monte Carlo samplers. *J. R. Stat. Soc. B* 68, 411–436 (2006).
- [9] Kalman, R. A New Approach to Linear Filtering and Prediction Problems. *ASME. J. Basic Eng.* 82, 35–45 (1960). doi:10.1115/1.3662552.
- [10] Kalman, R. & Bucy, R. New Results in Linear Filtering and Prediction Theory. *ASME. J. Basic Eng.* 83, 95–108 (1961). doi:10.1115/1.3658902.
- [11] Gordon, N., Salmond, D., & Smith, A. Novel approach to nonlinear/non-Gaussian Bayesian state estimation. *IEE Proc. F Radar Signal Process.* 140, 107 (1993). doi:10.1049/ip-f-2.1993.0015.
- [12] Doucet, A. & Johansen, A.M. A Tutorial on Particle Filtering and Smoothing : Fifteen years later. In D. Crisan & B. Rozovsky, editors, *Oxford Handb. Nonlinear Filter.*, (Oxford University Press2009).
- [13] Reich, S. & Cotter, C. *Probabilistic Forecasting and Bayesian Data Assimilation*, (Cambridge University Press, Cambridge2015).
- [14] Acevedo, W., De Wiljes, J., & Reich, S. Second-order accurate ensemble transform particle filters. *SIAM J SCI Comput* 39, A1834–A1850 (2017). doi:10.1137/090750688.
- [15] Reich, S. A non-parametric ensemble transform method for Bayesian inference. *SIAM J. Sci. Comput.* 35, A2013–A2024 (2013). doi:10.1137/130907367.
- [16] Kloft, C., Wallin, J., Henningsson, A., Chatelut, E., & Karlsson, M.O. Population Pharmacokinetic-Pharmacodynamic Model for Neutropenia with Patient Subgroup Identification: Comparison across Anticancer Drugs. *Clin. Cancer Res.* 12, 5481–5490 (2006). doi:10.1158/1078-0432.CCR-06-0815.
- [17] Netterberg, I., Nielsen, E.I., Friberg, L.E., & Karlsson, M.O. Model-based prediction of myelosuppression and recovery based on frequent neutrophil monitoring. *Cancer Chemother. Pharmacol.* 80, 343–353 (2017). doi:10.1007/s00280-017-3366-x.
- [18] Netterberg, I. Kloft\_2006\_myelosuppression\_docetaxel. *DDMODEL00000224* (2016).
- [19] Friberg, L.E., Henningsson, A., Maas, H., Nguyen, L., & Karlsson, M.O. Model of chemotherapy-induced myelosuppression with parameter consistency across drugs. *J. Clin. Oncol.* 20, 4713–4721 (2002). doi:10.1200/JCO.2002.02.140.

- [20] Henrich, A. et al. Semimechanistic Bone Marrow Exhaustion Pharmacokinetic/Pharmacodynamic Model for Chemotherapy-Induced Cumulative Neutropenia. *J. Pharmacol. Exp. Ther.* 362, 347–358 (2017). doi:10.1124/jpet.117.240309.
- [21] Bruno, R., Vivier, N., Vergniol, J.C., De Phillips, S.L., Montay, G., & Sheiner, L.B. A Population Pharmacokinetic Model for Docetaxel (Taxotere®): Model Building and Validation. *J. Pharmacokinet. Biopharm.* 24, 153–172 (1996). doi:10.1007/BF02353487.
- [22] Joerger, M. et al. Evaluation of a pharmacology-driven dosing algorithm of 3-weekly paclitaxel using therapeutic drug monitoring: A pharmacokinetic-pharmacodynamic simulation study. *Clin. Pharmacokinet.* 51, 607–617 (2012). doi:10.2165/11634210-000000000-00000.
- [23] Dansirikul, C., Silber, H.E., & Karlsson, M.O. Approaches to handling pharmacodynamic baseline responses. *J Pharmacokinet Pharmacodyn* 35, 269–283 (2008). doi:10.1007/s10928-008-9088-2.