

# Assessment of Prediction Uncertainty Quantification Methods in Systems Biology

Alejandro F. Villaverde, Elba Raimúndez, Jan Hasenauer, and Julio R. Banga

**Abstract**—Biological processes are often modelled using ordinary differential equations. The unknown parameters of these models are estimated by optimizing the fit of model simulation and experimental data. The resulting parameter estimates inevitably possess some degree of uncertainty. In practical applications it is important to quantify these parameter uncertainties as well as the resulting prediction uncertainty, which are uncertainties of potentially time-dependent model characteristics. Unfortunately, estimating prediction uncertainties accurately is nontrivial, due to the nonlinear dependence of model characteristics on parameters. While a number of numerical approaches have been proposed for this task, their strengths and weaknesses have not been systematically assessed yet. To fill this knowledge gap, we apply four state of the art methods for uncertainty quantification to four case studies of different computational complexities. This reveals the trade-offs between their applicability and their statistical interpretability. Our results provide guidelines for choosing the most appropriate technique for a given problem and applying it successfully.

**Index Terms**—Computational methods, Dynamic models, Nonlinear systems, Observability, Prediction error methods, State estimation, Uncertainty.

## 1 INTRODUCTION

THE dynamics of many biological systems can be described by nonlinear ordinary differential equations (ODEs). As these models usually have a number of unknown parameters, it is necessary to calibrate them using experimental data [1]. Most models are partially observed, i.e. only a subset of their state variables – or functions thereof – are measured. In practice, these measurements are noisy, which complicates the model calibration task. Once parameter estimates are available, the dynamic behavior of the biological systems can be simulated by integrating the ODEs. This calculation yields the time courses of the model states, which are usually concentrations or other measures of abundance of biochemical species.

Calibrated models are frequently employed in prediction tasks [2], [3]. A prediction can in principle be any quantity or result derived from model simulations, including state variables. As parameter uncertainties can result in prediction uncertainties, a comprehensive uncertainty quantification is essential. Figure 1 provides a graphical illustration of the problem.

Uncertainty quantification is related to the concepts of observability and identifiability. Broadly speaking, a model is observable (respectively, identifiable) if it is possible to infer its state trajectories (respectively, parameters) from

knowledge about the time course of its observables. Structural and practical observability (identifiability) analysis methods are available for the assessment of this characteristic [4]. The former take into account only the model equations, and inform about the theoretical possibility of determining the unknowns. Therefore, these methods can reveal deficiencies of the model structure, such as symmetries. The latter also take into account the data available for model calibration. Hence, these methods inform about the confidence/credibility of parameter estimates and predictions. In this manuscript, we consider practical observability problems arising from practical identifiability issues, respectively prediction uncertainties resulting from parameter uncertainties.

Uncertainty quantification for a specific combination of model and datasets is a non-trivial task. The extent to which the uncertainty is propagated from parameters to predictions depends not only on the practical identifiability, but also on the sensitivity of the model to its parameters. Key challenges are nonlinearity and dimensionality of models considered in the field of computational biology, as well as the computational complexity required for simulation. Several numerical uncertainty quantification approaches have been proposed and a general overview is provided, e.g., in [5]. Some methods for the assessment of prediction uncertainty are discussed in [6], [7], [8]. Kaltenbach et al. [6] explicitly mention lack of scalability as one of the open issues.

While various approaches are available and applied, a systematic assessment of strengths and weaknesses of the state of the art methods is missing. In a previous study we analysed three methods [9]: a local method based on the Fisher Information Matrix (FIM) [10], a Bayesian method based on Markov chain Monte Carlo sampling (SAM) [11], and an ensemble modelling approach (ENS) based on optimization results [12]. We applied these methods to two case

- A. F. Villaverde is with CITMAga, 15782 Santiago de Compostela, Galicia, Spain, and with the Universidade de Vigo, Department of Systems and Control Engineering, 36310 Vigo, Galicia, Spain.  
E-mail: afvillaverde@uvigo.gal
- E. Raimúndez and J. Hasenauer are with Universität Bonn, 53113 Bonn, Germany, with Technische Universität München – Center for Mathematics, Garching, Germany, and with the Helmholtz Zentrum München, 85764 Neuherberg, Germany.  
E-mails: {elba.raimundez,jan.hasenauer}@uni-bonn.de
- J. R. Banga is with the Computational Biology Lab, Biological Mission of Galicia (Spanish National Research Council, MBG-CSIC), 36143 Pontevedra, Galicia, Spain.  
E-mail: j.r.banga@csic.es

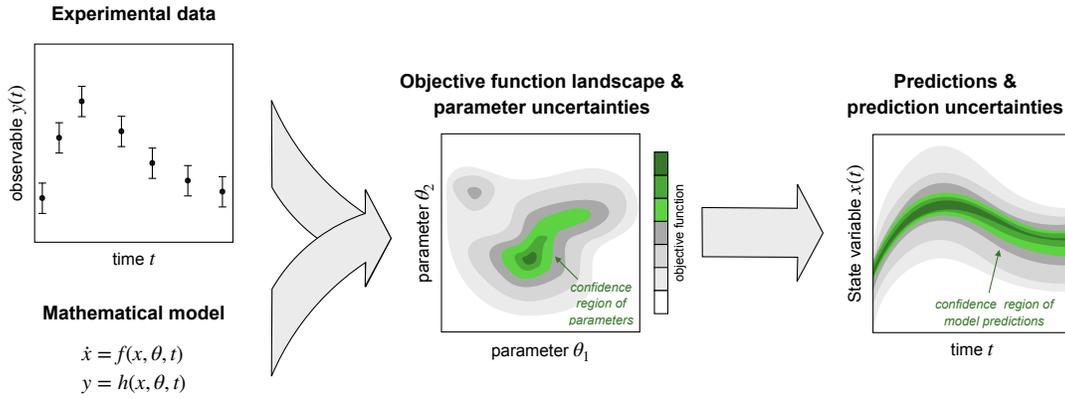


Fig. 1. Illustration of the general uncertainty quantification concept. (Left:) A mathematical model of a biological system usually has unknown parameters  $\theta$ . (Center:) These parameters are estimated by fitting the model to experimental measurements of functions  $y(t)$  of a subset of the states,  $x(t)$ . The resulting uncertainties in the estimated parameters can be described by confidence regions. (Right:) In turn, these uncertainties translate to uncertainties in the model predictions  $z$ ; in this work we focus on predictions about the time course of the state variables, i.e.  $z = x(t)$ .

studies with a small number of parameters. In the present work we address two important limitations of the aforementioned paper: (1) we consider prediction profile likelihood (PPL) [13], and (2) we enlarge the set of case studies to allow for a deeper evaluation of the performance of the methods, revealing the trade-offs between computational cost and statistical rigor. We are particularly interested in identifying methods which provide reliable estimates of the prediction uncertainty and scale well with model size (both in number of parameters and states). Based on these requirements, several types of existing methods were ruled out, including basic Monte Carlo sampling methods [6] and methods based on polynomial chaos expansions [14], which follow a rigorous approach with a sound mathematical basis, but currently can only be applied to small problems. We remark that in this work we consider the prediction uncertainty that results from propagation of parametric uncertainty. We do not take into account the possible uncertainty about the model structure. However, the methods considered here can also account for this type of uncertainty, by encoding alternative topologies as parameterized relationships in the ODEs.

The article is structured as follows: In Section 2, we describe the four selected methods, the metrics used for their evaluation, and the implementation details. In Section 3, we apply the methods to four case studies of increasing complexity, and evaluate their performance. In Section 4, we discuss the most relevant methodological aspects in view of the results. Finally, we present the conclusions of the comparison and provide guidelines for the application of the methods in Section 5.

## 2 METHODS

### Modeling framework and notation

We consider ODE models,

$$\begin{aligned} \dot{x} &= f(x, \theta, t), \quad x(t_0) = x_0(\theta), \\ y &= g(x, \theta, t), \end{aligned} \quad (1)$$

in which  $x(t) \in \mathbb{R}^{n_x}$  is the vector of state variables at time  $t$ ,  $y(t) \in \mathbb{R}^{n_y}$  is the vector of observables at time  $t$ , and

$\theta \in \mathbb{R}^{n_\theta}$  is the vector of unknown parameters. The vector field  $f : \mathbb{R}^{n_x} \times \mathbb{R}^{n_\theta} \times \mathbb{R} \mapsto \mathbb{R}^{n_x}$  and the mappings  $g : \mathbb{R}^{n_x} \times \mathbb{R}^{n_\theta} \times \mathbb{R} \mapsto \mathbb{R}^{n_y}$  and  $x_0 : \mathbb{R}^{n_\theta} \mapsto \mathbb{R}^{n_x}$  are possibly nonlinear.

The calibration of ODE models requires the estimation of  $\theta$  from measurements of  $y(t)$  at  $n_t$  times,  $t_i = t_1, t_2, \dots, t_{n_t}$ . The number of measurements is  $n_t \times n_y$ . In the application examples, the measurement noise follows a normal distribution,  $\epsilon_{k,i} \sim \mathcal{N}(0, \sigma_{k,i}^2)$ , where  $k = 1, \dots, n_y$  and  $\sigma_k(t_i)$  is the standard deviation. Thus, a noise-corrupted measurement of the  $k^{\text{th}}$  observable is  $\tilde{y}_{k,i} = y_k(t_i) + \epsilon_{k,i}$ . We denote the set of all measurement data as  $D$ . The maximum likelihood estimate of the vector of unknown parameters for a given dataset  $D$  can be found by minimizing the negative log-likelihood function:

$$J_{\text{nil}} = \frac{1}{2} \sum_{k=1}^{n_y} \sum_{i=1}^{n_t} \left[ \log(2\pi\sigma_{k,i}^2) + \left( \frac{\tilde{y}_{k,i} - y_k(t_i)}{\sigma_{k,i}} \right)^2 \right] \quad (2)$$

For the maximum a posterior estimate, the negative log-posterior is minimized,  $J_{\text{nlp}} = J_{\text{nil}} - J_{\text{np}}$ , in which the  $J_{\text{np}}$  denotes the negative logarithm of the prior evaluated at  $\theta$ . The search space for  $\theta$  is usually constrained, e.g. by lower and upper bounds, yielding  $\theta^L \leq \theta \leq \theta^U$ .

Predictions of the models are denoted by

$$z = h(x, \theta, t), \quad (3)$$

in which  $h : \mathbb{R}^{n_x} \times \mathbb{R}^{n_\theta} \times \mathbb{R} \mapsto \mathbb{R}^{n_z}$  represents a possibly nonlinear mapping. In this study, we focus on the assessment of the uncertainties in the time-dependent state variable, meaning that  $z = x$ . Yet, the considered analysis approaches are more flexible, which is why we provide the equations for generic functions  $h$ .

In Sections 2.1–2.4, we describe the four considered methods for the quantification of prediction uncertainties (Figure 2). For each method, we provide a definition of its prediction,  $x_j^p(t_i)$ , and of its uncertainty estimates,  $e_j^p(t_i)$ . Here,  $j$  indexes for the state variables ( $j = 1, \dots, n_x$ ) and  $i$  indexes the time points ( $i = 1, \dots, n_t$ ) (which do not necessarily have to be aligned with the time points of the observations). In Section 2.5, we define the metrics used to

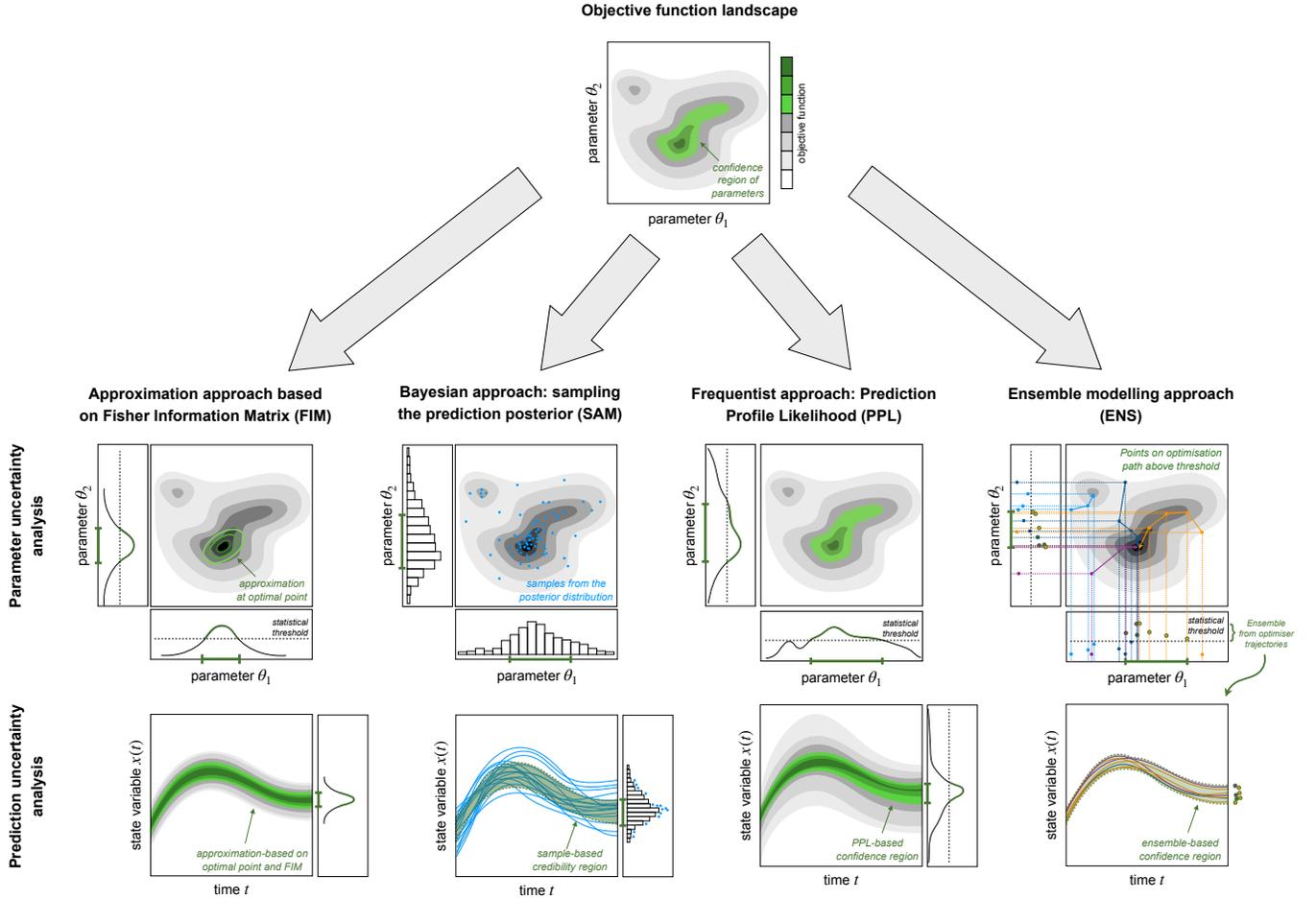


Fig. 2. Illustration of the employed uncertainty analysis methods. All methods seek to estimate the uncertainty in the time courses of state variables that results from uncertainties in the parameter estimations. Each method quantifies uncertainty in a different way, as defined by equations (8), (12), (15) and (18). Briefly, FIM approximates the prediction uncertainty as the standard deviation calculated from the square root of the prediction covariance matrix (8). SAM considers the credibility region as the distance between the 0.5th- and the 99.5th-percentile of the prediction samples, which are obtained by integration with the parameter samples drawn from the posterior distribution (12). PPL also approximates the confidence region as the width of the 99th percentile, and obtains the upper and lower levels by solving an optimization problem constrained by the prediction values (15). ENS adopts a similar approach, but builds the confidence region with those vectors explored during parameter estimation that yield an objective function below a certain threshold (18).

quantify the performance of the methods. In Section 2.6, we provide descriptions on the implementation.

## 2.1 Approximation approach based on Fisher Information Matrix (FIM)

For many decades parameters uncertainties were predominantly analyzed using methods based on asymptotic statistics. These methods are based on the assessment of the variability of the parameter estimates  $\hat{\theta}$ , given different replicates of the measurements. For globally identifiable models, the variability of the point estimates is described by the Fisher information matrix (FIM),

$$\text{FIM}(\theta) = \sum_{i=1}^{n_t} \sum_{k=1}^{n_y} \frac{1}{\sigma_{k,i}^2} \left( \frac{\partial y_k(t_i)}{\partial \theta} \right) \left( \frac{\partial y_k(t_i)}{\partial \theta} \right)^T, \quad (4)$$

in which  $\frac{\partial y_k(t_i)}{\partial \theta}$  denotes the sensitivity of the  $k$ -th observable with respect to the parameters  $\theta$ , evaluated at the measurement time point  $t_i$  and the parameter estimates  $\hat{\theta}$ .

As the sensitivity is the first order derivative, it provides information about the effect of small changes.

The Cramér-Rao theorem [15] states that, if  $\hat{\theta}$  is an unbiased estimate of  $\theta$  (i.e.  $E(\hat{\theta}) = \bar{\theta}$ ), the inverse of the FIM provides a lower bound of the covariance matrix,

$$\text{Cov}[\hat{\theta}] = E \left[ (\hat{\theta} - \bar{\theta}) (\hat{\theta} - \bar{\theta})^T \right] \geq \text{FIM}^{-1}(\hat{\theta}), \quad (5)$$

in which  $\bar{\theta}$  is the true parameter vector. The parameter covariance matrix informs about the individual and pairwise variability of parameter estimates along different realizations of the experimental data.

The parameter covariance matrix (and hence the FIM) can be used to estimate the uncertainty in predictions. To this end, the first order Taylor series expansion of the mapping  $h$  is used to propagate variability from the parameters to the predictions [16]. This yields the prediction covariance

matrix:

$$\text{Cov}[z(t)] = \left( \frac{\partial h}{\partial x} \frac{\partial x}{\partial \theta} + \frac{\partial h}{\partial \theta} \right) \text{Cov}(\hat{\theta}) \left( \frac{\partial h}{\partial x} \frac{\partial x}{\partial \theta} + \frac{\partial h}{\partial \theta} \right)^T, \quad (6)$$

with all derivatives being evaluated at time  $t$  for parameter  $\hat{\theta}$  and corresponding state  $x(t)$ .

The covariance matrices of  $\theta$  and  $z(t)$  can only be approximated using the inverse of the FIM if the FIM is invertible. If a single parameter is locally non-identifiable, this is not the case. In principle, this would preclude the application of this approach to unidentifiable models. A solution is to approximate the inverse with the Moore-Penrose pseudoinverse, as e.g. in [17].

The assessment of the predictions using the FIM yields for  $z_j$  at time  $t$ , the point estimate

$$z_j^p(t) = h_j(x(t, \hat{\theta}), \hat{\theta}, t), \quad (7)$$

which is the evaluation of model simulations with the optimal parameter vector  $\hat{\theta}$ . The uncertainty of the prediction as measured by the standard deviation is

$$z_j^p(t) = \sqrt{\text{Cov}_{jj}[z(t)]}, \quad (8)$$

in which  $\text{Cov}_{jj}[x^p(t_i)]$  is the  $j^{\text{th}}$  diagonal element of  $\text{Cov}[x^p(t_i)]$ . This approximate assessment is here denoted as FIM-based approach, but was in previous studies also referred to as Linear Covariance Analysis (LCA) (see [10]).

## 2.2 Bayesian approach: sampling the posterior predictive distribution (SAM)

In systems biology, measurements are often scarce and the application of asymptotic approaches arguable. Therefore, a broad spectrum of Frequentist and Bayesian uncertainty quantification methods have been introduced. In Bayesian statistics, the uncertainty of parameters is studied using the posterior distribution

$$p(\theta|D) = \frac{p(D|\theta)p(\theta)}{p(D)}, \quad (9)$$

in which  $p(D|\theta)$  denotes the likelihood of the data  $D$  given the parameters  $\theta$ ,  $p(\theta)$  denotes the prior distribution of  $\theta$ , and  $p(D)$  denotes the marginal probability.

The posterior  $p(\theta|D)$  encodes the available information about the parameters. Hence, it also describes the uncertainty about the parameters  $\theta$  taking into account the available experimental data,  $D$ , and the prior belief,  $p(\theta)$ . Likewise, one can define the posterior predictive distribution  $p(z|D)$ , which is obtained by integrating over the latent variables, simply speaking:

$$p(z|D) = \int p(z|\theta)p(\theta|D)d\theta. \quad (10)$$

The posterior distributions are usually not available in closed-form. In most cases their properties are assessed using sampling procedures such as Markov chain Monte Carlo methods (MCMC). Since these procedures are computationally expensive, it is crucial to use an efficient sampling technique. The adaptive parallel tempering algorithm combines the sampling from tempered posterior distributions with a local adaptation to improve sampling efficiency.

The algorithm provides  $n_s$  samples from the posterior distribution for the parameters,  $\{\theta^{(s)}\}_{s=1}^{n_s}$ , which can be used to quantify parameter uncertainties. The corresponding samples from the posterior predictive distribution are obtained by simulating the model for the sampled parameters,  $\{z^{(s)} = h(x(\theta^{(s)}, t), \theta^{(s)}, t)\}_{s=1}^{n_s}$ . These samples are then used to calculate the mean predictions and their associated uncertainties. The computationally demanding step is the sampling of the parameter posterior distribution. Once it has been computed, the calculation of the parameter and prediction uncertainties is efficient.

The mean prediction for  $z_j$  at time point  $t_i$  is

$$z_j^p(t_i) = \frac{1}{n_s} \sum_{s=1}^{n_s} z_j^{(s)}(t_i), \quad (11)$$

yet in many applications the prediction obtained for the maximum a posterior estimate might be preferred. To quantify the prediction uncertainty, the sample-based approximation of the marginal distribution is considered, common choices are the highest posterior density interval and the equal-tailed interval. Here, we choose the latter for a credibility level of 99%, meaning that the prediction uncertainty is the distance between the 0.5th- and the 99.5th-percentile of the samples of the prediction,  $z_j^{0.5}(t_i)$  and  $z_j^{99.5}(t_i)$ , which yields

$$z_j^p(t_i) = z_j^{99.5}(t_i) - z_j^{0.5}(t_i). \quad (12)$$

## 2.3 Frequentist approach: Prediction Profile Likelihood (PPL)

In contrast to sampling-based methods used in Bayesian statistics, frequentist approaches for uncertainty quantification often use profile likelihoods. Profile likelihoods provide a maximum projection of the likelihood on the parameter of interest,

$$\text{PL}_{\theta_l}(c) = \max_{\substack{\theta^L \leq \theta \leq \theta^U \\ \theta_l = c}} p(D|\theta). \quad (13)$$

The profile likelihood value  $\text{PL}_{\theta_l}(c)$  is the maximum value of the likelihood function attainable for  $\theta_l = c$ . The definition of statistical significance levels, e.g. based on the likelihood ratio test (relating to the  $\chi^2$ -distribution), yields the parameter confidence intervals.

Following this concept, Kreutz et al. [13] introduced the concept of prediction profile likelihoods (PPLs),

$$\text{PPL}_{x(\theta)}(c) = \max_{\substack{\theta^L \leq \theta \leq \theta^U \\ x(\theta) = c}} p(D|\theta). \quad (14)$$

Conceptually, the PPL provides the highest possible likelihood value for a specific value of the (parameter dependent) prediction. This translates to a optimization problem with an equality constraint for the value of the prediction. Yet, to avoid problems related to nonlinear equality constraints, a reformulation that relies on artificial data points has been presented [13]. Additionally, integration based techniques that are more efficient for the analysis of uncertainties of state trajectories have also been proposed [18].

For PPLs to provide information about the prediction uncertainties, it is necessary to define a statistical threshold.

We choose a confidence level of  $\alpha = 0.01$  related to the likelihood ratio test, which yields the uncertainty measure

$$e_j^p(t_i) = z_j^{\max}(t_i) - z_j^{\min}(t_i), \quad (15)$$

in which  $z_j^{\min}(t_i)$  and  $z_j^{\max}(t_i)$  are the minimal and maximal values of  $c$  for which  $\text{PPL}_{z_j(t_i)}(c)/p(D|\hat{\theta}) > \exp(-\Delta_\alpha/2)$ , respectively. The threshold parameter  $\Delta_\alpha$  is the  $\alpha$  percentile of the  $\chi^2$ -distribution. As in the FIM-based method, the PPL state prediction is the model simulation with the optimal parameter vector  $\hat{\theta}$ ,

$$z_j^p(t_i) = h_j(x(t, \hat{\theta}), \hat{\theta}, t) \quad (16)$$

## 2.4 Ensemble modelling approach (ENS)

The frequentist and Bayesian methods have more recently been complemented by ensemble modelling approaches. These approaches exploit that parameter optimization has already explored the parameter space [12]. The parameter vectors encountered during the optimization process which meet a certain quality criteria, e.g. reasonable log-likelihood values compared to the optimal point, are considered as an ensemble.

In practice, a diverse ensemble of parameter vectors is obtained by performing several optimizations with different random seeds and initial points. From the set of parameter vectors found during the optimizations, those with an objective function value smaller than a given threshold are included in the ensemble. Statistical interpretability is ensured by using a threshold according to the likelihood ratio test (as for parameter and prediction profile likelihoods). We use a confidence level of  $\alpha = 0.01$ . The resulting ensemble provides an envelope for parameter and model predictions. As the envelope is not pushed towards the boundaries (as done in profile calculation), the resulting parameter and prediction envelopes provide inner approximations compared to profile likelihood-based approaches.

The ensemble prediction is the average prediction of the predictions by the ensemble,

$$z_j^p(t_i) = \frac{1}{n_m} \sum_{m=1}^{n_m} z_j^m(t_i), \quad (17)$$

with  $z_j^m(t_i)$  denoting the prediction  $z_j$  at time  $t_i$  for the  $m^{\text{th}}$  model parameterisation in the ensemble, and  $n_m$  is the number of parameter vectors in the ensemble. While in principle the full range of the ensemble-based envelop could be used, previous works filter extreme results. Similar to the Bayesian method, the width of the 99th-percentile interval of the ensemble (12) was employed,

$$e_j^p(t_i) = z_j^{99.5}(t_i) - z_j^{0.5}(t_i). \quad (18)$$

with  $z_j^{\text{per}}(t_i)$  denoting the *per*-th percentile of the ensemble simulations.

## 2.5 Performance metrics

In this work we perform a comparative study based on published parameter problems. To facilitate a comprehensive assessment we used synthetic data with the same characteristics as the published datasets for the considered

problems. Furthermore, we used as predictions of interest the complete set of state variables.

To assess the performance of the methods we used the three metrics:

### Computational cost

We use the CPU time of all the calculations performed to solve a particular problem. We set a computational budget, i.e. a maximum time for each problem per method, of 1350 hours (somewhat less than two months).

### Agreement between predictions and true states

We assessed the agreement of changes in the predicted value  $x_j^p(t_i)$  and true value  $x_j(t_i)$  of the state variable  $x_j$  at time  $t_i$ . Therefore, we subtracted from both their averages over all time points,  $\overline{x_j^p}$  and  $\overline{x_j}$ . To obtain a dimensionless quantity between -1 and 1, we normalize with the respective variability. This yields the performance metric

$$\rho_{x_j} = \frac{\sum_{i=1}^{n_t} (x_j^p(t_i) - \overline{x_j^p})(x_j(t_i) - \overline{x_j})}{\sqrt{\sum_{i=1}^{n_t} (x_j^p(t_i) - \overline{x_j^p})^2 \sum_{i=1}^{n_t} (x_j(t_i) - \overline{x_j})^2}}, \quad (19)$$

for which a value of 1 indicates that the prediction and true values are perfectly aligned up to an offset and a scaling constant. The time grid used for the predictions was defined so as to resemble the original experimental data points, whenever available. The overall value of the performance metric for a model,  $\rho_x$ , is the average of the performance metrics of the state variables,  $\rho_x = \frac{1}{n_x} \sum_{j=1}^{n_x} \rho_{x_j}$ .

The equation of the performance metric is identical to the equation of the Pearson's correlation coefficient, yet, a statistical interpretation is potentially problematic.

### Agreement between uncertainty estimates and observed error

We assessed the agreement of changes in the predicted uncertainty  $e_j^p(t_i)$  and the actual error  $e_j(t_i) = |x_j^p(t_i) - x_j(t_i)|$  for state variable  $x_j$  at time  $t_i$ . Therefore, we subtracted from both their averages over all time points,  $\overline{e_j^p}$  and  $\overline{e_j}$ . To obtain a dimensionless quantity between -1 and 1, we normalize with the respective variability. This yields the performance metric

$$\rho_{e_j} = \frac{\sum_{i=1}^{n_t} (e_j^p(t_i) - \overline{e_j^p})(e_j(t_i) - \overline{e_j})}{\sqrt{\sum_{i=1}^{n_t} (e_j^p(t_i) - \overline{e_j^p})^2 \sum_{i=1}^{n_t} (e_j(t_i) - \overline{e_j})^2}}. \quad (20)$$

As before, a value of 1 indicates a good agreement of predicted uncertainties and actually errors, meaning that in the case of large predicted uncertainties also the error is large. The overall performance metric for a model,  $\rho_e$  is the average of the performance metrics of the state variables,  $\rho_e = \frac{1}{n_x} \sum_{j=1}^{n_x} \rho_{e_j}$ .

## 2.6 Implementation and availability

Parameter optimization tasks were performed using the MATLAB version of the enhanced scatter search (eSS) method included in the MEIGO optimization toolbox [19], combined with the MATLAB version of AMICI [20] for model simulation. The initial parameters were generated using latin hypercube uniform sampling.

For the PPL method we used an optimization-based approach [21], and for SAM we used an adaptive parallel tempering algorithm [11]. We used the implementation of both methods available in the MATLAB toolbox `DataDynamics` [22]. For numerical integration, `DataDynamics` and `AMICI` rely on the `SUNDIALS` solver `CVODES` [23]. For initialization of the algorithms, we used the optimal parameter vector found by the `eSS` method.

We implemented in-house MATLAB scripts for the FIM and ENS analyses, as well as for the evaluation of the MCMC samples to obtain the prediction posterior samples. In the FIM analyses of the unidentifiable models (all but the  $\alpha$ -pinene) we used the Moore-Penrose pseudoinverse as a replacement of the inverse of the FIM. When computing the pseudoinverse of a matrix  $A$ , singular values smaller than a threshold given by  $\max(\text{size}(A)) * \text{eps}(\text{norm}(A))$  were treated as zero, where `eps` is Matlab's floating point accuracy. The complete implementations of the case studies analysed in this work (including the MATLAB code, computational results and the respective version of the used toolboxes) are available on ZENODO (doi:10.5281/zenodo.5995941).

FIM and ENS results were obtained in a multi-core PC running Windows 7 64-bit with 16 GB RAM and 12 cores, Intel Xeon 2.30 GHz with MATLAB version R2017b. SAM and PPL results were obtained in a multi-core HPC cluster running Fedora 25 64-bit in a computing node using up to 16 cores allocating 300 MB of memory per core with MATLAB version R2017b. Both environments have similar – but not identical – computing power. To allow a fair comparison, we have scaled their CPU times according to the result of the LINPACK 100 benchmark<sup>1</sup>, which is a common measure of computational performance. To this end we used the C version of the LINPACK benchmark. Executing this test on both computers yielded 2301.46 MFLOPS in the Windows system and 3922.43 in the HPC cluster, i.e. a ratio of 1.70. Thus, we divided the CPU times of the Windows system by 1.70 and reported them in Fig. 3.(A).

### 3 CASE STUDIES AND RESULTS

To assess the performance of the four uncertainty analysis methods described in the preceding section, we applied them to four case studies of increasing complexity. They possess in the order of  $10^1$ ,  $10^2$ ,  $10^3$ , and  $10^4$  predictions; these numbers represent number of state variables times number of time points.

The main characteristics of the four case studies are shown in Table 1. In the following, we describe the application of the methods to each case study.

#### 3.1 Isomerization of $\alpha$ -pinene

As a first case study and a sanity check of the different methods, we considered a fully observed model for isomerization of  $\alpha$ -pinene [24]. It describes the thermal isomerization of  $\alpha$ -pinene to dipentene and allo-ocimene, which in turn yields pyronene and a dimer. There are thus five state variables in the model, all of which can be measured at nine time points, including the initial conditions. Assuming first order

TABLE 1

Main features of the case studies used in this paper: number of unknown parameters ( $n_\theta$ ), state variables ( $n_x$ ), measured observables ( $n_y$ ), time points ( $n_t$ ), and predictions ( $n_p = n_x \times n_t$ ).

	$n_\theta$	$n_x$	$n_y$	$n_t$	$n_p$
$\alpha$ -pinene	5	5	5	9	45
EGF	48	28	6	12	336
JAK/STAT	27	25	20	128	3 200
BM1	383	104	5	105	10 920

kinetics, the model has five rate constants that are the unknown parameters.

All parameters of this model were practically identifiable and the FIM could be inverted. For the ENS method we built an ensemble with 4000 parameter vectors through optimization. We found that this number could be obtained with a low computational cost, and adding more vectors did not alter the results. For the SAM method we created a Monte Carlo chain with 100 000 samples. Identifiability led to low dispersion in the parameter values included in the ensemble and in the SAM samples, as seen in Fig. 4.

All approaches were applicable and computation times were in the order of minutes for all methods (Fig. 3A). Furthermore, all methods achieved good agreement of predicted and true state (Fig. 3B) as well as predicted uncertainty and error (Fig. 3C). The prediction uncertainties were relatively small (Fig. 5).

#### 3.2 EGF signaling

As a second case study, we considered a model [25] that describes the nerve growth factor (NGF)-induced differentiation of neuronal cells. It models the effect of two growth factors, NGF and the mitogenic epidermal growth factor (EGF), in rat pheochromocytoma (PC12) cells. NGF and EGF phosphorylate extracellular regulated kinase (Erk) through different signaling pathways. The resulting model has 28 states, six of which are measured at twelve time points, and 48 unknown parameters. Thus, this model was larger than the model for isomerization of  $\alpha$ -pinene, and only partially observed. This hampered parameter identifiability, which led to a large uncertainty in the values of the parameters. Indeed, for some parameters a wide range of values allowed for a good fit to the data (Fig. 4). The results for the ENS method are based in 2 400 parameter vectors and the results for the SAM method on 38 000 samples.

For the experimental designs used in [25], the considered model for EGF signalling was locally practically non-identifiable. Due to this, the FIM is not invertible; hence, as mentioned in Subsection 2.1, we calculated the FIM-based uncertainties using the Moore-Penrose pseudoinverse. Non-identifiability, in turn, led to a decrease in the accuracy of the predictions. That said, all methods were able to obtain good predictions (Fig. 3B); however, accuracy of the estimate of the prediction errors decreased for most methods (Fig. 3C). The performance degradation was particularly notable in FIM, for which the agreement decreased roughly from 0.8 to 0.5. While the decrease of PPL was not as pronounced, this method only calculated uncertainty estimates for approximately 35% of the predictions before reaching the compu-

1. <https://netlib.org/utk/people/JackDongarra/faq-linpack.html>

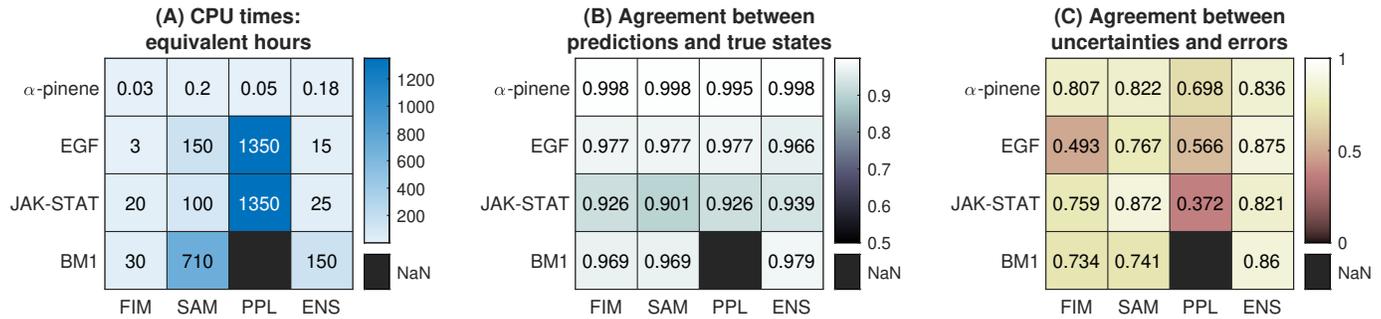


Fig. 3. **(A)** Approximate computation times, in hours, needed by each method and case study. Since the results were obtained using two different computing environments, the CPU times reported for FIM and ENS have been scaled to ensure a fair comparison, as explained in Section 2.6. For the FIM method, the computation time corresponds to the optimization used to obtain the optimum. For the ENS method it includes all the optimization runs used to obtain the parameter vectors in the ensemble. For SAM it includes the sampling time, which was performed with adaptive parallel tempering. We set a maximum CPU time per problem of 1350 hours (slightly less than two months); calculations with PPL hit this limit for all but the smallest model: for EGF and JAK-STAT, PPL completed 35% and 2% of the calculations in the allowed time, respectively. For the largest case study (BM1) PPL was deemed as not applicable (therefore shown in black). **(B–C)**: Agreement between predictions and true states (19) and between prediction uncertainty quantified by each state and actual prediction error (20), for all methods and case studies. The prediction uncertainties are those of the 95% percentile. Results for 99% and 68% were also calculated but are not shown here, since they do not vary for the FIM method and the differences are relatively small for SAM and ENS, showing that the metric is robust with respect to the choice of the confidence level. The values shown are the mean  $\pm$  the standard deviation for each state. For all but the smallest model ( $\alpha$ -pinene) the PPL method only produced results for a fraction of the predictions before exceeding the computation time limit. The percentage of predictions that could be calculated is shown with an asterisk. For the most computationally expensive model (BM1) this method was not applicable, which is noted as N/A.

tation time limit. Fig. 6 shows the results of the different methods for this case study for a representative subset of 10 of the 28 model state variables.

### 3.3 JAK/STAT signalling

As a third case study, we considered a model for JAK2/STAT5 signalling [26]. The purpose of the model is to elucidate the role of two transcriptional feedback regulators in erythropoiesis. The response to erythropoietin stimulation first activates receptor and JAK2 phosphorylation, and then phosphorylates the latent transcription factor STAT5. This model has 25 states and a relatively large number of outputs (20), but only a few of them are direct measurements of state variables, while most of them are functions of a subset of the states. Therefore, they do not provide as much information as it might seem at first sight. This model shared many characteristics with the model for EGF signalling, but the number of experimental conditions was substantially higher, resulting also in a ten-fold larger number of predictions. As before, the FIM was not invertible and the Moore-Penrose pseudoinverse was applied for FIM-based uncertainty quantification. Furthermore, the PPL calculation finished within the considered time constraints only for 2% of predictions.

The computations times for FIM, SAM and ENS methods were comparable (Fig. 3A). Furthermore, all methods achieved a good agreement between predicted and true state variables ( $\rho_x > 0.9$ ), but lower than for the two preceding case studies (Fig. 3B). The agreement between predicted uncertainty and actual error were  $\rho_e > 0.75$  except for PPL, which falls to 0.372. Fig. 7 depicts the predictions and uncertainties estimated by all methods for a representative subset of states.

### 3.4 Insulin signaling (BM1)

As a fourth case study, we considered a model for insulin signaling in mice [27]. It also considers the interaction of

insulin signaling with oxidative stress, and it includes transcriptional feedback through the FOXO transcription factor, which controls long-term adaptation. Thus, the model consists of several interconnected modules, with over one hundred states and almost four hundred parameters. However, only five states are measured. This was by far the largest and most computationally demanding model. Indeed, the PPL method – which already struggled with the two previous problems – did not complete any calculations within the allowed computation time. The performances of the other three methods were comparable to their performances on the two previous problems, albeit with a higher computational cost. Fig. 3 and Fig. 8 show the results of the different methods for this case study.

## 4 DISCUSSION OF METHODOLOGICAL ASPECTS

### 4.1 Approximation approach based on Fisher Information Matrix (FIM)

The FIM-based method is computationally cheap, but it has a number of limitations. First, it is strictly local, being calculated from a single parameter vector. In the presence of non-identifiability, the true vector can be very different from the estimated (optimal) one, affecting the results. Second, the confidence intervals estimated from the FIM are always symmetric, which might violate constraints (e.g. positivity bounds). Third, this method relies on a linearisation, and can be overly optimistic if nonlinearities are present. For the finite sample case, it is expected to give inaccurate results in the presence of strong nonlinearities. Our computational results have confirmed this theoretical expectation. Finally, if the model has non-identifiability issues, the FIM cannot be inverted and a pseudoinverse has to be used [17]. This was the case for three of the four case studies analysed here, and it is a very common scenario in systems biology models. However, to calculate the pseudoinverse it is necessary to specify a threshold, which may affect the results.

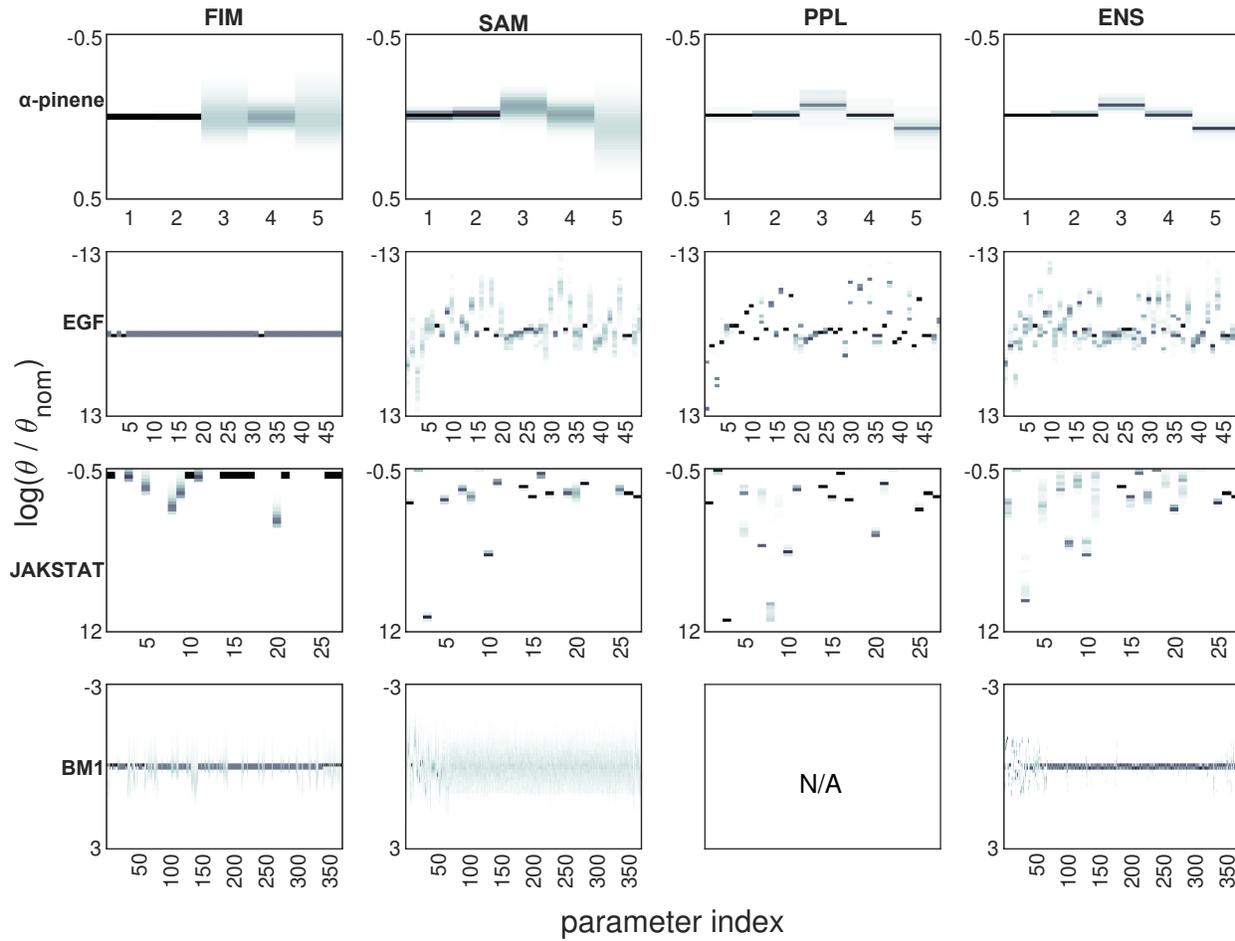


Fig. 4. Dispersion of the parameter values. For ENS and MCMC, we show the values of the parameter vectors included in the ensemble and in the MCMC samples, respectively. For FIM, we draw parameters from a normal distribution using the standard deviation calculated from the inverse of the FIM. For PPL, we show the optimal parameter vectors calculated at each step of the profile.

## 4.2 Bayesian approach: sampling the Posterior predictive distribution (SAM)

The Bayesian approach enables the assessment of uncertainties in a comprehensive manner. Yet, the posterior distribution needs to be approximated at first. This step tends to be computationally challenging, especially for increasingly complex models where MCMC methods tend to suffer from convergence issues. For the model of JAK/STAT signalling we found that the predictions for the state variables were “shifted” when compared to the ground truth (Fig. 7). This probably implies that the MCMC chain did not properly sample from the posterior distribution. Yet, this was not the case for the other models. Therefore, one should make sure that the samples are converged to the posterior distribution before further analyses.

## 4.3 Frequentist approach: Prediction Profile Likelihood (PPL)

Unlike the Bayesian approach, the use of PPLs require individual calculations for each model prediction. This renders the methods computationally demanding if a large number of predictions needs to be assessed. This has been reflected in the EGF, JAK-STAT and BM1 model, as only a fraction of the predictions were covered in the computation

time limit considered. In this regard, it must be noted that the high performance computing infrastructure used in the present study had a time limit of 48 hours for each job. Possibly, longer run times for individual jobs could allow more calculations to finish. We considered the use of advanced integration-based PPL calculation implemented in Data2Dynamics, but did not succeed. Hence, while providing stringent statistical guarantees, the use of PPLs is challenging.

## 4.4 Ensemble modelling approach (ENS)

As the ENS method exploits the results of parameter optimization, it is applicable even for high-dimensional models. Furthermore, if proper thresholds are used, the results can be interpreted as inner approximations to the confidence intervals obtained using PPLs. Yet, we encountered a number of open questions. First, the appropriate size for the ensemble is unclear. Unlike in the other methods, there is no clear-cut criterion to determine the number of parameter vectors that should be included in it. A practical solution is (1) allocate an affordable computational budget to the optimizations used to obtain the parameter vectors, (2) build test ensembles of increasing size, and (3) stop when adding more vectors does not change the ensemble predictions. If

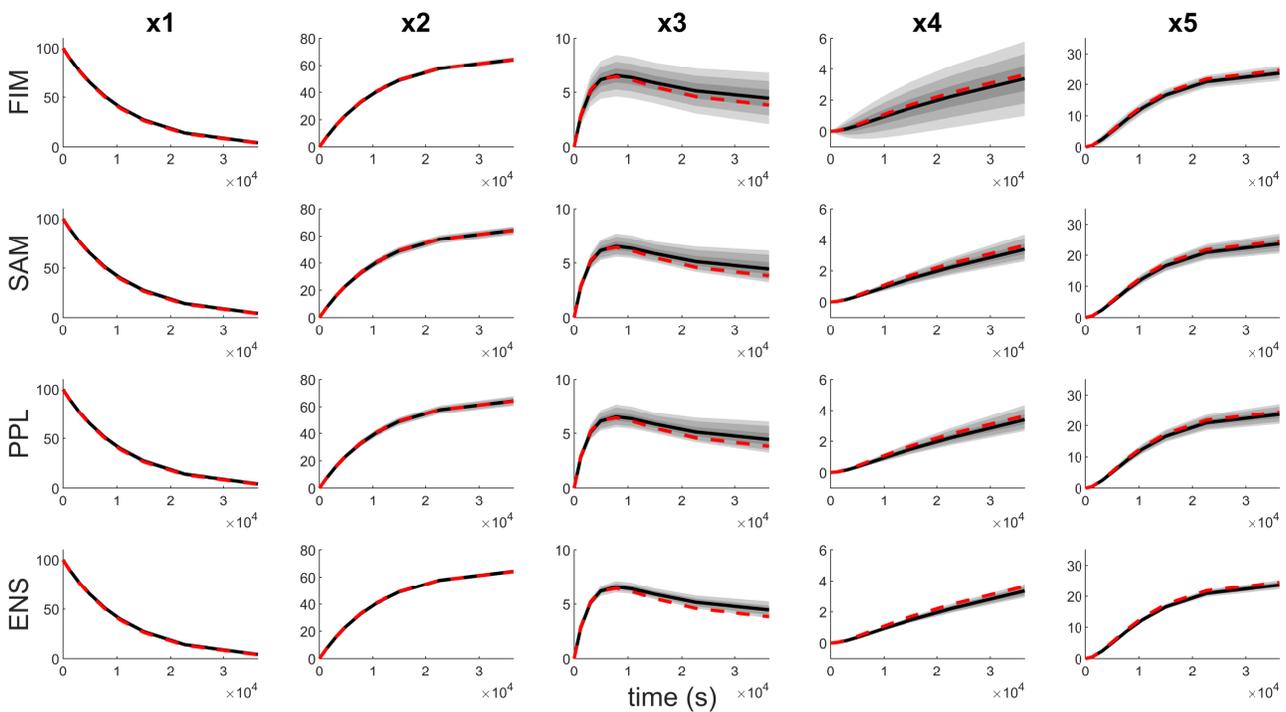


Fig. 5. Results of the different approaches for the  $\alpha$ -pinene example. The solid black lines are the predictions of the state trajectories, and the dashed red lines the true states. The grey areas show the percentiles of the predictions calculated with each method (dark grey: 68.27%, light grey: 95.45%, lighter grey: 99%).

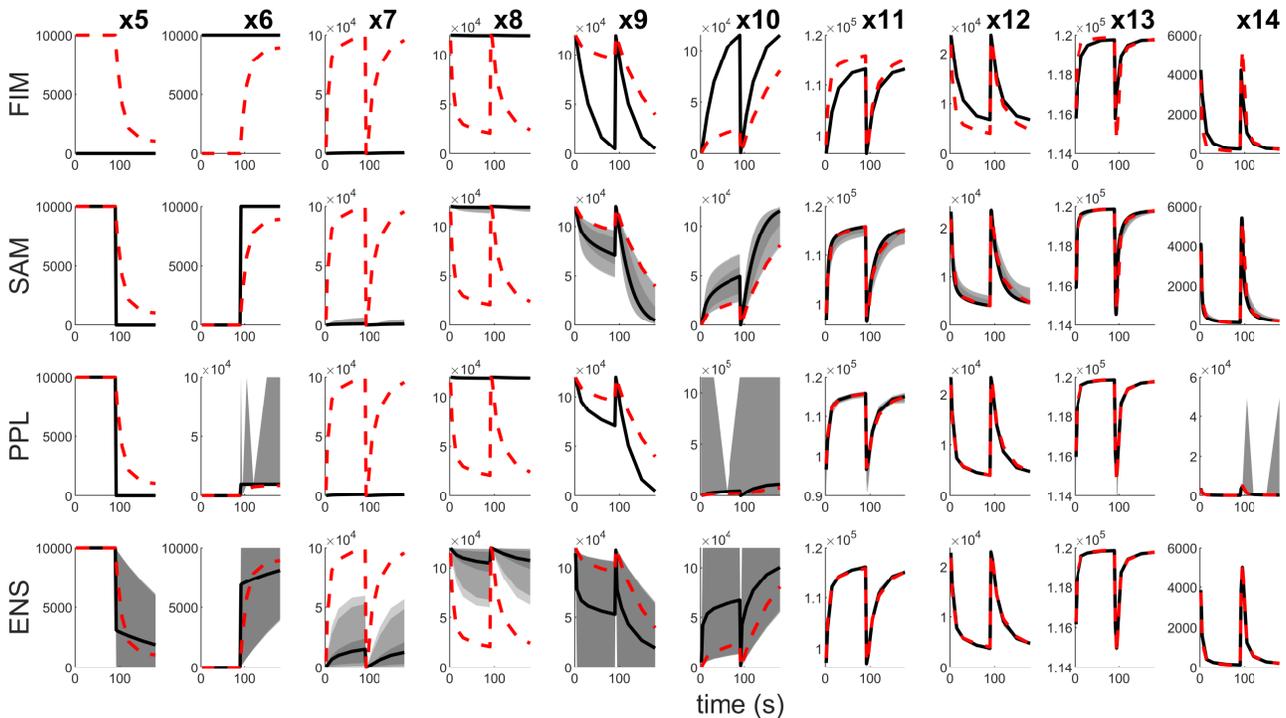


Fig. 6. Results of the different approaches for 10 representative states of the EGF signaling pathway example. The solid black lines are the predictions of the state trajectories, and the dashed red lines the true states. The grey areas show the percentiles of the predictions calculated with each method (dark grey: 68.27%, light grey: 95.45%, lighter grey: 99%). Note that, for this case study, PPL only produced results for a subset of the states within the allowed computation time.

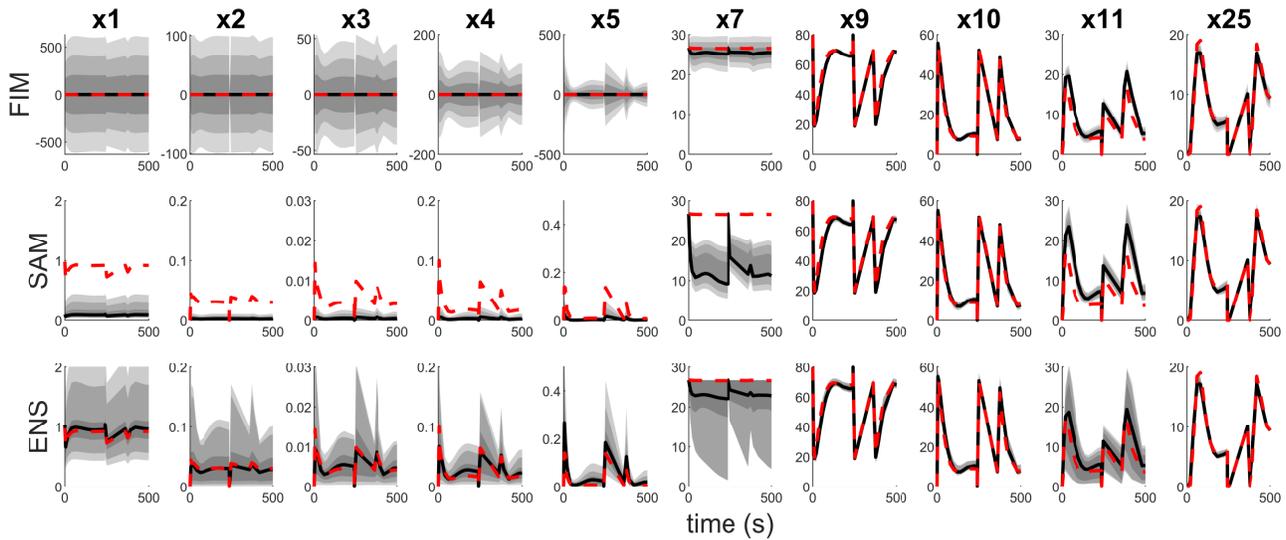


Fig. 7. Results of the different approaches for 10 representative states of the JAK-STAT example. Only the simulations corresponding to the first three experimental conditions are shown, to improve visualization. The solid black lines are the predictions of the state trajectories, and the dashed red lines the true states. The grey areas show the percentiles of the predictions calculated with each method (dark grey: 68.27%, light grey: 95.45%, lighter grey: 99%). Note that in the  $x_1$  case the Y axis scale is different for each method.

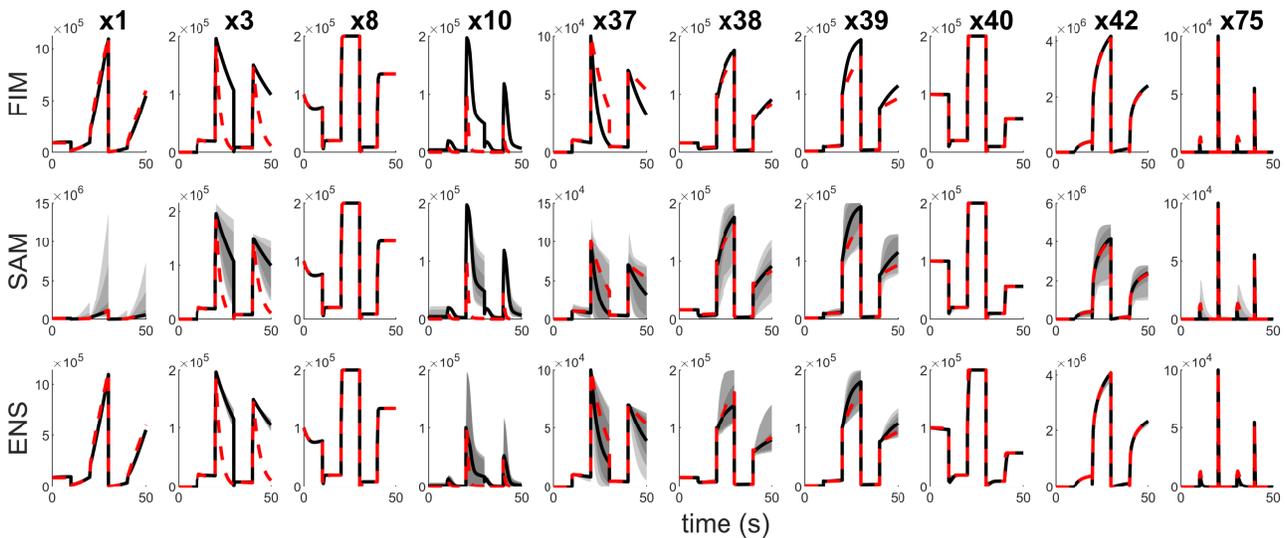


Fig. 8. Results of the different approaches for 10 representative states of the BM1 example. The solid black lines are the predictions of the state trajectories, and the dashed red lines the true states. The grey areas show the percentiles of the predictions calculated with each method (dark grey: 68.27%, light grey: 95.45%, lighter grey: 99%).

the maximum ensemble size attainable in this way is too small, more optimizations can be performed. This is the approach followed in the present work. Second, the criteria for including a parameter vector in the ensemble is unclear. It would be particularly interesting to know if a relaxation of the threshold would provide better approximations to the confidence intervals obtained using PPL methods (while sacrificing that a strict inner bound is obtained). Third, it remains open if additional criteria should be used (in the optimization process) to enforce diversity of the vectors in the ensemble and how this would effect the results. If the ensemble does not contain sufficient diversity, there is a risk of underestimating uncertainty. The way in which these

choices affect the results is worthy of further investigation.

## 5 CONCLUSIONS

In this paper we have compared four different approaches for uncertainty quantification in dynamic biological models: FIM, SAM, ENS, and PPL. These four methods estimate the uncertainty of the time-dependent state variables. We found that several factors should be taken into account when choosing a method for a specific problem.

In regard to *accuracy* of the uncertainty estimates, the four methods showed good agreement for the  $\alpha$ -pinene model. This is the simplest of the case studies considered, since it is the smallest one, linear, and fully observed. For

the larger and more complex (nonlinear) models, more differences appeared, as expected. For those three case studies, SAM and ENS yielded more accurate estimations of the uncertainty of the predictions than FIM. Yet, the confidence intervals often did not cover the true trajectory. This might be due to conceptual limitations or technical problems (e.g. convergence of the MCMC sampler for SAM). PPL calculations did not outperform FIM due to computation time limitations.

In regard to *statistical interpretability*, Bayesian and Frequentist approaches have arguably the most rigorous foundations. ENS is arguably the technique with less theoretical justification, although, since we have used it with an uncertainty metric defined in the same way as that of SAM, it could be regarded as a low-cost approximation of a Bayesian approach. ENS may also be considered as an inner approximation of PPL, which provides a lower bound on the uncertainty estimates, since by construction its envelopes are narrower than those obtained with a working PPL method.

Another key consideration is *computational cost*. The FIM-based method is the cheapest one, since it only requires one successful optimization in order to find an optimal parameter vector. The most expensive one is the PPL approach, which can become very expensive – and even inapplicable – for large models. The computational costs of ENS and SAM are in the same order of magnitude (although ENS is generally cheaper than SAM) and they are intermediate between FIM and PPL.

*Parallelization* is a way of reducing the wall clock time of the computations. In this regard, it should be noted that PPL is easily parallelizable, while other approaches such as FIM and ENS are not. Yet, the most computationally demanding step of these methods is the parameter optimization, which can be performed with different techniques. In principle, parallelizable strategies such as multistart optimization could be used to this end [28]; however, in the present work this step was performed with a metaheuristic optimization method that is less amenable to parallelization. SAM methods are generally more difficult to parallelize, but there are also approaches to exploit computation resources [29].

Our results suggest a *trade-off* between computational scalability, on the one hand, and accuracy and statistical rigor on the other. At one end of the trade-off there is the FIM-based method, which should be chosen only if the other approaches are computationally too expensive for the problem under consideration. At the other end there is the PPL method, whose computational cost hampers its application to high-dimensional problems. ENS and SAM lie between both extremes; while ENS has a lower computational cost, SAM provides a clearer statistical interpretation.

In this studies we did not assess the *flexibility* of the different uncertainty analysis approaches. While the four methods are generally applicable to every nonlinear ODE model, a difference may exist if not only the parameter values but also the model structure is uncertain. While in the present work we have not considered this possibility, such uncertainty can easily be taken into account in the ENS and the SAM framework by building an ensemble of models with different structures. For PPL and the FIM-based ap-

proach one could encode the existence of different possible structures using additional “on/off” parameters, turning the parameter estimation to a mixed-integer optimization problem. In general the topic is related to model averaging.

We note that in this study a specific set of state-of-the-art optimization, profile calculation and sampling methods was used. This selection influences the results. Yet, we made an effort to select efficient and robust approaches within the respective classes, based on previous benchmarking studies (e.g. [28], [30]) and own experiences. We expect that the qualitative findings are robust to the choice of the methods as well as models. Accordingly, we expect this assessment to be of broad relevance.

## ACKNOWLEDGMENTS

This research has received funding from the European Union’s Horizon 2020 research and innovation program under grant agreement No 686282 (CanPathPro). ER and JH acknowledge funding from the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) - TRR 333/1 - 450149205 and under Germany’s Excellence Strategy EXC-2047/1 - 390685813 and EXC 2151 - 390873048. AFV acknowledges funding from grant ED431F 2021/003 funded by Consellería de Cultura, Educación e Ordenación Universitaria, Xunta de Galicia; and from grant RYC-2019-027537-I funded by MCIN/AEI/ 10.13039/501100011033 and by “ESF Investing in your future”. JRB acknowledges funding from MCIN/AEI/ 10.13039/501100011033 through grant PID2020-117271RB-C22 (BIODYNAMICS)

## REFERENCES

- [1] A. F. Villaverde, D. Pathirana, F. Fröhlich, J. Hasenauer, and J. R. Banga, “A protocol for dynamic model calibration,” *Briefings in Bioinformatics*, p. bbab387, 2021.
- [2] J. Bondia, S. Romero-Vivo, B. Ricarte, and J. L. Diez, “Insulin estimation and prediction: A review of the estimation and prediction of subcutaneous insulin pharmacokinetics in closed-loop glucose control,” *IEEE Control Systems Magazine*, vol. 38, no. 1, pp. 47–66, 2018.
- [3] M. Castro, S. Ares, J. A. Cuesta, and S. Manrubia, “The turning point and end of an expanding epidemic cannot be precisely forecast,” *Proceedings of the National Academy of Sciences*, vol. 117, no. 42, pp. 26 190–26 196, 2020.
- [4] F.-G. Wieland, A. L. Hauber, M. Rosenblatt, C. T’onsing, and J. Timmer, “On structural and practical identifiability,” *Current Opinion in Systems Biology*, vol. 25, pp. 60–69, 2021.
- [5] S. Streif, K.-K. Kim, P. Rumschinski, M. Kishida, D. E. Shen, R. Findeisen, and R. D. Braatz, “Robustness analysis, prediction, and estimation for uncertain biochemical networks: An overview,” *Journal of Process Control*, vol. 42, pp. 14–34, 2016.
- [6] H.-M. Kaltenbach, S. Dimopoulos, and J. Stelling, “Systems analysis of cellular networks under uncertainty,” *FEBS letters*, vol. 583, no. 24, pp. 3923–3930, 2009.
- [7] G. Cedersund, “Prediction uncertainty estimation despite unidentifiability: an overview of recent developments,” in *Uncertainty in Biology*. Springer, 2016, pp. 449–466.
- [8] J. Vanlier, C. Tiemann, P. Hilbers, and N. van Riel, “Parameter uncertainty in biochemical models described by ordinary differential equations,” *Mathematical Biosciences*, vol. 246, pp. 305–314, 2013.
- [9] A. F. Villaverde, E. Raimúndez, J. Hasenauer, and J. R. Banga, “A comparison of methods for quantifying prediction uncertainty in systems biology,” *IFAC-PapersOnLine*, vol. 52, no. 26, pp. 45–51, 2019.
- [10] R. Gutenkunst, F. Casey, J. Waterfall, C. Myers, and J. Sethna, “Extracting falsifiable predictions from sloppy models,” *Reverse Engineering Biological Networks: Opportunities and Challenges in Computational Methods for Pathway Inference Annals of the New York Academy of Sciences*, vol. 1115, no. 203–211, pp. 1115–203, 2007.

- [11] B. Miasojedow, E. Moulines, and M. Vihola, "An adaptive parallel tempering algorithm," *Journal of Computational and Graphical Statistics*, vol. 22, no. 3, pp. 649–664, 2013.
- [12] A. F. Villaverde, S. Bongard, K. Mauch, D. Müller, E. Balsa-Canto, J. Schmid, and J. R. Banga, "A consensus approach for estimating the predictive accuracy of dynamic models in biology," *Computer methods and programs in biomedicine*, vol. 119, no. 1, pp. 17–28, 2015.
- [13] C. Kreutz, A. Raue, and J. Timmer, "Likelihood based observability analysis and confidence intervals for predictions of dynamic models," *BMC Systems Biology*, vol. 6, no. 1, p. 120, 2012.
- [14] S. Yang, F. Xiong, and F. Wang, "Polynomial chaos expansion for probabilistic uncertainty propagation," *Uncertainty Quantification and Model Calibration*, p. 13, 2017.
- [15] H. Cramér, *Mathematical Methods of Statistics (PMS-9)*. Princeton university press, 2016, vol. 9.
- [16] F. Geier, G. Fengos, F. Felizzi, and D. Iber, "Analyzing and Constraining Signaling Networks: Parameter Estimation for the User," in *Computational Modeling of Signaling Networks*, ser. Methods in Molecular Biology, X. Liu and M. D. Betterton, Eds. Totowa, NJ: Humana Press, 2012, vol. 880, pp. 23–40.
- [17] A. Shahmohammadi and K. B. McAuley, "Sequential model-based  $\alpha$ -optimal design of experiments when the fisher information matrix is noninvertible," *Industrial & Engineering Chemistry Research*, vol. 58, no. 3, pp. 1244–1261, 2019.
- [18] H. Hass, C. Kreutz, J. Timmer, and D. Kaschek, "Fast integration-based prediction bands for ordinary differential equation models," *Bioinformatics*, vol. 32, no. 8, pp. 1204–1210, 2016.
- [19] J. A. Egea, D. Henriques, T. Cokelaer, A. F. Villaverde, A. MacNamara, D.-P. Danciu, J. R. Banga, and J. Saez-Rodriguez, "Meigo: an open-source software suite based on metaheuristics for global optimization in systems biology and bioinformatics," *BMC bioinformatics*, vol. 15, no. 1, p. 136, 2014.
- [20] F. Fröhlich, D. Weindl, Y. Schälte, D. Pathirana, L. Paszkowski, G. T. Lines, P. Stapor, and J. Hasenauer, "Amici: High-performance sensitivity analysis for large ordinary differential equation models," *Bioinformatics*, 04 2021, btab227.
- [21] C. Kreutz, A. Raue, D. Kaschek, and J. Timmer, "Profile likelihood in systems biology," *FEBS Journal*, vol. 280, no. 11, pp. 2564–2571, May 2013. [Online]. Available: <https://doi.org/10.1111/febs.12276>
- [22] A. Raue, B. Steiert, M. Schelker, C. Kreutz, T. Maiwald, H. Hass, J. Vanlier, C. Tönsing, L. Adlung, R. Engesser *et al.*, "Data2dynamics: a modeling environment tailored to parameter estimation in dynamical systems," *Bioinformatics*, vol. 31, no. 21, pp. 3558–3560, 2015.
- [23] R. Serban and A. C. Hindmarsh, "CVODES: The sensitivity-enabled ODE solver in SUNDIALS," in *Volume 6: 5th International Conference on Multibody Systems, Nonlinear Dynamics, and Control, Parts A, B, and C*. ASME/EDC, Jan. 2005. [Online]. Available: <https://doi.org/10.1115/detc2005-85597>
- [24] G. Box, W. Hunter, J. MacGregor, and J. Erjavec, "Some problems associated with the analysis of multiresponse data," *Technometrics*, vol. 15, no. 1, pp. 33–51, 1973.
- [25] K. Brown, C. Hill, G. Calero, C. Myers, K. Lee, J. Sethna, and R. Cerione, "The statistical mechanics of complex signaling networks: nerve growth factor signaling," *Physical biology*, vol. 1, pp. 184–195, 2004.
- [26] J. Bachmann, A. Raue, M. Schilling, M. E. Böhm, C. Kreutz, D. Kaschek, H. Busch, N. Gretz, W. D. Lehmann, J. Timmer *et al.*, "Division of labor by dual feedback regulators controls jak2/stat5 signaling over broad ligand range," *Molecular Systems Biology*, vol. 7, no. 1, p. 516, 2011.
- [27] G. R. Smith and D. P. Shanley, "Computational modelling of the regulation of insulin signalling by oxidative stress," *BMC systems biology*, vol. 7, no. 1, p. 41, 2013.
- [28] A. F. Villaverde, F. Froehlich, D. Weindl, J. Hasenauer, and J. R. Banga, "Benchmarking optimization methods for parameter estimation in large kinetic models," *Bioinformatics*, vol. 35, no. 5, pp. 830–838, 2019.
- [29] R. C.P., E. V., T. N., and W. C., "Accelerating mcmc algorithms," *Wiley Interdiscip. Rev. Comput. Stat.*, vol. 10, no. 5, p. e1435, 2018.
- [30] B. Ballnus, S. Schaper, F. J. Theis, and J. Hasenauer, "Bayesian parameter estimation for biochemical reaction networks using

region-based adaptive parallel tempering," *Bioinformatics*, vol. 34, no. 13, pp. i494–i501, July 2018.



Alejandro F. Villaverde received the M.Eng. degree in electrical engineering in 2004, and the Ph.D. degree in control engineering in 2009, both from the Universidade de Vigo, Vigo, Galicia, Spain. From 2009 to 2015, and from 2016 to 2020, he was a Postdoctoral Researcher at the Bioprocess Engineering Group, IIM-CSIC. From 2015 to 2016 he was an I2C Postdoctoral Fellow with the Universidade de Vigo. He is currently a Ramón y Cajal Fellow at the Department of Systems Engineering and Control, Universidade de Vigo, and an affiliated researcher with CITMaga, the Galician Center for Mathematical Research and Technology. His research focuses on the identification, control, and optimization of nonlinear biosystems. Dr. Villaverde is an editor of PLOS ONE, Complexity, and Symmetry.



Elba Raimúndez received the B.Sc. degree in biochemistry from Universidad Autónoma de Madrid, Madrid, Spain, in 2014, and the M.Sc. degree in bioinformatics for health sciences from Universitat Pompeu Fabra, Barcelona, Catalonia, Spain. From 2017 to 2019, she was with the group of Prof. Hasenauer at the Institute of Computational Biology at the Helmholtz Zentrum München, Munich, Germany. She transferred with Prof. Hasenauer to the University of Bonn, Bonn, Germany, where she is currently a research assistant. Her research interests include the modelling, optimization and uncertainty quantification of biological processes.



Jan Hasenauer studied technical cybernetics at the University of Stuttgart and the University of Wisconsin, Madison. After his studies, he received his PhD in Systems Biology in February 2013. A few months later, he became team leader at the Institute of Computational Biology at Helmholtz Zentrum München. In 2015, he was awarded an independent junior research group at Helmholtz Zentrum München. Since October 2018, he is professor for Mathematics & Life Sciences at the University of Bonn. The research of Prof. Hasenauer focuses on the development of methods for data-driven modelling of biological processes. These methods enable model-based integration of different data sets, critical evaluation of available information, comparison of different biological hypotheses and tailor-made selection of future experiments.



Julio R. Banga received a PhD in Chemical Engineering in 1991 from the University of Santiago de Compostela (Spain). He is currently a Research Professor at the Spanish National Research Council (CSIC), leading the Computational Biology Lab at MBG-CSIC (Pontevedra, Spain). Prof. Banga has extensive experience in computational systems biology, with emphasis on reverse engineering, optimal experimental design and dynamic optimization of biological systems. He is a member of the Editorial Board of BMC Bioinformatics, and of the IFAC Technical Committee on Control of Biotechnological Processes.