Available online at www.sciencedirect.com

ScienceDirect

IFAC Papers Online CONFERENCE PAPER ARCHIVE

IFAC PapersOnLine 52-26 (2019) 45–51

A Comparison of Methods for Quantifying Prediction Uncertainty in Systems Biology ⋆

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

* *Bioprocess Engineering Group, IIM-CSIC, Vigo 36208, Galicia, Spain (e-mails: {afvillaverde, julio}@ iim.csic.es).*
** *Helmholtz Zentrum München, 85764 Neuherberg, Germany*
*** *Center for Mathematics, Chair of Mathematical Modeling of Biological Systems, Technische Universitt München, Garching 85748, Germany (e-mail: elba.raimundez@helmholtz-muenchen.de)*
**** *Universität Bonn, 53113 Bonn, Germany (e-mail: jan.hasenauer@uni-bonn.de)*

Abstract: The parameters of dynamical models of biological processes always possess some degree of uncertainty. This parameter uncertainty translates into an uncertainty of model predictions. The trajectories of unmeasured state variables are examples of such predictions. Quantifying the uncertainty associated with a given prediction is an important problem for model developers and users. However, the nonlinearity and complexity of most dynamical models renders it nontrivial. Here, we evaluate three state-of-the-art approaches for prediction uncertainty quantification using two models of different sizes and computational complexities. We discuss the trade-offs between applicability and statistical interpretability of the different methods, and provide guidelines for their application.

Keywords: Computational methods, Dynamic models, Nonlinear systems, Observability, Prediction error methods, State estimation, Uncertainty

1. INTRODUCTION

Many biological processes are modelled using systems of ordinary differential equations (ODEs). These ODE models usually have unknown parameters that can be estimated from data. If parameter estimates are available, it is possible to simulate the model and thus obtain the time courses of its state variables, e.g., concentrations of biochemical species.

Model simulations allow for the generation of a broad spectrum of model predictions, e.g. the response of a biological process to perturbations. However, if the parameter values are estimated and subject to uncertainties, the model predictions are also uncertain. The assessment of the prediction uncertainties is crucial and a variety of methods have been proposed (see (Cedersund, 2016; Kaltenbach et al., 2009) for an overview). However, for non-linear ODE models with dozens or hundreds or parameters, the quantification of prediction uncertainties is still challenging.

In this study, we compare available methods for the assessment of prediction uncertainties. We consider non-linear ODEs with partial observations which are subject to substantial measurement noise. For this challenging problem class, we assess the prediction uncertainties of state variables. As state variables encode the information about the dynamics of the systems, the prediction of them is of key interest.

The manuscript is structured as follows: In Section 2, we describe three methods for quantifying prediction uncertainty and the metrics used for their evaluation. In Section 3, we assess their suitability to dynamic systems biology models. Finally, we discuss the results and provide some guidelines in Section 4.

2. METHODS

In this work we consider dynamic models described by nonlinear ODEs of the following form:

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

in which $x(t) \in \mathbb{R}^{n_x}$ is the state vector at time t with initial conditions $x_0(\theta)$, $y(t) \in \mathbb{R}^{n_y}$ is the output vector at time t, f and g are possibly nonlinear functions, and $\theta \in \mathbb{R}^{n_\theta}$ is the unknown parameter vector. To calibrate model (1) we need to estimate θ.

We assume that the measurements of the outputs y are noise-corrupted, $\bar{y}_k(t_i) = y_k(t_i) + \varepsilon_k(t_i)$ for $k = 1, \dots n_y$ and $i = 1, \dots n_t$, with normally distributed measurement

⋆ This research has received funding from the European Unions Horizon 2020 research and innovation program under grant agreement No 686282 (CanPathPro) and the German Ministry of Education and Research (BMBF) under the grant agreement No 01ZX1310B (SYS-Stomach) and No 01ZX1705A (INCOME).

Peer review under responsibility of International Federation of Automatic Control.
10.1016/j.ifacol.2019.12.234

noise $\epsilon_k(t_i) \sim \mathcal{N}(0, \sigma_k^2(t_i))$. The maximum likelihood estimation of the parameters is than obtained by minimizing the negative log-likelihood function:

$$J_{\text{nll}} = \frac{1}{2} \sum_{k=1}^{n_y} \sum_{i=1}^{n_t} \left[\log \left(2\pi(\sigma_k^2(t_i)) \right) + \left(\frac{\tilde{y}_k(t_i) - y_k(t_i)}{\sigma_k(t_i)} \right)^2 \right] \tag{2}$$

subject to the bounds $\theta^L \leq \theta \leq \theta^U$. Here n_t is the number of measurement time points, $y_k(t_i)$ is the k^{th} component of the model output vector at the i^{th} time point, and $\tilde{y}_k(t_i)$ and $\sigma_k(t_i)$ are the corresponding measurement and standard deviation, respectively.

Subsections 2.1–2.3 describe the uncertainty prediction methods considered in this paper. Subsection 2.4 defines the metrics used for comparing the performance of the methods.

2.1 Uncertainty propagation with the Fisher Information Matrix (FIM)

The Fisher information matrix (FIM) provides information about the parameter uncertainty and practical identifiability. For a set of n_t measurements it can be calculated as

$$\text{FIM}(\theta) = \sum_{i=1}^{n_t} \left(\frac{\partial y(t_i)}{\partial \theta} \right) W^{(i)} \left(\frac{\partial y(t_i)}{\partial \theta} \right)^T \tag{3}$$

where $\frac{\partial \mathbf{y}(t_i)}{\partial \theta}$ are the sensitivities of the observables with respect to the parameters, and $W^{(i)}$ is a diagonal matrix with $W_{kk}^{(i)} = 1/\sigma_k^2(t_i)$. The Cramér-Rao theorem (Cramér, 2016) states that, if $\hat{\theta}$ is an unbiased estimate of θ (i.e. $E(\hat{\theta}) = \bar{\theta}$), the inverse of the FIM is a lower bound estimate of the covariance matrix,

$$\text{Cov}(\hat{\theta}) \geq \text{FIM}^{-1}(\hat{\theta}) \tag{4}$$

The covariance matrix provides information about variability of estimates of individual parameters and of parameter pairs across different realizations of the experimental data. It is defined as:

$$\text{Cov}(\hat{\theta}) = E \left[\left(\hat{\theta} - \bar{\theta} \right) \left(\hat{\theta} - \bar{\theta} \right)^T \right] = \tag{5}$$

$$= \begin{bmatrix} \text{var}(\hat{\theta}_1) & \cdots & \text{cov}(\hat{\theta}_1, \hat{\theta}_{n_\theta}) \\ \vdots & \ddots & \vdots \\ \text{cov}(\hat{\theta}_{n_\theta}, \hat{\theta}_1) & \cdots & \text{var}(\hat{\theta}_{n_\theta}) \end{bmatrix} \tag{6}$$

If the FIM is invertible, it is possible to approximate the uncertainty in the state trajectories by error propagation from the parameter estimates (Geier et al., 2012) as

$$\text{Cov}[x(t)] = \frac{\partial x(t)}{\partial \theta} \text{Cov}(\hat{\theta}) \frac{\partial x(t)}{\partial \theta}^T \tag{7}$$

In the FIM-based method the predicted value of the j^{th} state at time t_i is obtained by simulating the model with the optimal parameter vector $\hat{\theta}$, that is

$$x_j^p(t_i) = x_j(t_i, \hat{\theta}) \tag{8}$$

where we have made explicit the dependency on $\hat{\theta}$. The predicted state vector is $x^p(t_i)$. In this work we performed parameter estimation using the MATLAB version of the MEIGO toolbox (Egea et al., 2014).

The estimate of the uncertainty of the prediction about state x_j at time t_i is the standard deviation,

$$e_j^p(t_i) = \sqrt{\text{Cov}_{jj}[x^p(t_i)]}, \tag{9}$$

with $\text{Cov}_{jj}[x^p(t_i)]$ denoting the jth diagonal element of the covariance matrix $\text{Cov}[x^p(t_i)]$.

This is the same approach that Gutenkunst et al. (2007) called Linear Covariance Analysis (LCA). Its caveats are that confidence intervals estimated from the FIM are always symmetric. This might violate constraints (e.g. positivity bounds) and can be overly optimistic if nonlinearities are present, since they rely on a linearisation of the models. Furthermore, if the FIM is not invertible, as happens if there are unidentifiable parameters, this approach cannot be applied in principle. To approximate the results, the Moore-Penrose pseudoinverse of the FIM (Shahmohammadi and McAuley, 2019) can be employed. This is the solution used in this work.

2.2 Prediction Posterior (PP)

In Bayesian parameter estimation, the (parameter) posterior

$$p(\theta|D) = \frac{p(D|\theta)p(\theta)}{p(D)} \tag{10}$$

is considered, in which $p(\theta)$ denotes the prior, $p(D|\theta)$ denotes the likelihood of the data D given the parameters and $p(D)$ denotes the marginal probability. The posterior $p(\theta|D)$ encodes the available information about the parameters θ given the experimental data D and the prior information $p(\theta)$. Accordingly, the information about a model prediction $h(\theta)$ is encoded in the prediction posterior,

$$p(h|D) = \int p(h|\theta)p(\theta|D)d\theta \tag{11}$$

As the posterior distributions are in general not available in closed-from, sampling methods are used to assess their properties. Most widely used are Markov chain Monte Carlo methods (MCMC) which construct a sequence of points. For this study, we employed an adaptive parallel tempering algorithm (Miasojedow et al., 2013) implemented in the MATLAB parameter estimation toolbox PESTO (Stapor et al., 2017). This algorithm combines the sampling from tempered posterior distributions with a local adaptation to enhance the sampling efficiency. The method yields samples from the posterior distribution, $\{\theta^{(k)}\}_{k=1}^S$, which can be used to assess parameter uncertainties. The evaluation of the predictions for the sampled parameters, $\{h^{(k)} = h(\theta^{(k)})\}_{k=1}^S$, yields a sample from the prediction posterior. This can be used to assess the mean prediction as well as the prediction uncertainties.

The mean prediction of the state variable x_j at time point t_i is

$$x_j^p(t_i) = \frac{1}{S} \sum_{k=1}^{S} h^{(i)} \tag{12}$$

for $h(\theta) = x_j(t_i, \theta)$. As a measure of the uncertainty, we consider the distance between the 0.5th- and the 99.5th-percentile of the sample of prediction, $x_j^{per=0.5}(t_i)$ and $x_j^{per=99.5}(t_i)$, yielding

$$e_j^p(t_i) = x_j^{per=99.5}(t_i) - x_j^{per=0.5}(t_i). \tag{13}$$

We note that the computationally demanding step is the sampling of the parameter posterior distribution. The evaluation of parameter and prediction uncertainties is very efficient.

2.3 Ensemble Consensus (ENS)

As an alternative to established approaches we consider an ensemble approach for uncertainty analysis as presented in (Villaverde et al., 2015), where an ensemble of models was built with different parameter vectors obtained from calibrations. By simulating the models in the ensemble an envelope of predictions is obtained, the mean of which is considered as the ensemble prediction. A measure of the dissensus among model outputs is used as a proxy for the prediction uncertainty.

The parameter vectors included in the ensemble are obtained from optimization runs. To this end it is necessary to store not only the final optimum, but also other good parameter vectors evaluated during the optimization. As in the FIM method, we used the MATLAB version of the MEIGO toolbox (Egea et al., 2014) for parameter estimation, performing several optimizations in order to obtain a diverse ensemble. A parameter vector is accepted in the ensemble if its objective function value falls below a given threshold. To facilitate a statistical interpretability, we consider the same threshold as used in profile calculation (Raue et al., 2009). This objective function threshold is the negative log-likelihood function at the optimal point plus 1/2 times the α percentile of the χ^2-distribution, $J_{\text{nll}}(\theta^*) + \Delta_\alpha/2$. We choose a confidence level of $\alpha = 0.01$.

The ensemble prediction and the uncertainty are mathematically defined as follows. Let us denote the prediction of the dynamic state x_j at time t_i made by the k^{th} model parameterisation in the ensemble as $x_j^k(t_i)$. The average prediction of all models is an array of dimensions $n_t \times n_x$, the elements of which are:

$$x_j^p(t_i) = \frac{1}{n_m} \sum_{k=1}^{n_m} x_j^k(t_i) \tag{14}$$

where n_m is the number of parameter vectors in the ensemble. This average, $x_j^p(t_i)$, is the ensemble prediction.

In (Villaverde et al., 2015) a dissensus-based metric was used for quantifying uncertainty. The dissensus measures the variability in the predictions made by the ensemble models about a state, and is thus an indication of the uncertainty of the prediction of said state. Here we use for that purpose the width of the 99th-percentile interval of the ensemble (13), which gives similar results as the dissensus-based metric.

2.4 Assessment metrics

In this work we perform a comparative study using identification problems with synthetic data, where the actual solutions are known. To quantify the agreement between the predictions made by each method and the actual states, we use the sample correlation coefficient:

$$\rho_{x_j} = \frac{\sum_{i=1}^{n_t}(x_j^p(t_i) - \overline{x_j^p})(x_j(t_i) - \overline{x}_j)}{\sqrt{\sum_{i=1}^{n_t}(x_j^p(t_i) - \overline{x_j^p})^2 \sum_{i=1}^{n_t}(x_j(t_i) - \overline{x}_j)^2}} \tag{15}$$

where $x_j^p(t_i)$ is the prediction of a particular method for state x_j at time t_i, $x_j(t_i)$ is the actual value, and $\overline{x_j^p}$ and $\overline{x}_j$ are their averages along all time points. The overall correlation coefficient for a model, ρ_x, is the average of the correlation coefficients of the states, $\rho_x = \frac{1}{n_x}\sum_{j=1}^{n_x} \rho_{x_j}$.

To quantify the agreement between the uncertainty estimated by each method and the actual error we use

$$\rho_{e_j} = \frac{\sum_{i=1}^{n_t}(e_j^p(t_i) - \overline{e_j^p})(e_j(t_i) - \overline{e}_j)}{\sqrt{\sum_{i=1}^{n_t}(e_j^p(t_i) - \overline{e_j^p})^2 \sum_{i=1}^{n_t}(e_j(t_i) - \overline{e}_j)^2}} \tag{16}$$

where $e_j^p(t_i)$ is the uncertainty regarding the value of state x_j at time t_i estimated by the method, $e_j(t_i)$ is the actual error made by the method, that is, the difference between the prediction and the actual state value, $e_j(t_i) = x_j^p(t_i) - x_j(t_i)$, and $\overline{e_j^p}$ and $\overline{e}_j$ are their averages along all time points. Again, as an aggregate metric we use the average, $\rho_e = \frac{1}{n_x}\sum_{j=1}^{n_x} \rho_{e_j}$.

3. CASE STUDIES AND RESULTS

In this manuscript we evaluate the uncertainty analysis methods using two case studies with synthetic data. The knowledge of the true parameter values and state trajectories facilitates an in-depth analysis of the different methods. Table 1 displays the main features of the models, table 2 summarizes the results using the assessment metrics defined in Subsection 2.4, and table 3 shows the computation times.

Table 1. Properties of the considered case studies: number of parameters (n_θ), state variables (n_x), and measured outputs (n_y).

	n_θ	n_x	n_y
α-pinene	5	5	5
JAK/STAT	27	25	20

Table 3. Approximate computation times, in hours, needed by each method and case study. 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 PP it includes the sampling time, which was performed with parallel tempering.

	FIM	ENS	PP
α-pinene	0.05	0.3	1.5
JAK/STAT	35	40	350

3.1 Model for isomerization of α-pinene

Our first example is taken from Box et al. (1973). It is a small and relatively simple problem that is intended to serve as a sanity check of the different methods. It should be noted that all the states in this model are measured, which is an unusual feature in biological applications.

All methods can be easily applied to this model. Since all parameters are identifiable, the FIM is invertible and the FIM-based method can be directly applied. To apply the

Table 2. Correlations of predictions and true states (ρ_x) and between prediction uncertainty quantified by each state and actual prediction error (ρ_e), for all methods and case studies. For the prediction uncertainties, correlations for the 99%, 95%, and 68% percentiles are shown; the results do not vary for the FIM method and the differences are relatively small for PP and ENS, showing that the metric is robust with respect to the choice of the confidence level. The values shown are the mean ± the standard deviation for each state.

		FIM	PP	ENS
α-pinene	ρ_x	0.9983 ± 0.0036	0.9983 ± 0.0036	0.9983 ± 0.0036
	$\rho_{e_{99}}$	0.8069 ± 0.2506	0.8264 ± 0.2068	0.8406 ± 0.1877
	$\rho_{e_{95}}$	0.8069 ± 0.2506	0.8235 ± 0.2118	0.8364 ± 0.2140
	$\rho_{e_{68}}$	0.8069 ± 0.2506	0.8227 ± 0.2139	0.8043 ± 0.2693
JAK/STAT	ρ_x	0.9256 ± 0.1216	0.9007 ± 0.1151	0.9392 ± 0.1111
	$\rho_{e_{99}}$	0.7586 ± 0.3068	0.8730 ± 0.1809	0.8314 ± 0.2642
	$\rho_{e_{95}}$	0.7586 ± 0.3068	0.8748 ± 0.1801	0.8206 ± 0.2694
	$\rho_{e_{68}}$	0.7586 ± 0.3068	0.8757 ± 0.1765	0.8095 ± 0.2518

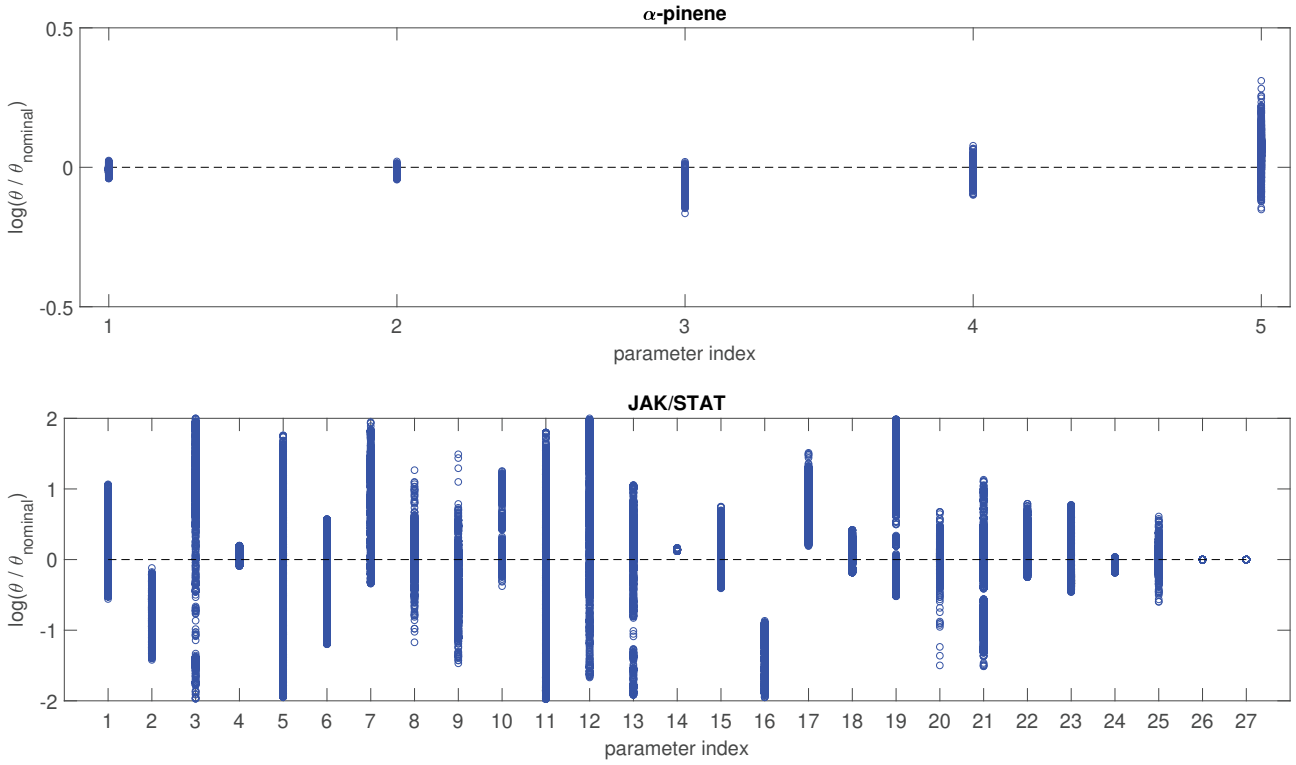

Fig. 1. Dispersion of the parameter values in the ensembles (ENS method).

ENS method we built an ensemble with 4000 parameter vectors obtained through optimization. The PP method was applied with 100000 samples. Identifiability leads to low dispersion in the parameter values included in the ensemble, as seen in Fig. 1, as well as in the PP samples (Fig. 2).

Fig. 3 shows the results of the FIM, PP, and ENS approaches, which are in general qualitative agreement. The prediction uncertainties are relatively small. Computation times range from a few minutes (FIM) to less than two hours (PP), as shown in table 3.

3.2 Model for JAK/STAT pathway

Our second example is the model of the JAK2/STAT5 signaling pathway presented by Bachmann et al. (2011) and included in the benchmark collection by Hass et al. (2019). It is larger than the α-pinene case study, and only partially observed, i.e. not all of its states are directly measured. This hampers parameter identifiability, which leads to a large dispersion in the values of the parameters. Fig. 1 shows the range of parameter values for the 5000 vectors included in the ensemble, and Fig. 2 shows the 40000 samples considered in the PP method. It can be seen that for some parameters a wide range of values is consistent with a good fit to the data. This in turn leads to a decrease in the accuracy of the predictions. As can be seen in table 2, the correlations between the actual time-courses of the states and the predictions made by the model are smaller than in the α-pinene case for all methods, although they are nonetheless still $\rho_x > 0.90$. The correlations between uncertainty and actual error

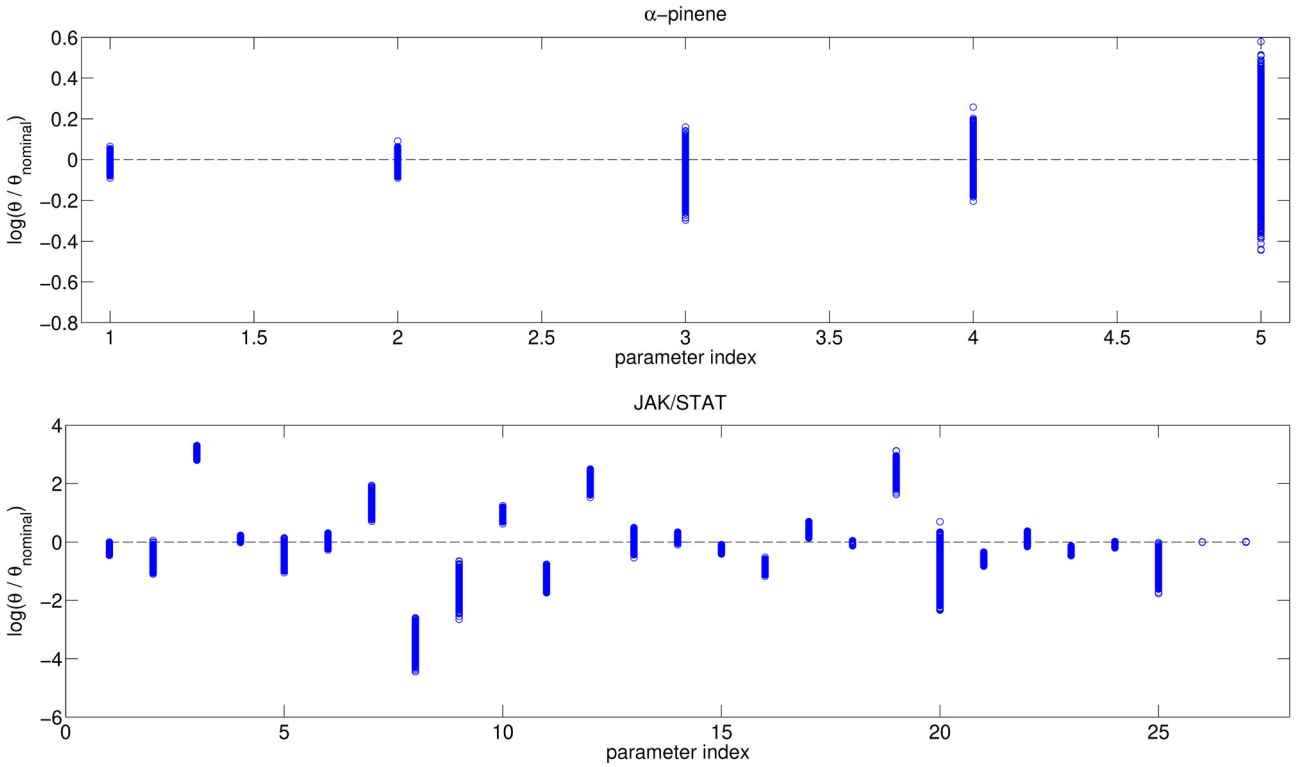

Fig. 2. Dispersion of the parameter values in the MCMC samples (PP method).

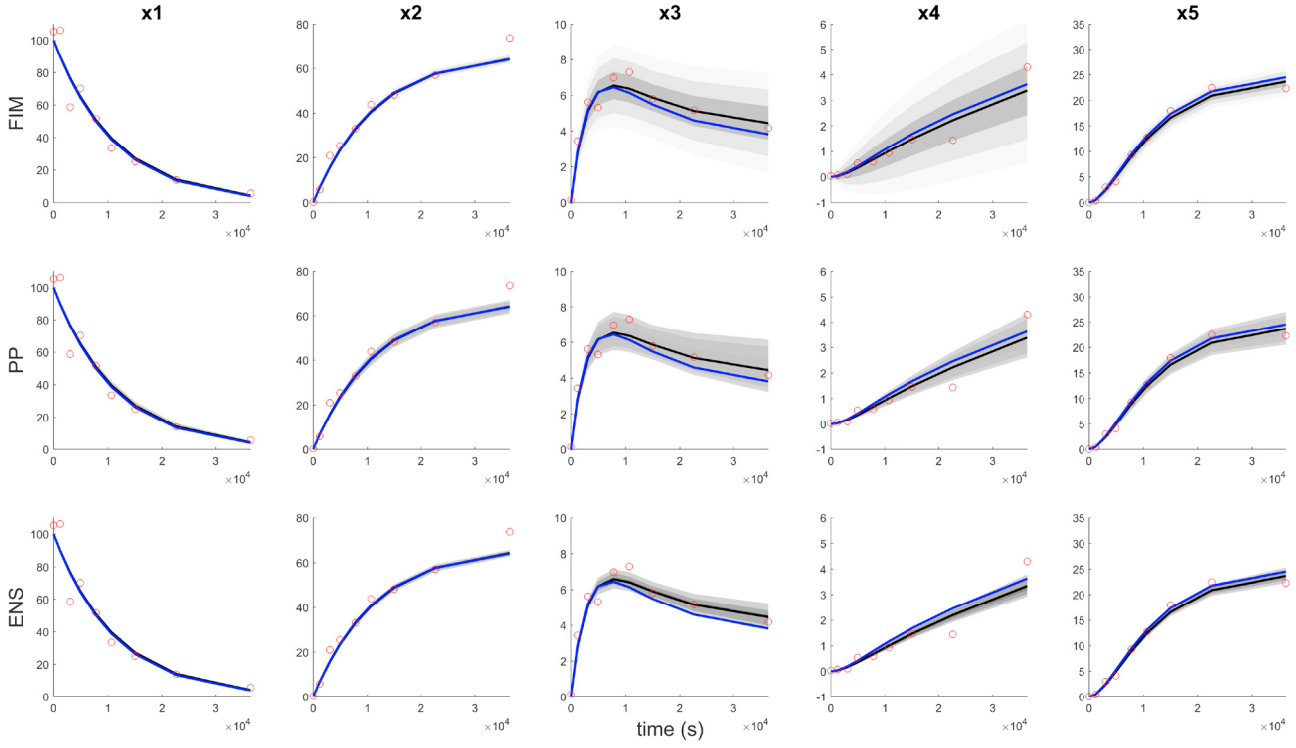

Fig. 3. Results of the different approaches for the α-pinene example. The black lines are the predictions, and the blue lines the true states. Experimental data are shown as red circles. The grey areas show the percentiles of the predictions calculated with each method (dark grey: 99%, light grey: 95.45%, lighter grey: 68.27%).

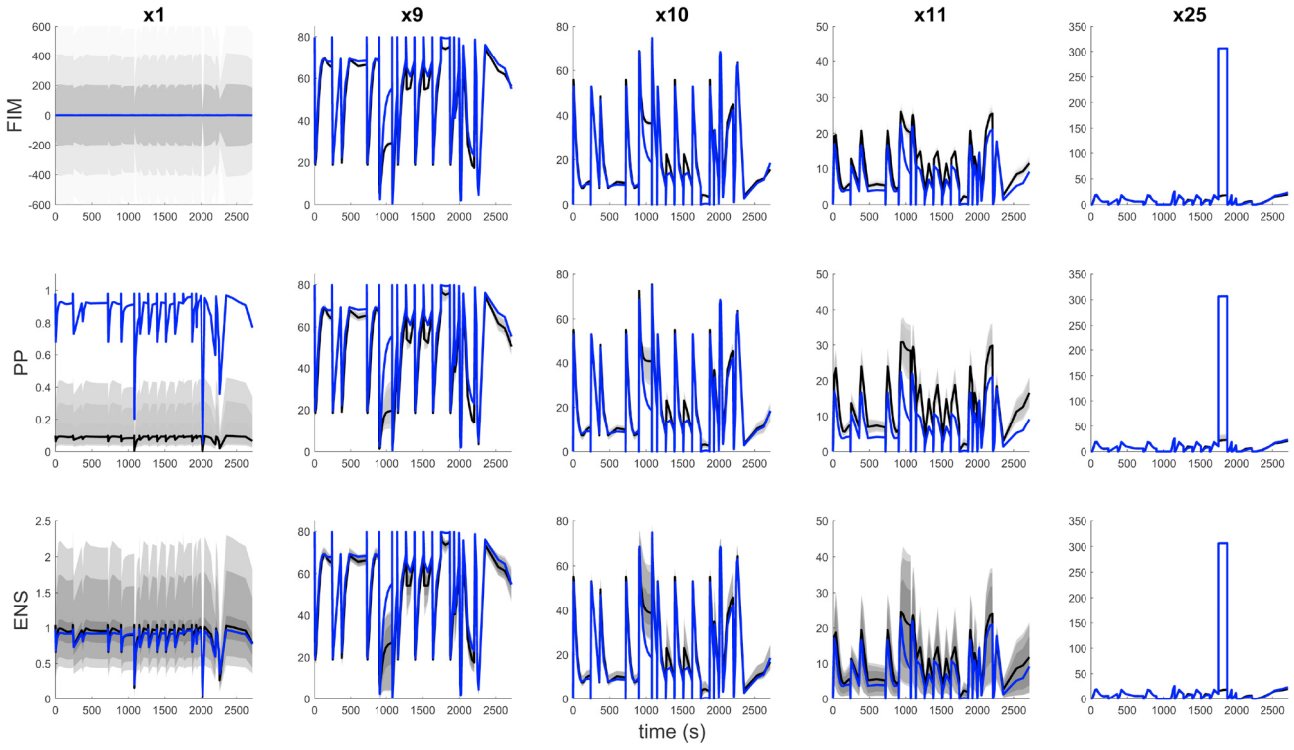

Fig. 4. Results of the different approaches for a representative subset of the states of the JAK-STAT example. The black lines are the predictions, and the blue lines the true states. The grey areas show the percentiles of the predictions calculated with each method (dark grey: 99%, light grey: 95.45%, lighter grey: 68.27%). Note that in the **x1** case the Y axis scale is different for each method.

are smaller than in the α-pinene example for the FIM method, similar for ENS, and higher for PP ($\rho_e > 0.75$ in all cases). Unidentifiability also entails that the FIM is not invertible; as mentioned in Subsection 2.1, we calculated the FIM-based uncertainties using the Moore-Penrose pseudoinverse (Shahmohammadi and McAuley, 2019). Table 3 shows the computation times, which are considerably larger (in the order of days) than for the α-pinene case study. Fig. 4 depicts the predictions and uncertainties estimated by all methods for a representative subset of states.

4. DISCUSSION AND CONCLUSIONS

In this paper we have compared three different approaches for uncertainty quantification in dynamic biological models. Specifically, the methods considered here estimate the uncertainty of the time-dependent state variables. We found that several factors can be taken into account when choosing a method for a specific problem.

In regard to accuracy, it should be noted that the FIM-based method is strictly local, and uses a linearization around a single point. It captures symmetric confidence intervals and is for the finite sample case expected to give inaccurate results in the presence of strong nonlinearities. Furthermore, if the model has unidentifiability issues the FIM cannot be inverted and a pseudoinverse has to be used. This was the case for the second case study analysed here, and is a very common scenario in systems biology models. To calculate the pseudoinverse it is necessary to specify a threshold, and the results can be affected by it. The PP and ENS approaches do not share these limitations of the FIM. In the case studies considered here, the three methods showed good agreement for the simplest example, the α-pinene model. For the other larger and more complex (nonlinear) model, more differences appeared. For this case study, PP 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 PP).

Another consideration is computational cost. In this regard, 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 PP approach, which can become very expensive for large models. The computational cost of ENS is intermediate between FIM and PP.

Overall, our results suggest a trade-off between computational scalability, on the one hand, and accuracy and rigour on the other. At one end of the trade-off there is the FIM 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 PP method. The ENS approach has a lower computational cost than PP, but PP provides a clearer statistical interpretation than ENS. The question of how to build the ensemble – i.e. how many vectors should be included in it, and which threshold should be chosen for their inclusion – is worthy of further investigation. If the ensemble does not contain sufficient diversity, there is a risk of underestimating uncertainty.

An alternative method for uncertainty quantification, is the prediction profile likelihood (Kreutz et al., 2012; Hass et al., 2016), which has not been considered in the present paper. We are currently assessing this method and we plan to present this ongoing work in the short term. Future work will also include the consideration of larger and computationally more demanding examples.

Finally, we would like to emphasize that in addition to the parameters often the model structure is also unknown. In this case, the uncertainty analysis should also take the topological uncertainties into account, using concepts such as model averaging (Fröhlich et al., 2019). A detailed assessment of this methods would also be helpful, but it is however beyond the scope of this contribution.

REFERENCES

Bachmann, J., Raue, A., Schilling, M., Böhm, M.E., Kreutz, C., Kaschek, D., Busch, H., Gretz, N., Lehmann, W.D., Timmer, J., et al. (2011). Division of labor by dual feedback regulators controls JAK2/STAT5 signaling over broad ligand range. *Molecular Systems Biology*, 7(1), 516.

Box, G., Hunter, W., MacGregor, J., and Erjavec, J. (1973). Some problems associated with the analysis of multiresponse data. *Technometrics*, 15(1), 33–51.

Cedersund, G. (2016). Prediction uncertainty estimation despite unidentifiability: an overview of recent developments. In *Uncertainty in Biology*, 449–466. Springer.

Cramér, H. (2016). *Mathematical Methods of Statistics (PMS-9)*, volume 9. Princeton university press.

Egea, J.A., Henriques, D., Cokelaer, T., Villaverde, A.F., MacNamara, A., Danciu, D.P., Banga, J.R., and Saez-Rodriguez, J. (2014). MEIGO: an open-source software suite based on metaheuristics for global optimization in systems biology and bioinformatics. *BMC Bioinformatics*, 15(1), 136.

Fröhlich, F., Loos, C., and Hasenauer, J. (2019). Scalable inference of ordinary differential equation models of biochemical processes. In G. Sanguinetti and V.A. Huynh-Thu (eds.), *Gene Regulatory Networks: Methods and Protocols*, volume 1883 of *Methods in Molecular Biology*, chapter 16, 385–422. Humana Press, 1 edition.

Geier, F., Fengos, G., Felizzi, F., and Iber, D. (2012). Analyzing and Constraining Signaling Networks: Parameter Estimation for the User. In X. Liu and M.D. Betterton (eds.), *Computational Modeling of Signaling Networks*, volume 880 of *Methods in Molecular Biology*, 23–40. Humana Press, Totowa, NJ.

Gutenkunst, R., Casey, F., Waterfall, J., Myers, C., and Sethna, J. (2007). Extracting falsifiable predictions from sloppy models. *Annals of the New York Academy of Sciences*, 1115(1), 203–211.

Hass, H., Kreutz, C., Timmer, J., and Kaschek, D. (2016). Fast integration-based prediction bands for ordinary differential equation models. *Bioinformatics*, 32(8), 1204–1210.

Hass, H., Loos, C., Raimúndez-Álvarez, E., Timmer, J., Hasenauer, J., and Kreutz, C. (2019). Benchmark problems for dynamic modeling of intracellular processes. *Bioinformatics*, btz020.

Kaltenbach, H.M., Dimopoulos, S., and Stelling, J. (2009). Systems analysis of cellular networks under uncertainty. *FEBS letters*, 583(24), 3923–3930.

Kreutz, C., Raue, A., and Timmer, J. (2012). Likelihood based observability analysis and confidence intervals for predictions of dynamic models. *BMC Systems Biology*, 6(1), 120.

Miasojedow, B., Moulines, E., and Vihola, M. (2013). An adaptive parallel tempering algorithm. *Journal of Computational and Graphical Statistics*, 22(3), 649–664.

Raue, A., Kreutz, C., Maiwald, T., Bachmann, J., Schilling, M., Klingmüller, U., and Timmer, J. (2009). Structural and practical identifiability analysis of partially observed dynamical models by exploiting the profile likelihood. *Bioinformatics*, 25(15), 1923–1929.

Shahmohammadi, A. and McAuley, K.B. (2019). Sequential model-based a-optimal design of experiments when the fisher information matrix is noninvertible. *Industrial & Engineering Chemistry Research*, 58(3), 1244–1261.

Stapor, P., Weindl, D., Ballnus, B., Hug, S., Loos, C., Fiedler, A., Krause, S., Hroß, S., Fröhlich, F., and Hasenauer, J. (2017). PESTO: Parameter estimation toolbox. *Bioinformatics*, 34(4), 705–707.

Villaverde, A.F., Bongard, S., Mauch, K., Müller, D., Balsa-Canto, E., Schmid, J., and Banga, J.R. (2015). A consensus approach for estimating the predictive accuracy of dynamic models in biology. *Computer Methods and Programs in Biomedicine*, 119(1), 17–28.