

Improving dynamic predictions with ensembles of observable models

Gemma Massonis¹, Alejandro F. Villaverde^{2,3*} and Julio R. Banga^{1*}

¹Computational Biology Lab, MBG-CSIC (Spanish National Research Council), 36143 Pontevedra, Galicia, Spain, ²CITMAga, 15782 Santiago de Compostela, Galicia, Spain, ³Universidade de Vigo, Department of Systems and Control Engineering, 36310 Vigo, Galicia, Spain

*To whom correspondence should be addressed.

Associate Editor: XXXXXXX

Received on XXXXX; revised on XXXXX; accepted on XXXXX

Abstract

Motivation: Dynamic mechanistic modelling in systems biology has been hampered by the complexity and variability associated with the underlying interactions, and by uncertain and sparse experimental measurements. Ensemble modelling, a concept initially developed in statistical mechanics, has been introduced in biological applications with the aim of mitigating those issues. Ensemble modelling uses a collection of different models compatible with the observed data to describe the phenomena of interest. However, since systems biology models often suffer from lack of identifiability and observability, ensembles of models are particularly unreliable when predicting non-observable states.

Results: We present a strategy to assess and improve the reliability of a class of model ensembles. In particular, we consider kinetic models described using ordinary differential equations (ODEs) with a fixed structure. Our approach builds an ensemble with a selection of the parameter vectors found when performing parameter estimation with a global optimization metaheuristic. This technique enforces diversity during the sampling of parameter space and it can quantify the uncertainty in the predictions of state trajectories. We couple this strategy with structural identifiability and observability analysis, and when these tests detect possible prediction issues we obtain model reparameterizations that surmount them. The end result is an ensemble of models with the ability to predict the internal dynamics of a biological process. We demonstrate our approach with models of glucose regulation, cell division, circadian oscillations, and the JAK-STAT signalling pathway.

Availability: The code that implements the methodology and reproduces the results is available at <https://doi.org/10.5281/zenodo.6782638>

Contact: j.r.banga@csic.es; avillaverde@uvigo.gal

Supplementary information: Supplementary data are available at *Bioinformatics* online.

1 Introduction

Modelling and analysis of cellular networks under uncertainty remains a fundamental challenge in systems biology, biotechnology, and bioengineering (Kaltenbach *et al.*, 2009; Mišković and Hatzimanikatis, 2011). Ensemble modelling, a concept initially developed in statistical mechanics (Brown and Sethna, 2003) that uses a collection of different models compatible with the observed data to describe the phenomena of interest, is a

suitable strategy to handle a model's parametric and structural uncertainty (Kuepfer *et al.*, 2007; Kirk *et al.*, 2015; Kremling *et al.*, 2018).

During the last two decades, the use of model ensembles has started to play an increasingly important role in the study of biological systems (Swigon, 2012), with applications in cell signalling (Brown and Sethna, 2003; Kuepfer *et al.*, 2007), metabolic networks (Tran *et al.*, 2008; Jia *et al.*, 2012; Saa and Nielsen, 2017; Hameri *et al.*, 2019), and gene expression and regulation (Ud-Dean and Gunawan, 2014; Samee *et al.*, 2015). However, building model ensembles is typically very computationally costly.

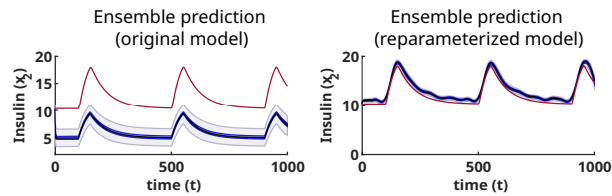


Fig. 1. Illustration of the core idea behind our method. The plots show ensemble predictions of the time course of insulin concentration, produced by the β IG model of glucose regulation (for details, see Section 3, Table 1, and the Supplementary Information). The true trajectory is shown as a red line, the ensemble prediction as a black line, and the darker and lighter blue shaded areas represent the 40% and 80% confidence percentiles respectively. The left hand plot shows the ensemble prediction of the original model, in which for the state variable x_1 , corresponding to insulin concentration, is unobservable. This model has three states (of which only one, glucose, is measured; the two unmeasured states are unobservable) and five parameters (two of which are unidentifiable). Due to the lack of structural identifiability and observability, the ensemble prediction does not reproduce the true trajectory. The right hand plot shows the ensemble prediction of the reparameterized model, which is fully observable. As a result of insulin becoming observable, the new simulations of its time course are much closer to the true trajectory, and the confidence intervals are more accurate representations of the prediction uncertainty. The NRMSE for the original model (unobservable) is 0.8516, while the NRMSE for the reparameterized model (observable) is 0.1455, a reduction of 82.91%.

Further, recent research has revealed that ensemble modelling can exhibit a number of important pitfalls (Stumpf, 2020), so ensembles must be carefully constructed in order to avoid them.

Here we present a strategy to assess and improve the reliability of a class of model ensembles. In particular, we consider kinetic models described using ordinary differential equations (ODEs) with a fixed structure. In previous work (Villaverde *et al.*, 2015) we developed a consensus-based technique, where the ensemble is built using the sampling from optimization runs of parameter estimation by means of a global optimization metaheuristic that enforces diversity during the sampling of parameter space. This method was successfully used to perform uncertainty quantification of state predictions of large kinetic models (Villaverde *et al.*, 2022).

Our new contribution is based on the observation that most models in systems biology suffer from lack of distinguishability, identifiability, and observability (Szederkényi *et al.*, 2011; Kreutz *et al.*, 2012; Wieland *et al.*, 2021). As a consequence, we expect ensembles of models to be particularly unreliable when predicting non-observable states. In order to surmount these difficulties, we present a new methodology that starts by analysing structural identifiability and observability. When these analyses reveal deficiencies in the model structure that could lead to prediction issues, our method searches for model reformulations that surmount those difficulties. Once a fully identifiable and observable model structure has been obtained, we perform parameter estimation and use the results to build an ensemble of models following a systematic procedure described in this paper. The resulting ensemble allows making predictions about the time course of internal (i.e. unmeasured) state variables, as well as quantifying their uncertainty. Figure 1 illustrates the core idea by means of the glucose regulation model, where the lack of identifiability and observability is surmounted by merging some of the non-identifiable parameters into new variables, yielding a fully observable model.

In summary, in this paper we address three key issues. The first one is the analysis of the role of structural identifiability and observability in the context of ensemble modelling, a previously overlooked topic. Second, the comparison of the uncertainty in the predictions made by ensembles of observable models versus unobservable models. Third, the development of a step-by-step procedure following a frequentist approach, where each step is clearly described and mathematically formulated.

2 Methods

The proposed methodology consists of the six main steps shown in Fig. 2. We describe them in the following subsections. Furthermore, we provide their pseudo-code as a Supplementary Information file.

2.1 Model definition

We consider dynamic models described by systems of deterministic ordinary differential equations (ODEs) of the form:

$$\mathcal{M} = \begin{cases} \dot{x}(t) = f(x(t), \theta, u(t), w(t)) & (1) \\ y(t) = g(x(t), \theta, u(t), w(t)) & (2) \\ x^0 = x(t_0, \theta) & (3) \end{cases}$$

where f and g are rational functions of the states, $x(t) \in \mathbb{R}^{n_x}$, known and unknown inputs, $u(t) \in \mathbb{R}^{n_u}$ and $w(t) \in \mathbb{R}^{n_w}$, respectively, and unknown constant parameters, $\theta \in \mathbb{R}^{n_\theta}$. The output, $y(t) \in \mathbb{R}^{n_y}$, represents the measurements as functions of model variables.

2.2 Structural identifiability and observability: analysis and reparameterization

Once a dynamic model is available, we analyze its *structural* identifiability and observability. This analysis reveals if it is possible to estimate the true values of the parameters θ (identifiability), state variables $x(t)$ (observability), and unknown inputs $w(t)$ (input observability) from output measurements, $y(t)$ (Distefano, 2015).

If this analysis yields a positive result, the model is said to have the FISPO property, which stands for Full Input-State-Parameter Observability (or Observable, when the acronym is used as an adjective). This property, along with a method for its analysis and an implementation in the Matlab toolbox STRIKE-GOLDD (<https://github.com/afvillaverde/strike-goldd>), was introduced by Villaverde *et al.* (2019b). Other tools can be used to analyse structural identifiability and observability, based on differential geometry or differential algebra. A critical comparison of the software tools currently available for this task, which analyses their capabilities and limitations, can be found in Rey-Barreiro and Villaverde (2022).

If the analysis yields a negative result, the model is unidentifiable and/or unobservable. This means that its outputs are invariant under certain modifications of some of its parameters and/or states. This can be explained mathematically as the existence of Lie symmetries among model variables (Merkt *et al.*, 2015). If it is possible to find and remove such symmetries by reparameterization, a FISPO model can be achieved. To this end we use AutoRepar (Massonis *et al.*, 2021), a method included in the aforementioned Matlab toolbox STRIKE-GOLDD, which computes model reformulations that seek to preserve the mechanistic character of selected model variables.

It should be noted that, even if a model is FISPO, it may still produce bad parameter estimates if its *practical* identifiability is poor. Practical identifiability analysis takes into account the quantity and quality of the available data (Wieland *et al.*, 2021). It can be performed with a number of methods, including the Fisher Information Matrix (FIM), profile likelihood, and sampling-based procedures. Practical unidentifiability may be surmounted by using additional experimental data, ideally obtained from optimal experimental designs.

Although it is not the focus of the present paper, the methodology could be extended so as to address practical identifiability issues. To this end, an additional procedure might be added between Steps 4 and 5. It would consist of an analysis of practical identifiability, possibly followed by an improved experiment design to obtain more time-point measurements and/or data of better quality, and the repetition of the procedure since the parameter estimation step.

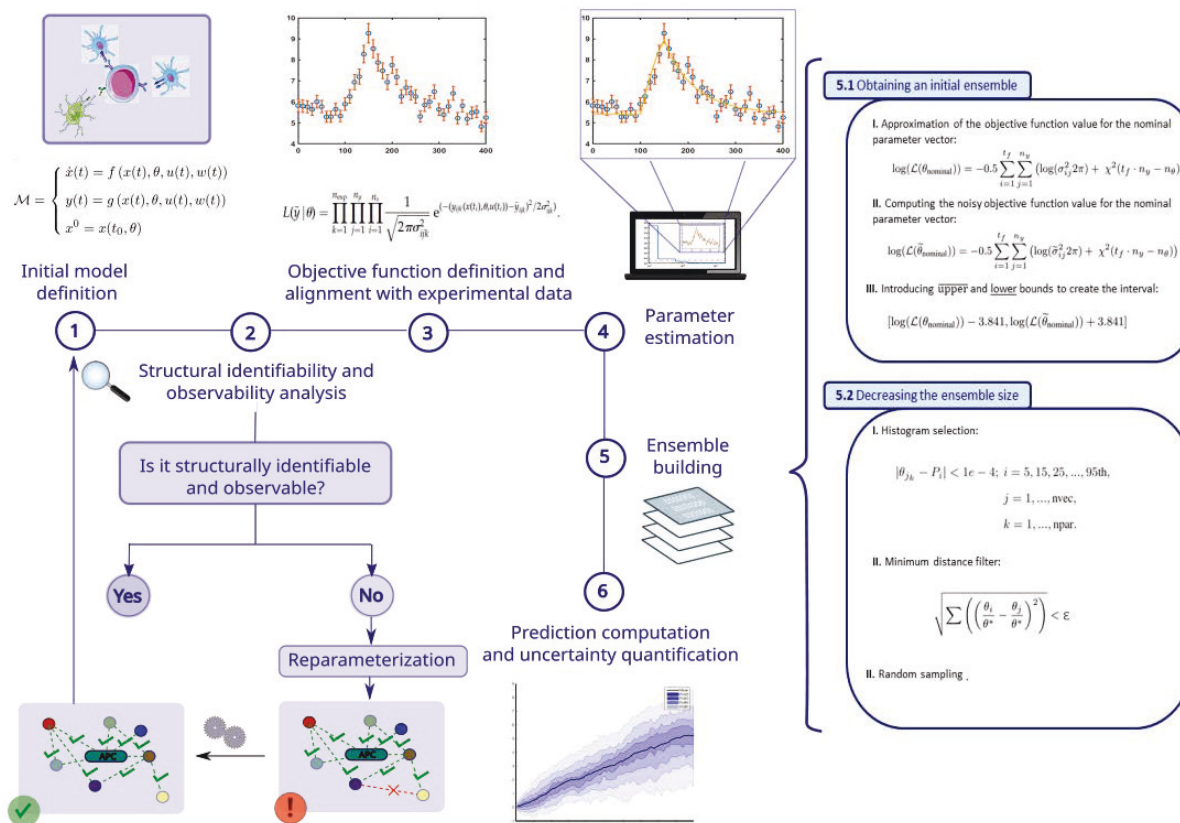


Fig. 2. Outline of the proposed ensemble modelling methodology.

2.3 Objective function definition

Parameter estimation, also known as model calibration, consists of finding the set of parameters that provide the optimal fit between model predictions and experimental data. It is performed by optimising an objective function that quantifies the distance between data and model output (Villaverde *et al.*, 2021).

The likelihood of observing the data D given parameters θ , assuming independent, normally distributed additive measurements noise with standard deviation $\sigma_{i,j}$, is

$$\mathbf{P}(D|\theta) = \prod_{i=1}^{t_f} \prod_{j=1}^{n_y} \frac{1}{\sqrt{2\pi}\sigma_{ij}} \exp\left(-\frac{1}{2} \left(\frac{\tilde{y}_{ij} - y_{ij}}{\sigma_{ij}}\right)^2\right),$$

which corresponds to a Gaussian probability function. Maximizing this expression is equivalent to minimizing the negative log-likelihood:

$$-\log(\mathcal{L}(\theta)) = \sum_{i=1}^{t_f} \sum_{j=1}^{n_y} \left(\log(\sigma_{ij} \sqrt{2\pi}) + \frac{1}{2} \left(\frac{\tilde{y}_{ij} - y_{ij}}{\sigma_{ij}}\right)^2 \right) \quad (4)$$

where \tilde{y} is the measured data, and $\sigma_{i,j}$ is the standard deviation of output j at the measurement time point i .

We use simulated data, so that the true solution is known and it is possible to assess the performance of our method. To this end we generate synthetic data by adding normally distributed noise to the model outputs:

$$\tilde{y} = y + \sigma = y + (\sigma_{\text{abs}} + \sigma_{\text{rel}} \cdot y) \cdot \mathcal{X}, \quad (5)$$

where \mathcal{X} is a normally distributed random number.

If the noise distribution is unknown, the terms σ_{abs} and σ_{rel} must be included in the set of parameters to be optimized. Otherwise they can be left out of the optimization.

2.4 Parameter estimation

Next we perform parameter estimation by minimising the objective function with numerical optimization methods. The goal of this step is not only to find the optimal solution of the parameter estimation problem (i.e. the optimal parameter vector and, in some cases, the σ_{rel} and σ_{abs}), but also to obtain additional parameter vectors that will be used to build the ensemble. To this end, it is necessary to store the parameter vectors explored during the optimization procedure along with their objective function value.

Since we wish to obtain an ensemble whose diversity represents the feasible ranges of parameter values, the optimization method must explore different regions of the parameter space. Purely local strategies are ill suited for this task, and global or hybrid optimization strategies should be used instead. In this work we use an hybrid optimization approach implemented in the MEIGO Matlab toolbox (Egea *et al.*, 2014). It is a meta-heuristic called enhanced scatter search (eSS), which is a population-based evolutionary optimization method that obtains new parameter vectors by pseudo-randomly exploring the parameter space, launching local searches from promising starting points, and recombining its members. As local search methods we used the nonlinear least-squares algorithm NL2SOL and the direct search method DHC (dynamic hill climbing). This hybrid approach has been shown to explore parameter spaces with multiple local

minima in a computationally efficient way (Villaverde *et al.*, 2019a). In particular, the combination of diversification (global search) with intensification (local search) ensures a balanced sampling of the parameter space with results comparable to other sampling techniques such as MCMC (Villaverde *et al.*, 2022), and it avoids the biased sampling found in multi-start gradient-based methods (Fröhlich *et al.*, 2014). The sampling of (4) obtained during the optimization is stored for its post-processing in the next step.

2.5 Ensemble building

From the sampling of parameter vectors obtained during the optimization, we want to select a representative subset of those producing outputs in close agreement with the observed data. To this end, we select those that are inside a predefined admissible range of objective function values.

2.5.1 Obtaining an initial ensemble

The goal of this step is to select parameter vectors with an objective function value similar to that of the true solution. Much higher values would indicate underfitting, and lower values overfitting. Since in practice the true value is unknown, we need to compute an approximation. Assuming that the errors are normally (Gaussian) distributed, the sum of squared residuals follows a χ^2 distribution of $t_f \cdot n_y - n_\theta$ degrees of freedom, corresponding to the second term of the objective function (4) (Geier *et al.*, 2012). To approximate the first term of (4) we use the estimated values of the absolute and relative standard deviations, σ_{abs^*} and σ_{rel^*} , obtaining

$$\log(\mathcal{L}(\theta_{\text{nominal}})) = -\frac{1}{2} \sum_{i=1}^{t_f} \sum_{j=1}^{n_y} (\log(\sigma_{ij}^2 2\pi) + \chi^2(t_f \cdot n_y - n_\theta)), \quad (6)$$

where $\sigma_{ij} = \sigma_{rel^*} \cdot y_{ij} + \sigma_{abs^*}$. Note that it is not necessary to know θ_{nominal} in order to calculate (6); instead, with this formula we obtain an approximation of the log likelihood value of the nominal parameter vector.

Next, we define an interval of values around this approximation. We set as lower bound the confidence region of $\log(\mathcal{L}(\theta_{\text{nominal}})) - \Delta_\alpha$, where Δ_α is $\Pr(\chi_1^2) < 0.05$ (i.e. 3.841). The term Δ_α is the result of applying the *Likelihood Ratio Test* when all the parameters are fixed except σ (Vanlier *et al.*, 2012; Villaverde *et al.*, 2022). We found empirically that setting the upper bound at the same distance from $\log(\mathcal{L}(\theta_{\text{nominal}}))$ as the lower bound would lead to narrow confidence intervals, which would not always include the experimental data. Hence we calculate the upper bound using a variant of equation (6), in which we modify the standard deviation $\tilde{\sigma}_{ij}$:

$$\log(\mathcal{L}(\tilde{\theta}_{\text{nominal}})) = -\frac{1}{2} \sum_{i=1}^{t_f} \sum_{j=1}^{n_y} (\log(\tilde{\sigma}_{ij}^2 2\pi) + \chi^2(t_f \cdot n_y - n_\theta)),$$

where

$$\tilde{\sigma}_{ij} = \sigma_{rel^*} \cdot y_{ij} \cdot \max(\sigma_{\cdot,j}) + \sigma_{abs^*}.$$

Using the notation $\Delta_{\log(\mathcal{L}(\theta))} = \log(\mathcal{L}(\tilde{\theta}_{\text{nominal}})) - \log(\mathcal{L}(\theta_{\text{nominal}}))$, we obtain the following interval:

$$[\log(\mathcal{L}(\theta_{\text{nominal}})) - 3.841, \log(\mathcal{L}(\theta_{\text{nominal}})) + \Delta_{\log(\mathcal{L}(\theta))} + 3.841] \quad (7)$$

As a result of this step, every parameter vector that produces an objective function value in the interval (7) is included in the ensemble. This interval is an approximation based on a statistical criterion, namely the Likelihood Ratio Test, as used by e.g. Vanlier *et al.* (2012) and Villaverde *et al.* (2022).

2.5.2 Decreasing the ensemble size

There is typically a very large number of parameter vectors within the range defined by (7). To avoid excessive redundancy while preserving the desired diversity, we select a representative subset of them as follows.

First we compute, for each parameter, all percentiles from 0 to 100 with a step of 5 (giving a total of 21 groups). In order to reduce the total number of vectors while preserving the underlying distribution, we eliminate some of these groups based on its position inside the percentiles. Specifically, we keep only those that are in an even position. In this way we discard the two extremes, leaving out the outliers while keeping values far from the mean, median and mode, which contain information not provided by measures of central tendency.

Second, to avoid almost identical vectors, we introduce a minimal distance criterion. To this end we use the relative Euclidean distance rescaled by the *best fit* (θ^*) between a candidate vector (θ_i) and all the others (θ_j). We exclude a parameter vector from the ensemble if this distance is less than a cut-off value. That is, a vector θ_i is removed from the ensemble if:

$$\sqrt{\sum \left(\left(\frac{\theta_i}{\theta^*} - \frac{\theta_j}{\theta^*} \right)^2 \right)} < \epsilon \quad (8)$$

Third, we perform a final reduction by random sampling. We have found empirically that ensembles of 1000 vectors are representative enough. Larger ensemble sizes do not increase the predictive power of the ensemble, as depicted in Fig. 3 and discussed below.

2.6 Ensemble prediction computation and uncertainty quantification

The ensemble prediction is defined as the median of the simulated model outputs using all the parameter vectors in the ensemble. The median is defined as the centrally placed value if these are ordered, or, equivalently, as the 50th percentile. It is an appropriate metric for unknown or non Gaussian distributions, since it is less susceptible to outliers (Gneiting, 2011; Rousseeuw, 1990). The main metric for evaluating the quality of the ensemble predictions will be the normalized root mean square error (RMSE), defined by the following expression:

$$\text{NRMSE}(y) = \frac{\text{RMSE}(y^f)}{\max(\tilde{y}) - \min(\tilde{y})} \quad (9)$$

where

$$\text{RMSE}(y^f) = \sqrt{\frac{\sum_{i=1}^{n_y} (y^f - \tilde{y})^2}{n_y}}. \quad (10)$$

denoting y^f as the forecast ensemble. It is also possible to compute the NRMSE of internal states when the nominal parameter vector is known substituting, in the above expression, y and \tilde{y} for x and \tilde{x} , respectively. Another variant of the above expression can be used to compute the NRMSE(θ) for the parameters. The NRMSE has the additional advantage of allowing the comparison of errors when model outputs, or parameters, have different orders of magnitude.

3 Results

In the present section we demonstrate the proposed methodology by applying it to the models defined in Table (1), in order to illustrate all the aspects of the method and show its applicability. Here we show only selected results of each case study; the accompanying repository (doi:10.5281/zenodo.6782638) includes the full results, along with Matlab scripts that apply every step in the methodology.

3.1 Case studies

The models chosen for this study are described in Table 1. Their equations, as well as additional information about them, can be found in the scripts provided in the accompanying repository.

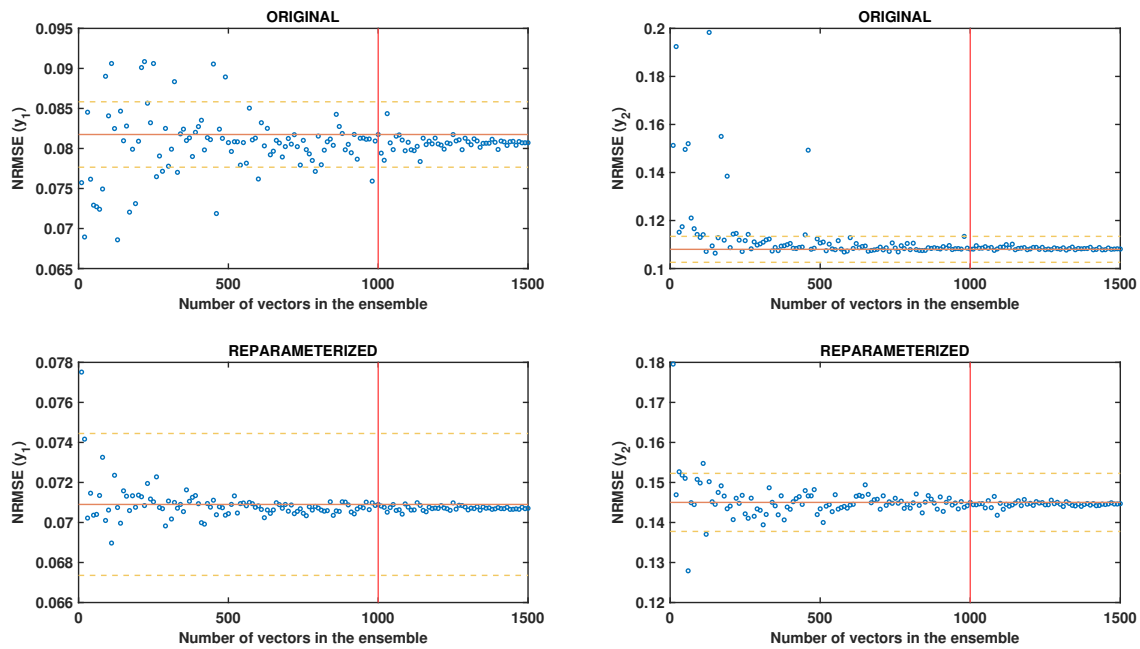


Fig. 3. JAK-STAT: NRMSE (y) depending on the number of vectors considered for the ensemble. The orange line is the NRMSE of the ensemble considering 1000 vectors, while the yellow lines represents a variation of $\pm 5\%$ of it.

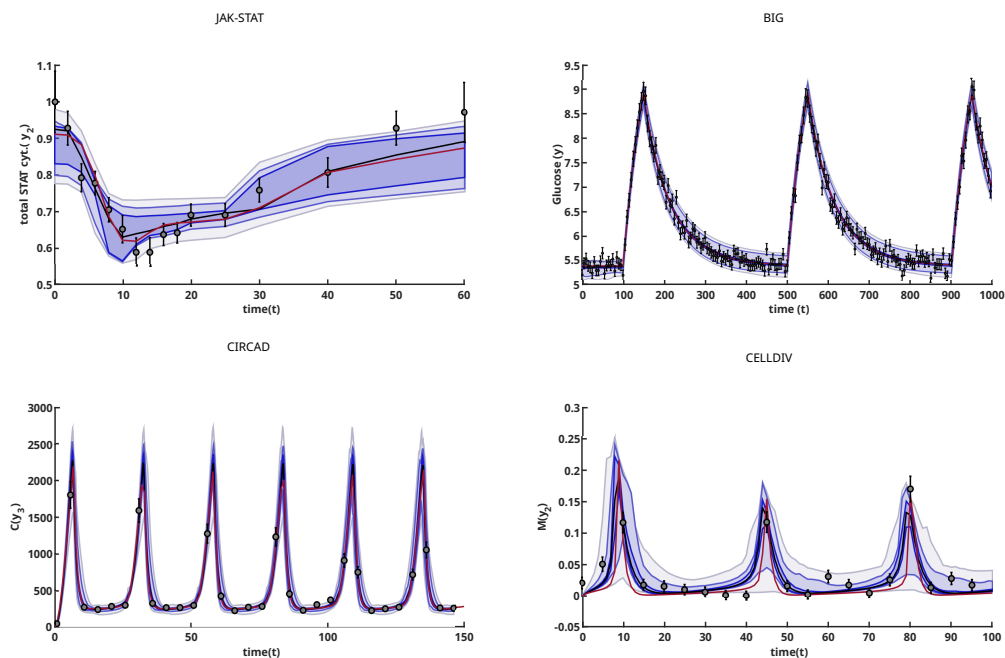


Fig. 4. Illustration of the ensemble predictions for one output of each case study. The true trajectory is shown as a red line, the ensemble prediction as a black line, and the darker and lighter blue shaded areas represent the 2.2%, 50% and 97.5% confidence percentiles respectively. Although the prediction accuracy varies among case studies, in all cases it can be noticed that the uncertainty envelope covers the experimental data.

3.2 Structural identifiability and observability analysis and reparameterization

The observability and identifiability analyses concluded that none of the models are observable nor identifiable. For the CELLDIV model it was not possible to find a fully identifiable and observable reparameterization.

In contrast, for the other three models (*BIG*, *CIRCAD* and *JAKSTAT*) the reparameterizations were as follows:

- The JAKSTAT model has 3 unidentifiable parameters and one unobservable state (x_1 , corresponding to the STAT variable). An identifiable reparameterization was found by transforming the unobservable variable so that it is multiplied by one of the unidentifiable parameters,

ID	β IG	CELLDIV	CIRCAD	JAKSTAT
Reference	Karin <i>et al.</i> (2016)	Tyson (1991)	Vilar <i>et al.</i> (2002)	Vanlier <i>et al.</i> (2012)
Description	Glucose regulation	Cell division	Circadian oscillations	Signalling pathway
Parameters	5	8	15	6
Dynamic states	3	6	9	4
Observed states	1	2	3	2
Data points	200	20	30	16
Data type	Simulated	Simulated	Simulated	Real
Noise level	$\sigma_{\text{abs}} = 2\%, \sigma_{\text{rel}} = 2\%$	$\sigma_{\text{abs}} = 2\% \sigma_{\text{rel}} = 10\%$	$\sigma_{\text{abs}} = 2\% \sigma_{\text{rel}} = 10\%$	Real

Table 1. Main features of the case studies.

p_2 . As a result, the following transformation was applied to x_1 : $x_1^* = x_1 \cdot p_2$.

- The CIRCAD model has 6 unidentifiable parameters and 2 unobservable states (M_A and M_R). An identifiable reparameterization was found by transforming the two unobservable state variables, multiplying each of them by a parameter that is thus removed from the model. As a result, the following transformation was applied to x_5 and x_7 : $x_5^* = x_5 / \alpha_{AP}$ and $x_7^* = x_7 / \alpha_{AR}$.
- The β IG model has 2 unobservable states (x_2 and x_3 , corresponding to the insulin and β cell mass) and 2 unidentifiable parameters. In this case, an identifiable and observable reparameterization could be found, which makes the insulin state variable observable without undergoing any transformation. Instead, the transformations involve the other state (x_3) and the unidentifiable parameters, which are removed from the model. The following reparameterization has been applied to x_3 and the input u : $x_3^* = x_3 \cdot p$ and $u^* = u - x_3 \cdot x_1 (s_i - x_3)$. As a result, the unobservable state x_2 became observable without undergoing any transformation.

3.3 Objective function definition

For each model we defined the objective function as in (4), using simulated data. The values of σ_{abs} and σ_{rel} can be found in Table 1. Further details can be found in the accompanying scripts.

3.4 Parameter estimation

We performed parameter estimation for the four original models and the three reparameterized ones, by minimising the objective function defined in the previous subsection, using a single-shooting approach (i.e. the dynamic model was simulated for each evaluation of the objective function). The optimizations were carried out using the eSS metaheuristic, as described in Section 2.4. The dynamic model simulations were performed with the AMICI toolbox (Fröhlich *et al.*, 2021) for all the cases except for the β IG model, for which the AMIGO2 toolbox (Balsa-Canto *et al.*, 2016) was used due to the need for estimating an unknown input.

3.5 Ensemble building

Following the procedure described in Section 2.5, we obtained ensembles of 1000 parameter vectors for each model. It should be noted that, while the optimizations had a fairly large computational cost – ranging from minutes for CELLDIV to days for CIRCAD (in this latter case, due to the challenging nature of parameter estimation in oscillatory models) – the process of ensemble building and exploitation took only a few minutes.

3.6 Prediction computation and uncertainty quantification: measured vs unmeasured states

The predictions of the *measured* state variables – i.e. the fits achieved from the optimization – obtained with both the original and the reparameterized model were of similar quality in all cases. For all models and versions the ensemble trajectories approximated the experimental data, and their envelopes (which represent the prediction uncertainty) matched the standard deviation of the data.

In contrast, the results for the *internal* – i.e. not measured – states were clearly divided in two groups: observable and unobservable. The ensemble forecasts for the observable states were generally in good agreement with the true trajectories, i.e. those obtained when simulating the model with the nominal parameter vector. The uncertainty bounds were also close to these trajectories. Illustrative examples are shown in the lower row of Fig. 5. For the unobservable states, on the other hand, the predicted trajectories usually showed large deviations from their true values (see the first rows of Fig. 5 and 6), and in some cases the prediction envelopes covered an overly large area (see the JAK-STAT example). For these cases, the ensemble predictions of the unobservable states were unreliable.

However, this limitation could be overcome by reparameterizing the models so that the unobservable states became observable (Fig 6, second row). The ensemble forecasts produced by the reparameterized models were in better agreement with the true trajectories than the original predictions, and their NRMSEs decreased significantly. Furthermore, the reduction of the uncertainty bands in almost all cases is an indication of increased confidence in the predictions. These results illustrate the benefits of working with observable models. When ensemble modelling is applied to non-observable and non-identifiable models, the bounds of the predictions are only limited by the propagation to the state space of the parameter bounds set in the parameter estimation step. However, when the models are fully observable and identifiable, the ensemble forecast will be more narrowly constrained, as can be seen in Figs. (5-6).

4 Discussion

In this work we have studied the relevance of structural identifiability and observability for building and exploiting ensembles of dynamic models. As our results have shown, the lack of these properties can compromise predictive power. To address this issue we have proposed an ensemble modelling framework that tests for the existence of such issues and takes the necessary actions to remedy them.

The procedure starts by analysing structural identifiability and observability; if the analysis of these properties reveals deficiencies in the model structure that prevent it from inferring key parameters or state variables, the method then searches for a suitable reparameterization. Once a fully identifiable and observable model structure is obtained, it is calibrated using

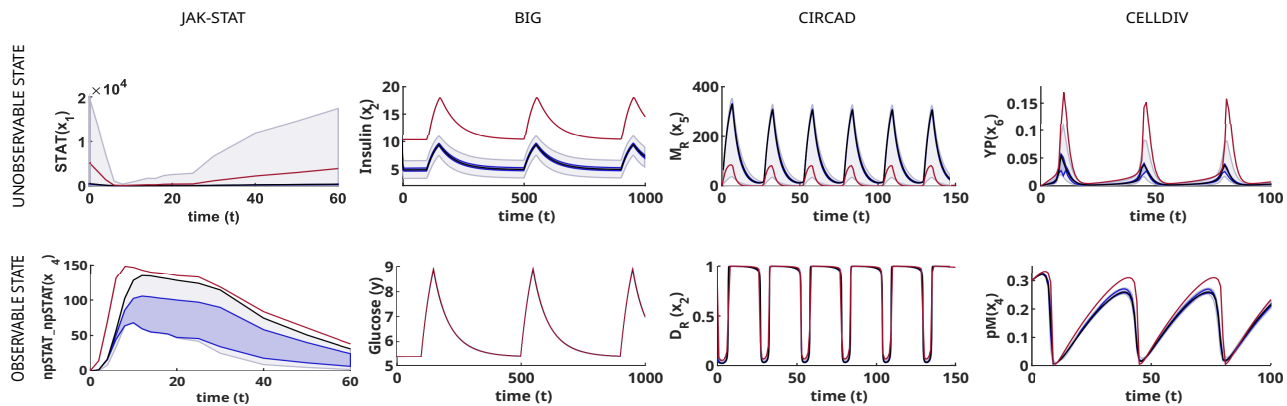


Fig. 5. Illustration of the differences between ensemble predictions for an observable versus an unobservable state. For each case study (i.e. column) the figure shows time course simulations of one of the unobservable states (upper row) and one of the observable states (lower row). The true trajectory is shown as a red line, the ensemble prediction as a black line, and the darker and lighter blue shaded areas represent the 40% and 80% confidence percentiles respectively. Although the prediction accuracy varies among case studies, in all cases it can be noticed that the predictions of unobservable states may be very far from reality, while predictions of observable states are better constrained and more precise.

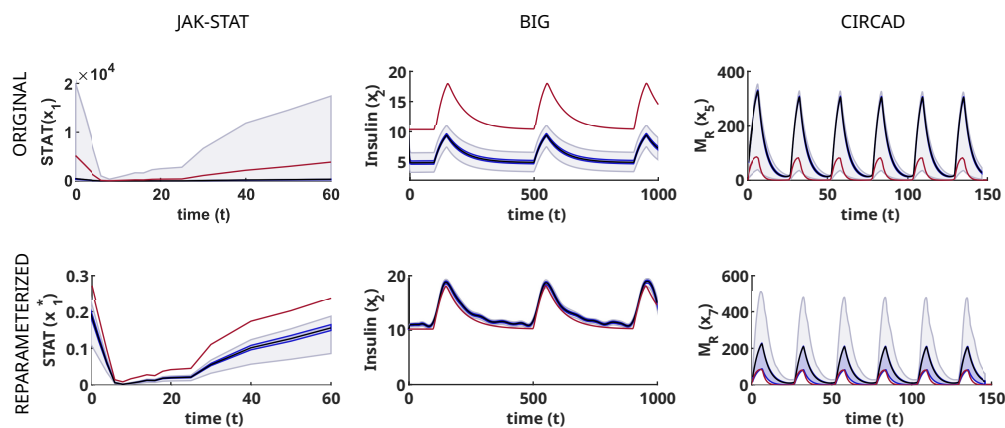


Fig. 6. Using reparameterization, we transform an unobservable state variable (upper row) into an observable one (lower row), thus improving the quality of the corresponding ensemble predictions. Each column corresponds to a case study (for the fourth case study, CELLDIV, no observable reparameterization was found). The true trajectory is shown as a red line, the ensemble prediction as a black line, and the darker and lighter blue shaded areas represent the 40% and 80% confidence percentiles respectively.

a global optimization procedure, that yields not only an optimal parameter vector but also an ensemble of other possible solutions. Our method exploits the information in these additional vectors to build an ensemble of models with different parameterizations. To this end, we have described how to select parameter vectors with appropriate objective function values, obtaining an ensemble of moderate size. The hybrid global optimization approach used here performs a balanced sampling of the parameter space; as a consequence, the median of the ensemble is a good approximation of the median of the model given parameter uncertainty. Furthermore, Villaverde *et al.* (2022) show how a parameter sampling similar to the one applied here yields good estimates of the uncertainty of the predictions.

The whole procedure can be performed systematically and is computationally efficient. To demonstrate its application we have used four case studies based on models of different sizes, all of which have unobservable parameters and states. For each model we created an ensemble that explains the available experimental data. However, when the ensembles are built from the initial models, their predictions of the unobservable internal states have high uncertainty. In contrast, if the ensembles are built from the reparameterized models, their predictions are better constrained. Importantly, in some cases the reparameterization requires transforming all the unidentifiable parameters and all the unobservable state variables, while in other cases it is possible for certain variables of interest to become

observable without being transformed. In the latter case the mechanistic meaning of the untransformed variables is preserved, as was shown for the variable representing insulin concentration in the β IG model example.

In regard to the scalability of the approach, its main bottleneck is currently the reparameterization step. It is a task that involves symbolic computations, and its computational cost increases rapidly with model complexity. Massonis *et al.* (2021) applied it to a NF- κ B model of 30 variables; applying it to larger and/or more complex models can be challenging. However, it should be noted that the algorithmic improvements included in recent versions of the STRIKE-GOLDD toolbox have reduced its computational cost (Díaz-Seoane *et al.*, 2022). The other steps in our procedure have better scalability, and they can be applied to models with hundreds of variables, as shown in Villaverde *et al.* (2022).

Acknowledgements

This work has received support from the Supercomputing Center of Galicia (CESGA).

Funding

This research has received support from grant PID2020-117271RB-C22 (BIODYNAMICS) funded by MCIN/AEI/ 10.13039/501100011033; from the CSIC intramural project grant PIE 202070E062 (MOE-BIUS); from grant PID2020-113992RA-I00 funded by MCIN/AEI/ 10.13039/501100011033 (PREDYCTBIO); 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”. The funding bodies played no role in the design of the study, the collection and analysis of data, or in the writing of the manuscript.

References

- Balsa-Canto, E., Henriques, D., Gábor, A., and Banga, J. R. (2016). AMIGO2, a toolbox for dynamic modeling, optimization and control in systems biology. *Bioinformatics*, **32**(21), 3357–3359.
- Brown, K. S. and Sethna, J. P. (2003). Statistical mechanical approaches to models with many poorly known parameters. *Phys. Rev. E*, **68**(2).
- Díaz-Seoane, S., Rey-Barreiro, X., and Villaverde, A. F. (2022). Strikegold4.0: user-friendly, efficient analysis of structural identifiability and observability. *Bioinformatics*.
- Distefano, J. (2015). *Dynamic Systems Biology Modeling and Simulation*. Elsevier Science.
- 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 Bioinform.*, **15**(1).
- Fröhlich, F., Theis, F. J., and Hasenauer, J. (2014). Uncertainty analysis for non-identifiable dynamical systems: Profile likelihoods, bootstrapping and more. In P. Mendes, J. O. Dada, and K. Smallbone, editors, *Computational Methods in Systems Biology*, pages 61–72. Cham: Springer International Publishing.
- Fröhlich, F., Weindl, D., Schälte, Y., Pathirana, D., Paszkowski, L., Lines, G. T., Stapor, P., and Hasenauer, J. (2021). Amici: high-performance sensitivity analysis for large ordinary differential equation models. *Bioinformatics*, **37**(20), 3676–3677.
- Geier, F., Fengos, G., Felizzi, F., and Iber, D. (2012). *Analyzing and Constraining Signaling Networks: Parameter Estimation for the User*, pages 23–39. Humana Press.
- Gneiting, T. (2011). Making and evaluating point forecasts. *J. Amer. Stat. Assoc.*, **106**(494), 746–762.
- Hameri, T., Boldi, M.-O., and Hatzimanikatis, V. (2019). Statistical inference in ensemble modeling of cellular metabolism. *PLoS Comput. Biol.*, **15**(12), e1007536.
- Jia, G., Stephanopoulos, G., and Gunawan, R. (2012). Ensemble kinetic modeling of metabolic networks from dynamic metabolic profiles. *Metabolites*, **2**(4), 891–912.
- Kaltenbach, H.-M., Dimopoulos, S., and Stelling, J. (2009). Systems analysis of cellular networks under uncertainty. *FEBS Lett.*, **583**(24), 3923–3930.
- Karin, O., Swisa, A., Glaser, B., Dor, Y., and Alon, U. (2016). Dynamical compensation in physiological circuits. *Mol. Syst. Biol.*, **12**(11), 886.
- Kirk, P. D. W., Babbie, A. C., and Stumpf, M. P. H. (2015). Systems biology (un)certainties. *Science*, **350**(6259), 386–388.
- Kremling, A., Geiselman, J., Ropers, D., and de Jong, H. (2018). An ensemble of mathematical models showing diauxic growth behaviour. *BMC Syst. Biol.*, **12**(1).
- Kreutz, C., Raue, A., and Timmer, J. (2012). Likelihood based observability analysis and confidence intervals for predictions of dynamic models. *BMC Syst. Biol.*, **6**(1), 1–9.
- Kuepfer, L., Peter, M., Sauer, U., and Stelling, J. (2007). Ensemble modeling for analysis of cell signaling dynamics. *Nat. Biotechnol.*, **25**(9), 1001–1006.
- Massonis, G., Banga, J. R., and Villaverde, A. F. (2021). AutoRepair: A method to obtain identifiable and observable reparameterizations of dynamic models with mechanistic insights. *Int. J. Robust Nonlin. Control*.
- Merk, B., Timmer, J., and Kaschek, D. (2015). Higher-order lie symmetries in identifiability and predictability analysis of dynamic models. *Phys. Rev. E*, **92**(1), 012920.
- Mišković, L. and Hatzimanikatis, V. (2011). Modeling of uncertainties in biochemical reactions. *Biotechnol. Bioeng.*, **108**(2), 413–423.
- Rey-Barreiro, X. and Villaverde, A. F. (2022). Benchmarking tools for a priori identifiability analysis. *ArXiv*, **abs/2207.09745**.
- Rousseeuw, P. J. (1990). Robust estimation and identifying outliers. *Handbook of statistical methods for engineers and scientists*, **16**, 16–1.
- Saa, P. A. and Nielsen, L. K. (2017). Formulation, construction and analysis of kinetic models of metabolism: A review of modelling frameworks. *Biotechnol. Adv.*, **35**(8), 981–1003.
- Samee, M. A. H., Lim, B., Samper, N., Lu, H., Rushlow, C. A., Jiménez, G., Shvartsman, S. Y., and Sinha, S. (2015). A systematic ensemble approach to thermodynamic modeling of gene expression from sequence data. *Cell Syst*, **1**(6), 396–407.
- Stumpf, M. P. H. (2020). Multi-model and network inference based on ensemble estimates: avoiding the madness of crowds. *J. Roy. Soc. Interface*, **17**(171), 20200419.
- Swigon, D. (2012). *2.1 Ensemble Modeling of Biological Systems*, pages 19–42. De Gruyter.
- Szederkényi, G., Banga, J. R., and Alonso, A. A. (2011). Inference of complex biological networks: distinguishability issues and optimization-based solutions. *BMC Syst. Biol.*, **5**, 177.
- Tran, L. M., Rizk, M. L., and Liao, J. C. (2008). Ensemble modeling of metabolic networks. *Biophys. J.*, **95**(12), 5606–5617.
- Tyson, J. J. (1991). Modeling the cell division cycle: cdc2 and cyclin interactions. *Proc. Natl. Acad. Sci. USA*, **88**(16), 7328–7332.
- Ud-Dean, S. M. M. and Gunawan, R. (2014). Ensemble inference and inferability of gene regulatory networks. *PLoS One*, **9**(8), e103812.
- Vanlier, J., Tiemann, C., Hilbers, P., and van Riel, N. (2012). An integrated strategy for prediction uncertainty analysis. *Bioinformatics*, **28**(8), 1130–1135.
- Vilar, J. M. G., Kueh, H. Y., Barkai, N., and Leibler, S. (2002). Mechanisms of noise-resistance in genetic oscillators. *Proc. Natl. Acad. Sci. USA*, **99**(9), 5988–5992.
- 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. *Comput. Meth. Progr. Biomed.*, **119**(1), 17–28.
- Villaverde, A. F., Fröhlich, F., Weindl, D., Hasenauer, J., and Banga, J. R. (2019a). Benchmarking optimization methods for parameter estimation in large kinetic models. *Bioinformatics*, **35**(5), 830–838.
- Villaverde, A. F., Tsiantis, N., and Banga, J. R. (2019b). Full observability and estimation of unknown inputs, states and parameters of nonlinear biological models. *J. Roy. Soc. Interface*, **16**(156), 20190043.
- Villaverde, A. F., Pathirana, D., Fröhlich, F., Hasenauer, J., and Banga, J. R. (2021). A protocol for dynamic model calibration. *Brief. Bioinform.*, **23**(1).
- Villaverde, A. F., Raimúndez, E., Hasenauer, J., and Banga, J. R. (2022). Assessment of prediction uncertainty quantification methods in systems biology. *IEEE/ACM Trans. Comput. Biol. Bioinform.*, pages 1–12.
- Wieland, F.-G., Hauber, A. L., Rosenblatt, M., Tönsing, C., and Timmer, J. (2021). On structural and practical identifiability. *Curr. Opin. Syst. Biol.*, **25**, 60–69.