Critical Assessment of Parameter Estimation Methods in Models of Biological Oscillators

Jake Alan Pitt, Lucian Gomoescu, Constantinos C. Pantelides, Benoît Chachuat, Julio R. Banga

(Bio)Process Engineering Group, IIM-CSIC (Spanish National Research Council), C/Eduardo Cabello 6, 36208 Vigo, Spain
RWTH-Aachen University Hospital, Joint Research Centre for Computational Biomedicine, 52074, Aachen, Germany
Centre for Process Systems Engineering, Imperial College London, London SW7 2AZ, UK
Process Systems Enterprise Ltd, London W6 7HA, UK

Abstract: Many biological systems exhibit oscillations in relation to key physiological or cellular functions, such as circadian rhythms, mitosis and DNA synthesis. Mathematical modelling provides a powerful approach to analysing these biosystems. Applying parameter estimation methods to calibrate these models can prove a very challenging task in practice, due to the presence of local solutions, lack of identifiability, and risk of overfitting. This paper presents a comparison of three state-of-the-art methods: frequentist, Bayesian and set-membership estimation. We use the Fitzhugh-Nagumo model with synthetic data as a case study. The computational performance and robustness of these methods is discussed, with a particular focus on their predictive capability using cross-validation.

Keywords: biological oscillators, model calibration, regularisation, overfitting, identifiability, frequentist estimation, Bayesian estimation, set-membership estimation

1. INTRODUCTION

Biological oscillators play a central role in many key physiological processes, including cell cycles, metabolism, signalling, and circadian rhythms (Novák and Tyson, 2008; Cao et al., 2016). Estimating the parameters in mathematical models of such oscillators can be a challenging task. Potential pitfalls are mainly due to a lack of identifiability (the existence of multiple parameter values giving the same output for a given input), multimodality (the existence of both global and local solutions in the cost function) and overfitting (fitting the measurement noise, rather than the signal) (Moles et al., 2003; Chen et al., 2010; Ljung and Chen, 2013; Villaverde and Banga, 2014; Gábor and Banga, 2015). Although these issues remain as open questions for general non-linear dynamic models, a number of techniques can be used to detect and mitigate them, including the use of structural identifiability analysis, global optimisation, regularisation, and uncertainty quantification.

The main objective of this paper is to compare three state-of-the-art estimation methods in terms of their ability to handle the aforementioned issues, namely:

- a frequentist estimation method relying on stochastic global optimisation and regularisation
- a Bayesian estimation method using Markov Chain Monte Carlo (MCMC) sampling
- a set-membership estimation method based on a rigorous complete search.

We analyse the strengths and shortcomings of each method based on a case study of the Fitzhugh-Nagumo model (Fitzhugh, 1961), a simplified version of the Hodgkin-Huxley model describing the activation and deactivation dynamics of a spiking neuron (Hodgkin and Huxley, 1952).

2. METHODOLOGY

We consider non-linear dynamic models of oscillators described by set of ordinary differential equations (ODEs). We use the following notation: \( \theta \) represents the parameter vector, \( y(t, \theta) \) (we omit \( t \) for notational simplicity) the observed model outputs, \( \bar{y} \) the output measurements, \( \sigma \) the standard deviations of the measurements, \( n_y \) the number of observables and \( n_t \) the number of time points observed.

2.1 Structural identifiability

The structural identifiability of a model should always be analysed before attempting to estimate its parameter values. This is in order to check that they can be uniquely estimated for a given set of outputs (DiStefano III, 2015),
either locally or globally. This structural identifiability test assumes noiseless, continuous output measurements. A lack of structural identifiability is an indication that a model may require reformulation (e.g. reducing the number of its parameters) or additional experimental information (e.g. measuring more outputs).

2.2 Frequentist estimation

In a frequentist framework, we consider the weighted non-linear least-squares criterion (1), as derived from a maximum-likelihood formulation, under the condition of having Gaussian measurement noise.

\[
Q_{\text{NLS}}(\theta) = \sum_{i=1}^{n} \sum_{j=1}^{n} \left( \frac{y_{i,j}(\theta) - \hat{y}_{i,j}}{\sigma_{i,j}} \right)^2
\]  

(1)

In order to avoid convergence to local solutions and over-fitting, we use a two-step optimisation procedure. While these methods can not guarantee convergence to the global optima, the use of robust global optimisation helps maximise the chances of doing so. Firstly, we use an exploratory global optimisation step to tune regularisation parameters. Secondly, we perform a regularised global estimation. The initial exploratory estimation is formulated as the following optimisation problem:

\[
\min_{\theta \in [\theta^L, \theta^U]} Q_{\text{NLS}}(\theta)
\]  

(2)

The results from this exploratory phase are used to tune the regularisation parameters \(\alpha\) and \(\theta_{\text{ref}}\). We then perform a second estimation including Tikhonov regularisation, by solving the following minimisation problem:

\[
\min_{\theta \in [\theta^L, \theta^U]} Q_{\text{NLS}}(\theta) + \alpha(\theta - \theta_{\text{ref}})^T W^T W (\theta - \theta_{\text{ref}})
\]  

(3)

where \(W\) is a scaling matrix: \(W_{i,i} = \frac{1}{\sigma_{i,i}}\) and \(W_{j\neq i,i} = 0\).

We solve both optimisation problems (2) and (3) using a single-shooting approach. This requires solving the embedded initial value problem iteratively, using AMICI (Fröhlich et al., 2016), a high-level wrapper of CVODES (Serban and Hindmarsh, 2005). In order to efficiently circumvent the issue of multimodality, we use a hybrid optimisation algorithm. We combine an efficient gradient-based local solver (NL2SOL; Dennis et al., 1981), with a robust stochastic global hybrid heuristic enhanced scatter search (eSS; Egea et al., 2010). We provide NL2SOL with accurate sensitivity information, computed using AMICI.

2.3 Bayesian estimation

Bayesian inference aims to determine a parameter probability distribution such as \(p(\theta|\tilde{y}, I_0)\) instead of a single parameter estimate \(\theta\). The available (prior) information \(I_0\) about the parameters is incorporated seamlessly into the inference problem as an a priori distribution \(p(\theta|I_0)\), and therefore the data likelihood under the given model is no longer the only relevant measure of the parameter values. The trade-off between likelihood and prior parameter distribution is given by Bayes’ theorem:

\[
p(\theta|\tilde{y}, I_0) = \frac{p(\tilde{y}|\theta, I_0) \cdot p(\theta|I_0)}{\int_{\theta} p(\tilde{y}|\theta, I_0) \cdot p(\theta|I_0)d\theta}
\]  

(4)

The posterior distribution \(p(\theta|\tilde{y}, I_0)\) may not be expressed in a closed form in general, but an approximate distribution can be computed using algorithms such as Markov Chain Monte Carlo (MCMC) sampling. Since Bayesian estimation aims at computing the parameter posterior distribution, rather than a single maximum-likelihood estimate, we expect a higher computational burden than with the frequentist approach.

We implement the ensemble MCMC algorithm Affine Invariant Stretch Move (AISM; J. Goodman, 2010; Foreman-Mackey, 2013). This algorithm has only two adjustable parameters, making it easy to tune. It also has a short mixing period, so that fewer iterations are needed for reaching the bulk of the target distribution. Herein, we use a uniform prior, an ensemble of 90 walkers, and a stretch parameter value of 3.

2.4 Set-membership estimation

Set-membership inference seeks to determine the set of all possible parameter values, for which a given model structure is consistent with a set of observations, subject to bounded errors:

\[
\Theta_{\text{SME}} = \left\{ \theta \in [\theta^L, \theta^U] \mid \begin{array}{c}
y_{i,j}(\theta) \leq \hat{y}_{i,j} + \eta \sigma_{i,j} \\
y_{i,j}(\theta) \geq \hat{y}_{i,j} - \eta \sigma_{i,j}
\end{array} \right\}
\]  

(5)

where \(\eta\) reflects the confidence level on the output measurements; e.g., \(\eta = 1\) for 1-sigma error bounds.

The fact that this approach does not require a statistical description of the observation errors, can be more realistic in practical applications, such as biological systems, where the measurements are often scarce and subject to large errors (Marvel and Williams, 2012). The binary yes-or-no answer provided by set-membership techniques can also be used for model inconsistency detection (Anderson and Papachristodoulou, 2009). One caveat however, is that the set of feasible parameter values may be empty if the error bounds are chosen incorrectly, as well as in the presence of outliers. Another key challenge is describing the feasible parameter set accurately, while remaining computationally tractable. Herein, we use a complete-search approach based on rigorous uncertainty propagation in the dynamic system, as available in the package CRONOS (Chachuat et al., 2015).

3. RESULTS AND DISCUSSIONS

3.1 Case study definition

We consider the Fitzhugh-Nagumo (FHN) model (6)–(9) below, which describes a small oscillatory system comprising 3 parameters. The wide parameter ranges increase the complexity of the parameter estimation, both in terms of computational effort and by exacerbating issues such as multimodality and overfitting.

\[
\frac{dV}{dt} = g \left( V - \frac{V^3}{3} + R \right), \quad V(0) = V_0
\]  

(6)

\[
\frac{dR}{dt} = -\frac{1}{g} \left( V - a + b \cdot R \right), \quad R(0) = R_0
\]  

(7)

\[
y(t_j|\theta) = V(t_j, \theta)
\]  

(8)

\[
a, b, g \in [10^{-5}, 10^5]
\]  

(9)
Synthetic data for the FHN model was generated using parameter values of $a = b = 0.2$ and $g = 3$. This data was generated via the addition of Gaussian noise with zero mean and a standard deviation of 10% of the nominal signal level and a detection threshold of 0.1.

The same data sets and prior knowledge are used in all three estimation methods. In order to assess the predictive capability of the calibrated model and detect possible overfitting, we use a training (calibration) step followed by a cross-validation step with a separate data set, obtained for different initial conditions.

3.2 Parameter estimation results

A structural identifiability analysis of the FHN model (6)–(9) using the package STRIKE-GOLDD [Villaverde et al. (2016)] confirms that all 3 parameters are globally structurally identifiable. No reformulation is therefore necessary in order to uniquely identify the parameters in this model.

This global identifiability property does not preclude the issue of multimodality nonetheless, i.e. the presence of local solutions in the parameter domain. With the least-squares criterion (1), a local-search method (e.g. Levenberg-Marquardt) may produce suboptimal fits, such as the ones presented in Fig. 1. In contrast, the global regularised method in Sec. 2.2 is able to arrive at a near-global solution. In Bayesian estimation, this issue of multimodality is reflected in a multimodal posterior distribution (4), and easily circumvented by retaining (one of) the maximum-likelihood estimate(s). In set-membership estimation, any local solution giving output predictions outside of the bounded-error observation is automatically eliminated from the feasible parameter region (5).

![Fig. 1. Comparison between the true solution and several local solutions of the weighted non-linear least-squares problem (2).](image)

Apart from the issue of multimodality, overfitting is another common pitfall in model identification. One approach to quantifying the goodness-of-fit (or cross-validation) is by means of the normalised-root-mean-square-error (NRMSE) criterion (10).

$$\text{NRMSE} (\theta) = \sqrt{\frac{\sum_{j=1}^{n_{y}} \sum_{i=1}^{n_{x}} (y_{i,j}(\theta) - \tilde{y}_{i,j})^2}{\max (\tilde{y}_i) - \min (\tilde{y}_i)}} \quad (10)$$

The comparison in Table 1 shows that both the regularised-frequentist and maximum-likelihood estimates have a lower NRMSE than the true solution itself, an indication that these values are fitting some of the measurement noise. Nevertheless, the use of a regularisation in the frequentist approach would appear to be mitigating the effect of overfitting, by enabling a better cross-validation performance. Such overfitting is indeed typical in biological systems, where measured data is sparse, and it is a serious impediment to the predictive capability of these models in practice.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Fitting</th>
<th>Cross-validation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Maximum-likelihood</td>
<td>0.0170</td>
<td>0.3593</td>
</tr>
<tr>
<td>Regularised-frequentist</td>
<td>0.0575</td>
<td>0.2511</td>
</tr>
<tr>
<td>True solution</td>
<td>0.0677</td>
<td>0.0382</td>
</tr>
</tbody>
</table>

![Fig. 2. Comparison of parameter estimates and parameter regions computed for the training experiment using the different methods.](image)

Such parameter regions can be exploited further for constructing envelopes of output trajectories in order to predict the possible outcomes of the cross-validation experiment. The trajectory envelopes corresponding to the parameter regions in Fig. 2 are shown in Fig. 3, and compared
with the regularised-frequentist, maximum-likelihood and true trajectories thereof. These envelopes reflect the level of uncertainty in the parameter values and could be computed for multiple confidence levels in order to inform a decision-maker.

![Comparison of predicted trajectories and envelopes](image)

**Fig. 3.** Comparison of predicted trajectories and envelopes for the cross-validation experiment using the different methods.

A final comparison of the three parameter estimation techniques is in terms of their computational requirements. For the frequentist estimation, the entire two-step procedure, including regularisation tuning takes approximately 60 CPU-seconds, with the final estimation itself converging in less than 3 CPU-seconds (Intel Xeon CPU @ 2.40GHz). The Bayesian estimation takes approximately 19 CPU-minutes to complete a total of 1,000 iterations, representing 90,000 likelihood evaluations, at which point the parameter posterior distribution does no longer change significantly (Intel Core i7-5600U CPU @ 2.60GHz, using 1 core only). Finally, the set-membership estimation takes approximately 315 CPU-minutes to run 10,000 partitions of the parameter domain, which results in a subpaving comprising over 7,200 interval boxes (Intel Core i7-6500U CPU @ 2.50GHz, using 1 core only).

While the computational efficiency of the frequentist approach allows for efficient and scalable parameter estimation methods, such methods are inherently reliant on some heuristics. Bayesian estimation allows the approximation of a posterior probability distribution for the model parameters and associated output trajectory envelopes, but it entails a larger computational effort. Set-membership estimation provides guarantees (rigorous enclosure) on the feasible parameter region and corresponding output envelopes, but it is sensitive to measurement noise and may become computationally intractable for estimation problems with more than a handful of parameters.

Our current work aims to extend this comparative assessment to larger-scale models of biological oscillators.

**REFERENCES**


