Uncertainty Analysis for Non-identifiable Dynamical Systems: Profile Likelihoods, Bootstrapping and More

Fabian Fröhlich\textsuperscript{1,2}, Fabian J. Theis\textsuperscript{1,2}, and Jan Hasenauer\textsuperscript{1,2}

\textsuperscript{1} Institute of Computational Biology, Helmholtz Zentrum München, 85764 Neuherberg, Germany
\{fabian.froehlich, fabian.theis, jan.hasenauer\}@helmholtz-muenchen.de
\textsuperscript{2} Department of Mathematics, Technische Universität München, 85748 Garching, Germany

Abstract. Dynamical systems are widely used to describe the behaviour of biological systems. When estimating parameters of dynamical systems, noise and limited availability of measurements can lead to uncertainties. These uncertainties have to be studied to understand the limitations and the predictive power of a model. Several methods for uncertainty analysis are available. In this paper we analysed and compared bootstrapping, profile likelihood, Fisher information matrix, and multi-start based approaches for uncertainty analysis. The analysis was carried out on two models which contain structurally non-identifiable parameters. We showed that bootstrapping, multi-start optimisation, and Fisher information matrix based approaches yield misleading results for parameters which are structurally non-identifiable. We provide a simple and intuitive explanation for this, using geometric arguments.

Keywords: parameter estimation, uncertainty analysis, bootstrapping, profile likelihood, identifiability.

1 Introduction

In systems and computational biology, mechanistic models are used to advance our understanding of a process of interest. In order to test whether the model can adequately reproduce measured behaviour, it is necessary to fit the model parameters to the measurement data. This process of inferring model parameters is usually termed parameter estimation.

In general, it is not possible to measure every biochemical component. Furthermore, measurements are noise corrupted. These two factors can result in a non-negligible uncertainty of the estimated parameters \cite{1}. These parameter uncertainties have to be studied, to determine limitations of the model and its predictive power. Moreover, the resulting uncertainties in parameter estimates are building blocks for subsequent investigation, such as model predictions or experimental design \cite{2}.

P. Mendes et al. (Eds.): CMSB 2014, LNBI 8859, pp. 61-75, 2014.
© Springer International Publishing Switzerland 2014
The model parameters are commonly estimated using maximum likelihood and maximum a posteriori estimators. For the analysis of the uncertainty of such estimators, several methods have been established: asymptotic analysis, bootstrapping, profile likelihoods and Bayesian statistics.

Asymptotic analysis based on the Fisher information matrix (FIM) is related to a local approximation of the objective function at the current optimum [4]. Bootstrapping exploits data resampling and estimation using the resampled data to construct confidence intervals [5]. The profile likelihood approach approximates the extent of super-level-sets of the likelihood function by constrained optimisation [4]. Some researchers also exploit uncertainty estimates derived from multiple optimiser starts [6]. However, these uncertainty estimates have no statistical foundation. There also exist Bayesian methods for uncertainty analysis, such as marginal densities [7] or profile posteriors [8]. Nonetheless, in the scope of this paper, we will only consider the frequentist perspective of uncertainty analysis, as the comparison between the frequentist perspective and the Bayesian perspective has already been covered in a recent study [8].

Several papers are available which show that bootstrapping, profile likelihood, and FIM derived confidence intervals are rather similar, given that parameters are identifiable [13][10]. Similar studies, in the presence of structural non-identifiability (cf. Definition 1), are however missing.

In this paper, we will compare the performance of several methods for uncertainty analysis based on their prediction of confidence intervals. The comparison will be carried out on two examples which both contain structurally non-identifiable parameters. Based on these two models, we will conclude that most methods for uncertainty analysis yield misleading results for parameters that are structurally non-identifiable.

This paper is organised as follows: In Section 2 we will formulate the general problem considered in parameter estimation and introduce methods for parameter estimation, as well as uncertainty analysis. In Section 3 we will compare the previously introduced methods for uncertainty analysis based on two models where structural non-identifiability of parameters is present. In Section 1 we will summarise the results and respective conclusions and give some advice on how to cope with the presented shortcomings of methods in practice.

2 Methods

In the following sections, we will describe the general problem of parameter estimation and methods for parameter estimation. Subsequently, we will introduce several methods for uncertainty analysis, namely the Fisher information matrix approximation (FIM), the Bootstrapping approach (BS), the Profile Likelihood approach (PL), and the Multi-start approach (MS). Eventually, we will present the concept of structural non-identifiability.
2.1 Problem Formulation

We consider ordinary differential equation models, in which the time-dependence of states \( \mathbf{x} \) of the biological system are captured by a set of differential equations and the observable components \( y \) are described by a mapping of these states:

\[
y(t; \theta) = h(x(t; \theta), \theta) \quad \text{with } \dot{x} = f(x, \theta),
\]

where \( f \) is the right hand side of the differential equation of the dynamical system, \( x \) is the state vector of the dynamical system and \( h \) is the function that maps states of the system to the observable components. Under the assumption of additive, independent noise, the measurement of the \( i \)-th observable component \( \hat{y}_i \) at time point \( t_k \) is

\[
\hat{y}_i(t_k) = y_i(t_k; \theta) + \varepsilon_{ik} \quad i = 1, \ldots, n_y, k = 1, \ldots, n_t,
\]

where \( n_y \) is the number of observable components, \( n_t \) is the number of observed time points and \( \varepsilon_{ik} \sim \mathcal{N}(0, \sigma^2_{ik}) \) is the measurement noise.

The maximum likelihood estimate of the parameters is obtained from measurement data \( D = \{(t_k, \hat{y}(t_k))\}_{k=1}^{n_t} \), by minimising the negative log-likelihood

\[
J_D(\theta) = -\log P(D|\theta) = \frac{1}{2} \sum_{i=1}^{n_y} \sum_{k=1}^{n_t} \left[ \log (2\pi \sigma^2_{ik}) + \left( \frac{\hat{y}_i(t_k) - y_i(t_k; \theta)}{\sigma_{ik}} \right)^2 \right].
\]

The likelihood of \( \theta \) is equal to the conditional probability \( P(D|\theta) \) to observe the data \( D \), given the parameter vector \( \theta \). In the following, the optimiser of the objective function for the respective measurement data will be denoted as

\[
\hat{\theta}^D = \arg \min_{\theta} J_D(\theta).
\]

In practice, the optimisation problem is often reformulated by transforming the parameters [11]. This can improve the numerical properties of optimisation algorithms, as well as its convergence.

2.2 Parameter Estimation

A multitude of different optimisation algorithms are available for the minimisation of the function \( J_D(\theta) \). Most commonly, global optimisation schemes like simulated annealing or particle swarm are used [12][13][14][15]. A recent study also suggests a good performance of multi-start local optimisation, when provided with high-quality gradients [11]. In this paper we employ the particle swarm method [15], as well as multi-start local optimisation. For the local optimisation scheme, the gradient \( \nabla J_D(\theta) \) is computed using sensitivity equations.

2.3 Uncertainty Analysis

In the frequentist perspective, the uncertainty of a parameter estimate is usually described in terms of confidence intervals. Confidence intervals to the level \( \delta \), will
contain the true parameter \( \theta^* \) in \((\delta \times 100)\%\) of the times, given that they are computed from realisations of the true parameter \( \theta^* \).

To facilitate the direct analysis of the extent of confidence intervals across a multitude of values for \( \delta \) we will study the function

\[
R_{\theta_i}(c) = \min_{\delta} \{ \delta : c \in C_{i, \delta} \},
\]

which describes the minimal confidence level \( \delta \) such that the parameter \( \theta_i = c \) is contained in the respective confidence interval \( C_{i, \delta} \).

**Profile Likelihood Approach (PL).** The confidence interval to the confidence level \( \delta \) for a parameter \( \theta_i \) can be interpreted as the sub-level-set to the level \( \log(\delta) \) of the objective function \( J_D(\theta) \). The extent of these sub-level-sets can be determined using the profile likelihood ratio

\[
R_{\theta_i}^{\text{PL}}(c) = \exp \left( \min_{\theta \neq \hat{\theta}_i} \left( J_D(\theta) - J_D(\hat{\theta}_D) \right) \right) \text{ s.t. } \theta_i = c,
\]

which internally uses the profile likelihood [1].

\[
PL_{\theta_i}(c) = \min_{\theta \neq \hat{\theta}_i} J_D(\theta).
\]

**Fisher Information Matrix Approximation (FIM).** The FIM approximation relies on a local quadratic approximation of \( J_D(\theta) \), which yields the approximation

\[
R_{\theta_i}^{\text{FIM}}(c) = \exp \left( \frac{-(\hat{\theta}_i^D - c)^2}{2(F^{-1})_{ii}} \right)
\]

where \( F \) is the Fisher information matrix [16].

**Bootstrapping Approach (BS).** For the bootstrapping approach, the model is fitted to the true experimental data, yielding a parameter estimate \( \hat{\theta}_D \). Subsequently \( n_D \) datasets \( D_k \) are generated by simulating the system \( y(\cdot, \hat{\theta}_D) \) and corrupting the simulation results with measurement noise. The parameter estimation is repeated for these artificial datasets by minimising \( J_{D_k}(\theta) \), yielding \( \hat{\theta}_D \). The uncertainty of parameter estimates is then derived from the spread of the estimates \( \hat{\theta}_D \).

Apparently, different parameter estimation methods can be used to determine the bootstraps. In this study we consider:

- a **single local** optimisation scheme initialised at \( \hat{\theta}_D^{\text{MS}} \) to obtain \( \{ \hat{\theta}_D^{\text{MS}}\}_k^{n_D} \) (BS-MS)
- a **multi-start local** optimisation scheme to obtain \( \{ \hat{\theta}_D^{\text{MS}}\}_k^{n_D} \) (BS-MS)
- a **particle swarm** based optimisation scheme to obtain \( \{ \hat{\theta}_D^{\text{PSO}}\}_k^{n_D} \) (BS-PSO)
Fig. 1. Illustration of the bootstrapping process. (a) Schematic illustration of the model for the turnover reaction introduced in Section 3.1. (b) Depiction of original dataset \( D_1 \) and bootstrapped dataset \( D_2 \). Data-points \( y_{i}^{k} \) are shown as red \( (i = 1) \) and blue \( (i = 2) \) circles. Simulations \( y(t; \theta_{MS}^{P}) \) of the system for corresponding optimal parameters are shown as solid lines of respective colour. (c) The continua of optimal parameters corresponding to \( D_1 \) are plotted as red \( (i = 1) \) and blue \( (i = 2) \) solid lines. The optimal parameter \( \theta_{MS}^{P} \) is shown as teal circle. The respective optimal parameter \( \theta_{LG}^{P} \) that was obtained by optimisation initiated at \( \theta_{MS}^{P} \) is shown as teal star. The difference in the location of the continuum of optimal points induced by bootstrapping is illustrated in black.

The normalised histograms over these three sets yield three approximations \( R_{p}^{BS-LG} \), \( R_{p}^{BS-MS} \), and \( R_{p}^{BS-PSO} \) to \( R_{p}(c) \). The histograms were normalised, such that the height of the bin with highest frequency is equal to 1.

Multi-start Approach (MS). Another approach, which is often used by practitioners, is to use the normalised histogram \( R_{p}^{MS}(c) \) over the best optima found by multi-start optimisation. At a first glance, MS approaches for uncertainty analysis may seem similar to bootstrapping approaches. However, for a problem where all parameters are identifiable and the optimiser converges always to the global optimum, the MS approach would suggest no parameter uncertainty.

2.4 Structural Identifiability

The aforementioned methods are used to assess the parameter uncertainties, given a certain set of experimental data \( D \). In addition, also the structural identifiability or non-identifiability might be assessed:

**Definition 1 (Structural Identifiability [17]).** A parameter \( \theta_i \), \( i = 1, \ldots, n_{\theta} \) is structurally identifiable, if for almost any \( \theta' \),

\[
y(\cdot, \theta) = y(\cdot, \theta') \Rightarrow \theta_i = \theta'_i.
\]
Except for pathological cases, structural non-identifiability of a parameter will lead to infinitely extended sets of parameters on which $y$ is invariant. In fact, in most cases the invariant set will also be infinitely extended. It is evident that $R_{\theta_{i}}(c)$ must be constant on the invariant set.

The assessment of structural identifiability and non-identifiability for large scale system is challenging and can yield inconclusive results [17]. In the following, we will study the practical identifiability of system. In particular, we will study the uncertainty intervals predicted by different methods, investigate whether the methods properly reflect the structural non-identifiabilities and investigate how the results depend on parameter transformations.

3 Results

In this section, we will compare the previously introduced methods for uncertainty analysis based on two example models. The first model is relatively simple and allows for an in-depth analysis and discussion of the results. The second model is more complex and describes a signal transduction pathway. For the second model, experimental data is available.

3.1 Example 1: Turnover Reaction

The first example we consider is a turnover reaction with synthesis rate $k_{1}$ and degradation rate $k_{2}$. The one observable of the system is $y = sx$. The initial value for the state $x$ is 0. This yields the following system of equations:

\[
\begin{align*}
\dot{x} &= k_{1} - k_{2} x \\
y &= sx.
\end{align*}
\]  

(10)
The analytical solution to this system is $y = \frac{1}{k_2} k_2 (1 - \exp(-k_2 t))$. Thus, the parameter $k_1$ and $s$ are structurally non-identifiable.

For this model, we generated the data $\mathcal{D}$ by simulating the system with parameters $k_1 = 0.75, k_2 = 0.25$, $s = 1$ for 30 equi-spaced time points $t_k$ in the interval $[0, 30]$ and adding i.i.d. noise $\epsilon_k \sim \mathcal{N}(0, 1)$.

We used multi-start local optimisation with $n_M = 10^5$ multi-starts in logarithmic coordinates in the hypercube $\Omega = [-3, 3]^3$. The multi-start optimisation was carried out both as constrained optimisation problem (MATLAB function `fmincon` with default parameters), with parameters restricted to the hypercube $\Omega$, as well as an unconstrained optimisation problem (MATLAB function `fminunc` with default parameters). For both algorithms the convergence of the algorithm to a local minimum was verified.

A multi-start local optimisation yields an ensemble $\{\hat{\theta}(k)\}_{k=1}^{n_M}$ of parameter estimates. The distribution of parameter estimates is depicted in Fig. 2 for the case where constrained and unconstrained optimisation is used. We find that for parameter $k_2$ all estimates are identical, which can be expected as this parameter is structurally identifiable. For $k_1$ and $s$ a triangular histogram shape is observed. This is surprising as the parameters are non-identifiable and a flat distribution should be expected.

The triangular shape of the histogram can be explained by studying the emergence of the distribution in more detail. Initial guesses for the parameters are drawn uniformly from the domain $\Omega$ and a local optimisation is performed. It is well known that the optimised parameters evolve roughly along the gradient of the objective function $\hat{\theta} = -\nabla_{\theta} J_D(\hat{\theta})$ (gradient descent) or a rescaled version of that $\hat{\theta} = -H^{-1}(J)\nabla_{\theta} J_D(\hat{\theta})$ (steepest descent). The optimisers stop as soon as a point with negligible gradient, i.e. the global optimum, a local optimum or
Fig. 4. Comparison of uncertainty analysis using different methods for the turnover model across logarithmic and linear parametrisation. The number of bootstraps was $10^5$ for the BS-LG method and $10^4$ for the BS-MS and BS-PSO methods. Number of multi-starts was 10 for the BS-MS method. Bin size is chosen according to optimal bin size from Scott's Rule [13].

a point on a non-identifiable manifold, is reached. Accordingly, the optimisation can be interpreted as an optimisation-method-dependent projection of starting points on the subset of points with zero gradient. In addition to the dependence on the optimisation method, the distribution of optimal parameters depends on the choice of the parameter domain $\Omega$. For the considered example, the vector field in the $k_1$ - $s$ plane (cf. Fig. 3), we find that if we draw random starting points from a square and follow the vector field, the resulting distribution corresponds to the observed triangular. The histograms for constrained and unconstrained optimisation are slightly different, as the projection is altered.

This simple example illustrates that multi-start local optimisation results are insufficient for uncertainty analysis, as they depend on the selected optimisation method and the parameter domain. Furthermore, for identifiable parameters and perfect convergence of the optimiser, a delta-distribution will be observed which also does not reflect the uncertainty correctly.

In a second step, bootstrapping methods using different optimisation schemes have been studied for logarithmic as well as linear parametrisation. For BS-MS and BS-LG we used the MATLAB function fmincon with default parameters
was used. For BS-PSO we used the MATLAB function pso, also with default parameters. The results were compared with profile likelihoods and FIM derived approximations and are shown in Fig. 1.

For the identifiable parameter $k_2$ all four methods yield similar predictions for the uncertainty. In contrast, for the structurally non-identifiable parameters $k_1$ and $s$ there are distinct differences. The PLs are flat and suggest structural non-identifiability of $k_1$ and $s$. In the case of logarithmic parameterisation, roughly the same is true for BS-PSO, while for linear parameterisation BS-PSO indicates identifiability. This might be due to differences in the distribution of the starting points and indicates the sensitivity of this method to the choice of the parameterisation. BS-MS in logarithmic parameterisation yields a triangular shape, indicating a large uncertainty, but no structural non-identifiability. For the linear parameterisation, the distribution is even tighter.

The most alarming result is revealed by comparing FIM and BS-LG derived uncertainties. In the literature, this comparison seems to be often used to validate the results of the uncertainty analysis [2]. However, we find that although both results agree, they are misleading as they indicate a small uncertainty for structurally non-identifiable problems.

In conclusion, we find already for this simple problem that results of bootstrapping approaches and MS methods depend significantly on the parameterisation, the parameter domain and the optimisation method. FIM-based methods only provide a local approximation. PL is seemingly the only approach to yield reasonable results.

3.2 Example 2: JAK/STAT Pathway

To analyse whether the same problems occur for a more realistic example, we consider the central module of the JAK-STAT signalling pathway [19]. The experimental data presented in [19] has already been studied in great detail and it is well established that the parameters $x_1(0)$, $s_1$ and $s_2$ are structurally non-identifiable [11,62,37,20]. A schematic of the model kinetics, as well as the experimental data is shown in Fig. 6. In this study we employed the D2B-Toolbox, which implements state-of-the-art simulation and estimation methods [11].

Figure 6 shows the resulting uncertainty analysis from BS-LG, the FIM approximation, and the PL approach. For the structurally identifiable parameters $p_1$, $p_2$, and $p_3$ all three methods yield similar uncertainty estimates.

For the three structurally non-identifiable parameters $x_1(0)$, $s_1$, and $s_2$ the profile likelihood ratio is flat up to some deviations at the border of the shown intervals. The FIM approximation for all three parameter is quite narrow, which indicates identifiability of the respective parameters.

The histograms for $x(0)$, $s_1$, and $s_2$ are bimodal. This is surprising and we expect that this is caused by the numerics of the problem. We can conclude that the results of the bootstrapping approach again do not properly reflect the uncertainty of structurally non-identifiable parameters.
Fig. 5. Model schematic, data and uncertainty analysis for the JAK-STAT model. (a) Schematic illustration of the states and reactions described in the JAK-STAT pathway. Figure taken from [19]. (b) Experimental measurement and respective best fit for the JAK-STAT model. Experimental data is shown as black stars, the optimal fit is shown as solid black line. (c) Uncertainty Analysis for the JAK-STAT model. The number of bootstraps was 10^3. For x₁(0), s₁, p₂, and s₂ there is evident difference in the approximation of R̂[c] between different methods. For p₁, p₂ and p₃ all methods yield comparable results.

4 Discussion

In this paper we investigated the effect of structural non-identifiabilities on the performance of frequentist methods for uncertainty analysis. We reviewed multi-start, bootstrapping, FIM, as well as the profile likelihood based methods for uncertainty analysis.

In Section 3.1 we considered a model, which is simple, but for which two of the three parameters are structurally non-identifiable. Despite the simplicity of the model, none of the methods for uncertainty analysis could indicate the structural non-identifiability of parameters, except for the profile likelihood approach. FIM, and BS-LG approaches even yielded finite confidence intervals and that parameters are identifiable. Similarly, for the model considered in Section
we observe that the profile likelihood approach also is the only method that properly identifies structural non-identifiability of parameters.

We provided detailed explanations for the emergence of the observed effects which allow for the generalisation of the results to other models and implementations. According to our results, a preceding investigation for structural non-identifiable parameters [17] is advisable in practice, as the emphasised shortcomings of the studied methods could otherwise give rise to misleading results regarding parameter identifiability and uncertainty when using FIM, multi-start and bootstrapping based uncertainty analysis.

Acknowledgements. This work was supported by a German Research Foundation (DFG) Fellowship through the Graduate School of Quantitative Biosciences Munich (QBM; F.F.), the Federal Ministry of Education and Research (BMBF) within the Virtual Liver project (Grant No. 0315766: J.H.), and the European Union within the ERC grant ‘Latent Causes’ (F.J.T.).

References


