Model identification by utilizing likelihood-based methods

Andreas Raue, J. Timmer

Angaben zur Veröffentlichung / Publication details:

Raue, Andreas, and J. Timmer. 2011. "Model identification by utilizing likelihood-based methods." In *Handbook of statistical systems biology*, edited by Michael Stumpf, David J. Balding, and Mark Girolami, 395–416. New York, NY: Wiley. https://doi.org/10.1002/9781119970606.ch20.

20

Model Identification by Utilizing Likelihood-Based Methods

Andreas Raue and Jens Timmer

Institute of Physics and Freiburg Institute for Advanced Studies (FRIAS) and Centre for Biological Systems Analysis (ZBSA), University of Freiburg, Germany

Biotechnological progress provided the possibility to decipher the human genome, yielding a multiplicity of genes representing specific functional components such as proteins or siRNA (International Human Genome Sequencing Consortium, 2004; Gerstein *et al.*, 2007). The functionality that can be observed in nature is mainly not arising from genes directly but from the interactions of these components. Therefore, knowing the whole genome gives an inventory of the functional components rather than a manual or explanation of functionality. The remaining task, putting together the henceforward known pieces to understand how functionality actually arises from the interplay of these components, is enormous. It is nevertheless essential to be able to understand how failures in this functionality, such as in cancer, arise and how they can be treated. To this end it is necessary to investigate the properties of reaction networks.

In this chapter, we want to elucidate how dynamical modeling and likelihood-based model identification strategies can be utilized to reveal properties of the underlying biological system. The purpose of modeling is not to rebuild the biological system in all of its complexity. A model is rather a tool that helps to understand and to handle the complexity of the processes occurring in biology. In this context, ordinary differential equations (ODEs) are often used as a mathematical description to investigate the dynamics of molecular processes. It is assumed that diffusion is fast compared with the reaction rates of the protein interactions and the spatial extent of the cell. Tools to build ODE models for complex reaction networks are available (Schmidt and Jirstrand, 2006; Maiwald and Timmer, 2008).

The advantage of building and utilizing a model is that all the available knowledge about the biological processes has to be formulated explicitly. The model can then be tested in a statistical framework by comparison with experimental data. There are two main applications:

Handbook of Statistical Systems Biology, First Edition. Edited by Michael P. H. Stumpf, David J. Balding and Mark Girolami.

(i) Concurring biological hypothesis such as presence or absence of a reaction can be formulated and yield different model structures. By comparing the agreement of the output of each candidate model with the experimentally measured data, using appropriate model selection techniques, some hypothesis may be falsified. For an application, see Swameye *et al.* (2003) where nuclear cycling of STAT protein is identified.

(ii) Once a promising candidate model is selected the model can be used to study the dynamical behavior of the system that cannot be accessed by experiments directly. For example, the estimated values of rate constants can be investigated, the dynamics of unobserved model components can be predicted or the model response to altered external conditions such as altered network structure or different external stimulations can be studied. For an application, see Becker *et al.* (2010) where the dynamics of erythropoietin receptor (EpoR) trafficking and interaction was investigated. Finally, systems properties such as robustness can be investigated (Kollmann *et al.* 2005).

Both applications require the reliable estimation of beforehand unknown model parameters such as rate constant or initial concentrations. The predicted dynamics depend intrinsically on these estimated parameters. At this point it is important to consider the uncertainties in the parameter estimation procedure: the measurement uncertainties translate to parameter uncertainties, and parameter uncertainties translate to uncertainties in the predicted model dynamics.

Due to technical limitations the reaction network under investigation is often only partly accessible by experiments. This means that not all molecular species incorporated in a model can be measured directly. For example Western blotting, a widely used technique to measure the dynamics of protein concentration, depends on the availability and specificity of anti-bodies for protein detection. Furthermore, the measurement errors of these techniques are usually large. Given a certain amount and quality of experimental data measured under specific experimental conditions, it is therefore not assured that model parameters can be estimated unambiguously. Frequently, experimental data are limited considering the size of the model which results in parameters that are *nonidentifiable* (Raue *et al.* 2009). Even identifiable parameters can only be determined within *confidence intervals*, which contain the true value of the parameter with a desired probability (Lehmann and Leo 1983). Consequently, if model parameters are not well determined the predicted model dynamics are also not well determined and the questions that should be answered by the model might not be addressable. Therefore, it is important to have detailed knowledge of parameter uncertainties to be able to assess the precision of model predictions realistically. On the other hand, the analysis of the uncertainties itself provides the potential to improve predictability of a model, i.e. by resolving parameter nonidentifiabilities using *experimental design* techniques.

In the following, the above mentioned concepts will be introduced from a data-based perspective, using likelihood-based methods and an illustrative example.

20.1 ODE models for reaction networks

Reaction networks can be modelled using a system of ODEs (Wolkenhauer 2008). To adequately describe the underlying biological process with this mathematical framework spatial effects have to be neglected by assuming that diffusion is fast compared with the reaction rates of the protein interactions and the spatial extent of a cell. Intrinsic stochasticity can be neglected, if the copy number of proteins is sufficiently large, i.e. for a protein copy number >10 (Taniguchi *et al.* 2010). Moreover, stochastic effects are suppressed, if the dynamics of protein concentration is modelled as average over a cell population. This applies for many measurement techniques such as obtained by Western blotting or by quantitative reverse transcription polymerase chain reaction (qRT-PCR).

A model $\mathcal{M}$ describing n species concentrations x_i in a reaction network by ODEs can be written as

$$\dot{\vec{x}}(t,\theta) = \vec{f}(\vec{x}(t,\vec{p}),\vec{u}(t),\theta) \tag{20.1}$$

$$\vec{x}(0,\theta) = \vec{w}(\theta) \tag{20.2}$$

$$\vec{y}(t,\theta) = \vec{g}(\vec{x}(t,\theta),\theta) + \vec{\epsilon} \tag{20.3}$$

with internal model states $\vec{x}$, an externally given stimulus $\vec{u}$, model parameters θ, a m-dimensional mapping $\vec{g}$ of the internal model states to the model outputs $\vec{y}$, possibly involving additional parameters θ such as scaling and offset parameters. The measurement noise $\vec{\epsilon} \sim \mathrm{N}(0,\sigma^2)$ is assumed to be known and normally distributed. Equation (20.1) represents the dynamics of protein concentration, where usually each component of the right-hand side is a sum of reaction fluxes $\vec{v}$ of several protein interactions, as will be introduced in the next section. The initial conditions of the dynamical system given in Equation (20.2) can also depend on the parameters θ. The model outputs $\vec{y}$ described by Equation (20.3) represent the experimentally measurable quantities. For partially observed models, the dimension m of the model outputs is smaller than the dimension n of the internal model states.

20.1.1 Rate equations

Consider a reaction network consisting of protein interactions r_j with $j = 1, ..., \ell$ and corresponding reaction flux $v_j(\vec{x},\vec{u},\theta)$ that can be dependent on species concentrations $\vec{x}$, the stimulus $\vec{u}$ and on parameters θ. In this context, the parameters θ involved in the dynamics, i.e. the reaction rate equations, are often called rate constants. Together with the stoichiometry matrix $\mathbf{N}$, Equation (20.1) can be expressed as

$$\dot{\vec{x}} = \begin{pmatrix} \dot{x}_1 \\ \dot{x}_2 \\ \vdots \\ \dot{x}_n \end{pmatrix} = \begin{pmatrix} N_{11} & N_{12} & \dots & N_{1\ell} \\ N_{21} & N_{22} & \dots & N_{2\ell} \\ \vdots & \vdots & \ddots & \\ N_{n1} & N_{n1} & & N_{n\ell} \end{pmatrix} \cdot \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_\ell \end{pmatrix} = \vec{f}(\vec{x},\vec{u},\theta).$$

One element N_{ij} of the stoichiometry matrix reflects whether and how often the species x_i occurs as reactant or product in the reaction r_j with flux v_j. For $N_{ij} = 0$, reaction r_j does not affect species x_i. For $N_{ij} > 0$, reaction r_j produces N_{ij} copies of species x_i and for $N_{ij} < 0$, reaction r_j consumes $|N_{ij}|$ copies of species x_i. In a graphical representation of a reaction network, each nonzero element of $\mathbf{N}$ corresponds to an edge connecting two molecular species. Some commonly used rate equations modeling the reaction fluxes $\vec{v}$ are:

- First-order mass-action kinetics for reactions $x_1 \rightarrow x_2$

$$v(x_1,\theta_1) = \theta_1 \cdot x_1 \tag{20.4}$$

  describing processes such as translocation of a species between compartments.
- Second-order mass-action kinetics for reactions $x_1 + x_2 \rightarrow x_3$

$$v(x_1,x_2,\theta_1) = \theta_1 \cdot x_1 \cdot x_2 \tag{20.5}$$

  describing processes such as complex formation.

- Michaelis–Menten kinetics for simplified enzymatic reactions $x_1 \xrightarrow{x_2} x_3$

$$v(x_1, x_2, \theta_1, \theta_2) = \frac{\theta_1 \cdot x_1}{\theta_2 + x_1} \cdot x_2 \tag{20.6}$$

describing saturation effects for rate-limiting enzyme concentration x_2. For high substate concentrations $x_1 \gg \theta_2$ the reaction rate saturates

$$v(x_2, \theta_1) = \theta_1 \cdot x_2. \tag{20.7}$$

If the substrate concentration $x_1 \ll \theta_2$, saturation can be neglected by approximating Equation (20.6) by

$$v(x_1, x_2, \theta_3) = \theta_3 \cdot x_1 \cdot x_2 \quad \text{with} \quad \theta_3 = \theta_1/\theta_2. \tag{20.8}$$

- An inhibitor x_2 can decrease the flux of a reaction $x_1 \rightarrow x_3$. The reaction flux can be approximated by

$$v(x_1, x_2, \theta_1, \theta_2) = \frac{\theta_1 \cdot x_1}{1 + \theta_2 \cdot x_2.} \tag{20.9}$$

- Similarly, x_2 increasing the flux of a reaction $x_1 \rightarrow x_3$ can be approximated by

$$v(x_1, x_2, \theta_1, \theta_2) = \theta_1 \cdot x_1 \cdot (1 + \theta_2 \cdot x_2). \tag{20.10}$$

For more details about enzyme kinetics and the modelling of inhibition, see Segel (1993).

20.2 Parameter estimation

The set of parameters θ fully specifies the dynamical behavior of an ODE model $\mathcal{M}$ and its link to the experimental data. Parameters in reaction networks such as rate constants, initial species concentrations and concentration scaling or offset factors are usually positive definite, $\theta \in \mathbb{R}^+$. To avoid the natural lower bound of zero, logarithmic parameter values should be used. The agreement of experimental data with the predicted model output is measured by an *objective function*, commonly the weighted sum of squared residuals

$$\chi^2(\theta) = \sum_{k=1}^{m} \sum_{l=1}^{d_k} \left(\frac{y_{kl}^{\dagger} - y_k(t_l, \theta)}{\sigma_{kl}} \right)^2 \tag{20.11}$$

where experimental data $y_{kl}^{\dagger}$ for each model output $k = 1, \ldots, m$ is measured at time points t_l with $l = 1, \ldots, d_k$. The measurement noise is assumed to be normally distributed and its magnitude σ_{kl} is assumed to be known. It is important to note that the noise of concentration measurements is usually log-normally distributed (Kreutz *et al.* 2007). Therefore, the experimental data $y_{kl}^{\dagger}$ and the predicted model output $y_k(t_l, \theta)$ for parameters θ should be log-transformed as well. The parameters can be estimated by

$$\hat{\theta} = \arg\min \left[\chi^2(\theta) \right], \tag{20.12}$$

which corresponds to maximum likelihood estimation (MLE) of θ for normally distributed measurement noise, because

$$\chi^2(\theta) = \text{const.} - 2 \cdot \log(L(\theta)) \tag{20.13}$$

where

$$L(\theta) = \prod_{k=1}^{m} \prod_{l=1}^{d_k} \frac{1}{\sqrt{2\pi\sigma_{kl}^2}} \exp\left(-\frac{1}{2}\left(\frac{y_{kl}^{\dagger} - y_k(t_l, \theta)}{\sigma_{kl}}\right)^2\right) \tag{20.14}$$

is the likelihood function. Due to the non linear nature of the ODE model of Equation (20.1) with respect to the model parameters, the minimization of $\chi^2(\theta)$ in Equation (20.12) cannot be solved analytically. Therefore, numerical algorithms have to be utilized to minimize/optimize the objective function $\chi^2(\theta)$. There are many different algorithmic approaches to solve the problem given in Equation (20.12).

Local optimizers. Beginning from an initially guessed parameter set θ_0, a local optimizer tries to reduce $\chi^2(\theta)$ successively, see for example the Levenberg–Marquardt algorithm in Press *et al.* (1990). For ODE models, suppling analytic derivatives that guide these algorithms to the minimum of χ^2 are essential. The numerical reasons for this and practical considerations for the calculation will be given in the next section.

If experimental data are insufficient considering the complexity of the model, and due to the non linearity of the model with respect to the parameters, the problem given in Equation (20.12) may not be well behaved, i.e. the objective function may exhibit several local minima. In this case, a local optimizer can get stuck and does not yield the best possible parameter set.

Global optimizers. In contrast, global optimizers try to keep track of several possible solutions or allow for steps in increasing direction of χ^2 temporarily (Egea *et al.* 2007). Despite the terminology 'global', it is important to note that there is no algorithm that can guarantee to find the global optimum.

Applying successive runs of a local optimizer from different initial parameter sets yields effectively a global optimizer.

20.2.1 Sensitivity equations

A local or effectively global optimization algorithm that efficiently estimates θ numerically by minimizing $\chi^2(\theta)$ in Equation (20.12), needs reliable estimates of the *sensitivities*, i.e. the derivative of the model outputs $\vec{y}$ with respect to the parameters θ,

$$\frac{\partial \vec{y}}{\partial \theta} = \frac{\partial \vec{g}}{\partial \vec{x}} \cdot \frac{\partial \vec{x}}{\partial \theta}. \tag{20.15}$$

where $\partial \vec{g}/\partial \vec{x}$ is given by Equation (20.3) analytically. In the case of ODE models, the analytical calculation of the remaining sensitivities of the internal model states $\partial \vec{x} \,/\, \partial \theta$ is hampered, because Equation (20.1) cannot, in general, be solved analytically. Usually numerical ODE solvers have to be applied. The most obvious way to calculate $\partial \vec{x} \,/\, \partial \theta$ numerically is by finite difference (FD) approximation

$$\frac{\partial \vec{x}}{\partial \theta_j} \approx \frac{\vec{x}(\theta) - \vec{x}(\theta + h \cdot e_j)}{h} \tag{20.16}$$

where e_j is the jth unit vector and h should be sufficiently small. In the context of ODE models, this approach is problematic because the integrator used to solve Equation (20.1) numerically might use different step sizes to obtain $\vec{x}(\theta)$ and $\vec{x}(\theta + h \cdot e_j)$ and only guarantees an approximate solution given a certain solver tolerance level. The numerical error especially introduced for small values of h can not be controlled in this way and leads to discontinuities of $\partial \vec{x} \,/\, \partial \theta$ along the time axis.

An elegant method to calculate the sensitivities $\partial \vec{x} \,/\, \partial \theta$ numerically is via integration of the sensitivity equations (SE). The calculation

$$\frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial \vec{x}}{\partial \theta} = \frac{\partial}{\partial \theta}\dot{\vec{x}} = \frac{\partial}{\partial \theta}\vec{f} = \frac{\partial \vec{f}}{\partial \vec{x}} \cdot \frac{\partial \vec{x}}{\partial \theta} + \frac{\partial f}{\partial \theta} \tag{20.17}$$

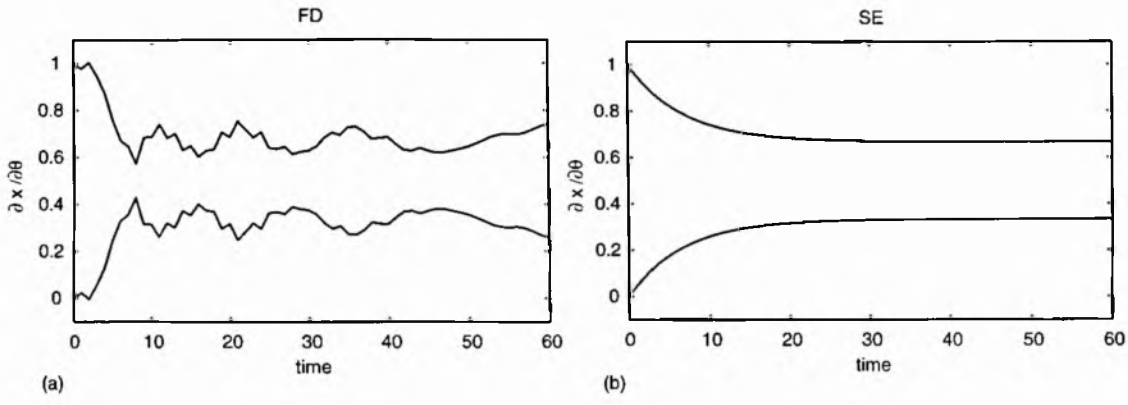

Figure 20.1 *Two sensitivities $\partial x_i / \partial \theta_j$ calculated by finite differences (FD) in (a) and by the sensitivity equations (SE) in (b) illustrating the numerical problems in the case of FD*

yields a differential equation for $\partial \vec{x} / \partial \theta$ where the right-hand side is given analytically, i.e. $\partial \vec{f} / \partial \vec{x}$ and $\partial \vec{f} / \partial \theta$ can be calculated from Equation (20.1) (Leis and Kramer 1988). These $n \cdot (n + \#\theta)$ additional differential equations can be simultaneously solved with the original ODE system. The additional initial conditions are given by

$$\frac{\partial \vec{x}}{\partial \theta}(0) = \frac{\partial \vec{w}}{\partial \theta} \tag{20.18}$$

using Equation (20.2). The resulting enlarged ODE system, including the original model dynamics and the dynamics of the sensitivities, can be efficiently solved because the Jacobian of the sensitivity subsystem is composed of the Jacobian of the original system, see Hindmarsh *et al.* (2005) for a computationally efficient implementation. Figure 20.1 shows a comparison of sensitivities calculated by FD [Figure 20.1(a)] and by the SE [Figure 20.1(b)] illustrating the numerical problems in the case of FD.

20.2.2 Testing hypothesis

Once the optimal parameter values, i.e. the best model fit, are obtained from the estimation procedure for all candidate models, one needs to judge whether the models are in agreement with experimental data and whether one of the models explains the experimental data significantly better. Various approaches exist, see Cox and Hinkley (1974) or Seber and Wild (2003). Here, two commonly used methods derived from the theory of MLE will be explained.

20.2.2.1 Pearson's χ^2-test

The limit distribution of the objective function given in Equation (20.11) is the χ^2_{df} distribution with degrees of freedom $df = \sum_{k=1}^{m} d_k - \#\theta$, where $\sum_{k=1}^{m} d_k$ is the total amount of experimental data and $\#\theta$ is the number of parameters. The expectation value and the variance of the χ^2_{df} distribution with df degrees of freedom is

$$\left\langle \chi^2_{df} \right\rangle = df \quad \text{and} \quad \text{Var}\left(\chi^2_{df}\right) = 2 \cdot df, \tag{20.19}$$

respectively, see Figure 20.2. Thus, a model that adequately represents the experimental data should have a value of the objective function not larger than a given quantile of the χ^2_{df} distribution, see the dashed line in Figure 20.2.

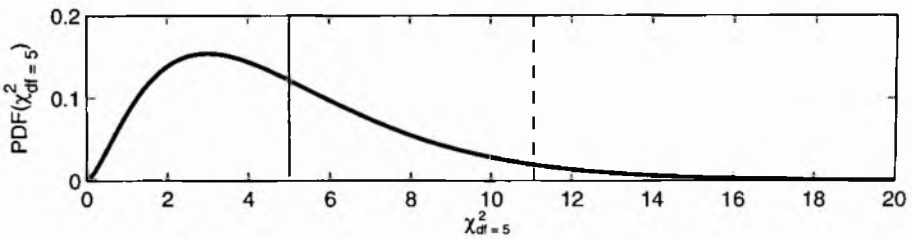

Figure 20.2 *Probability density function of the χ^2_{df} distribution with $df = 5$ degrees of freedom. The solid line indicates the expectation value. The dashed line indicates the 95% quantile that is utilized for a test with a significance level of $\alpha = 0.05$*

Please note, that this test yields a reliable conclusion only if the magnitude of the measurement noise σ^2 is assessed realistically. Furthermore, the actual degrees of freedom consumed by a nonlinear model might be slightly different from $\#\theta$. This can be checked by simulation studies.

20.2.2.2 Likelihood-ratio test

Consider a model $\mathcal{M}^*$ and an extended model $\mathcal{M}$ containing df additional parameters. For this test the null hypothesis is that an enlarged model does not fit the experimental data significantly better. The models need to be nested, $\mathcal{M}^* \subset \mathcal{M}$. If the additional df parameters lead to a significant decrease of the objective function, the hypothesis can be rejected indicating that the larger model $\mathcal{M}$ is preferable to the small model $\mathcal{M}^*$. The test statistic is given by

$$\chi^2(\mathcal{M}^*) - \chi^2(\mathcal{M}) =: \Delta\chi^2 \quad \sim \quad \chi^2_{df}. \tag{20.20}$$

Here, the decrease of the objective function $\Delta\chi^2$ due to the additional parameters is χ^2_{df} distributed. The hypothesis is rejected with a significance level of α, if

$$p_{\text{val}} = \int_{\Delta\chi^2}^{\infty} \text{PDF}(\chi^2_{df})\, \mathrm{d}\chi^2 \quad \leq \quad \alpha \tag{20.21}$$

where $\text{PDF}(\chi^2_{df})$ denotes the probability density function of the χ^2_{df} distribution. This indicates that the smaller model $\mathcal{M}^*$ is dismissed in favour of the enlarged model $\mathcal{M}$ due to a significant decrease of the objective function. If $p_{\text{val}} > \alpha$ the hypothesis cannot be rejected and the smaller model cannot be dismissed. A significance level $\alpha = 0.05$ for the χ^2_{df} distribution with degrees of freedom $df = 5$ is illustrated by the dashed line in Figure 20.2.

20.2.3 Confidence intervals

For brevity, $\chi^2(\theta)$ will be termed likelihood in the following.

Assume a model $\mathcal{M}$ that sufficiently describes the available experimental data. A confidence interval $[\sigma_i^-, \sigma_i^+]$ of a parameter estimate $\hat{\theta}_i$ to a confidence level $1 - \alpha$ signifies that the true value θ_i^* is located within this interval with probability $1 - \alpha$. In the following, *asymptotic* and *finite sample* confidence intervals will be introduced.

20.2.3.1 Asymptotic confidence intervals

Confidence intervals can be derived from the curvature of the likelihood, e.g. using the Hessian matrix

$$\mathbf{H} = \nabla^{\mathrm{T}} \nabla \chi^2|_{\hat{\theta}_i}. \tag{20.22}$$

Using the covariance matrix of the parameter estimates $\mathbf{C} = 2 \cdot \mathbf{H}^{-1}$, asymptotic confidence intervals are given by

$$\sigma_i^{\pm} = \hat{\theta}_i \pm \sqrt{\mathrm{Q}(\chi^2_{df},\ 1-\alpha) \cdot \mathbf{C}_{ii}} \tag{20.23}$$

where $\mathrm{Q}(\chi^2_{df},\ 1-\alpha)$ is the $1-\alpha$ quantile of the χ^2_{df} distribution with df degrees of freedom (Press *et al.* 1990), see Figure 20.2. The choice of df yields two different types of confidence intervals: $df = 1$ gives *pointwise* confidence intervals that hold individually for each parameter, $df = \#\theta$ being the number of parameters giving *simultaneous* confidence intervals that hold jointly for all parameters.

Asymptotic confidence intervals are a good approximation of the uncertainty of $\hat{\theta}_i$, if the amount of experimental data is large compared with $\#\theta$ and/or the amount of measurement noise is small. They are exact if the model outputs $\bar{y}$ depend linearly on θ. Even for the simplest reaction network described by Equations (20.1–20.3), the model outputs $\bar{y}$ depend nonlinearly on θ. Note that for a ODE system consisting of mass-action reactions, only the right-hand side of the ODE system is linearly dependent on the parameters, its solution $\bar{x}$ and hence the model outputs $\bar{y}$ are not. Furthermore, the amount and quality of experimental data is often limited, e.g. protein concentration measurements by Western blotting. Therefore, asymptotic confidence intervals might not be appropriate (Joshi *et al.* 2006). The ellipsoid in Figure 20.3 indicates asymptotic confidence intervals and illustrates its discrepancy from the actual shape of the likelihood.

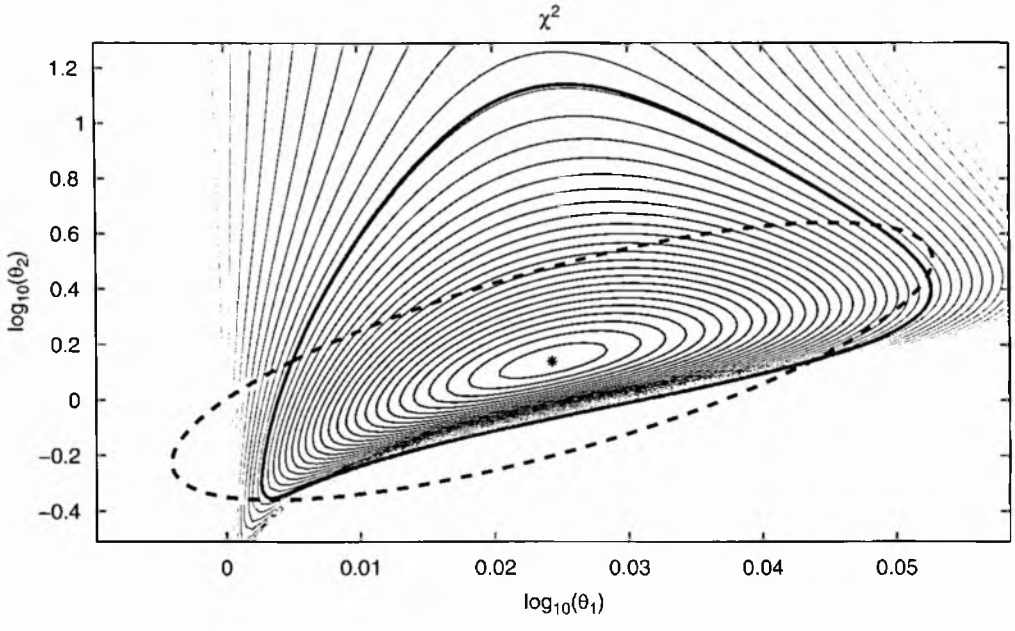

Figure 20.3 *Illustrative contour plot of $\chi^2(\theta)$ in logarithmic parameter space. Shaded contour lines from black to white correspond to low to high values of $\chi^2(\theta)$. The dashed ellipsoid indicates asymptotic confidence intervals; the thick solid line indicates likelihood-based confidence intervals*

20.2.3.2 *Finite sample confidence intervals*

Confidence intervals can also be derived using a threshold in the likelihood. These so-called *likelihood-based* confidence intervals define a confidence region

$$\{\theta \mid \chi^2(\theta) - \chi^2(\hat{\theta}) < \Delta_\alpha\} \quad \text{with} \quad \Delta_\alpha = Q(\chi^2_{df},\, 1-\alpha) \tag{20.24}$$

whose borders represent confidence intervals (Meeker and Escobar 1995). The threshold Δ_α is the $1-\alpha$ quantile of the χ^2_{df} distribution and represents with $df = 1$ and $df = \#\theta$ point-wise, respectively, simultaneous confidence intervals to a confidence level $1-\alpha$, see Figure 20.2. Likelihood-based confidence intervals are considered superior to asymptotic confidence intervals for finite samples (Neale and Miller 1997). This is illustrated in Figure 20.3. In the asymptotic case, both approaches are equivalent.

20.2.3.3 *Coverage*

In order to evaluate the appropriateness of confidence intervals, usually, coverage rates are studied. The coverage rate CR for a parameter signifies how often its true value θ_i^* is actually covered by the confidence interval $[\sigma_i^-,\ \sigma_i^+]$. To this end data is simulated assuming a set of true parameter values. If confidence intervals are appropriate, the coverage rate should reflect the desired level of confidence $CR \approx 1-\alpha$. If the coverage rate is larger than the desired level the confidence intervals tend to be more conservative than required, if it is smaller the actual uncertainty in the estimates is underestimated.

For likelihood-based confidence intervals that are based on Equation (20.24), the difference $\chi^2(\theta^*) - \chi^2(\hat{\theta})$ corresponds to the amount of over-fitting for the estimated parameters $\hat{\theta}$. For nonlinear models and small data samples the actual distribution of $\chi^2(\theta^*) - \chi^2(\hat{\theta})$ may differ from the χ^2_{df} distribution (Press *et al.* 1990). For instance the distribution can be shifted, if the actual degrees of freedom df consumed by the nonlinear model differs from the number of model parameters. Since the deviation of the distribution is dependent on the specific application, the distribution of $\chi^2(\theta^*) - \chi^2(\hat{\theta})$ should always be verified by simulation studies. If deviations are observed, the threshold Δ_α should be adjusted according to the generated distribution to obtain appropriate coverage rates for the confidence intervals.

20.3 Identifiability

A parameter θ_i is identifiable, if the confidence interval $[\sigma_i^-,\ \sigma_i^+]$ of its estimate $\hat{\theta}_i$ is finite. Two phenomena accounting for parameters to be nonidentifiable will be discussed here. *Structural nonidentifiability* is related to the model structure, independent of experimental data. This is intensively discussed in the literature (Cobelli and DiStefano III 1980). In contrast *practical nonidentifiability* also takes into account the amount and quality of experimental data that was used for parameter calibration (Raue *et al.* 2009). In the following, identifiability will be introduced from the perception of parameter estimation, i.e. in a data-based way.

20.3.1 Structural nonidentifiability

A structural nonidentifiability arises from the structure of the model itself, independent of measurement noise $\vec{\epsilon}$. Assume a model defined by Equations (20.1–20.3), where the functions $\vec{f}$, $\vec{w}$ and $\vec{g}$ describing the model dynamics and the model outputs are given, the external stimulation $\vec{u}$ is known and perfect measurements with $\sigma^2 = 0$ are assumed. The crucial question is whether the model parameters are uniquely identifiable from the experimentally observed quantities $\vec{y}$ in the given setting.

The formal solution of $\vec{y}$ may contain a redundant parameterization, due to a dimension reducing mapping $\vec{g}$ of internal model states $\vec{x}$ to model outputs $\vec{y}$ in Equation (20.3). The redundancy can be characterized by

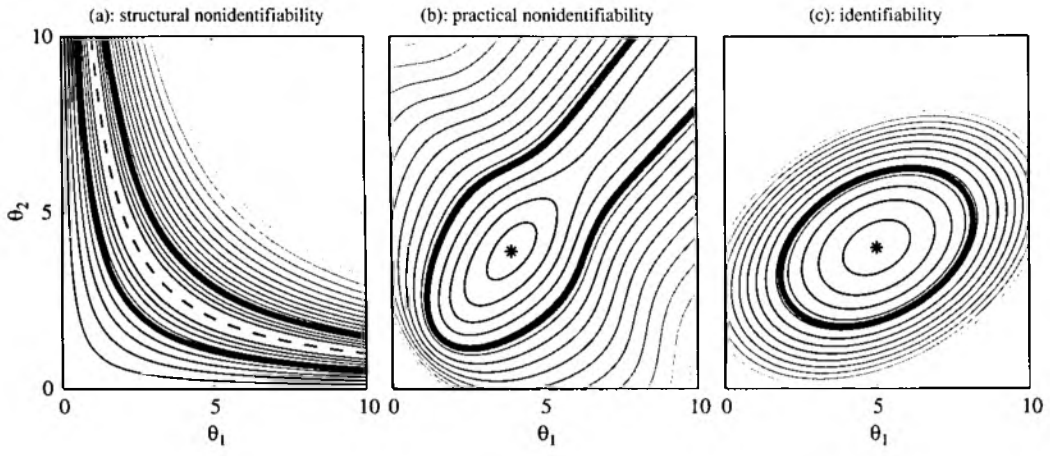

Figure 20.4 *Contour plots of $\chi^2(\theta)$ for a two-dimensional parameter space, shown in nonlogarithmic scale for illustrative purposes. Shaded contour lines from black to white correspond to low to high values of $\chi^2(\theta)$. Thick solid lines display likelihood-based confidence regions and asterisks the optimal parameter estimates $\hat{\theta}$. (a) A structural nonidentifiability along the functional relation $h(\theta) = \theta_1 \cdot \theta_2 - 10 = 0$ (dashed line). The likelihood-based confidence region is infinitely extended. (b) A practical nonidentifiability. The likelihood-based confidence region is infinitely extended for $\theta_1 \to +\infty$ and $\theta_2 \to +\infty$, and lower confidence bounds can be derived. (c) Both parameters are identifiable. Adapted from Raue et al. (2009)*

functional relations $\vec{h}(\theta_{\text{sub}}) = 0$ between a subset of parameters $\theta_{\text{sub}} \subset \theta$. These redundant parameters θ_{sub} are structurally nonidentifiable and may be varied according to the functional relations $\vec{h}$ without changing the model outputs $\vec{y}$. In terms of the objective function $\chi^2(\theta)$ defined by Equation (20.11), a structural nonidentifiability therefore manifests as iso–χ^2 manifold

$$\left\{\theta \mid \vec{h}(\theta_{\text{sub}}) = 0\right\} \quad \Rightarrow \quad \chi^2(\theta) = \text{const.} \tag{20.25}$$

Consequently, the parameter estimates $\hat{\theta}_{\text{sub}}$ and, respectively, the internal model states $\vec{x}$ affected by these parameters are not uniquely identified by measurements of $\vec{y}$. Confidence intervals of a structurally nonidentifiable parameter $\theta_i \in \theta_{\text{sub}}$ are infinite, $[-\infty, +\infty]$ in logarithmic parameter space considered here. Hence, θ_i cannot be estimated at all. A direct detection of a redundant parameterization in the analytic form of $\vec{y}$ is hampered, because Equation (20.1) cannot, in general, be solved analytically.

For a two-dimensional parameter space, $\chi^2(\theta)$ can be visualized as a landscape. A structural nonidentifiability results in a perfectly flat valley, infinitely extended along the corresponding functional relation, as illustrated in Figure 20.4(a).

Since a structural nonidentifiability is independent of the accuracy of available experimental data, it cannot be resolved by a refinement of existing measurements. The only remedy is a qualitatively new measurement which alters the mapping $\vec{g}$, usually by increasing the number of observed species. A parameter is structurally identifiable, if a unique minimum of $\chi^2(\theta)$ with respect to θ_i exists, see Figure 20.4(b) and (c).

20.3.2 Practical nonidentifiability

A parameter that is structurally identifiable may still be practically nonidentifiable. This can arise due to insufficient amount and quality of experimental data, manifesting in a confidence interval that is infinite. Please note, that the asymptotic confidence interval of a structurally identifiable parameter estimate may be

large, but is always finite because $\mathbf{C}_{ii} > 0$ [see Equation (20.23)]. Therefore, it is not possible to infer practical nonidentifiability using asymptotic confidence intervals.

According to Raue *et al.* (2009), a parameter estimate $\hat{\theta}_i$ is practically nonidentifiable, if the likelihood-based confidence region is infinitely extended in the direction of θ_i, although the likelihood has a unique minimum for this parameter. This means that the increase in $\chi^2(\theta)$ stays below the threshold Δ_α for a desired confidence level $1 - \alpha$ in the direction of θ_i. Similarly as for structural nonidentifiability, this flattening out of the likelihood can continue along a functional relation. The confidence interval of a practically nonidentifiable parameter is not necessarily extended infinitely to both sides. There can be a finite upper or lower bound of the confidence interval $[\sigma_i^-, \sigma_i^+]$, but either $\sigma_i^- = -\infty$ or $\sigma_i^+ = +\infty$ in logarithmic parameter space is considered here.

For a two-dimensional parameter space, a practical nonidentifiability can be visualized as a relatively flat valley, which is infinitely extended. The distance of the valley bottom to the lowest point $\hat{\theta}$ never exceeds Δ_α, as illustrated in Figure 20.4(b).

Along a practical nonidentifiability the model outputs $\vec{y}$ change only negligibly and remain compliant with the given measurement accuracy. Nevertheless, model behavior in terms of internal states $\vec{x}$ might vary strongly. Improving the detection of typical dynamical behavior by increasing the amount and quality of measured data and/or the choice of measurement time points will ultimately cure a practical nonidentifiability, yielding finite likelihood-based confidence intervals, see Figure 20.4(c). Inferring how to decrease confidence intervals most efficiently is the subject of experimental design, which will be discussed in Section 20.4.1.

20.3.3 Connection of identifiability and observability

The uncertainty of parameter estimates $\hat{\theta}$ directly translate to uncertainty of model trajectories. Nonidentifiability describes the phenomenon that parameters might not be determined. Consequently, non observability indicates that model trajectories might not be determined due to nonidentifiability of model parameters. As mentioned in Section 20.3.1, a structural nonidentifiability only affects the observability of the internal model states $\vec{x}$. The model outputs $\vec{y}$ are reproduced exactly for all parameter combinations on the manifold that corresponds to the structural nonidentifiability [see Equation (20.25)]]. As mentioned in Section 20.3.2, a practical nonidentifiability affects the model outputs $\vec{y}$ only negligibly. This means that the model outputs stay in agreement with the measurement accuracy of the experimental data. Some internal model states $\vec{x}$ might nevertheless be affected strongly by a practical nonidentifiability and hence might be nonobservable. Also, confidence intervals of parameter estimates translate to confidence intervals of model trajectories. Similar to coverage rates for confidence intervals of parameter estimates, coverage rates for confidence intervals of model trajectories can be assessed.

From a different point of view, identifiability can be stated in terms of observability. Additional internal model states x_{n+i} corresponding to parameters θ_i where $i = 1, \ldots, \#\theta$ can be introduced. The ODE system of Equation (20.1) is then enlarged by equations $\dot{x}_{n+i} = 0$ and every incidence of θ_i is replaced by x_{n+i} in both Equations (20.1) and (20.3). Consequently, showing observability of x_{n+i} implies identifiability of θ_i.

20.4 The profile likelihood approach

In the remainder of this chapter the profile likelihood approach (Raue *et al.* 2009) will be presented, which facilitates the numerical implementation of the concepts for data-driven modeling introduced so far. The approach enables to reveal which parameters are identifiable and thus allows to infer which model predictions are feasible. Provided that parameters are identifiable the question that follows is how large their confidence intervals are, indicating how reliable a model prediction is.

The approach is able to detect both structurally and practically nonidentifiable parameters and simultaneously calculates confidence intervals. Since large models are under consideration, it is important that the approach is computationally feasible and its output is interpretable even if the issue depends on a high-dimensional parameter space. Furthermore, the approach can be used for experimental design, to suggest additional measurements that efficiently reduce parameter uncertainties, and for *model reduction*, to tailor the model complexity to the information content provided by the experimental data. Finally, the usage and benefits of the approach will be illustrated by applying it to the core model of erythropoietin (Epo) and EpoR interaction and trafficking with the corresponding subset of the experimental data from Becker *et al.* (2010).

The idea of the approach is to explore the parameter space for each parameter in the direction of least increase in $\chi^2(\theta)$. For a structurally nonidentifiable parameter this means following the functional relations $\vec{h}(\theta_{\text{sub}}) = 0$. In the case of a practically nonidentifiable parameter, the aim is to detect directions where the likelihood flattens out.

A useful concept for this task is the profile likelihood χ^2_{PL} (Venzon and Moolgavkar 1988; Murphy and van der Vaart 2000). It can be calculated for each parameter individually by

$$\chi^2_{PL}(\theta_i) = \min_{\theta_{j\neq i}} \left[\chi^2(\theta)\right] \tag{20.26}$$

meaning re-optimization of $\chi^2(\theta)$ with respect to all parameters $\theta_{j\neq i}$, for each fixed value of the parameter θ_i. Hence, the profile likelihood keeps $\chi^2(\theta)$ as small as possible alongside θ_i. Figure 20.5 displays the profile likelihood for the three typical cases introduced in Figure 20.4 showing that the likelihood is explored in the desired way to detect nonidentifiabilities.

Structural nonidentifiable parameters are characterized by a flat profile likelihood [see Equation (20.25)]. The profile likelihood of a practically nonidentifiable parameter has a minimum but does not exceed a threshold Δ_α for increasing and/or decreasing values of θ_i, see Section 20.3.2. In contrast, the profile likelihood of an identifiable parameter exceeds Δ_α for both increasing and decreasing values of θ_i. The points of passover represent likelihood-based confidence intervals as defined in Equation (20.24) (Royston 2007). Figure 20.5(**b**) shows the profile likelihood approaching the borders of the likelihood-based confidence region in the desired way. By following the change of parameters $\theta_{j\neq i}$ along $\chi^2_{PL}(\theta_i)$, the functional relations $\vec{h}(\theta_{\text{sub}}) = 0$ corresponding to a structural nonidentifiability can be recovered, see Figure 20.5(a). An algorithm to calculate χ^2_{PL} was described in Raue *et al.* (2009).

20.4.1 Experimental design

To improve the certainty of a specific model prediction, it would be valuable to suggest additional measurements that efficiently cure nonidentifiability and narrow the confidence interval of a parameter θ_i affecting this issue. The set of trajectories along the profile likelihood of θ_i reveals spots where the uncertainty of θ_i has the largest impact on the model. Additional measurements at spots of largest variability of the trajectories efficiently accomplish this task. The amplitude of variability of the trajectories at these spots allows to assess the necessary precision of a new measurement, to provide adequate data that are able to improve parameter identification.

The impact of new measurements can be evaluated by Monte Carlo simulations. With this aim, the described analysis of the profile likelihood is repeated, taking into account additional simulated data. The resulting change of the profile likelihood and correspondingly the resolution of nonidentifiability and the narrowing of the likelihood-based confidence intervals, justifies the use of new measurements to gain a more confident model prediction.

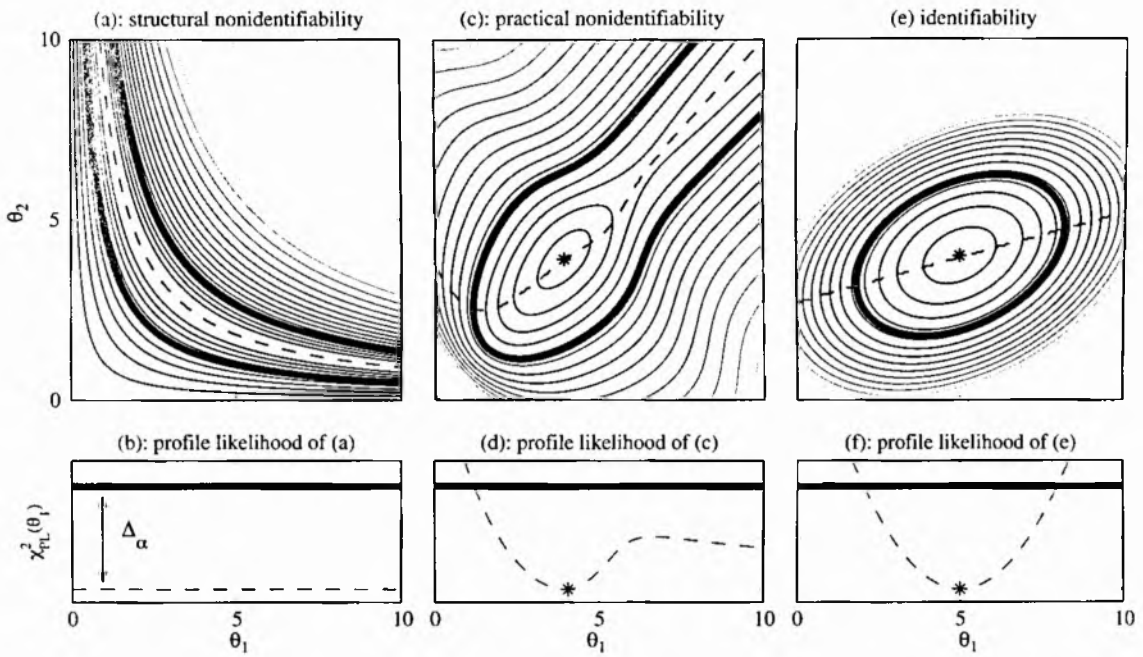

Figure 20.5 *Assessing the identifiability of parameter θ_1 by the profile likelihood $\chi^2_{PL}(\theta_1)$. (a) A structural nonidentifiability along the functional relation $h(\theta_1, \theta_2) = \theta_1 \cdot \theta_2 - 10 = 0$ manifesting in a flat profile likelihood in (b). (c) A practical non identifiability manifesting in a flattening out of the profile likelihood for $\theta_1 \to \infty$ in (d). A lower confidence bound can be assessed by the point where $\chi^2_{PL}(\theta_1)$ exceeds Δ_α. (e) An identifiable parameter θ_1. The profile likelihood approaches a parabola shape yielding a finite confidence interval. (a,c,e) Contour lines shaded from black to white correspond to low to high values of $\chi^2(\theta)$. Thick contour lines indicate likelihood-based confidence regions and asterisks correspond to the optimal parameters $\hat{\theta}$. Dashed lines indicate the trace of the profile likelihood for parameter θ_1 in terms of parameter θ_2. (b,d,f) Dashed lines indicate the profile likelihood χ^2_{PL} of parameter θ_1. The thick lines display the threshold Δ_α utilized to assess likelihood-based confidence regions for a confidence level α*

20.4.2 Model reduction

The approach can be used for model reduction by considering a threshold Δ_α corresponding to point-wise confidence intervals using $df = 1$ [see Equation (20.24)]. Assume a parameter θ_i is practically nonindentifiable for decreasing parameter value. Consider a reduced model $\mathcal{M}^*$ with simplified kinetics concerning θ_i, e.g. for mass-action kinetics by removing the reaction. The reduced and the original model are nested $\mathcal{M}^* \subset \mathcal{M}$. In this case, the threshold Δ_α corresponds to a likelihood-ratio test of the reduced model $\mathcal{M}^*$ against the original model $\mathcal{M}$ to a significance level α, as introduced in Section 20.2.2. Falling below this threshold for θ_i being nonindentifiable, the profile likelihood indicates that it is not possible to dismiss the reduced model $\mathcal{M}^*$ in favor of $\mathcal{M}$, based on the available experimental data. Hence, for practical reasons, the smaller model is favorable.

20.4.3 Observability and confidence intervals of trajectories

As mentioned in Section 20.3.3, identifiability of model parameters has a direct impact on the observability of model trajectories. Nonobservability can be investigated by plotting both the internal model trajectories

$\vec{x}$ and the model outputs $\vec{y}$ for parameter values sampled along the profile likelihood of a nonidentifiability. The variations reveal which trajectories are nonobservable due to nonidentifiability. In the case of identifiable parameters, the same procedure allows to assess confidence intervals of model trajectories by plotting trajectories for all parameter values sampled along the profile likelihood that are inside the confidence region defined by Equation (20.24) and taking the maximum and minimum trajectory for each time point. Here, it is sufficient to consider a threshold corresponding to point-wise confidence intervals of the parameter estimates, see Section 20.2.3, if a confidence interval for each time point separately is desired.

20.4.4 Application

This section is based on Raue *et al.* (2010).

To demonstrate the usage of identifiability and observability analysis for experimental design, the profile likelihood approach will be applied to the core model of Epo and EpoR interaction and trafficking with the corresponding subset of the experimental data. The network representation of the model is shown in Figure 20.6. Using mass-action kinetics the corresponding ODE system is given by:

$$\begin{aligned}
\dot{[\text{Epo}]} &= -k_{on} \cdot [\text{Epo}] \cdot [\text{EpoR}] + k_{on} \cdot k_D \cdot [\text{Epo_EpoR}] + k_{ex} \cdot [\text{Epo_EpoR_i}] \\
\dot{[\text{EpoR}]} &= -k_{on} \cdot [\text{Epo}] \cdot [\text{EpoR}] + k_{on} \cdot k_D \cdot [\text{Epo_EpoR}] + k_t \cdot B_{\max} - \\
&\quad -k_t \cdot [\text{EpoR}] + k_{ex} \cdot [\text{Epo_EpoR_i}] \\
\dot{[\text{Epo_EpoR}]} &= +k_{on} \cdot [\text{Epo}] \cdot [\text{EpoR}] - k_{on} \cdot k_D \cdot [\text{Epo_EpoR}] - k_e \cdot [\text{Epo_EpoR}] \\
\dot{[\text{Epo_EpoR_i}]} &= +k_e \cdot [\text{Epo_EpoR}] - k_{ex} \cdot [\text{Epo_EpoR_i}] - k_{di} \cdot [\text{Epo_EpoR_i}] - \\
&\quad -k_{de} \cdot [\text{Epo_EpoR_i}] \\
\dot{[\text{dEpo_i}]} &= +k_{di} \cdot [\text{Epo_EpoR_i}] \\
\dot{[\text{dEpo_e}]} &= +k_{de} \cdot [\text{Epo_EpoR_i}]
\end{aligned}$$

where $[\cdot]$ indicates internal model states $\vec{x}$ corresponding to species concentration.

The experimental set-up and the biological background for the model is described in detail in Becker *et al.* (2010). Briefly, in erythroid progenitor cells the dynamical properties of EpoR determine how signals encoded in the concentration of the ligand Epo are processed at the receptor level and how subsequently downstream signaling cascades such as the JAK2-STAT5 pathway are activated. This leads to cellular responses such as differentiation and proliferation of erythrocytes. A mathematical model is used to infer the dynamical characteristics of ligand binding and ligand as well as receptor trafficking, because unoccupied EpoR is not directly accessible by experiments.

20.4.4.1 Initial set-up

In a first experiment, experimental data were recorded in triplicates in two compartments: in the extracellular medium (y_1) and bound to EpoR on the cell membrane (y_2)

$$\begin{aligned}
y_1 &= scale \cdot ([\text{Epo}] + [\text{dEpo_e}]) \\
y_2 &= scale \cdot [\text{Epo_EpoR}],
\end{aligned}$$

see Figure 20.7(a). After parameter estimation, a good agreement of model output and experimental data yielding a value of the objective function $\chi^2 = 6.55$ for 16 data points and 10 free parameters is obtained, see Section 20.2.2. To investigate the uncertainty of the parameter estimates the profile likelihood of each parameter was evaluated, as displayed in Figure 20.8(a). The calculation takes about 30 seconds per parameter on a normal office computer.

The flatness of the profile likelihood reveals that parameters $B_{\max}$, Epo_0, k_D, k_{on} and *scale* are structurally nonidentifiable. The change of the other parameters along the profile likelihood of one of these parameters

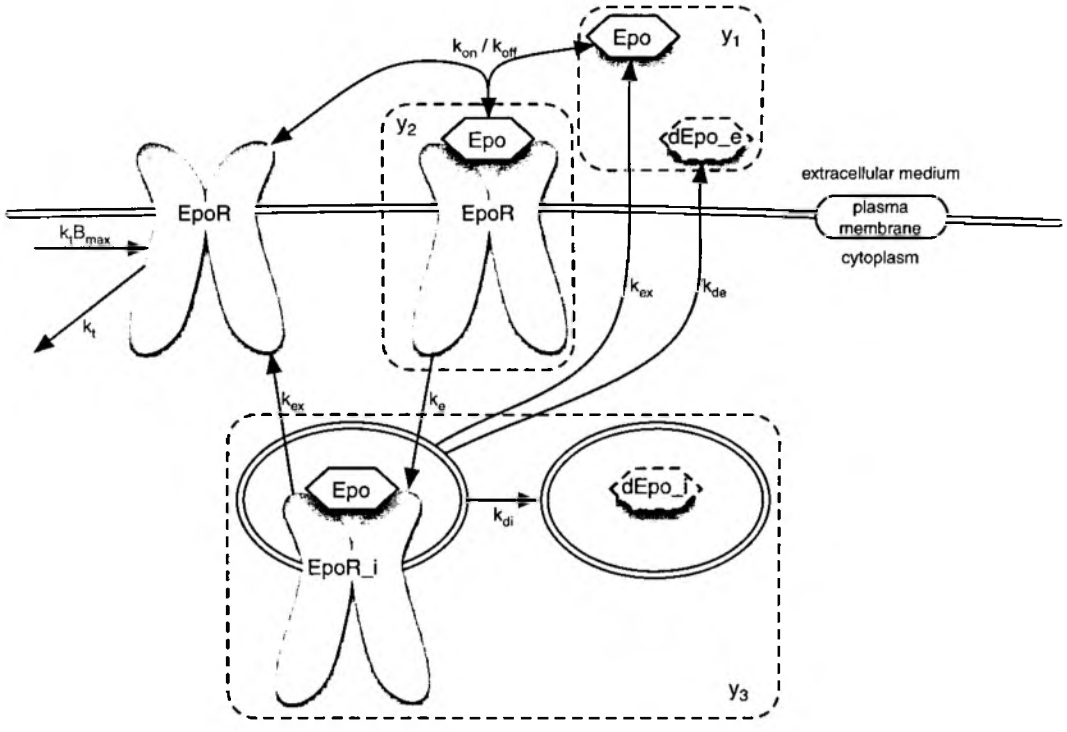

Figure 20.6 *Model of Epo and EpoR interaction and trafficking. Dashed boxes correspond to the quantities accessible by measurements*

reveals a functional relation $\vec{h}$ linking the five structurally nonidentifiable parameters, see e.g. $\chi^2_{PL}(k_{on})$ in Figure 20.9.

The corresponding effect on the model trajectories $\vec{x}$ and $\vec{y}$ of this concerted change in the parameters can be illustrated by plotting the model trajectories for parameter values along the profile likelihood of one of these parameters, see e.g. $\chi^2_{PL}(k_{on})$ in Figure 20.7(a). As expected, the model outputs $\vec{y}$ are not affected, but the trajectories of internal model states $\vec{x}$ are shifted by a common factor. Hence, in this case the structural nonidentifiability represents a freedom in the choice of the concentration scale and is a result of missing information about absolute concentration in the experimental set-up. Consequently, all parameters containing concentration in their unit are affected. Similar results were obtained in Raue *et al.* (2009) for a model published by Swameye *et al.* (2003).

20.4.4.2 Including absolute concentrations

In order to resolve the structural nonidentifiability, the observability analysis presented in Figure 20.7(a) suggests including information about the absolute concentration of one of the species. For the experiment, the cells were treated with 2100 ± 210 pM of ^{125}I-Epo. Therefore, a prior distribution $Epo_0 \sim \mathrm{N}(2100, 210^2)$ of the parameter describing the initial Epo concentration is assumed by penalizing the objective function of (20.11) with an additional summand $(Epo_0 - 2100)^2/210^2$. Recalculating the profile likelihood verifies that the structural nonidentifiability is resolved, see Figure 20.8(b). Nevertheless, the parameter k_{ex} remains practically nonidentifiable. Its upper confidence bound is determined at $\sigma^+ = -2.230$ but its lower confidence bound is not feasible.

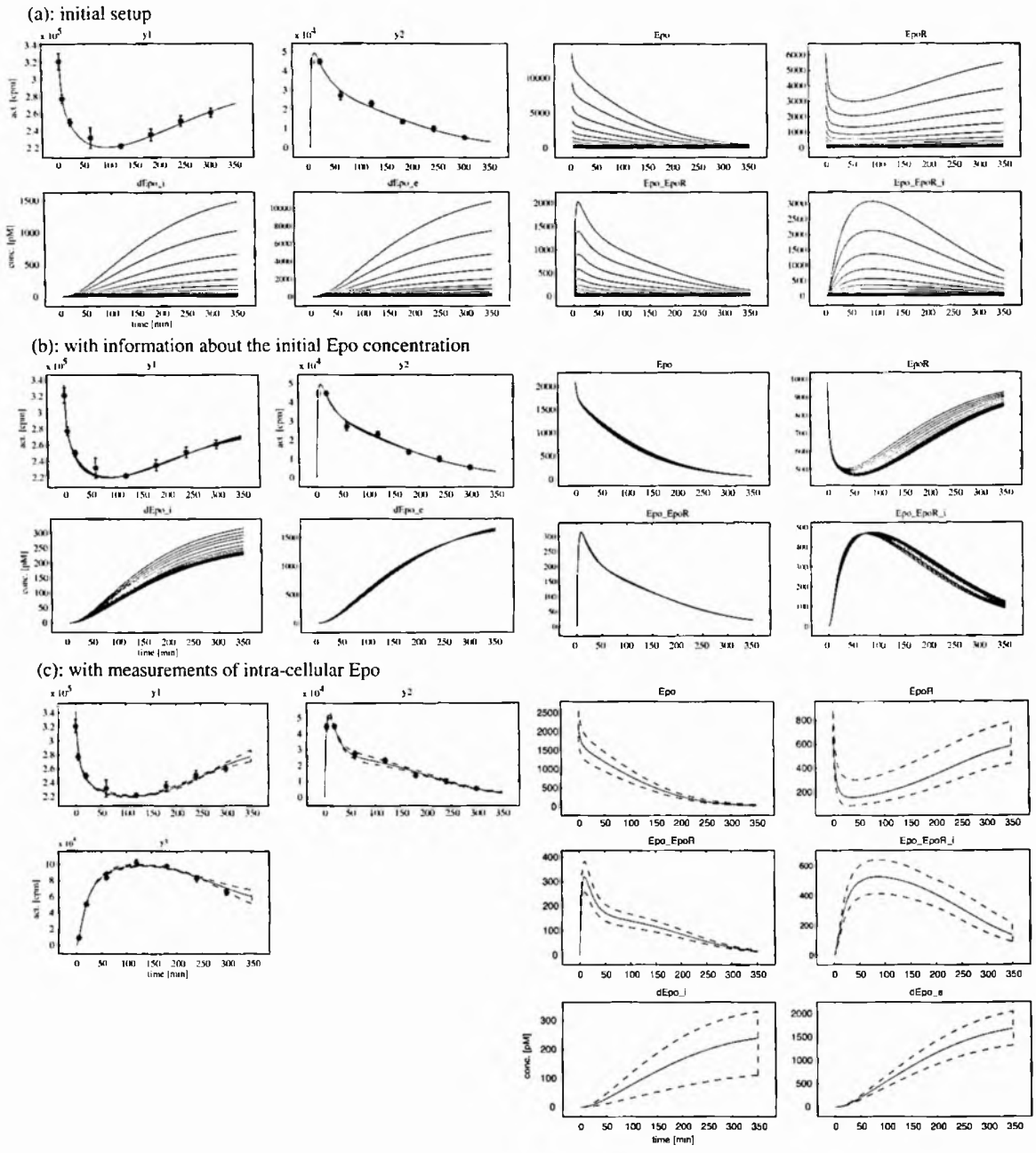

Figure 20.7 *Dependence of the trajectories of the model outputs $\vec{y}$ and of internal model states $\vec{x}$ on uncertainties in the parameter estimates. (a) The concerted change of parameters along the structural nonidentifiability of k_{on} does not affect the model outputs but shifts the trajectories of the internal model states by a common factor. (b) The practical nonidentifiability of k_{ex} only slightly affects the model outputs $\vec{y}$, staying in agreement with the measurement precision of the experimental data. Nevertheless the trajectories of EpoR, Epo_EpoR_i and dEpo_i are affected. (c) The remaining uncertainties of the identifiable parameters translate to confidence intervals of the model trajectories*

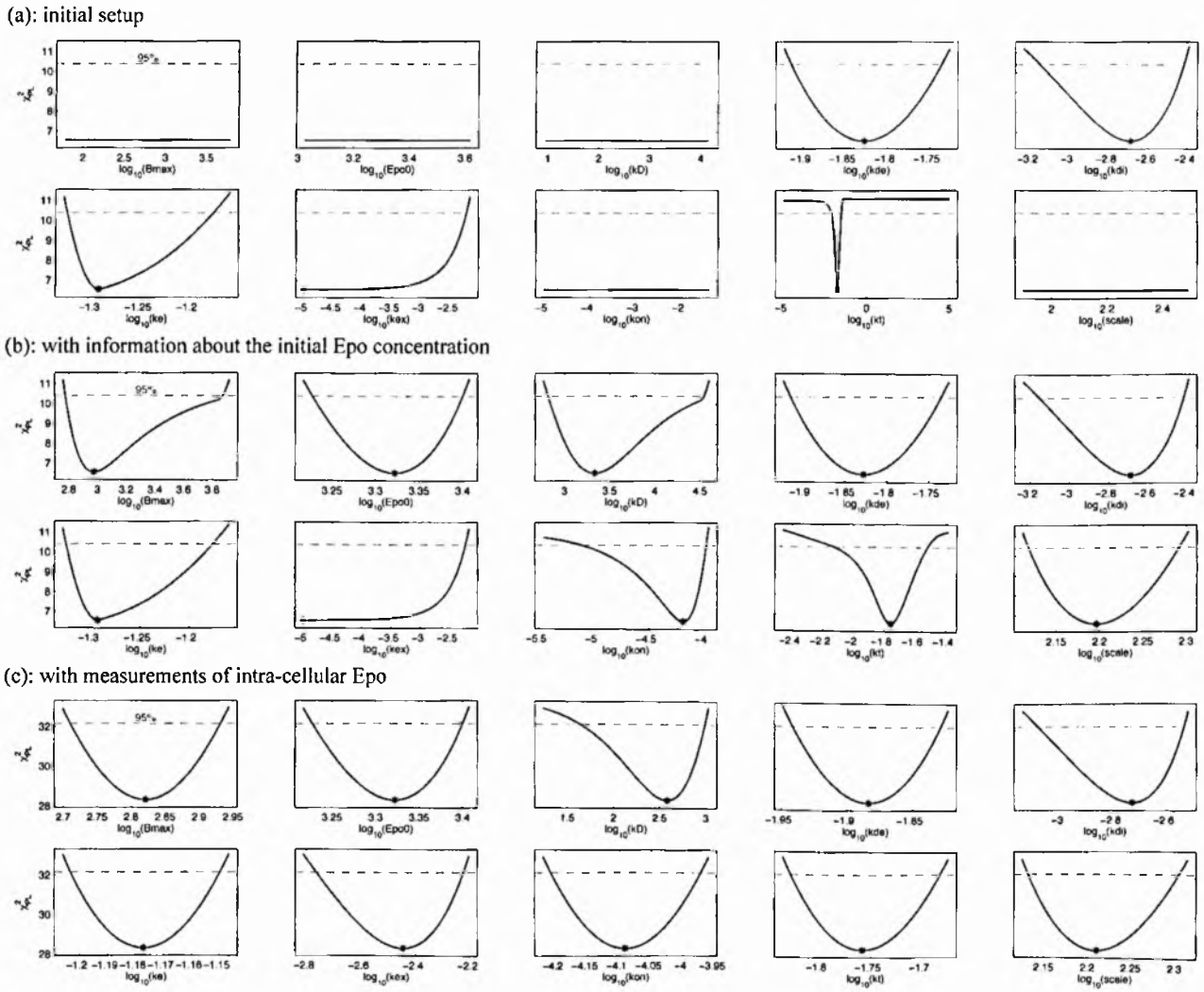

Figure 20.8 *Profile likelihood χ^2_{PL} of the model parameters displayed in combination with the thresholds $\Delta_{0.95}$ yielding with $df = 1$ confidence intervals that hold for each parameter individually. The optimal parameter value is indicated by an asterisk, if unique. (a) The flatness of the profile likelihood reveals that five parameters are structurally nonidentifiable, given the initial set-up. (b) By including information about the initial Epo concentration the structural nonidentifiability can be resolved. Parameter k_{ex} remains practically nonidentifiable. (c) By including measurements of intracellular Epo all parameters are made structurally and practically identifiable*

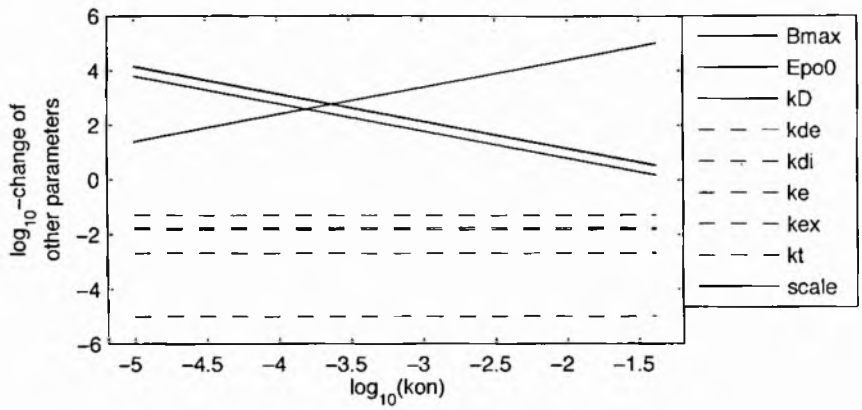

Figure 20.9 *Initial set-up. The change of the other parameters along the profile likelihood $\chi^2_{Pl}(k_{on})$ indicates functional relations between all five structurally nonidentifiable parameters (indicated by the solid lines)*

20.4.4.3 Including measurement of intracellular Epo

In order to efficiently resolve the practical nonidentifiability of k_{ex}, additional measurements need to be planned. Therefore, the variability of the model trajectories along the profile likelihood $\chi^2(k_{ex})$ is investigated, as displayed in Figure 20.7(b). The available model outputs y_1 and y_2 show only slight variations and hence are not suitable for refined measurements. However, the trajectories of Epo_EpoR_i and dEpo_i show much larger variations and suggest an additional measurement of ^{125}I-Epo inside the cell

$$y_3 = \text{scale} \cdot ([\text{Epo\_EpoR\_i}] + [\text{dEpo\_i}]),$$

that was recorded in triplicates as well. After including the additional data, the profile likelihood of all model parameters indicate their structural and practical identifiability, as displayed in Figure 20.8(c).

20.4.4.4 Confidence intervals

The resulting confidence intervals of the now identifiable parameters are finite and their values are given in Table 20.1. Finally, the remaining uncertainty of the now structurally and practically identifiable parameters can be translated to confidence intervals of the model trajectories. Therefore the trajectories corresponding to all acceptable parameter values according to the profile likelihood are evaluated. The resulting upper and lower confidence band are displayed in Figure 20.7(c) for the thresholds $\Delta_{0.95}$ that is also indicated in Figure 20.8(c) for the derivation of confidence intervals of the model parameters. Based on a confidence threshold with $df = 1$ the confidence bands have to be interpreted point-wise for each time point individually.

In order to ensure the appropriateness of the derived confidence intervals, we performed a simulation study by assuming that the estimated parameter values given in Table 20.1 are the true values. Using these values and the same model outputs, measurement time points and measurement noise as in the original experimental dataset, 450 independent datasets were generated. After estimating the parameters for each of these datasets, the resulting distribution of the over-fitting is in line with the χ^2_{df}-distribution with $df = 10$, the number of estimated parameters, see Figure 20.10(a). This verifies that the threshold Δ_α utilized to derive the likelihood-based confidence intervals was assessed correctly.

Table 20.1 *Individual confidence intervals $[\sigma^-, \sigma^+]$ of the model parameter to a confidence level of 95%. Values are given on a $\log_{10}$ scale. The coverage rates CR of the estimates are in line with the expected values (93.33 and 96.66%)*

Name	$\hat{\theta}$	σ^-	σ^+	unit	*CR*
B_{max}	+2.821	+2.710	+2.932	pM	95.78%
Epo_0	+3.322	+3.227	+3.400	pM	92.44%
k_D	+2.583	+1.641	+2.993	pM	95.78%
k_{de}	-1.884	-1.941	-1.829	1/min	95.78%
k_{di}	-2.730	-3.083	-2.535	1/min	95.56%
k_e	-1.177	-1.203	-1.150	1/min	97.33%
k_{ex}	-2.447	-2.764	-2.225	1/min	96.44%
k_{on}	-4.091	-4.208	-3.973	1/(pM·min)	95.56%
k_t	-1.758	-1.828	-1.683	1/min	97.33%
scale	+2.210	+2.133	+2.305	cpm/pM	92.89%

20.4.4.5 Coverage rates

In order to compute coverage rates for the parameter estimates, for each of the simulated data sets 95% individual confidence intervals were calculated. According to the 0.05 and 0.95 quantiles of the binomial distribution, the coverage rates CR for 450 simulated datasets and a 95% confidence level are expected to be between 93.33% and 96.66%. The coverage rates are in line with this expectation, see in Table 20.1. Coverage rates can also be assessed for the time point-wise confidence bands on the model trajectories shown in Figure 20.7(c), see Figure 20.10(b). The coverage rates are, in most cases, in line with their expected values described by the 0.05 and 0.95 quantiles of the binomial distribution. Too low coverage rates are observed for species Epo at $t \approx 20$ minutes, for Epo_EpoR at $t > 300$ minutes and for Epo_EpoR_i between 150 and 250 min. Here, increased attention is indicated when interpreting the results. The deviations occur because the high dimensional confidence region in parameter space was not sampled densely but approximated by the profile likelihood.

20.5 Summary

The iterative cycle between mathematical modeling and experimentation in systems biology facilitates a better understanding of the highly complex processes that occur in cell biology. We elucidated a likelihood-based framework for model identification that allows to systematically proceed from the building of an ODE model, to the practical aspects of parameter estimation, to the analysis of the uncertainty of parameter estimates in terms of identifiability and confidence intervals, to the corresponding effect of the uncertainty on the model trajectories in terms of observability, and finally to the design of new experiments that efficiently improve model predictability until the relevant biological question can be assessed reliably. Our argumentation was based on the profile likelihood approach and was illustrated by a realistic example of ligand and receptor interaction and trafficking.

The profile likelihood is a powerful concept to infer parameter uncertainties in a high-dimensional parameter space. Since it is a systematic and directed exploration of the parameter space it has less computational cost than sampling parameter space randomly, which becomes intractable for high dimensions. The profile likelihood can be calculated for each parameter separately. Thereby it is possible to restrict the analysis to the parameters relevant for the specific question. Moreover, this allows to perfectly parallelize the approach, which is a major

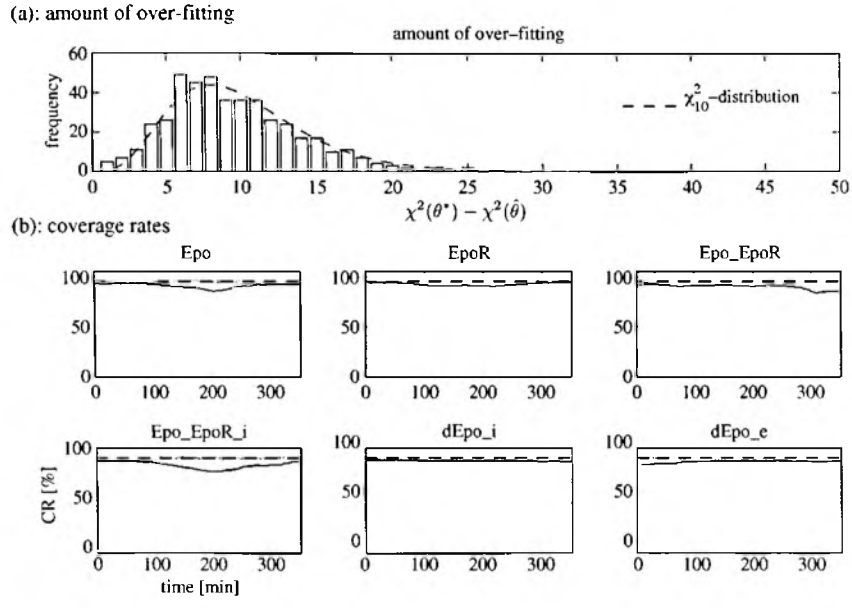

Figure 20.10 *The appropriateness of confidence intervals was assessed by a simulation study with 450 generated data sets. (a) Comparison of the amount of over-fitting to the expected χ^2_{df}-distribution with $df = 10$. (b) The solid lines indicate the coverage rates CR for the time point-wise confidence bands on the model trajectories shown in Fig. 20.7c. The dashed lines are the desired 95% level of confidence, the gray shades are the expected coverage rates.*

benefit for its scalability. The approach can be applied to any parameter estimation problem, where a likelihood or a similar objective criterion is available, e.g. partial differential equations, stochastic differential equations or a Bayesian framework.

The approach results in easily interpretable plots of profile likelihood versus parameter. It can be automated, but an explicit advantage is that the output can be evaluated visually. This gives insight into a complex and high-dimensional parameter space. Structural nonidentifiabilities, originating from incomplete observation of the internal model states can also be detected. Arising from a limited amount and quality of experimental data, practical nonidentifiabilities can a be inferred. Bridging the gap between identifiability and confidence intervals, the profile likelihood allows to derive likelihood-based confidence intervals for each parameter. Functional relations between parameters occurring due to nonidentifiabilities can be recovered. The results of the approach can, on the one hand, be used to design new experiments that efficiently resolve non identifiability and narrow confidence intervals, and on the other hand be used for model reduction. Whether a model that is not well determined should be reduced or additional data should be measured depends on the issue being addressed. Thus, identifiability analysis ensures that the model complexity is tailored to the information content given by the experimental data.

Acknowledgements

We thank Clemens Kreutz, Daniel Kaschek, Seong-Hwan Rho, Thomas Maiwald, Verena Becker, Marcel Schilling, Julie Bachmann and Ursula Klingmüller for their support.

This work was supported by the German Federal Ministry of Education and Research (Virtual Liver, LungSys 0315415E, FRISYS 0313921); the European Union (CancerSys EU-FP7 HEALTH-F4-2008-223188); the Initiative and Networking Fund of the Helmholtz Association within the Helmholtz Alliance on Systems Biology (SBCancer DKFZ II.1) and the Excellence Initiative of the German Federal and State Governments (EXC 294).

References

Becker V, Schilling M, Bachmann J, Baumann U, Raue A, Maiwald T, Timmer J and Klingmueller U 2010 Covering a broad dynamic range: information processing at the erythropoietin receptor. *Science* **328**(5984), 1404–1408.

Cobelli C and DiStefano III J 1980 Parameter and structural identifiability concepts and ambiguities: a critical review and analysis. *American Journal of Physiology-Regulatory, Integrative, and Comparative Physiology* **239**(1), 7–24.

Cox D and Hinkley D 1974 *Theoretical Statistics*. Cambridge University Press.

Egea J, Rodríguez-Fernández M, Banga J and Martí R 2007 Scatter search for chemical and bio-process optimization. *Journal of Global Optimization* **37**(3), 481–503.

Gerstein M, Bruce C, Rozowsky J, Zheng D, Du J, Korbel J, Emanuelsson O, Zhang Z, Weissman S and Snyder M 2007 What is a gene, post-encode? History and updated definition. *Genome Research* **17**(6), 669–81.

Hindmarsh A, Brown P, Grant K, Lee S, Serban R, Shumaker D and Woodward C 2005 Sundials: suite of nonlinear and differential/algebraic equation solvers. *ACM Transactions on Mathematical Software* **31**(3), 363–396.

International Human Genome Sequencing Consortium 2004 Finishing the euchromatic sequence of the human genome. *Nature* **431**(7011), 931–945.

Joshi M, Seidel-Morgenstern A and Kremling A 2006 Exploiting the bootstrap method for quantifying parameter confidence intervals in dynamical systems. *Metabolic Engineering* **8**(5), 447–455.

Kollmann M, Løvdok L, Bartholomé K, Timmer J and Sourjik V 2005 Design principles of a bacterial signalling network. *Nature* **438**(7067), 504–507.

Kreutz C, Rodriguez MMB, Maiwald T, Seidl M, Blum HE, Mohr L and Timmer J 2007 An error model for protein quantification. *Bioinformatics* **23**(20), 2747–2753.

Lehmann E and Leo E 1983 *Theory of Point Estimation*. John Wiley & Sons, Ltd.

Leis J and Kramer M 1988 The simultaneous solution and sensitivity analysis of systems described by ordinary differential equations. *ACM Transactions on Mathematical Software* **14**(1), 45–60.

Maiwald T and Timmer J 2008 Dynamical modeling and multi-experiment fitting with potterswheel. *Bioinformatics* **24**(18), 2037–2043.

Meeker W and Escobar L 1995 Teaching about approximate confidence regions based on maximum likelihood estimation. *The American Statistician* **49**(1), 48–53.

Murphy S and van der Vaart A 2000 On profile likelihood. *Journal of the American Statistical Association* **95**(450), 449–485.

Neale M and Miller M 1997 The use of likelihood-based confidence intervals in genetic models. *Behavior Genetics* **27**(2), 113–120.

Press W, Teukolsky S, Flannery B and Vetterling W 1990 *Numerical Recipes: FORTRAN*. Cambridge University Press.

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.

Raue A, Becker V, Klingmüller U and Timmer J 2010 Identifiability and observability analysis for experimental design in non-linear dynamical models. *Chaos* **20**(4), 045105.

Royston P 2007 Profile likelihood for estimation and confidence intervals. *Stata Journal* **7**(3), 376–387.

Schmidt H and Jirstrand M 2006 Systems biology toolbox for matlab: a computational platform for research in systems biology. *Bioinformatics* **22**(4), 514–515.

Seber G and Wild C 2003 *Nonlinear Regression*. John Wiley & Sons, Ltd.

Segel I 1993 *Enzyme Kinetics: Behavior and Analysis of Rapid Equilibrium and Steady-state Enzyme Systems*. Wiley-Interscience.

Swameye I, Müller T, Timmer J, Sandra O and Klingmüller U 2003 Identification of nucleocytoplasmic cycling as a remote sensor in cellular signaling by databased modeling. *Proceedings of the National Academy of Sciences of the United States of America* **100**(3), 1028–1033.

Taniguchi Y, Choi P, Li G, Chen H, Babu M, Hearn J, Emili A and Xie X 2010 Quantifying e. coli proteome and transcriptome with single-molecule sensitivity in single cells. *Science* **329**(5991), 533.

Venzon D and Moolgavkar S 1988 A method for computing profile-likelihood-based confidence intervals. *Applied Statistics* **37**(1), 87–94.

Wolkenhauer O 2008 *Systems Biology*. Portland Press.