Profile likelihood in systems biology

Clemens Kreutz, Andreas Raue, Daniel Kaschek, Jens Timmer

Angaben zur Veröffentlichung / Publication details:

Kreutz, Clemens, Andreas Raue, Daniel Kaschek, and Jens Timmer. 2013. "Profile likelihood in systems biology." *The FEBS Journal* 280 (11): 2564–71. https://doi.org/10.1111/febs.12276.

Profile likelihood in systems biology

Clemens Kreutz[1,2], Andreas Raue[1,3], Daniel Kaschek[1] and Jens Timmer[1,2,4,5]

1 Physics Department, University of Freiburg, Germany
2 Freiburg Institute for Advanced Studies (FRIAS), University of Freiburg, Germany
3 Institute of Bioinformatics and Systems Biology, Helmholtz Center Munich, Neuherberg, Germany
4 Freiburg Center for Biosystems Analysis, University of Freiburg, Germany
5 BIOSS Centre for Biological Signalling Studies, University of Freiburg, Germany

Keywords
parameter estimation; prediction; profile likelihood; propagation of errors; uncertainty

Correspondence
C. Kreutz, Physics Department, University of Freiburg, Hermann-Herder Str. 3, 79104 Freiburg, Germany
Fax: +49 761 203 5754
Tel: +49 761 203 8533
E-mail: ckreutz@fdm.uni-freiburg.de

doi:10.1111/febs.12276

Inferring knowledge about biological processes by a mathematical description is a major characteristic of Systems Biology. To understand and predict system's behavior the available experimental information is translated into a mathematical model. Since the availability of experimental data is often limited and measurements contain noise, it is essential to appropriately translate experimental uncertainty to model parameters as well as to model predictions. This is especially important in Systems Biology because typically large and complex models are applied and therefore the limited experimental knowledge might yield weakly specified model components. Likelihood profiles have been recently suggested and applied in the Systems Biology for assessing parameter and prediction uncertainty. In this article, the profile likelihood concept is reviewed and the potential of the approach is demonstrated for a model of the erythropoietin (EPO) receptor.

Introduction

A major aim of Systems Biology is the establishment of mathematical models of biological processes like signal transduction, metabolism, or gene regulation in order to gain insight in these nonlinear dynamical systems. As an initial step, an appropriate model structure has to be identified, i.e. the relevant molecular compounds and the nature and characteristic of their interactions. Then, the model's parameters like concentrations of compounds and rate constants are estimated from experimental data to calibrate the model. For this calibration step, an objective function assessing the goodness of fit can be optimized, e.g. the parameters are chosen to minimize deviations between measurements and model. A very efficient and flexible objective function for this purpose is the so-called *likelihood* which coincides with the least-squares criterion in typical Systems Biology applications.

An essential task of the modelling procedure is the assessment of uncertainty, e.g. by calculating confidence intervals for parameters and predictions. In the classical regression setting, this is typically accomplished by so-called *standard errors*, i.e. by propagating the measurement uncertainty using the *Gaussian law of error propagation* which is based on linearization of the model.

In Systems Biology, the models are typically *mechanistic*, i.e. the components of the models have counterparts in the biological process. Therefore, the mathematical models are typically nonlinear and more

Abbreviations
CI, confidence interval; dEpo, degraded EPO; EPO, erythropoietin; EpoR, EPO receptor; icdf, inverse cumulative density function; LL, log-likelihood function; ODE, ordinary differential equation; PCI, prediction confidence interval; PL, profile likelihood; PPL, prediction profile likelihood; SD, standard deviation; VCI, validation confidence interval.

complex than in a regression setting. Frequently, ordinary differential equations are used to describe the dynamics of biochemical interactions. For such models, the likelihood is nonlinear and therefore confidence regions for model parameters can exhibit complex shapes. This renders classical approaches as rough approximations in the finite sample case. Sometimes, they are even infeasible, e.g. if structurally non-identifiable parameters are present.

In contrast, the *profile likelihood* approach [1,2] results in confidence intervals which are invariant under parameter transformations [3] and therefore not affected by nonlinear distortions of the likelihood landscape. The profile likelihood is a one-dimensional representation of the likelihood indicating which values of a single parameter component are in statistical agreement with the available measurements. In the Systems Biology setting, the parameter profile likelihood has been proposed for the calculation of confidence intervals and in addition for the investigation of parameter identifiability [4,5]. It is increasingly applied in recent years [6–12].

For the more general setting of a model prediction, a respective theoretical concept was established decades ago [13,14]. However, the classical calculation of a prediction profile likelihood requires analytical formulas which are only available for trivial ODE models. To circumvent this hurdle, the prediction profile likelihood approach was presented in the context of differential equation models [15,16]. Subsequently, this concept and its use for investigating practical non-observability were rephrased in [17], but without making reference to earlier literature.

The suggested calculation procedure in [15,16] derives the prediction profiles likelihood either based on numerical constraint- or on penalized optimization. In these publications, it has been demonstrated by Monte-Carlo simulations, that the resulting confidence intervals have desired statistical properties like proper coverage. Moreover, the prediction profile likelihood has been utilized for a data-based observability analysis and for experimental design considerations. Within this concept, sampling of the parameter space is replaced by optimization which constitutes the most efficient way to numerically evaluate the parameter space.

In the following, the potential of likelihood profiles in Systems Biology is discussed and illustrated. For this purpose, a model of the EPO receptor is used [6].

Methodology

Experimental observations are always compromised by measurement errors. A general goal of statistical analyses is to evaluate feasible conclusions despite this uncertainty. For this purpose, experimental data y is described by a probability density $\rho(y|\theta)$ with parameters θ. A mathematical model of a biological process typically describes the relationship $\rho(y|\theta)$ between parameters and data, comprising experimental conditions like time or treatment. For biochemical reactions in the cell, as an example, the dependency can be described by ordinary differential equations

$$\dot{x}(t) = f(x(t), u(t), \theta) \tag{1}$$

[Equation 1 was corrected on 19 July 2013 after original online publication]

for the concentrations x of molecular compounds. $u(t)$ denotes the input to the system, e.g. a treatment or stimulation. f is given by rate equations like the law of mass action or the Michaelis–Menten rate law [18]. The time course $x(t)$ of the concentrations is calculated by integration of Eqn (1). For comparing the model with experimental data, the dynamic variables x are mapped to the experimentally observed quantities

$$y(t) = g(x(t), \theta) + \varepsilon(t). \tag{2}$$

by the so-called *observation function* g. Typically, the noise ε is additive either on the nominal or on the logarithmic scale [19] although this is not required for the presented formalism. The parameter vector θ comprises the kinetic parameters of f, like rate constants or Hill coefficients, as well as the initial concentrations $x(0)$, and additional offset or scaling parameters for the observations contained in g. Equations (1) and (2) comprise the effect of the parameters, of time, and treatment on the studied system and the expected outcome of an experiment and is referred as the *state space model* in literature.

EPO receptor model

Figure 1A shows the EPO receptor model and experimental data as published in [6] which is used for demonstration purpose in the following. Briefly, EPO can bind to its membrane receptor (EpoR). The Epo_EpoR complex activates the downstream signaling, e.g. the JAK2/STAT5 signaling cascade [7]. The Epo_EpoR complex can be internalized (Epo_EpoR_i) and degraded. Degraded EPO can accumulate inside (dEpo_i) or outside (dEpo_e) of the cell. Unoccupied receptors EpoR are constantly transported to the cell membrane and degraded with turnover rate k_t. Translating the depicted interactions using mass action kinetics yields a system of six ordinary differential equations with nine kinetic parameters which are complemented by one parameter

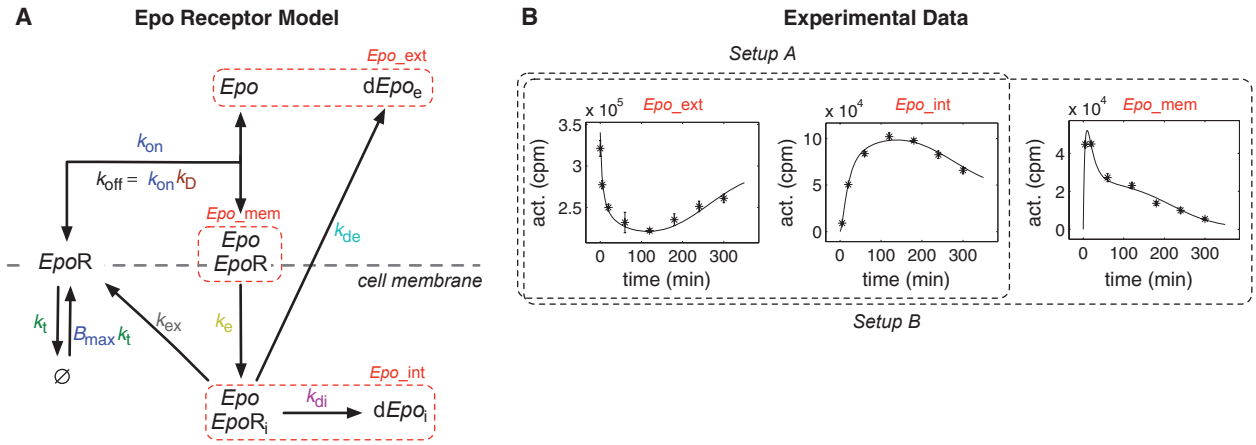

Fig. 1. Structure of molecular interactions and experimental data. (A) EPO receptor model. EPO can bind to its membrane receptor (EpoR). The Epo_EpoR complex activates the downstream signaling. The Epo_EpoR complex can be internalized (Epo_EpoR_i) and degraded. Degraded Epo can accumulate inside ($dEpo_i$) or outside ($dEpo_e$) of the cell. (B) Experimental data obtained by labeled EPO in different compartments. In the experimental *Setup A*, time-course data of EPO in the extracellular medium (Epo_ext) and of intracellular EPO (Epo_int) attached are available. In the experimental *Setup B*, additionally data of the receptor amounts on the cell membrane (Epo_mem) are available as well as estimates of B_{max} and k_D from Scatchard analysis.

for the observations, see in [6] for details. Two different stages of experimental setup will be used in the following. In the basic experimental *Setup A*, time-course data of EPO in the extracellular medium $Epo\_ext = scale \cdot (Epo + dEpo_e)$ and of intracellular EPO, $Epo\_int = scale \cdot (Epo\_EpoR_i + dEpo_i)$ are available. The parameter *scale* accounts for the unknown absolute physical unit of the data. In the comprehensive experimental *Setup B*, a Scatchard analysis was performed yielding further data for parameters B_{max} and k_D. Additionally time-course data of the receptor attached to the cell membrane $Epo\_mem = scale \cdot Epo\_EpoR$ are available, see in Fig. 1B. This setup is identical to the extended experimental setup investigated in [5,7]. In both setups, the amount of stimulating EPO in the medium is assumed to be known without error. In order to match the model's observables with the experimental data, the parameters have to be estimated as discussed in the following.

Parameter estimation

Estimation of parameters from measurements can be accomplished by calculating the likelihood $L(y|\theta)$ which denotes the probability of the measured data y, given a model with parameters θ. For statistically independent additive noise, the likelihood is given by the product

$$L(y|\theta) = \prod_i \rho(y_i|\theta) \tag{3}$$

and the *maximum likelihood estimator (MLE)*

$$\hat{\theta} = \arg\max_{\theta} L(y|\theta) \tag{4}$$

is the parameter vector maximizing the likelihood. Maximum likelihood estimation is widely applied in statistics because of its beneficial properties like efficiency and consistency [20]. For additive Gaussian noise $\varepsilon \sim N(0, \sigma^2)$ with known variance σ^2, MLE is equivalent to least squares estimation

$$\hat{\theta} = \arg\min_{\theta} \sum_i (y_i - g(t_i, u, \theta))^2 / \sigma^2. \tag{5}$$

The right hand side of Eqn (5) is proportional to minus two times the log-likelihood, $-2LL$, and is called the χ^2 or goodness of fit statistic in literature [21]. $-2LL$ is usually easier to interpret than the likelihood L because it typically has the same order of magnitude as the number of data points if the model is appropriate. Since maximization of the likelihood L and minimization of $-2LL$ is equivalent, the discussion will be focused on the least square setting in the following without loss of generality.

Parameter profile likelihood

The impact of the value of a parameter component for fitting the model to the data can be assessed by the *profile likelihood*

$$PL_j(p) = \max_{\theta \in \{\theta | \theta_j = p\}} LL(y|\theta), \tag{6}$$

i.e. the log-likelihood is evaluated as a function of the values p of a parameter component θ_j while all other

parameters θ_i, $i = j$ are reoptimized. Confidence intervals

$$\mathrm{CI}_{j,\alpha}(y) = \left\{ p| - 2\ \mathrm{PL}_j(p) \leq \min_{\theta} -2\ LL(y|\theta) + \Delta(\alpha) \right\} \tag{7}$$

for the estimation of the j'th parameter component are given by a threshold $\Delta(\alpha)$ according to the confidence level α [3]. Asymptotically, i.e. for a sufficiently large number of data points, the threshold

$$\Delta(\alpha) = \mathrm{icdf}(\chi^2_1, \alpha) \tag{8}$$

is given by the α-quantiles of a χ^2 distribution with one degree of freedom. These quantiles are given by the inverse cumulative density function denoted by icdf.

In general, a flat profile likelihood indicates an infinite size of the confidence interval for all confidence levels α which corresponds to a *structural non-identifiability*. In such a case, changing the parameter component has no impact on the likelihood, i.e. the effect can be compensated by adjusting other parameters. Therefore, the data provides no information about the respective parameter component.

If the profile likelihood has a unique minimum but does not exceed the threshold in at least one direction, e.g. exhibits a plateau below the threshold, the parameter is termed *practically non-identifiable* [4]. In such a case, the data contain information about the parameter, but in terms of significance, the supposed parameter range is not restricted towards small and/or large values.

Figure 2 shows the profile likelihood for all parameters of the EPO receptor model for the two experimental setups. For the basic experimental *Setup A*, plotted in the upper panel, three profiles do not exceed the threshold and therefore indicate practical non-identifiability. The profiles of k_{ex} and k_D are monotonically decreasing towards small values, the profile likelihood for k_{di} exhibits a flat plateau below the 95% confidence threshold.

In general, additional experiments have to be performed to resolve such identifiability issues. In the lower panel of Fig. 2, the outcome is plotted for the comprehensive experimental *Setup B*. Here, all likelihood profiles indicate identifiability because the threshold is exceeded in upward- and in downward direction. In such circumstances, the respective confidence intervals cover only a finite range.

Prediction profile likelihood

The parameter profile likelihood yields the dependency of the likelihood on a single parameter component. This idea can be generalized by a more general constraint optimization of the likelihood, i.e. instead of fixing a single parameter component like in Eqn (6), a constraint for a prediction F is introduced [15,16]. This yields the *prediction profile likelihood* which is given by

$$\mathrm{PPL}_F(z) = \max_{\theta \in \{\theta | F(\theta) = z\}} \mathrm{LL}(y|\theta). \tag{9}$$

Here, maximization is performed only for the subset of parameters with model response $F(\theta)$ equals to z.

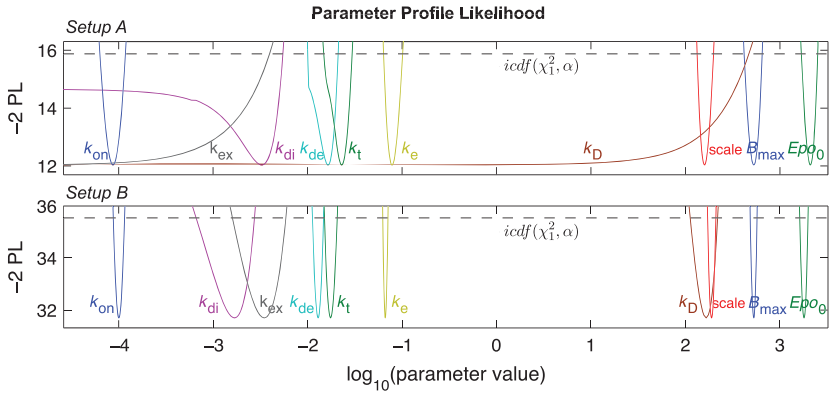

Fig. 2. Parameter profile likelihood. Likelihood profiles for all parameters for two experimental setups, i.e. the basic experimental *Setup A* (upper panel) and the comprehensive experimental *Setup B* (lower panel). In the experimental *Setup A*, there are three practically non-identifiable parameters. Two parameters are have flat profiles towards lower values, namely the Michaelis constant k_D for binding of EPO to the receptor as well as the rate for externalization k_{ex}, i.e. recycling of the receptor to the membrane. For the degradation rate k_{di} in the cytoplasm, there is a unique minimum but the profile flattens out on a plateau below the 95% confidence threshold. In the comprehensive experimental *Setup B*, all parameters are identifiable, i.e. the profiles exceed the threshold yielding confidence intervals of finite size and the minima, i.e. the maximum likelihood estimates, are unique.

In analogy to Eqn (7), the prediction confidence interval is given by

$$\mathrm{PCI}_{F,\alpha}(y) = \left\{ z \,|\, -2\,\mathrm{PPL}_F(z) \leq \min_{\theta} -2\,\mathrm{LL}(y|\theta) + \Delta(\alpha) \right\}. \tag{10}$$

The prediction or response F could be any characteristics of a model which may serve as a constraint. Typical examples comprise the concentrations of the compounds occuring as dynamic variables but also more complex features like concentration ratios, steady states, minimal or maximal abundances, or the position and height of a peak. This flexibility emphasizes the relevance of the predictions profile likelihood.

As argued in [15,16], there is a strong relationship between the parameter and the prediction profile likelihood. On the one hand, the value of a parameter can be seen as special kind of prediction. On the other hand, a reparametrization of the model could be performed in a way that the prediction is unambiguously given by a single parameter. Then, the parameter profile likelihood for such a parameter coincides with the respective prediction profile likelihood. Due to this equivalence, the threshold $\Delta(\alpha)$ for the parameter- and prediction profile likelihood approaches coincides.

Figure 3 shows the prediction profile likelihood for predicting the concentration of degraded EPO receptors in the cytoplasm (dEpo_i). For this illustration purpose, the concentration is predicted for three time points. Applying the threshold (Eqn 8) yields the respective predictions confidence intervals. For the basic experimental *Setup A* (red lines), the dEpo_i concentration is practically non-observable which is indicated by flat prediction profiles. These profiles show that the lower boundary of the concentration of degraded receptors in the cytoplasm is not specified by the data in the basic setup. In contrast, the comprehensive experimental *Setup B* yields almost quadratic prediction profiles indicating observability. For plotting purpose, the minimum of −2LL has been subtracted so that the 95% prediction confidence intervals are given the intersection of the profiles with the threshold $\Delta(95\%) = 3.81$.

Although non-identifiability and non-observability are not independent, the relationship is typically non-trivial. In general, it only holds that non-observability requires weakly specified parameters and that a non-identifiable parameter induces some weakly specified model predictions. In our illustration, predictions of the dynamic variables have been considered. Such predictions are of primary interest in terms of

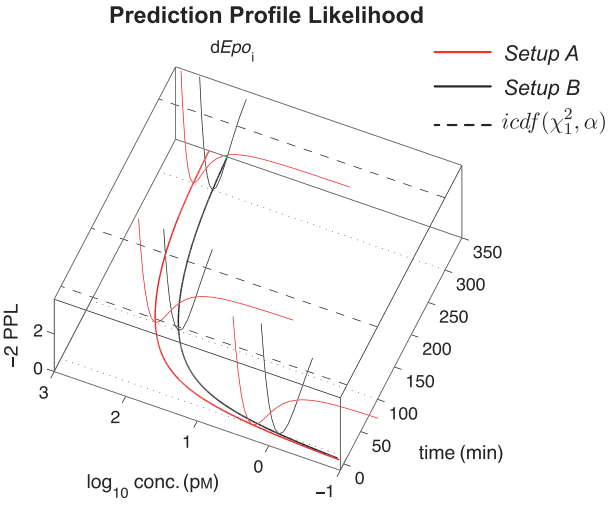

Fig. 3. Prediction likelihood profiles. Prediction profile likelihood for the dynamics of degraded EPO receptors in the cytoplasm, dEpo_i, at time points 10, 100, and 300 min. In the experimental *Setup A*, the dynamics is practically non-observable indicated by the flat profiles (red vertical lines) whereas in the comprehensive experimental *Setup B* (black vertical lines) dEpo_i is observable.

observability. The term practically non-observability has been introduced in [15] for indicating the inability of making predictions with finite size confidence intervals based on the available data. In our example, there are three practically non-identifiable parameters, but there is only a single practically non-observable dynamic state, namely dEpo_i. This practical non-observability is due to the practical non-identifiability of the parameter k_{di} controlling the production of dEpo_i. In contrast, neither the non-identifiability of the export rate of unoccupied receptors causes non-observability of membrane bound receptors, nor does the non-identifiability of k_{D} induce non-observability of receptor-ligand complexes. Because of the complex relationship between identifiability and observability, the prediction likelihood profiles provides insight which is not directly given by parameter profiles.

Profiles for validation data

A prediction confidence interval can be used to indicate uncertainty of the systems' behaviour for a condition of a new validation experiment. However, the prediction confidence intervals covers only the uncertainty of the model, i.e. the restricted knowledge about the true underlying process but not the limited accuracy of the new measurement. Depending on the noise level of a new data point, such a new validation measurement can exhibit an increased dispersion.

To account for the this effect prediction, the prediction confidence intervals have been generalized in [15,16] for the validation setting. Let z denote a potential value of a new data point with standard deviation SD of the measurement error, the validation profile likelihood is the maximized joint likelihood

$$VPL^{\mathrm{SD}}(\mathrm{z}|\mathrm{y}) = \max_{\theta} \mathrm{LL}(z, y|\theta) \tag{11}$$

of the existing data y and new data point z read as a function of the new measurement z. Again, *validation confidence intervals* are asymptotically given the set of measurements

$$\mathrm{VCI}_{\alpha}^{\mathrm{SD}}(y) = \Big\{ z| - 2\ VPL^{\mathrm{SD}}(z|y) \leq \min_{\theta} -2\ LL(z, y|\theta) + \Delta(\alpha) \Big\}. \tag{12}$$

using the same threshold as before. Validation confidence intervals are always larger than the respective prediction confidence intervals. In the limit SD $\rightarrow$ 0, both confidence intervals coincide.

Implementation

Likelihood profiles for parameters, predictions, or validation measurements are one-dimensional representations of the likelihood ratio statistic. A fundamental theorem in statistics, the so-called *Neyman-Pearson lemma* states, that the likelihood ratio is the most powerful statistic to test hypothesis related to specific model components [22]. This lemma elucidates the widespread use of likelihood ratio based methods in statistical literature and theoretically corroborates the capability of likelihood profiles.

An analytical calculation of the profile likelihood requires an explicit formula for the maximum likelihood estimate. Usually, such formulas are not available because ordinary differential equations (1) cannot be integrated analytically in general. Therefore, likelihood profiles have to be calculated numerically. Implementing Eqn (9) constitutes an optimization problem which is nonlinear with respect to the parameters and has a nonlinear equality constraint. In addition, further constraints like upper and lower boundaries for the parameter may exist. There are several numerical techniques for solving such optimization problems, e.g. summarized in [23]. Since it is usually not feasible to explicitly account for the constraint, so-called *indirect methods* can be applied, i.e. the unconstrained problem is iteratively solved, to approximate the constrained solution, e.g. by projection of the gradient on the linearised constraint.

As an alternative, it has been shown in [15,16] that the prediction profile likelihood can be calculated from the validation profile likelihood (Eqn 11) since the additional term can be interpreted as a penalty which can be subtracted after the validation profile calculation. This constitutes an elegant way to calculate the prediction and validation profiles in parallel without reformulating or impeding the optimization problem. This approach has been used in this article, the ODEs have been solved by the CVODES algorithm [24] and the trust-region method LSQNONLIN from MATLAB was used for numerical optimization. Since any continuous set of solutions of a penalized optimization problem can be adjusted to be interpreted as a solution of the constraint optimization problem as shown in [15,16], any penalization term can be utilized to find prediction profiles. A prominent class of penalties are so-called $l1$-penalties which are proportional to the absolute value of the constraint violation and are more appropriate than quadratic penalties to guarantee that constraints are exactly satisfied [25].

These numerical computations of likelihood profiles for ordinary differential equation models typically requires optimization of the likelihood for each value of the profiled parameter. Alternatively, approximate profiles can be obtained by an integration method based on the Lagrange multiplier formulation. Let $l(\theta) = -2\mathrm{LL}(y|\theta)$ and $G(\theta) = F(\theta) - F(\hat{\theta})$ be the negative log-likelihood and the constraint function, respectively. For each constraint value $G = \Delta z$, there are parameter values $\hat{\theta}_{\Delta z}$ and a Lagrange multiplier value $\hat{\theta}_{\Delta z}$ such that

$$\left.\frac{\partial l}{\partial \theta}\right|_{\hat{\theta}_{\Delta z}} + \hat{\lambda}_{\Delta z} \left.\frac{\partial G}{\partial \theta}\right|_{\hat{\theta}_{\Delta z}} = 0, \tag{13}$$

$$G(\hat{\theta}_{\Delta z}) = \Delta z, \tag{14}$$

i.e. $\hat{\theta}_{\Delta z}$ is optimal and satisfies the constraint. For smooth l and G, both equations depend smoothly on Δz and can be derived with respect to Δz resulting in an ordinary differential equation for $\hat{\theta}_{\Delta z}$ and $\hat{\lambda}_{\Delta z}$. Hence, the likelihood profile $\Delta z \rightarrow l(\hat{\theta}_{\Delta z})$ can be obtained by numerical integration instead of optimization. According to [2,26], this differential equation can be efficiently approximated using only sensitivity information, $\frac{\partial f_i}{\partial \theta_j}$, and gradient information, $\left(\frac{\partial l}{\partial \theta_j}, \frac{\partial G}{\partial \theta_j}\right)$.

This integration approach allows a considerable reduction of function evaluations compared to the optimization approach. It can be used for both, parameter profiles and prediction profiles.

Two-dimensional profiles

If the likelihood is optimized with two constraints

$$PPL_{F_1,F_2}(z_1, z_2) = \max_{\theta \in \{\theta | F_1(\theta)=z_1, F_2(\theta)=z_2\}} LL(y|\theta) \quad (15)$$

a two-dimensional profile likelihood is obtained. Such two-dimensional profiles can be used to calculate common confidence intervals for two predictions. For the special case of predicting two parameters, $PPL_{F1,F2}$ indicates the combination of values of the two parameters which are able to explain the data. This outcome is more valuable for understanding which model components are weakly specified by available experiments than one-dimensional profiles. In [6, Supplementary Fig. S13], such two-dimensional parameter profiles have been used to identify combinations of k_{on} and k_{off} rates of the EPO receptor which are in statistical agreement with the measurements. These combinations are then interpreted in terms of the trade-off between bioavailability and bioactivity of Epo-stimulating agents.

Summary

Likelihood profiles generalize traditional concepts for confidence interval calculation like standard errors or the Fisher Information to the nonlinear and finite sample setting as it is typically realized in Systems Biology applications. In addition, likelihood profiles enable the investigation of practical identifiability of parameters as well as practical observability of model predictions. As long as optimization is feasible, the method is asymptotically exact, i.e. the probability that the true parameter or the prediction for true parameters is in the confidence interval is properly controlled by the confidence level α. If the asymptotic assumption is violated due to insufficient amount of data, adapting the threshold recovers this desired property [15,16].

In this article, the methodology related to the profile likelihood has been summarized briefly. Moreover, the profile likelihood method has been demonstrated for a model of EPO receptor and the interpretations of likelihood profiles with respect to identifiability and observability have been shown.

Acknowledgements

The authors acknowledge financial support provided by the BMBF-grants 0315766-VirtualLiver, 0316042G-LungSysII as well as SBCancer DKFZ V.2 by the Helmholtz Society. The article processing charge was funded by the German Research Foundation (DFG) and the Albert Ludwigs University Freiburg in the funding program 'Open Access Publishing'.

References

1 Murphy SA & Van Der Vaart AW (1998) On profile likelihood. *J Am Stat Assoc* **95**, 449–485.
2 Venzon DJ & Moolgavkar SH (1988) A method for computing profile-likelihood-based confidence intervals. *Appl Stat* **37**, 87–94.
3 Meeker W & Escobar L (1995) Teaching about approximate confidence regions based on maximum likelihood estimation. *Am Stat* **49**, 48–53.
4 Raue A, Kreutz C, Maiwald T, Bachmann J, Schilling M, Klingmüller U & Timmer J (2009) Structural and practical identifiability analysis of partially observed dynamical models by exploiting the profile likelihood. *Bioinformatics* **25**, 1923–1929.
5 Raue A, Becker V, Klingmüller U & Timmer J (2010) Identifiability and observability analysis for experimental design in non-linear dynamical models. *Chaos* **20**, 45105.
6 Becker V, Schilling M, Bachmann J, Baumann U, Raue A, Maiwald T, Timmer J & Klingmüller U (2010) Covering a broad dynamic range: information processing at the erythropoietin receptor. *Science* **328**, 1404–1408.
7 Bachmann J, Raue A, Schilling M, Böhm ME, Kreutz C, Kaschek D, Busch H, Gretz N, Lehmann WD, Timmer J *et al.* (2011) Division of labor by dual feedback regulators controls jak2/stat5 signaling over broad ligand range. *Mol Syst Biol* **7**, 516.
8 Maiwald T, Blumberg J, Raue A, Hengl S, Schilling M, Sy SK, Becker V, Klingmüller U & Timmer J (2012) *In silico* labeling reveals the time-dependent label half-life and transit-time in dynamical systems. *BMC Syst Biol* **6**, 13.
9 Schaber J, Baltanas R, Bush A, Klipp E & Colman-Lerner A (2012) Modelling reveals novel roles of two parallel signalling pathways and homeostatic feedbacks in yeast. *Mol Syst Biol* **8**, 622.
10 Schaber J (2012) Easy parameter identifiability analysis with COPASI. *Biosystems* **110**, 183–185.
11 Steiert B, Raue A, Timmer J & Kreutz C (2012) Experimental design for parameter estimation of gene regulatory networks. *PLoS ONE* **7**, e40052.
12 Schelker M, Raue A, Timmer J & Kreutz C (2012) Comprehensive estimation of input signals and dynamics in biochemical reaction networks. *Bioinformatics* **28**, i529–i534.
13 Hinkley D (1979) Predictive likelihood. *Ann Stat* **7**, 718–728.
14 Bjornstad JF (1990) Predictive likelihood: a review. *Stat Sci* **5**, 242–254.

15 Kreutz C, Raue A & Timmer J (2011) Likelihood based observability analysis and confidence intervals for predictions of dynamic models. *BMC Syst Biol* **6**, 120.
16 Kreutz C, Raue A & Timmer J (2012) Likelihood based observability analysis and confidence intervals for predictions of dynamic models. *BMC Syst Biol* **6**, 120.
17 Cedersund G (2012) Conclusions via unique predictions obtained despite unidentifiability–new definitions and a general method. *FEBS J* **279**, 3513–3527.
18 Wolkenhauer O, Ullah M, Wellstead P & Cho K-H (2005) The dynamic systems approach to control and regulation of intracellular networks. *FEBS Lett* **579**, 1846–1853.
19 Kreutz C, Bartolome-Rodriguez MM, Maiwald T, Seidl M, Blum HE, Mohr L & Timmer J (2007) An error model for protein quantification. *Bioinformatics* **23**, 2747–2753.
20 Cox DR & Hinkley DV (1994) Theoretical Statistics. Chapman & Hall, London.
21 D'Agostino RB & Stephens MA (1986) Goodness of Fit Techniques. Marcel Dekker Inc., New York.
22 Neyman L & Pearson ES (1933) On the problem of the most efficient tests of statistical hypotheses. *Phil Trans Roy Soc A* **231**, 289–337.
23 Gould N & Leyffer S (2003) An Introduction to Algorithms for Nonlinear Optimization. Springer, Berlin, pp. 109–197.
24 Hindmarsh AC, Brown PN, Grant KE, Lee SL, Serban R, Shumaker DE & Woodward CS (2005) SUNDIALS: suite of nonlinear and differential/algebraic equation solvers. *ACM Trans Math Softw* **31**, 363–396.
25 Fletcher R & Leyffer S (1998) Numerical experience with lower bounds for miqp branch–and–bound. *SIAM J Optim* **8**, 604–616.
26 Chen J-S & Jennrich RI (2002) Simple accurate approximation of likelihood profiles. *J Comput Graph Stat* **11**, 714–732.