Solving the DREAM6 parameter estimation challenge

C. Kreutz+ [1,2], B. Steiert+ [1,2], A. Raue+ [1,2], and J. Timmer[1–5]
[1] *Institute for Physics, University of Freiburg, Germany*
[2] *Freiburg Center for Systems Biology (ZBSA), University of Freiburg, Germany*
[3] *Freiburg Institute for Advanced Studies (FRIAS), University of Freiburg, Germany*
[4] *BIOSS Centre for Biological Signalling Studies, University of Freiburg, Germany*
[5] *Department of Clinical and Experimental Medicine, Linköping University, Sweden*

+ contributed equally

The goal of the DREAM6 parameter estimation challenge is to perform experimental design considerations to estimate parameters of gene regulatory networks and to be able to extrapolate the systems' behavior, i.e. time courses of dynamic variables are predicted under perturbed conditions.

This document contains a short description about the methods applied to solve the DREAM6 parameter estimation challenge. Moreover, our experimental design strategies are summarized.

1. METHODOLOGY

1.1. Error model and likelihood

The data in the challenge is simulated by adding noise to the time courses of the dynamic variables x. More precisely, the organizers report the error model

$$y = \max[0, x + \varepsilon_{\text{abs}} + \varepsilon_{\text{rel}}(x)], \tag{1}$$

where the absolute error $\varepsilon_{\text{abs}} \sim N(0, \sigma^2_{\text{abs}})$ and the relative error $\varepsilon_{\text{rel}}(x) \sim N(0, \sigma_{\text{rel}}(x)^2)$ are Gaussian random variables with standard deviations $\sigma_{\text{abs}} = 0.1$ and $\sigma_{\text{rel}}(x) = 0.2 \cdot x$.

In general, a sum $\varepsilon_{\text{total}} = \varepsilon_{\text{abs}} + \varepsilon_{\text{rel}}$ of two independent Gaussian random variables ε_{abs} and ε_{rel} is again normally distributed $\varepsilon_{\text{total}} \sim N(0, \sigma^2_{\text{total}})$ with variance

$$\sigma^2_{\text{total}}(x) = \sigma^2_{\text{abs}} + \sigma^2_{\text{rel}}(x) \tag{2}$$

Therefore, the likelihood of measurements y is

$$L(y|\theta) = \prod_i \frac{1}{\sqrt{2\pi}\sigma_{\text{total}}} e^{-\frac{(y_i - x_i)^2}{2\,\sigma^2_{\text{total}}}} \tag{3}$$

with $\sigma_{\text{total}} = \sigma_{\text{total}}(x)$ and $x = x(\theta)$. The log-likelihood is denoted as LL in the following.

Clipping negative data realizations to zero can be accounted in the likelihood by

$$L_{cut}(y|\theta) = \prod_{i|y_i>0} L(y_i|\theta) \cdot \lambda \prod_{i|y_i\equiv 0} \int_{-\infty}^{-x} \frac{1}{\sqrt{2\pi}\sigma_{\text{total}}} e^{-\frac{(y'-x_i)^2}{2\,\sigma^2_{\text{total}}}}\, dy' \tag{4}$$

where $L(y_i|\theta)$ denotes the contribution of a single data point to the likelihood in Eq. (3). The integral corresponds to the probability of negative data realizations for given $x_i(\theta)$. The product of both terms contributes additively to the log-likelihood, i.e. the proportionality constant λ is an offset at the log-scale. Because there is no natural way to weight both term and the choice has neither an impact on the parameter estimates nor on the confidence intervals, we set $\lambda = 1$.

The results of maximum likelihood estimation based on the likelihood for Gaussian noise (3) have been compared with parameter estimates using the exact likelihood (4) but only minor differences of the obtained parameter estimates have been observed. Therefore, we used (3) during the experimental planning stage to have better interpretable numbers of the objective function and applied (4) only for the final parameter estimation step.

1.2. Parameter space

To account for strictly positive parameter values, all the analyses have been performed in a logarithmic parameter space. Moreover, this accounts for the fact that changes of parameter values usually contribute multiplicatively rather than additively, i.e. usually changing a parameter by a factor a or $1/a$ have a similar impact, but adding a constant a has mostly a qualitatively different effect than the corresponding subtraction.

In our initial analyses, we restricted the parameter space to $\theta_{Hill} \in [1, 10]$ for the Hill-coefficients and $\theta_{\notin Hill}[10^{-2}, 10^2]$ for the other parameters. This restriction was weakened if parameter estimation yields parameters at the boundary of the parameter space and the profile likelihood indicate minima outside the parameter domain. For the final analyses we used a parameter domain of $[1, 4]$ for most Hill-coefficient and $[10^{-3}, 10^2]$ the other parameters. For three Hill coefficients of model $\mathcal{M}3$ we allowed a domain $[1, 8]$.

1.3. Bias for estimating Hill coefficients

A Monte-Carlo study showed that Hill coefficients in general suffer from biased estimation, i.e. there are many noise realizations yielding to estimates at the upper boundary of the parameter domain. Therefore, we

had to carefully investigate for which Hill coefficients we allow an upper boundary of the domain larger than four. This was done by visually inspecting the fits. Moreover, we counted the frequency of the occurrence of the bias in the Monte-Carlo study for each individual Hill coefficient. As an example for $v4h$ there was only a weak indications for an increase of the domain up to a value of 8 by the fits, however the Monte-Carlo study indicated that $v4h$ does not suffer from the bias in around $15-20\%$ of cases. Therefore we decided to increase the domain up to 8.

After careful evaluations, we decided to allow Hill coefficients up to 8 only in model $\mathcal{M}3$ for v_4,v_7, and v_{11}.

1.4. Parameter identifiability

The profile likelihood

$$\mathrm{PL}(\theta_i) = \max_{\theta_{j\neq i}} \mathrm{LL}(\theta|y) \tag{5}$$

for a parameter θ_i given the data y is utilized to assess parameter identifiability [9] and to calculate parameter confidence intervals

$$\mathrm{CI}_\alpha(\theta_i|y) = \left\{\theta_i \mid -2\mathrm{PL}(\theta_i) \leq -2\mathrm{LL(y)}^* + icdf(\chi^2_1, \alpha)\right\} \tag{6}$$

for the estimation of a single parameter. Here, α is the confidence level and $icdf(\chi^2_1, \alpha)$ denotes the α quantile of the chi-square distribution with one degree of freedom which is given by the respective inverse cumulative density function. LL^* is the maximum of the log-likelihood function after all parameters are optimized. In (5), the optimization is performed for all parameters except θ_i.

1.5. Performance of perturbation experiments

There are three types of perturbations, namely siRNA, gene deletion/knockout (KO), and ribosomal binding site activity enhancement ("rbs") which can be applied to all gene/protein pairs. For model $\mathcal{M}1$, there are 18 perturbations, for $\mathcal{M}2$ there are 21, and for $\mathcal{M}3$ there are 27 perturbations. A design D specifies the perturbation, the observables, as well as the time points of the measurements.

The expected information of a perturbation experiment is closely related to the uncertainty of model predictions for the respective perturbation setting. Therefore, we evaluated the model predictions for all parameters which have been calculated during profile likelihood estimation as introduced in [9]. This experimental design approach utilizes an approximate but efficient sampling of the parameter subspace which is in sufficient agreement with the experimental data.

To account for the error model, i.e. to assess the spread $\mathrm{Var}(\hat{\rho}(x|y))$ of the model predictions relative to the accuracy of experimental data, ratios

$$R(D) = \frac{\mathrm{Var}(\hat{\rho}(x(D)|y))}{\sigma^2_{total}(x(D)|\hat{\theta})} \tag{7}$$

have been calculated and perturbations/observables to obtain a ranking of the designs D. In addition to (7), we evaluated alternative strategies to relate the prediction variance with the data variance. However, (7) turned out to yield to most reasonable suggestions.

1.6. Monte-Carlo design evaluation

To confirm the efficiency of a candidate design, we applied Monte-Carlo experimental planning as published in [4]. Here, we generated simulated data according to error model with the current parameter estimates. Then, we checked the impact of such an additional data set on the parameter confidence intervals, i.e. profile likelihood, or on the accuracy of the demanded *model extrapolation*, i.e. the time course predictions which are scored by the organizers.

The Monte-Carlo approach does not rely on any restrictive assumptions. However, because Monte-Carlo is numerically demanding, it was only applied at specific points in our experimental planning procedure where we wanted to confirm whether a suggested experiment indeed provides the intended information.

1.7. Experimental planning in the asymptotic setting

An asymptotic setting is obtained if the experiments provide sufficient information about the parameters. Then, the profile likelihood can be approximated by a quadratic function. In such a situation, the Fisher-Information matrix

$$F_{jk}(\theta|y) = \frac{\partial^2 \mathrm{LL}}{\partial\theta_j \partial\theta_k} = 2H_{jk}(\theta|y) \tag{8}$$

which is the inverse of the covariance matrix of the parameter estimates $F = C(\hat{\theta})^{-1}$, can be utilized to assess the accuracy of parameter estimation and to optimize the design of new experiments [4]. The Hessian H is approximated

$$\frac{\partial^2 \mathrm{LL}(y|\theta)}{\partial\theta_j \partial\theta_k} \approx \sum_i \frac{\partial \mathrm{LL}_i}{\partial\theta_j} \frac{\partial \mathrm{LL}_i}{\partial\theta_k} \tag{9}$$

using the sensitivity equations [5] of the second derivatives [8].

Because the scoring scheme evaluates the accuracy of each parameter component independently, the diagonal elements of the parameter covariance matrix $C(\hat{\theta}|y)$ was evaluated to relate the Fisher information matrix to the

expected score, i.e. the sum of the squared diagonal elements

$$Score_\theta(C) = \sum_i C_{ii}^2 \tag{10}$$

was used to assess the Fisher information matrix and the corresponding covariance matrix $C(\hat{\theta}|y)$.

Eqs. (8, 9) can be used to calculated the Fisher information $F^{(new)}$ for existing data y, as well as for a new data set y^{new}. The comprehensive information for the existing and new data set is given by the sum

$$F(\theta|\{y, y^{new}\}) = F(\theta|y) + E(F(\theta|y^{new})) \ . \tag{11}$$

Here, it has been assumed that calculation of the expectation E can be exchanged with calculation of the Fisher information, i.e. $E(F(\theta|y^{new})) = F(\theta|E(y^{new}))) = F(\theta|x^{new})$ which is a good assumption in an asymptotic setting.

The covariance matrix of the parameter estimates

$$C(\theta|\{y, y^{new}\}) = F(\theta|\{y, y^{new}\})^{-1} \tag{12}$$

after performing a new experiment can be translated into a covariance matrix for a model prediction, e.g. for the demanded extrapolation via

$$C_x(x(t_i), x(t_l)) = \sum_{jk} \frac{\partial x_i}{\partial \theta_j} \left(C(\theta|\{y, y^{new}\})\right)_{jk} \frac{\partial x_l}{\partial \theta_k} \tag{13}$$

The expected score for the extrapolation is then given by the sum of the squared diagonal elements

$$Score_x = \sum_{t_i \in \{0, 0.5, \ldots, 20\}} C_x(x(t_i), x(t_i))^2 \ . \tag{14}$$

For model $\mathcal{M}1$ an almost asymptotic setting, characterized by small confidence intervals and almost quadratic shape of the likelihood could be obtained by the performed experiments. Although models $\mathcal{M}2$ and $\mathcal{M}3$ could be brought in an almost asymptotic setting after purchasing experiments, they were not in an asymptotic setting during the experimental planning stage which limited the applicability of the Fisher Information based approach.

1.8. Latin hypercube sampling

The existence of local minima has been checked by latin hypercube sampling of the initial guesses of the parameters used to perform optimization [7]. This means, that for N_{lhs} samples, the parameter domain of each parameter component is subdivided into N_{lhs} equally-sized intervals. The samples are drawn so that each interval for each component is chosen exactly once. With the intervals, the parameter components are drawn uniformly distributed.

To decide whether a local minimum sufficiently agrees with the data, we used a threshold given by the log-likelihood of the best fit plus the 95%-quantile of the chi-square distribution with degrees of freedom equal to the number of parameters. This corresponds to a threshold utilized to calculate common confidence intervals for all parameters.

Experimental design considerations to discriminate between local minima have been performed by evaluating the model predictions for the local minima and correlate the predicted $x(t)$ with the locally optimal parameter values.

1.9. Parameter post-processing

During our analyses we gained confidence in the fact that all Hill-coefficients are natural numbers. This insight was obtained by the purchased parameter values but also from the confidence intervals of the Hill-coefficients which were estimated from the data. In fact, all Hill-coefficients for models $\mathcal{M}1$ and $\mathcal{M}2$ are in agreement with $\theta_{Hill} \in \{1, 2, 3, 4\}$.

After spending the whole budget, this prior knowledge was incorporated by rounding all hill-coefficients whose confidence intervals are very close to natural numbers. However, to be sure that this does not introduce error we applied this rounding step conservatively, i.e. only in cases where the confidence interval uniquely indicate the natural number. Another reason for being conservative, is that the gain by rounding into the correct direction in terms of the final score is minor in comparison to a potential rounding in the wrong direction. The time-course predictions were performed before this parameter value post-processing to be able to make the prediction for the maximum likelihood estimates.

Because deviations from the true parameters are scored quadratically, the expectation of the posterior density minimize the expected score. For parameters with asymmetric profile likelihood shape, the expectation does not coincide with the maximum likelihood estimate. In our analyses, predominantly $\mathcal{M}3$ showed such distortions. However, a one-dimensional profile likelihood can only be interpreted in this manner, if there is no impact from other parameters. To account for such correlations between the estimates of the hill-coefficients and the corresponding K_D parameters, we calculated the two-dimensional profile likelihood and the respective two dimensional posterior densities. The expected performance of the Maximum Likelihood estimate $\hat{\theta} = \arg\max_\theta \rho(\theta|y)$ in comparison to the expectation of the posterior distribution

$$\theta_E = \int_\theta \rho(\theta'|y)\theta' \, d\theta' \tag{15}$$

was evaluated by calculating the Euclidean norm

$$E(Score_\theta) = \int_{\theta_{h,K_D}} \rho(\theta'_{h,K_D}|y)||\theta'_{h,K_D} - \hat{\theta}_{h,K_D}||_2 \, d\theta'_{h,K_D} \quad (16)$$

for each pair of hill-coefficient and K_D parameter $\theta_{h,K_D} := (\theta_h, \theta_{K_D})$ with respect to the two dimensional posterior densities $\rho(\theta_{h,K_D}|y)$.

1.10. Experimental design

The optimization of the experimental design has been performed on the basis of the arguments provided in this section. The abbreviations are later used to indicate the reasons for iteratively purchase data for the three models.

(WT) Wild-type measurements provide a quite large amount of data for less credits.

(WT-P) As an initial step, purchasing data for all proteins is advantageous to have measurements for all dynamic variables. This reduces non-identifiability issues and provides a minimal amount of information about the parameters allowing more detailed experimental design considerations.

(hd-MA) If there are fast processes, high-density time resolution is favorable in comparison to low-density measurements, which are provided by the start-up microarray.

(P>mRNA) The measurement of two proteins provides 80 data points for 400 credits, a high-density microarray experiments charges 1000 credits and yields $20 \cdot n_{proteins}$ data points, i.e. 120, 140, and 180 for the three models. The ratio of the number of data points per credits is 0.2 for proteins and $\{0.12, 0.14, 0.18\}$ for RNA/microarray data. If the proteins measurements have smaller relative error, then protein data is even more informative.

(MAlarge) The larger the models, the more informative are microarray experiments in comparison to the information provided by the measurement of two proteins. Moreover, there is no decision required about which compounds should be measured. This makes the design more robust.

(E-para) In optimal experimental design theory for linear systems, the optimal design is known to be an extreme strategy. Although we analyze nonlinear models, the extreme strategy of buying as much parameters as possible for the given budget could be optimal.

(E-data) Similarly, the extreme strategy of buying as many data points as possible could be optimal.

(Budget) Sometimes, experiments are advantageous because the remaining credits allow a more flexible planning or the budget can be spent more comprehensively.

(siRNA) siRNA experiments are the cheapest perturbations and provide often qualitatively the same information as a knockout.

(OptPerPL) Perturbation experiments have been considered as maximally informative, if the score R in Eq. (7) is optimal.

(OptPerF) In the asymptotic setting, the optimally informative perturbation experiments have been chosen on the basis of the improvement of the Fisher-Information of the parameters as well as for the time course extrapolations.

(MC) A Monte-Carlo evaluation of a perturbation confirmed the guess.

(ID) There was a practical non-identifiabilty issue, i.e. a parameter confidence interval was (semi-)infinite over whole domain on a logarithmic scale.

(BI) The model shows qualitatively different dynamics, e.g. bistability, and a perturbation is able to switch the model's behavior.

(Extra) The experiment or the parameter values are important for improving the accuracy of the demanded model extrapolation.

(1st) A protein was not yet measured at all.

(Ixyz) Informative for parameters "x, y, z".

(Pred) Model predictions indicate an informative dynamic behavior.

(Module) The parameters to be bought is in a sub-module of bad estimates and therefore there is hope to improve identifiability of the whole module.

(LocMin) If several local minima have been detected with similar agreement to the data, designs have been chosen which optimally discriminate between the local minima.

1.11. Implementation of the numerical methods

The ODE system was solved by the CVODES algorithm [1]. For numerical optimization the trust-region method [3] implemented in LSQNONLIN from MATLAB was applied to perform Maximum Likelihood estimation. For efficiency, the local gradient and curvature information was considered. Gradient-based optimization relies on the accuracy of first order sensitivities $\frac{\partial x}{\partial \theta}$. For ODE systems, finite difference are not appropriate to approximate sensitivities [2, 6]. Therefore the sensitivity equations [5] are solved simultaneously by the ODE solver CVODES.

We discovered numerical issues in solving the ODEs and sensitivity equations for $\mathcal{M}3$ for initial conditions equals to zero and therefore we set the initial values of the initially absent compounds to $1e-16$.

Step	Action	Arguments	Remaining credits
1	WT protein data of p1-p6	**(WT-P)**,(WT),(E-data)	8800
2	WT high-density MA	(WT),(hd-MA),(E-data)	7800
3	$v1\_h$, $v1\_K_D$	**(ID)**,(E-para)	6200
4	$v3\_h$, $v3\_K_D$	**(ID)**,(E-para)	4600
5	siRNA pp5, measurement of p2&p4	**(OptPerPL)**,(siRNA),(Budget)	3850
6	$v2\_h$, $v2\_K_D$	**(ID)**	2250
7	rbs of p4	(OptPerPL), (BI), (Extra)	1400
8	siRNA pp5 high-density MA	**(OptPerPL)**, (Budget)	50

TABLE I: Summary of the decision to spend the budget for model $\mathcal{M}1$. The arguments are provided in the order of their priority. If an argument dominated, it is display in bold-face.

Step	Action	Arguments	Remaining credits
1	WT protein data of p1-p3,p5-p7	**(WT-P)**,(WT),(E-data)	8800
2	$v3\_h$, $v3\_K_D$	**(ID)**,(Extra)	7200
3	siRNA pp6, measurement of p4&p7	(1st),(Iv7),(Pred),(siRNA)	6450
4	$v4\_h$, $v4\_K_D$	**(ID)**,(MC)	4850
5	$v8\_h$, $v8\_K_D$	**(ID)**	3250
6	siRNA pp1, measurement of p3&p4	**(OptPerPL)**,(siRNA)	2500
7	siRNA pp2, measurement of p4&p6	**(OptPerF)**,(siRNA)	1750
8	$v10\_h$, $v10\_K_D$	**(Iv10)**	150

TABLE II: Summary of the decision to spend the budget for model $\mathcal{M}2$. The arguments are provided in the order of their priority. If an argument dominated, it is display in bold-face.

2. RESULTS

The first step in our analyses of the three models was purchasing protein data for the wild-type setting for all proteins. Thereby, we followed the arguments (WT), (WT-P), and (E-data), as provided in Sec. 1.10.

After purchasing data for p1-p6 for the WT setting of model $\mathcal{M}1$, we discovered that the error model provided by the organizers did not fit to the data. On the one hand, the noise realization contained too many zeros than it would be expected. On the other hand, the relative noise level σ_{rel} of the protein data was larger than 10%. In contrast, with a common relative noise level of 20% for mRNA and protein measurement the χ^2, i.e. the sum of squared standardized residuals was in perfect agreement with the expected value. Therefore, we assumed a relative noise level of 20% also for the protein data in the following. For the start-up data "v2", we assumed that there is no absolute error contributing to the noise and a precision of the data due to rounding of 10^{-3}. However, because it turned out at the end of the challenge, that the "v2" data does not agree with an absent absolute error, we decided to use this data set conservatively by assuming an absolute error of 10^{-3}.

Models $\mathcal{M}2$ and $\mathcal{M}3$ had an uneven number of proteins, i.e. 7 and 9. Because only sets of two proteins could be bought, we initially bought all proteins except a single one. For $\mathcal{M}2$, we decided to omit measurement of p4 because p4 had only few interactions with other proteins and according to the mRNA data it seemed to be expressed at a low level. Moreover, it is neither self-enhancing nor self-inhibiting. For $\mathcal{M}3$ the same arguments guided us to skip an initial measurement of p1.

After the third decision for $\mathcal{M}2$, the version three ("v3") start-up data was provided by the organizers. However, for $\mathcal{M}1$ the "v3"start-up data was exactly the same noise realization as the purchased subset of the WT high-density microarray experiment, i.e. we didn't obtain additional information for $\mathcal{M}1$ by the "v3" start-up data.

Tab. I provides a summary of the decisions to be made during experimental planning as well as the arguments behind our choices. Tab. II summarizes the decisions for $\mathcal{M}2$, and Tab. III for model $\mathcal{M}3$. In most cases, the next experiment was no unambigously given. Then we utilized further aspects or arguments to decide the next step.

3. SUMMARY

The most important approaches we applied to solve the challenge were local optimization methods based on sensitivity equation based gradients in combination with latin hypercube sample of the initial parameter guess, and profile likelihood based identifiability analyses. Due to the limiting time to solve the challenge, Monte-Carlo could only be applied to check the most important decisions. Because we found a bias for the estimation of hill-coefficients, we had to carefully choose the upper boundaries of the parameter domain for the hill-coefficients.

Our experimental design strategy was guided by the following major aspects:

Step	Action	Arguments	Remaining credits
1	WT protein data of p2-p9	**(WT-P)**,(WT),(E-data)	8400
2	pp9 KO high-density MA	(BI),(hd-MA),(MAlarge)	6600
3	$v12\_h$, $v12\_K_D$	**(Extra)**,(LocMin),(ID)	5000
4	$v8\_h$, $v8\_K_D$	**(LocMin)**,(ID),(Module)	3400
5	$v9\_h$, $v9\_K_D$	**(ID)**,(LocMin),(Extra)	1800
6	$v6\_h$, $v6\_K_D$	**(ID),(Iv5)**	200

TABLE III: Summary of the decision to spend the budget for model $\mathcal{M}3$. The arguments are provided in the order of their priority. If an argument dominated, it is display in bold-face.

1. Obtaining a sufficient amount of cheap data, i.e. WT measurements, to have some knowledge about the underlying system.
2. Several similarly likely local minima have to be discriminated by new experiments.
3. Improving the knowledge of practically non-identifiable parameters.
4. Performing the experiments which reduces the bulk of the variance in the demanded extrapolation settings.

For the smaller models, we always chose protein measurements for the perturbation experiments because of the argument (P>mRNA) provided above. Only for the large model $\mathcal{M}3$, microarray measurements after a perturbation seemed reasonable from our point of view, because only for a large number of species the benefits dominate the disadvantage of having a worse data quality compared to protein data.

In general, it was difficult to assess whether it is better to buy exact parameter values or whether it is superior to make informative experiments. At the end we slightly preferred buying exact values because perturbations mostly couldn't resolve correlations of parameter estimates which are observed if several parameters have related impact on model predictions. Such correlated estimates were found for many triplets composed of the parameters of a single rate, i.e. the hill-coefficient, the K_D value, and production-strength ("pro_strength") of a transcription factor binding site.

Acknowledgments

The authors acknowledge financial support by the grants [0315766-VirtualLiver] and [0313921-FRISYS] provided by the German Federal Ministry of Education and Research (BMBF), as well as the FP7 EU project CancerSys [HEALTH-F4-2008-223188] .

[1] A.C. Hindmarsh, P.N. Brown, K. G. and Woodward, C. (2005). SUNDIALS: Suite of nonlinear and differential/algebraic equation solvers. *ACM Transactions on Mathematical Software*, 31:363–396.

[2] A.R. Conn, K. S. and Vicente, L. (2009). Introduction to derivative-free optimization. *Society for Industrial Mathematics*.

[3] Coleman, T. and Li, Y. (1996). An interior, trust region approach for nonlinear minimization subject to bounds. *SIAM Journal on Optimization*, 6:418–445.

[4] Kreutz, C. and Timmer, J. (2009). Systems biology: experimental design. *FEBS J*, 276(4):923–942.

[5] Leis, J. and Kramer, M. (1988). The simultaneous solution and sensitivity analysis of systems described by ordinary differential equations. *ACM T Math Software*, 14(1):45–60.

[6] Lohmann, T., Bock, H., and Schlöder, J. (1992). Numerical methods for parameter estimation and optimal experimental design in chemical reaction systems. *Ind. Eng. Chem. Res.*, 31:54–57.

[7] McKay, M. D., Beckman, R. J., and Conover, W. J. (1979). A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics*, 21(2):pp. 239–245.

[8] Press, W., Flannery, B., Saul, S., and Vetterling, W. (1992). *Numerical recipes*. Cambridge University Press, Cambridge.

[9] 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*, 25:1923–1929.