MINIREVIEW

Systems biology: experimental design

Clemens Kreutz and Jens Timmer

Physics Department, University of Freiburg, Germany

Keywords
confounding; experimental design; mathematical modeling; model discrimination; Monte Carlo method; parameter estimation; sampling; systems biology

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

(Received 8 April 2008, revised 13 August 2008, accepted 11 September 2008)

doi:10.1111/j.1742-4658.2008.06843.x

Experimental design has a long tradition in statistics, engineering and life sciences, dating back to the beginning of the last century when optimal designs for industrial and agricultural trials were considered. In cell biology, the use of mathematical modeling approaches raises new demands on experimental planning. A maximum informative investigation of the dynamic behavior of cellular systems is achieved by an optimal combination of stimulations and observations over time. In this minireview, the existing approaches concerning this optimization for parameter estimation and model discrimination are summarized. Furthermore, the relevant classical aspects of experimental design, such as randomization, replication and confounding, are reviewed.

Introduction

The development of new experimental techniques allowing for quantitative measurements and the proceeding level of knowledge in cell biology allows the application of mathematical modeling approaches for testing and validation of hypotheses and for the prediction of new phenomena. This approach is the promising idea of systems biology.

Along with the rising relevance of mathematical modeling, the importance of experimental design issues increases. The term 'experimental design' or 'design of experiments' (DoE) refers to the process of planning the experiments in a way that allows for an efficient statistical inference. A proper experimental design enables a maximum informative analysis of the experimental data, whereas an improper design cannot be compensated by sophisticated analysis methods.

Learning by experimentation is an iterative process [1]. Prior knowledge about a system based on literature and/or preliminary tests is used for planning. Improvement of the knowledge based on first results is followed by the design and execution of new experiments, which are used to refine such knowledge (Fig. 1A). During the process of planning, this sequential character has to be kept in mind. It is more efficient to adapt designs to new insights than to plan a single, large and comprehensive experiment. Moreover, it is recommended to spend only a limited amount of the available resources (e.g. 25% [2]) in the first experimental iteration to ensure that enough resources are available for confirmation runs.

Experimental design considerations require that the hypotheses under investigation and the scope of the study are stated clearly. Moreover, the methods intended to be applied in the analysis have to be specified [3]. The dependency on the analysis is one reason

Abbreviation
AIC, Akaike Information Criterion.

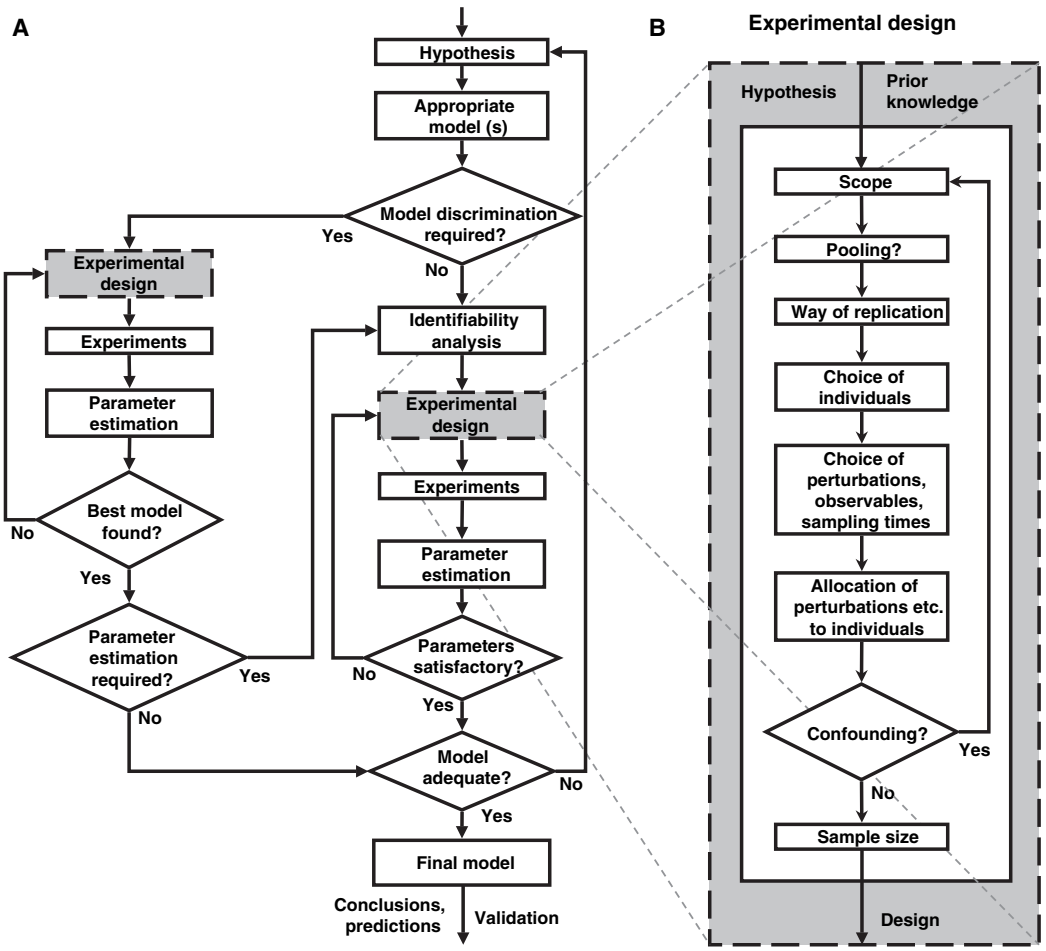

Fig. 1. (A) Overview of an usual model building process. Both loops, with and without model discrimination, require experimental planning (highlighted in gray). (B) The most important steps in experimental planning for systems biological applications.

for the wide range of experimental design methodologies in statistics.

In this minireview, we provide theoreticians with a starting point into the experimental design issues that are relevant for systems biological approaches. For the experimentalists, the minireview should give a deeper insight into the requirements of the experimental data that should be used for mathematical modeling. The aspects of experimental planning discussed here are shown in Fig. 1B. One of the main aspects when studying the dynamics of biological systems is the appropriate choice of the sampling times, the pattern of stimulation and the observables. Moreover, an overview about the design aspects that determine the scope of the study is provided. Furthermore, the benefit of pooling, randomization and replication is discussed.

Experimental design issues for the improvement of specific experimental techniques are not discussed. Microarray specific issues are discussed elsewhere [4–9]. Experimental design topics in proteomics are discussed by Eriksson and Feny [10]. Improvement of quantitative 'real-time polymerase chain reaction' is given elsewhere [11–13]. Design approaches for qualitative models, i.e. Boolean network models, semi-quantitative models or Bayesian networks, are also given elsewhere [14–18].

A review from a more theoretical point of view is given by Atkinson *et al*. [19]. A review with focus on optimality criteria and classical designs is also given by Atkinson *et al*. [20]. An early review containing a detailed bibliography until 1969 is provided by Herzberg and Cox [21]. The literature on *Bayesian experimental design* has been reviewed previously [22]. The contribution of R. A. Fisher, one of the pioneers in the field of design of experiments, has also been reviewed previously [23]. A review of the methods of experimental design with respect to applications in microbiology can be found elsewhere [24].

Apart from bringing quantitative modeling to biology, systems biology bridges the cultural gap between experimental an theoretical scientists. An efficient experimental planning requires that, on the one hand, theoreticians are able to appraise experimental feasibility and efforts and that, on the other hand, experimenters know which kind of experimental information is required or helpful to establish a mathematical model.

Table 1 constitutes our attempt to condense general theoretical aspects in planning experiments for the establishment of a dynamic mathematical model into some rules of thumb that can be applied without advanced mathematics. However, because the needs on experimental data depend on the questions under investigation, the statements cannot claim validity in all circumstances. Nevertheless, the list may serve as a helpful checklist for a wide range of issues.

Table 1. Some aspects in the design of experiments for the purpose of mathematical modeling in systems biology.

In comparison to classical biochemical studies, establishment of mechanistic mathematical models requires a relative large amount of data
Measurements obtained by experimental repetitions have to be comparable on a *quantitative* not only on a qualitative level
A measure of confidence is required for each data point
The number of measured conditions should clearly exceed the number of all unknown model parameters
Validation of *dynamic* models requires measurements of the *time* dependency after external perturbations
Perturbations of a single player (e.g. by knockout, over-expression and similar techniques) provide valuable information for the establishment of a mechanistic model
Single cell measurements can be crucial. This requirement depends on the impact of the occurring cell-to-cell variations to the considered question, and on the scope and generality of the desired conclusions
The biochemical mechanisms between the observables should be reasonably known
The predictive power of mathematical models increases with the level of available knowledge. It could therefore be preferable to concentrate experimental efforts on well understood subsystems
If the modeled proteins could not be observed directly, measurements of other proteins that interact with the players of interest, can be informative. The amount of information from such additional observables depends on the required enlargement of the model
The velocity of the underlying dynamics indicates meaningful sampling intervals Δt. The measurements should seem relatively smooth. If the considered hypothesis are characterized by a different dynamics, this difference determines proper sampling times
Steady-state concentrations provide useful information
The number of molecules per cell or the total concentration is a very useful information. The order of magnitude of the number of molecules (i.e. tens or thousands) per cellular compartment has to be known
Thresholds for a qualitative change of the system behavior, i.e. the switching conditions, are insightful information
Calibration measurements with known protein concentrations are advantageous because the number of scaling parameters is reduced
The specificity of the experimental technique is crucial for quantitative interpretation of the measurements
For the applied measurement techniques, the relationship between the output (e.g. intensities) and the underlying truth (e.g. concentrations) has to be known. Usually, a linear dependency is preferable
Known sources of noise should be controlled

General aspects

Sampling

Any biological experiment is conducted to obtain knowledge about a population of interest, e.g., about cells from a certain tissue. 'Sampling' refers to the process of the selection of experimental units, e.g. the cell type, to study the question under consideration. The aim of an appropriate sampling is to avoid systematic errors and to minimize the variability in the measurements due to inhomogeneities of the experimental units. Adequate sampling is a prerequisite for drawing valid conclusions. Moreover, the finally selected subpopulation of studied experimental units and the biochemical environment defines the scope of the results. If, as an example, only data from a certain phenotype or of a specific cell culture are examined then the generalizability of any results for other populations is initially unknown.

In cell biology, there is usually a huge number of potential features or 'covariates' of the experimental units with an impact on the observations. In principle, each genotype and each environmentally induced varying feature of the cells constitutes a potential source of variation. Further undesired variation can be caused by inhomogeneities of the cells due to cell density, cell viability or the mixture of measured cell types. Moreover, systematic errors can be caused by changes in the physical experimental conditions such as the pH value or the temperature.

The initial issue is to appraise which covariates could be relevant and should therefore be controlled. These interfering covariates can be included in the model to adjust for their influences. However, this yields often an undesired enlargement of the model [see example (3) in Fig. 2].

An alternative to extending the model is controlling the interfering influences by an appropriate

Example: Interfering covariates

The dynamics of protein activation is studied by measurements at different times after stimulation. It is assumed that there is a biological variability between cells obtained from different individuals and that circadian rhythms have an impact. There are three main strategies to account for the impact of these two 'covariates' or 'factors' 'individual' and 'circadian state'.

(1) Limiting the scope
Here, the impact of both factors is controlled by fixing their levels. Only one individual would be measured and all probes would be obtained at the same time of day. Then, all data points are affected by the same amount by the two covariates. However, this strategy would limit the scope. The generalization to the population of all individuals and to other points in the circadian rhythm in unknown.

(2a) Blocking/Stratification
Here, the measurements are grouped according to both factors (i.e. only measurements from one individual at a certain circadian state are considered as replicates). Each group is analyzed independently and the results are averaged. This yields mean estimates of the model parameters. However, this strategy could lead to a large experimental effort. If n_i denotes the number of regarded individuals, n_c the number of circadian states and n_t the number of time points, at least full factorial designs, $n_i \times n_c \times n_t$ measurements would be required.

(2b) Latin square blocking
Here, the same number of factor levels $n_i = n_c = n_t$ are investigated. The combinations of the three effects are chosen so that each time point is measured once for each individual and once for each circadian state. Then only $n_i \times n_c$ measurements are required and the impact of the two covariates is equal, in the average, for each point in time. The limitation is that, the biological variability as well as the impact of the circadian clock cannot be estimated.

(3) Expanding the model
Here, additional model parameters are introduced to adjust for the individual as well as the circadian impact. However, it has to be known in which way the model has to be expanded (e.g. if both factors can be accounted by different offset parameters, by different kinetic constants, etc). In addition, the increase in the number of parameters would require more data points to obtain the same precision as in a non-expanded model.

Fig. 2. An example of how the impact of two sources of variation can be accounted for in time course measurements.

sampling [25]. This is achieved by choosing a fixed 'level' of the influencing covariates or 'factors'. However, this restricts the scope of the study to the selected level.

Another possibility is to ensure that each experimental condition of interest is affected by the same amount on the interfering covariates. This can be accomplished by grouping or 'stratify' the individuals according to the levels of a factor. The obtained groups are called 'blocks' or 'strata'. Such a 'blocking strategy' is frequently applied, when the runs cannot be performed at once or under the same conditions. In a 'complete block design' [26], any treatment is allocated to each block. The experiments and analyses are executed for each block independently [Fig. 2, (2a)]. Merging the obtained results for the blocks yields more precise estimates because the variability due to the interfering factors is eliminated. 'Paired tests' [27] are special cases of such complete block designs.

In 'full factorial designs', all possible combinations of the factor levels are examined. Because the number of combinations rapidly increases with the number of regarded covariates, this strategy results in a large experimental effort. One possibility to reduce the number of necessary measurements is a subtle combination of the factorial influences. 'Latin square sampling' represents such a strategy for two blocking covariates. A prerequisite is that the number of the considered factor levels are equal to the number of regarded experimental conditions. Furthermore, latin square sampling assumes that there is no interaction between the two blocking covariates, i.e. the influence of the factors to the measurements are independent from each other; e.g. there are no cooperative effects.

A latin square design for elimination of two interfering factors with three levels is illustrated in Fig. 3 (2a). Here, three different conditions, e.g. times after a stimulation t_1,t_2,t_3, are measured for three individuals A, B, C at three different states c_1, c_2 and c_3 within the circadian rhythm. The obtained results are unbiased with respect to biological variability due to different individuals and due to the circadian effects.

Frequently, the covariates with a relevant impact on the measurements are unknown or cannot be controlled experimentally. These covariates are called 'confounding variables' or simply 'confounders' [28]. In the presence of confounders, it is likely that

		Individual		
		A	**B**	**C**
Circadian state	c_1	t_1	t_2	t_3
	c_2	t_2	t_3	t_1
	c_3	t_3	t_1	t_2

Fig. 3. Latin square experimental design for three individuals A, B, C measured at three states of the circadian rhythms c_1, c_2, c_3. Because each time t_1, t_2, t_3 is influenced by the same amount by both interfering factors, the average estimates are unbiased.

ambiguous or even wrong conclusions are drawn. This occurs if some confounders are over-represented within a certain experimental condition of interest. In an extreme case, for all samples within a group of replicates, one level of a confounding variable would be realized. Over-representation of confounders is very likely for small number of repetitions. In Fig. 4, the probabilities are displayed for the occurrence of a confounding variable for which the same level is realized for any repetition in one out of two groups. It is shown that there is a high risk of over-representation if the number of repetitions is too small.

An adequate amount of replication is a main strategy to avoid unintended confounding. This ensures that significant correlations between the measurements and the chosen experimental conditions are due to a causal relationship. However, especially in studies based on high-throughput screening methods, three or even less repetitions are very common. Consequently,

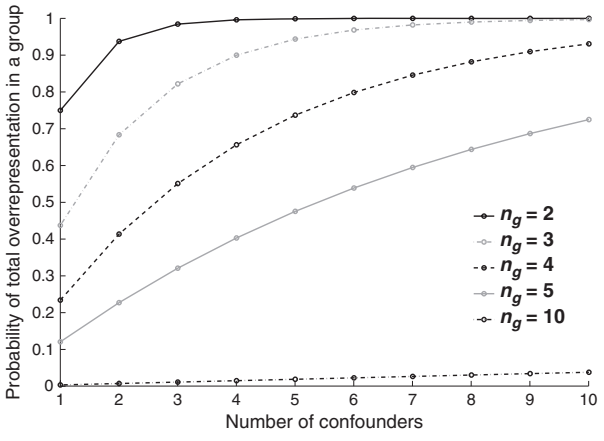

Fig. 4. The probability of a totally over-represented confounder, i.e. the chance of the occurrence of a confounding variable for which the same level is realized all n_g repetitions in a group. In this example, confounding variables are assumed to have two levels with equal probabilities.

without the use of prior knowledge, the obtained results are only appropriate as a preliminary test for the detection of interesting candidates.

In systems biology, measurements of the dynamic behavior after a stimulation is very common. Here, confounding with systematic trends in time can occur, e.g. caused by the cell cycle or by circadian processes. It has always be ensured that there is no systematic time drift. The issue of designing experiments that are robust against time trends is discussed elsewhere [29,30].

Another basic strategy to avoid systematic errors is 'randomization'. Randomization means both, a random allocation of the experimental material and a random order in which the individual runs of the experiment are performed. Randomization minimizes that the risk of unintended confounding because any systematic relationship of the treatments to the individuals is avoided. Any nonrandom assignment between experimental conditions and experimental units can introduce systematic errors, leading to distorted, i.e. 'biased', results [31]. If, as an example, the controls are always measured after the probes, a bias can be introduced if the cells are not perfectly in homeostasis. For immunoblotting, it has been shown that a chronological gel loading causes systematic errors [32,33]. A randomized, nonchronological gel loading is recommended to obtain uncorrelated measurement errors.

'Pooling' of samples constitutes a possibility to obtain measurements that are less affected by biological variability between experimental units without an increase in the number of experiments [34]. Pooling is only reasonable when the interest is not on single individuals or cells but on common patterns across a population. If the interest is in the single experimental unit, e.g. if a mathematical model for a intracellular biochemical network such as a signaling pathway has to be developed, pooled measurements obtained from a cell population are only meaningful, if the dynamics is sufficiently homogeneous across the population. Otherwise, e.g. if the cells do not respond to a stimulation simultaneously, only the average response can be observed. Then the scope of the mathematical model is limited to the population average of the response and does not cover the single cell behavior.

Pooling can cause new, unwanted biological effects, e.g. stress responses or pro-apoptotic signals. Therefore, it has to be ensured that these induced effects do not have a limiting impact on the explanatory power of the results. However, if pooling is meaningful, it can clearly decrease the biological variability and the

risk of unwanted confounding, especially for a small number of repetitions.

Replication

One purpose of 'replication' is the minimization of the risk of unintended confounding. Furthermore, repeated measurements allow for the estimation of the variability of the data. This enables the computation of error bars as a measure of confidence for each data point.

An additional advantage of replication is the improvement in the precision and power of the analyses. There is no generally valid rule for the amount of improvement if the sample size is enlarged. However, the estimation of any parameters is typically carried out by averaging over the replicate measurements. Because of the 'central limit theorem' of statistics, a sum over identically distributed random variables is normally distributed if standard conditions are fulfilled. Therefore the 'confidence interval' or 'standard error' of an estimate obtained after averaging over n repetitions decreases proportional to $1/\sqrt{n}$. Figure 5 shows, as an example, that the standard error σ_{μ_i} of the sample mean μ in an experimental condition i is equal to $\sigma/\sqrt{n}$ where σ denotes the standard deviation of a single data point. In the example, the two sample means constitute two population parameters that are estimated from experimental data. Additional information obtained from repeated measurements increases the precision in the parameter estimates.

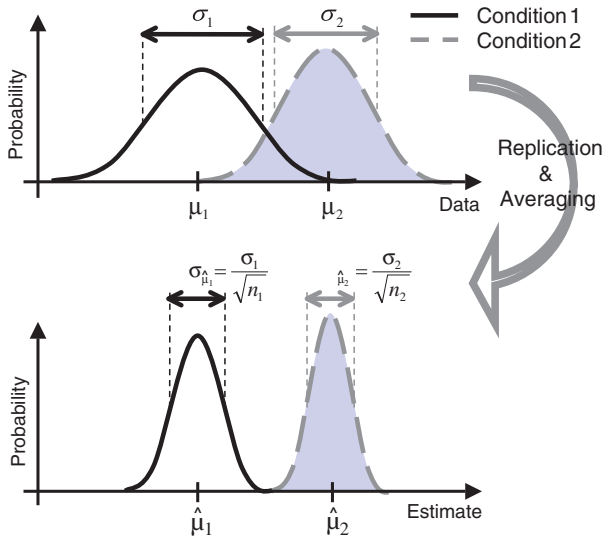

Fig. 5. The precision of experimental results can be improved by increasing the number of experimental repetitions. In this example, despite overlapping distributions of the measurements of two experimental conditions, the difference is unraveled after averaging of repeated observations. The spread of the distributions after averaging is quantified by the standard error $\sigma_{\hat{\mu}_i}$ of the estimated mean $\hat{\mu}_i$ of condition i, which is proportional to $1/\sqrt{n}$.

The $1/\sqrt{n}$ dependency of standard errors of estimated parameters could be regarded as an optimistic rule of thumb if experiments are planned efficiently [35]. By contrast, for statistical tests, the power of a design, i.e. the sensitivity to detect any effects, depends on the separation of distributions observed under the null and under the alternative hypothesis. There is a relationship between (a) the power of a statistical test; (b) the true underlying effect size, i.e. the distance of the two distributions; (c) the desired confidence, i.e. the significance level as the threshold for a rejection of the null hypothesis; (d) the amount of noise; and (e) the number of replications. Therefore, if (a)–(d) are given, the required sample size (e) can be calculated. Such a 'sample size calculation' [4,36,37] can be performed analytically or via simulations. Reviews about sample size calculations with focus on clinical studies are provided elsewhere [38,39].

If some experimental conditions play a special role in the analysis, e.g. as a common reference, these data points have a prominent impact on the results. In this case, it could be advantageous to measure the special condition more frequently to obtain a more precise estimate. Otherwise, if no experimental condition plays a special role and the noise level is equal, 'balanced' designs, i.e. designs with the same number of replicates in each group, have optimal power.

The manner in which the replicates are obtained is crucial for the scope of the results. Technical replication limits the scope of any results to the investigated biological unit because the obtained confidence intervals does not contain the biological variability. By contrast, biological replicates observed in different experimental runs lead to confidence intervals that reflect the inter-individual and inter-experimental variability. This leads to more general results and extends the scope of the study. If the interesting biological effects are small, the inter-individual variability can be eliminated by a blocking strategy. Appropriate replication and its pitfalls are discussed elsewhere [35,40,41].

The design problem

The discussion in the preceding section concerns qualitative aspects of experimental planning that are related to the scope and validity of the results. For planning at a quantitative level, i.e. for the proposal of optimally informative observables, perturbations or measurement times, the design problem has to be stated mathematically.

The mathematical models

In this minireview, it is assumed that the biological process is modeled by a system of 'ordinary differential equations'

$$\dot{x}(t) = f(x(t), u(t), p_x) \quad (1)$$

where p_x is a vector containing the dynamic parameters of the model and u represents the externally controlled inputs to the system as stimulation by ligands. Typically, the state variables x correspond to concentrations. Initial concentrations $x(0)$ have usually also to be considered as system parameters. The level of detail, i.e. the number of equations and parameters, depends on the hypotheses under investigation. The system dynamics, i.e. the function f, is often derived from the underlying biochemical mechanisms. These models are called 'mechanistic models'.

The discussed principles and mathematical formalism of experimental design also hold for 'partial differential equations, delay differential equations and differential algebraic equations'. Indeed, all the discussed principles hold for any deterministic relationship between the state variables and also for steady states. By contrast, models containing stochastic relations, e.g. as described via 'stochastic differential equations', would require a more general mathematical formalism at some points.

The definition of the dynamics $x(t)$ in Eqn (1) is the biologically relevant part of a mathematical model. Statistical inference requires an additional component

$$y(t_i) = g(x(t_i), p_y) + \varepsilon(t_i), \varepsilon(t_i) \sim N(0, \sigma^2) \quad (2)$$

linking the dynamical variables $x(t_i)$ to the measurements $y(t_i)$. Here, independently and identically distributed additive Gaussian noise is assumed, although the following discussion is not restricted to this type of observational noise. The vector p_y contains all parameters of the observational functions g, e.g. scaling parameters for relative data, and parameters for further 'effects' corresponding to experimental parameters, which account for interfering covariates. For simplicity, we introduce $p \in \mathbb{P}$ as the parameter vector containing all n_p model parameters p_x and p_y.

An experimental design $\mathcal{D}$ specifies the choice of the external perturbations u, the choice of the observables g and the number and time points t_i of measurements. The way of stimulation as well as the times of measurement can usually be controlled by the experimenter. Therefore, they are called 'independent variables'. By contrast, the measured variables y are called 'dependent variables' because the realizations depend on the design and on the system behavior. Note, that in the models, Eqns (1,2) only the dependent variables y are affected by noise. It is assumed that the independent variables, e.g. the sampling times, can be controlled exactly.

External perturbations

In systems biology, an important independent variable is the treatment. Such a stimulation, e.g. by hormones or drugs, can be time varying and is in this case modeled as continuous 'input function' $u(t)$. Up- or down-regulation of genes, i.e. by 'constitutive overexpression' or by 'knockouts', can also be regarded as external perturbations of the studied system.

A design can be optimized with respect to the chosen perturbations $u \subset \mathbb{U}$. This includes the choice of the applied treatments or treatment combinations as well as stimulation strength and the temporal pattern, e.g. permanent or pulsatile stimulation. $\mathbb{U}$ denotes the set of all experimentally applicable perturbations. For numerical optimization, the input functions has to be parameterized. A common approach is the 'control vector parameterization' [42,43] or using stepwise constant input functions.

Previously [1,44,45], a stepwise constant input function was optimized for a given number of switching times. More complex input functions have also been optimized [46–48]. A benchmark problem [49] has also been provided for model identification of a biochemical network in so called 'fed batch experiments'. Here, the externally controlled input function is the feed rate and feed concentration in the bioreactor. Inputs have been designed [45,50] for discrimination of models for growth of *Escherichia coli* and *Candida utilis*. An experimental design for the same growth models for the purpose of both, parameter estimation and model selection has also been proposed [51].

Measurement times

The choice of the sampling times, i.e. the times of measurement $t \subset \mathbb{T}$, is crucial if the dynamics of a system is studied by mechanistic models. On the one hand, the sampling interval Δt_i should be small enough to capture the fastest processes. On the other hand, the duration $t_{\max} - t_{\min}$ of observation should be appropriate to capture the long-term behavior of the studied system. Because of limitations in experimental resources, this trade-off has to be solved reasonable by experimental planning. This requires, however, some knowledge about the time scale of the studied dynamic processes.

It has been shown previously [52] how the sampling times could be chosen optimally to maximize the precision in parameter estimation. A model of enzymatic activation is used as an illustration. An example from process engineering with two state variables was also previously used [1] for optimization of the sampling times for a given number of measurements.

Observables

The output of an experiment y is represented in the model by observational functions g and the noise ε. The experimenter has the freedom to choose which measurement technique will be applied and which system players, e.g. proteins, will be measured. Thereby, it is possible to select the most informative observables $g \subset \mathbb{G}$ from the set of all available observational functions $\mathbb{G}$, which are determined by experimental feasibility.

In practice, such experimental design considerations are very helpful, if, for example, new antibodies have to be generated or experimental techniques have to be established in a laboratory. Another reason for the importance of the choice of the observables is that this step determines the expected amount of observational noise.

A sensitivity analysis was previously applied [53] to a model of the nuclear factor kappa B (NFκB) signal transduction pathway to determine proteins that are sensitive to changes in important model parameters. The measurement of these proteins provides the maximal amount of information for parameter estimation.

Experimental constraints

In cell biology, there are usually much more experimental restrictions than in more technically orientated disciplines such as engineering or physics. Often, only a small fraction of the dynamic variables can be measured. The feasible external perturbations are usually very limited, e.g. it is often impossible to define the stimulation in the frequency domain, which is a natural approach in engineering.

Experimental constraints are accounted by the definition of the 'design region' $\mathbb{D}$, i.e. the set of all practically applicable designs. During the optimization, $\mathbb{D}$ is considered as the domain, i.e. only designs $\mathcal{D} \in \mathbb{D}$ are allowed. If there are only separate experimental constraints for the domains $\mathbb{U}$,$\mathbb{G}$ and $\mathbb{T}$, then $\mathbb{D}$ corresponds to the set of all combinations

$$\mathbb{D} = \mathbb{U} \times \mathbb{G} \times \mathbb{T} \tag{3}$$

of possible perturbations, observations and measurement times. An example for commonly occurring constraints is a lower boundary for the sampling interval Δt or that only a limited number of measurements can be obtained from one experimental unit.

After the definition of a 'utility' (or 'loss') 'function' $V(\mathcal{D})$, the design can be optimized over the design region

$$\mathcal{D}^* = \arg\max_{\mathcal{D}\in\mathbb{D}} V(\mathcal{D}) \tag{4}$$

to identify the optimal design $\mathcal{D}^*$ as the solution of the design problem. The utility function, also called 'design criterion' V, reflects the purpose of the experiments. If, for example, parameters are estimated, the utility function could be a measure for the expected accuracy of the estimated parameters. If the discrimination between competing models for the description of a phenomenon is regarded, the design criterion measures the difference in the model predictions. The most commonly used utility functions are introduced below.

Prior knowledge

In general, besides the dependency on the design, the utility function depends on the true underlying parameters p and on the realization of the observational noise $V(\mathcal{D}) \to V(\mathcal{D},p,\varepsilon)$. Therefore, in the general case, the determination of an optimal design requires some prior knowledge about the parameters [54]. The accuracy of the predicted optimal designs is limited by the precision of the provided prior knowledge. Such knowledge, e.g. the order of magnitude or physiological meaningful ranges, could be obtained from preliminary experiments. The expected utility function

$$\bar{V}(\mathcal{D}) = \int_{\mathbb{P}} \int_{-\infty}^{\infty} \rho(\varepsilon)\rho(p)V(\mathcal{D},p,\varepsilon)\mathrm{d}\varepsilon\,\mathrm{d}p \tag{5}$$

is obtained by averaging over the parameter space $\mathbb{P}$ and over all possible realizations of the observational noise. By using a prior distribution $\rho(p)$, the parameter space is weighted according to its relevance. $\rho(\varepsilon)$ denotes the distribution of the observational noise.

In the case of an unknown model structure, i.e. for the purpose of model discrimination, an additional weighting with the prior probabilities $\pi(\mathcal{M})$ of different reasonable models $\mathcal{M}$ is required. Then Eqn (5) becomes

$$\bar{V}(\mathcal{D}) = \sum_{\mathcal{M}} \pi(\mathcal{M}) \int_{\mathbb{P}} \int_{-\infty}^{\infty} \rho(\varepsilon)\rho^{(\mathcal{M})}(p)V^{(\mathcal{M})}(\mathcal{D},p,\varepsilon)\mathrm{d}\varepsilon\,\mathrm{d}p \tag{6}$$

where $\rho^{(\mathcal{M})}(p)$ denotes the parameter prior for model $\mathcal{M}$.

After the analysis of new experimental data, the parameter prior as well as the model prior are updated to account for new insights. Bayes' formula yields to posterior probabilities

$$\pi'(\mathcal{M}) = \frac{\pi(\mathcal{M}) \int \rho(y|p^{(\mathcal{M})}) \rho^{(\mathcal{M})}(p) \mathrm{d}p}{\sum_m \pi(\mathcal{M}_m) \int \rho(y|p^{(\mathcal{M}_m)}) \rho^{(\mathcal{M}_m)}(p) \mathrm{d}p} \quad (7)$$

for the considered models and

$$\rho^{(\mathcal{M})'}(p) = \frac{\rho^{(\mathcal{M})}(p) \rho^{(\mathcal{M})}(y|p)}{\int \rho^{(\mathcal{M})}(p') \rho^{(\mathcal{M})}(y|p') \mathrm{d}p'} \quad (8)$$

for the model parameters. In turn, these refinements yield more precise experimental planning.

The iterative gain of knowledge about the studied system is displayed in Fig. 6. At the beginning, an initial prior knowledge is used for experimental planning. After execution and analysis of an experiment, posterior probabilities Eqns (7,8) are calculated, which serve as new prior knowledge for the design of the subsequent experiment.

Determination of optimal designs

After planning with respect to confounding and scope of the study, the model structure, the design region and the prior knowledge are defined mathematically, as described in the previous section. Then, the independent experimental variables can be chosen optimally. For this purpose, different utility functions are introduced in this section. Furthermore, techniques are introduced for the calculation of optimal designs.

The utility function or design criterion is used for numerical optimization, which yields optimal sampling time points, observational functions and external perturbations. The choice of the design criterion reflects the issues to be studied. Therefore, an important preliminary need for experimental design considerations is the exact formulation of the question under investigation [55]. Figure 7 shows a simple example where slight variations in the hypothesis lead to other optimal designs [56]. In systems biology, the hypotheses are usually answered by discrimination between different mathematical models [57] and/or the estimation of model parameters [58–60].

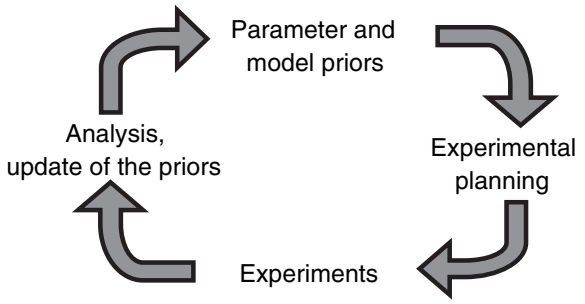

Fig. 6. Iterative cycle of the gain of knowledge about a system. For initial planning, a model and parameter prior has to be defined. This knowledge is updated and refined after any experimental result is obtained.

Example: Dependency on the hypothesis [56]
The effects θ_1, θ_2, θ_3 of three different treatments are investigated and the optimal design for six measurements has to be determined.

Issue 1: θ_1, θ_2, θ_3 are to be estimated.
The optimal design leading to best precision would be to measure each treatment twice.

Issue 2: The hypothesis $\theta_1 = \theta_2 = \theta_3$ is tested.
It would be optimal to measure two randomly chosen treatments three times.

Issue 3: The hypothesis $\theta_1 = \theta_2 = \theta_3 = 0$ is tested.
It would be optimal to measure one randomly selected treatment with six repetitions.

Fig. 7. A simple example showing how a slight variation in the question under investigation can change the optimal design. Additional details, e.g. of the underlying assumptions, are provided elsewhere [56].

Usually, the differential equations Eqn (1) cannot be solved analytically. In this case, an optimal design can only be determined by numerical techniques. By means of 'Monte Carlo' simulations, synthetic data are generated including their stochasticity [61,62]. By analyzing the simulated data in exactly the same way as intended for the analysis of the measurements, it is possible to evaluate and compare the possible outcomes (the utility functions obtained for different designs). Repeated simulations are then used to calculate the expected utility function. This expectation can be used for numerical optimization.

The disadvantage of Monte Carlo approaches is the high numerical effort. This drawback can be minimized by introducing reasonable approximations. The benefit of Monte Carlo simulations is their great flexibility. In principle, every source of uncertainty can be included by drawing from a corresponding prior distribution. Furthermore, nonlinear dependencies of the observations on the parameters or on the states does not constitute a limitation of the Monte Carlo methods.

In the next two sections, Monte Carlo procedures for optimization with respect to parameter estimation and model discrimination are described.

Experimental design for parameter estimation

An important step in the establishment of a mathematical model is the determination of the model

parameters. Besides initial protein concentrations and kinetic rate constants, parameters of the observational functions have to be estimated.

In the 'maximum likelihood' approach [43,63] the likelihood function, i.e. the probability $\rho(y|p)$ of the measurements y given a parameter set p, is maximized to obtain optimal model parameters $\hat{p}$. This probability is determined by the distribution of the observational noise. In the case of independently normally distributed noise Eqn (2), the log-likelihood function corresponds to the well known standardized residual sum of squares $\sum_i (y_i - g_i)^2/\sigma_i^2$.

'Fisher information' is defined as the expectation of the second derivative of the log-likelihood with respect to the change in the parameters [52,64,65]. If the observational noise is normally distributed, the 'Fisher information matrix'

$$F_{mn}(\mathcal{D}) = \sum_i \sum_j \frac{1}{\sigma_{ij}^2} \frac{\partial^2 g_j(t_i, \hat{p})}{\partial p_m \partial p_n} \tag{9}$$

contains second order derivatives of the model's observational functions g around estimated parameters $\hat{p}$ [66]. σ_{ij}^2 denotes the variance of the observational noise of observable g_j at time t_i. The summation extends the chosen design $\mathcal{D}$. The inverse of F is the covariance matrix of the estimated parameters. The standard errors of the estimated parameters are the diagonal elements of the matrix F^{-1}.

For optimization, a scalar utility function is required. There are several design criteria derived from the Fisher information matrix [67]. An alphabetical nomenclature for the different criteria was introduced by Kiefer [56].

Often, the determinant

$$V(\mathcal{D}) = \det(F(\mathcal{D})) = \prod_i \lambda_i(\mathcal{D}) \tag{10}$$

is maximized. λ_i denote the eigenvalues of F. The obtained optimal design is called 'D-optimal' [68]. Maximization of Eqn (10) corresponds to minimization of the 'generalized variance' of the estimated parameters, i.e. minimization of the volume of the confidence ellipsoid [69].

An 'A-optimal' design is obtained by maximizing the sum of eigenvalues

$$V(\mathcal{D}) = \sum_i \lambda_i(\mathcal{D}) \tag{11}$$

of the Fisher information matrix, i.e. minimizing the average variance of the estimated parameters.

Similarly, the 'E-optimal' design is obtained by maximization of the smallest eigenvalue

$$V(\mathcal{D}) = \lambda_{\min}(\mathcal{D}) \tag{12}$$

This is equivalent to minimization of the largest confidence interval of the estimated parameters.

A graphical illustration of the different design criteria is provided elsewhere [44]. Further design criteria have also been described [70]. Some equivalences to the above introduced criteria Eqns (10–12) have been demonstrated [71]. A parameterization has been introduced [72] that allows for a continuous change between the above introduced three criteria.

In systems biology, the number of unknown parameters is often large compared to the available amount of measurements. This raises the problem of 'non-identifiability' [73–76]. 'Structural' non-identifiability refers to a redundant parameterization of the model. 'Practical' non-identifiability is due to limited amount of experimental information.

The above mentioned criteria are only meaningful if all model parameters are identifiable. Otherwise, the Fisher information matrix is singular. In this situation, a regularization techniques could be applied [70], i.e. a small number is added to all matrix entries of F.

In the case of a diagonal Fisher information matrix, the parameters of the model are called 'orthogonal'. Then, the precision of all parameters can be optimized independently.

In the more general case, not all parameters, but only s linear combinations Ap of the parameters could be of interest. Here, A denotes an $s \times n_p$ matrix. Often, only the kinetic parameters p are of interest in contrast to the parameters k of the observational function. The covariance matrix of such linear combinations is $AF^{-1}(\mathcal{D})A^T$.The inverse can be interpreted as a new Fisher information matrix, which can be used to define new utility functions to optimize the design for the estimation of the linear combinations. The corresponding D-optimal design is called 'D_A-optimal' [77].

A similar criterion is 'D_S-optimality' [78,79]. Here, the Fisher information matrix is arranged and then partitioned

$$F = \left(\begin{array}{c|c} B_{11} & B_{12} \\ \hline B_{12}^T & B_{22} \end{array}\right) \tag{13}$$

into four blocks. Block B_{11} contains second derivatives with respect to the interesting parameters and block B_{22} contains the corresponding derivatives with respect to the unimportant or 'nuisance parameters'. By maximization of

$$V(\mathcal{D}) = \det\left(B_{11} - B_{12}B_{22}^{-1}B_{12}^{T}\right) \quad (14)$$

the variance of the nuisance parameters is only considered if they are correlated to the parameter estimates of interest.

If a model is linear in the parameters, the Fisher information matrix becomes independent on the true underlying parameters. In this case, a global optimal design can be achieved. Otherwise, the proposed design depends on the prior knowledge of the parameters. *D*-optimal designs usually have the number different experimental conditions equal to the number of model parameters. Such designs are often very sensitive to parameter assumptions. Robustness of the designs with respect to the presumed underlying parameters is discussed elsewhere [80–84] and in the next section.

In a Monte Carlo approach, robust designs for parameter estimation are obtained by computing the expected utility function $\bar{V}(\mathcal{D})$ from the parameter prior distribution according to Eqn (5). Figure 8 provides an overview of the Monte Carlo approach.

Experimental design has been applied in systems biology in different contexts. Polynomial input functions [66] have been optimized for parameter estimation of the MAP-kinase signaling pathway. Optimal experiments for the estimation of unknown parameters in EGF receptor signaling have also been proposed in [85]. The estimation of model parameters of thiamine degradation is improved by appropriate designs [69]. Here, it is shown that optimization of the temperature profile as input to the system requires half of the experimental effort. Optimal input functions for a fed batch experiment for parameter estimation for a metabolic model have been determined [86]. An additional iterative approach to model identification of biological networks has been developed [87]. The authors applied their approach for parameter estimation in a mechanistic model of caspase activation in apoptosis.

```
Initialize model M
Initialize parameter prior ρ(β)
Initialize design D
Optimize D by evaluation of φ̂(D):
    For a parameter set β
        For a noise realization ε
            y(D) = Simulate (D,M,β,ε)
            [β̂(D), F(D)] = Estimate_Parameters (y(D), M)
            φ(D,β,ε) = Evaluate_OPT_Criterion (F(D))        Eqns (10)-(12)
    φ̂(D) = Weighted_Average (φ(D,β,ε), ρ(β), ρ(ε))          Eqn (5)
```

Fig. 8. Schematic overview of a Monte Carlo approach to optimize a design for parameter estimation.

Experimental design for model discrimination

The structure of a mathematical model for describing the studied system is initially unknown. 'Model discrimination' or 'model selection' is the statistical procedure to decide, on the basis of experimental data, which model is the most appropriate [88–90].

The accordance of the data and the model is examined by evaluation of the maximum likelihood function $\rho(y|\hat{p}^{(\mathcal{M})})$ for a model $\mathcal{M}$ obtained after parameter estimation. A well-established criterion for model discrimination is the 'Akaike Information Criterion (AIC)' [91,92]

$$\mathrm{AIC}^{(\mathcal{M})}(\mathcal{D}) = -2\log\rho(y|\hat{p}^{(\mathcal{M})}) + 2n_p^{(\mathcal{M})} \quad (15)$$

A model with a small AIC, i.e. with a low number of parameters $n_p^{(\mathcal{M})}$ and a large likelihood, is preferable. If two models are compared, the signum of the difference

$$\Delta\mathrm{AIC}^{(\mathcal{M}_m,\mathcal{M}_n)} = \log\frac{\rho(y|\hat{p}^{(\mathcal{M}_n)})}{\rho(y|\hat{p}^{(\mathcal{M}_m)})} + \left(n_p^{(\mathcal{M}_m)} - n_p^{(\mathcal{M}_n)}\right) \quad (16)$$

indicates the superior model. Here, model $\mathcal{M}_m$ would be preferred for negative $\Delta\mathrm{AIC}^{(\mathcal{M}_m,\mathcal{M}_n)}$.

Besides some further variants of the AIC, there are other related criteria such as the 'Bayes Information Criterion' [93], or the 'Minimum Description Length' [94], which can also be applied for the purpose of model discrimination. They are mathematically derived under slight different assumptions. Here, only the application of the AIC is discussed. Nevertheless, the AIC can be replaced if another model assessment criterion is desired.

The advantage of these model discrimination criteria is the general applicability. However, these criteria do not allow any conclusions concerning statistical significance. This is enabled by statistical tests, i.e. by a 'likelihood ratio test', [95,96]. Here, *p*-values are computed under the additional assumption that the considered models are 'nested', i.e. the parameter space of one model is a submanifold of the parameter space of the other model. Often, the submanifold can be obtained by setting some parameters to zero. The nested model can be considered as a special case of the other, more general model. If $\mathcal{M}_m$ denotes the submodel, it holds $\rho(y|\hat{p}^{(\mathcal{M}_m)}) \leq \rho(y|\hat{p}^{(\mathcal{M}_n)})$ for the two likelihood functions. Furthermore, if, $\mathcal{M}_m$ is appropriate, the advantage of $\mathcal{M}_n$ is only due to overfitting. In this case, it can be shown that under standard assumptions [97] the likelihood ratio

$$\mathrm{LR}^{(\mathcal{M}_m,\mathcal{M}_n)}(\mathcal{D}) = 2\log\left(\frac{\rho(y|\hat{p}^{(\mathcal{M}_n)})}{\rho(y|\hat{p}^{(\mathcal{M}_m)})}\right) \tag{17}$$

is χ^2_{df}-distributed. The degree of freedom (df) is given by the difference in the number of parameters. If the likelihood ratio obtained from the experimental data is larger, as one would expect according to the χ^2 distribution, the small model is rejected.

If the observational noise is independently, normally distributed, the likelihood ratio Eqn (17) becomes

$$\Delta\mathrm{RSS}^{(\mathcal{M}_m,\mathcal{M}_n)}(\mathcal{D}) = \sum_{d\in\mathcal{D}}\left(\frac{y(d) - g^{(\mathcal{M}_m)}(d,\hat{p}^{(\mathcal{M}_m)})}{\sigma(d)}\right)^2 - \sum_{d\in\mathcal{D}}\left(\frac{y(d) - g^{(\mathcal{M}_n)}(d,\hat{p}^{(\mathcal{M}_n)})}{\sigma(d)}\right)^2 \tag{18}$$

which is equal to the difference of the two standardized residual sum of squares. Here, $d \in \mathcal{D}$ denotes the design points, i.e. the set of chosen experimental conditions. For models that are linear in the parameters, the expectation of Eqn (18) is

$$V^{(\mathcal{M}_m,\mathcal{M}_n)}(\mathcal{D}) = \sum_{d\in\mathcal{D}}\left(\frac{g^{(\mathcal{M}_m)}(d,\hat{p}^{(\mathcal{M}_m)}) - g^{(\mathcal{M}_n)}(d,\hat{p}^{(\mathcal{M}_n)})}{\sigma(d)}\right)^2 \tag{19}$$

and therefore asymptotically (for large sample size) independent of the noise realization [98]. Therefore, numerical optimization does not require averaging over the observational noise.

In the analysis of experimental data, the first step is always a parameter estimation procedure to obtain the maximum likelihood function. Subsequently, computation of a model discrimination criterion for pairs of rival models is performed.

A Monte Carlo approach that imitates exactly these steps is schematically displayed in Fig. 9. Here, the expectation of a model discrimination criterion $V(\mathcal{D})$ is calculated by drawing numerous realizations from the model and from the parameter priors as well as from the distribution of the observational noise. Each realization of simulated data is analyzed exactly in the same way as it is intended for the experimental data, yielding a realization of the model discrimination criterion. The expectation is then used to optimize the design.

```
For all considered models M_m
    Initialize model prior π(M_m)
    Initialize parameter prior ρ^(M_m)(β)

Initialize design D
Optimize D by evaluation of φ̂(D):
    For a noise realization ε
        For a 'true' model M'
          For a parameter set β^(M') for M'
            y(D) = Simulate (D, M', β^(M'), ε)
            For all models M_m
              β̂^(M_m) = Estimate_Parameters (y(D), M_m)
              φ(D, M', β^(M'), ε) = Evaluate_OPT_Criterion (y(D), M_m, β̂^(M_m))
    φ̂(D) = Weighted_Average (φ(D, M', β^(M'), ε), π(M'), ρ^(M')(β), ρ(ε))
```

Fig. 9. Schematic overview of a general Monte Carlo approach to optimize a design for model discrimination.

This Monte Carlo approach is very general because there are no restrictive assumptions and every kind of prior knowledge can be included. On the other hand, such an approach is very expensive in terms of computational time.

There, are some approaches for the optimization of experimental designs for model discrimination that constitutes approximations of the general Monte Carlo approach (Fig. 9). Most algorithms are based on Eqn (19). In Hunter and Reiner [98]

$$V^{(\mathcal{M}_m,\mathcal{M}_n)}(\mathcal{D}) = \sum_{d\in\mathcal{D}}\left(\frac{g^{(\mathcal{M}_m)}(d,\langle p^{(\mathcal{M}_m)}\rangle) - g^{(\mathcal{M}_n)}(d,\hat{p}^{(\mathcal{M}_n)})}{\sigma(d)}\right)^2 \tag{20}$$

is optimized. Here, the expected response $g^{(\mathcal{M}_m)}(d,\langle p^{(\mathcal{M}_m)}\rangle)$ of the 'true' model $\mathcal{M}_m$ at design points d is computed for the expected parameters $\langle p^{(\mathcal{M}_m)}\rangle$ according to the parameter prior. The parameters $\hat{p}^{(\mathcal{M}_n)}$ of the other models are obtained by parameter estimation. A similar approach was used previously [99] to find the optimal design for two rival regression models. The obtained design is called 'T-optimal'. The case of more than two competing models is discussed elsewhere [100].

A criticism of both approaches is that uncertainty in the expected response due to parameter uncertainty is not considered. An example was provided previously [101] this uncertainty depends strongly on the design points. In an improved approach [102,103], the covariance matrices of the parameter prior distributions are propagated to the model response after linearization of the model. This leads to optimization of

$$V^{(\mathcal{M}_m,\mathcal{M}_n)}(\mathcal{D}) = \sum_{d\in\mathcal{D}}\frac{\left(g^{(\mathcal{M}_m)}(d,\langle p\rangle) - g^{(\mathcal{M}_n)}(d,\hat{p})\right)^2}{n_M\sigma^2(d) + \sum_{m'}\sigma^2_{m'}(d)} \tag{21}$$

where $\sigma^2_{m'}$ are the covariance matrices of the responses due to parameter uncertainty.

In Hsiang and Reilly [104], an approach is introduced in which also higher order moments are propagated. Here, a representative group of parameters sets $\{\tilde{p}_1^{(\mathcal{M})}, \tilde{p}_2^{(\mathcal{M})}, \ldots\}$ is drawn from the prior distribution of the parameters for each model. For these groups of parameters, the models are evaluated. This yields an expected response

$$\hat{g}^{(\mathcal{M})}(d) = \sum_i g^{(\mathcal{M})}(d, \tilde{p}_i^{(\mathcal{M})}) \rho^{(\mathcal{M})}(\tilde{p}_i^{(\mathcal{M})}) \quad (22)$$

for model $\mathcal{M}$ and

$$V^{(\mathcal{M}_m, \mathcal{M}_n)}(\mathcal{D}) = \sum_{d \in \mathcal{D}} \left(\frac{\hat{g}^{(\mathcal{M}_m)}(d) - \hat{g}^{(\mathcal{M}_n)}(d)}{\sigma(d)} \right)^2 \quad (23)$$

as a utility function for the comparison of two models. Here, the linearization of the model is avoided by computing the expectation after evaluation of the model response g.

In Eqns (20–23), model $\mathcal{M}_m$ is assumed as the true underlying model. The averaging over all pairwise comparison of the models accounting for model uncertainty yields:

$$V(\mathcal{D}) = \sum_{m,n \neq m} \pi(\mathcal{M}_m) \pi(\mathcal{M}_n) V^{(\mathcal{M}_m, \mathcal{M}_n)}(\mathcal{D}) \quad (24)$$

An alternativeis optimization of the worst case, i.e. maximization of the difference between the two most similar models

$$V(\mathcal{D}) = \min_{m,n \neq m} V^{(\mathcal{M}_m, \mathcal{M}_n)}(\mathcal{D}) \quad (25)$$

The introduced approaches are reasonable in the case of normally distributed noise. In a more general setting, the expected likelihood ratio

$$V_{\mathrm{LR}}^{(\mathcal{M}_m, \mathcal{M}_n)}(\mathcal{D}) = \sum_{m,n \neq m} \pi(\mathcal{M}_m) \pi(\mathcal{M}_n) \mathrm{LR}^{(\mathcal{M}_m, \mathcal{M}_n)}(\mathcal{D}) \quad (26)$$

or, for non-nested models, the expected difference

$$V_{\mathrm{AIC}}^{(\mathcal{M}_m, \mathcal{M}_n)}(\mathcal{D}) = \sum_{m,n \neq m} \pi(\mathcal{M}_m) \pi(\mathcal{M}_n) \Delta\mathrm{AIC}^{(\mathcal{M}_m, \mathcal{M}_n)}(\mathcal{D}) \quad (27)$$

in the Akaike Information can be used, instead.

A Bayesian methodology for optimal experimental design was introduced previously [101,105]. In this 'exact entropy approach', the entropy

$$S = -\sum_m \pi(\mathcal{M}_m) \ln \pi(\mathcal{M}_m) \quad (28)$$

is used to quantify the amount of information, i.e. the certainty about the true underlying model. A linearization of the model response is used to propagate the covariance matrices of the prior distributions. By this way, the expected change

$$V(\mathcal{D}) = S'(\mathcal{D}) - S \quad (29)$$

in the entropy is calculated which has to be optimized in the experimental planning. Equations for the expected entropy $S'(\mathcal{D})$ after a new experiment are provided elsewhere [101].

A comparison of both the Bayesian approach and the more frequentist approach are given elsewhere [106]. Only slight differences in the proposed designs were found. Another comparison of the published approaches is provided elsewhere [107].

Despite the importance of model selection, there are still few applications of the discussed experimental design procedures in the field of systems biology. Feng and Rabitz [108] introduced a concept called 'optimal identification' to estimate model parameters and discriminate between different models. Their algorithm is illustrated by a simulation study for a tRNA proofreading mechanism. The criteria in Eqn (21) were used previously [50] to calculate the optimal input for model selection between different dynamical models for a yeast fermentation in a bioreactor. Computer simulations [107] have also been used to check the applicability of model discrimination methods to modeling of polymerization reactions in organic chemistry. Here, some of the discussed design optimization approaches also were applied and compared. An overview about model selection and design aspects in engineering applications provided elsewhere [109].

An appropriate design for model selection is not necessarily advantageous for parameter estimation. An example of where the optimal design for discrimination between two regression models cannot be used to estimate the parameters of the true model has been described [70]. If both, parameter estimation and model discrimination is required, different design criteria, i.e. D-optimality and T-optimality, have to be combined [70].

Illustration by examples

In this section, the optimization of an experimental design is illustrated by some examples. Here, the sampling times are optimized. Analogical strategies could be applied for the optimization of the chosen observables, perturbations or the total number of measurements.

Figure 10 shows as an example a protein P and an enzyme E, which are produced with a common rate p_1. The enzyme is degraded with rate p_2 and promotes the degradation of the protein with parameter p_3. The time dependency of the protein concentration $x_P(t)$ and enzyme concentraton $x_E(t)$ is then given in model $\mathcal{M}_1$ by

$$\begin{aligned} \mathcal{M}_1 : \dot{x}_E(t) &= p_1 - p_2 x_E(t) \\ \dot{x}_P(t) &= p_1 - p_3 x_E(t) x_P(t) \end{aligned}$$

with $x_P(0) = x_E(0) = 0$. Initially, $p_1 = 2$, $p_2 = 1$ and $p_3 = 1$ are assumed as the true underlying parameters.

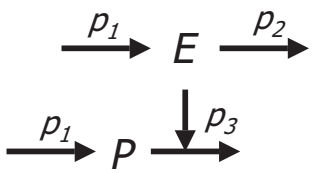

Fig. 10. In our example, a protein P and an enzyme E are produced with a common rate p_1. The enzyme is degraded with rate p_2 and promotes the degradation of the protein with rate p_3.

Furthermore, it is assumed that the protein concentration

$$y(t) = x_P(t) + \varepsilon, \varepsilon \sim N(0, 0.05) \quad (30)$$

is measured in absolute concentrations with a signal to noise ratio of approximately 5%.

First, the calculation of the optimal sampling times is exemplified for the estimation of the three rates p_1, p_2 and p_3, with an initial measurement at time t_1 followed by nine subsequent equidistant measurements in time. In this case, two design parameters, the point in time t_1 of the first measurement and the sampling interval Δt, have to be optimized. For this purpose, the D-optimality criterion according to Eqn (10) is applied.

The design region, i.e. the set of feasible and experimentally reasonable values of t_1 and Δt, can be restricted as an example to $t_1 > 0$ and $\Delta t > 0.25$. Another prerequisite could be that the measurements have to be executed within the first 10 min, leading to a further constraint $t_1 + 9\Delta t \leq 10$ if the time unit is minutes.

Because the model $\mathcal{M}_1$ is nonlinear in the parameters, the performance of a design, i.e. the expected accuracy of the parameter estimates, depends on the true underlying parameters and on the realization of the noise.

To examine the impact of the noise realizations, a hundred data sets $y(t) = x_P(t) + \varepsilon(t), t = t_1, t_1 + \Delta t, \ldots, t_1 + 9\Delta t$ for the same parameter set p_1, p_2, p_3 have been simulated for different t_1 and Δt. For each realization, the parameters have been (re)estimated and the covariance matrices of the parameter estimates have been calculated to determine $V = \det(F) = \det(\mathrm{Cov}(\hat{p}_i, \hat{p}_j)^{-1})$ according to Eqn (10). Figure 11 shows the expected performance, as well as the 25%, 50% (median) and 75% quantiles of $V(t_1, \Delta t)$.

Usually, the impact of different noise realization is neglected [44,51,69] and the performance is optimized for a single realization, namely the expected measurements $y(t) = x_P(t), t = t_1, t_1 \Delta t, \ldots, t_1 + 9\Delta t$. Figure 12 shows $V(t_1, \Delta t)$ for this approximation. The most informative design is obtained for $t_1^* = 0.52$ and $\Delta t^* = 0.56$, which is in accordance with Fig. 11, where the average and quantiles of the performance are displayed when many noise realizations are considered.

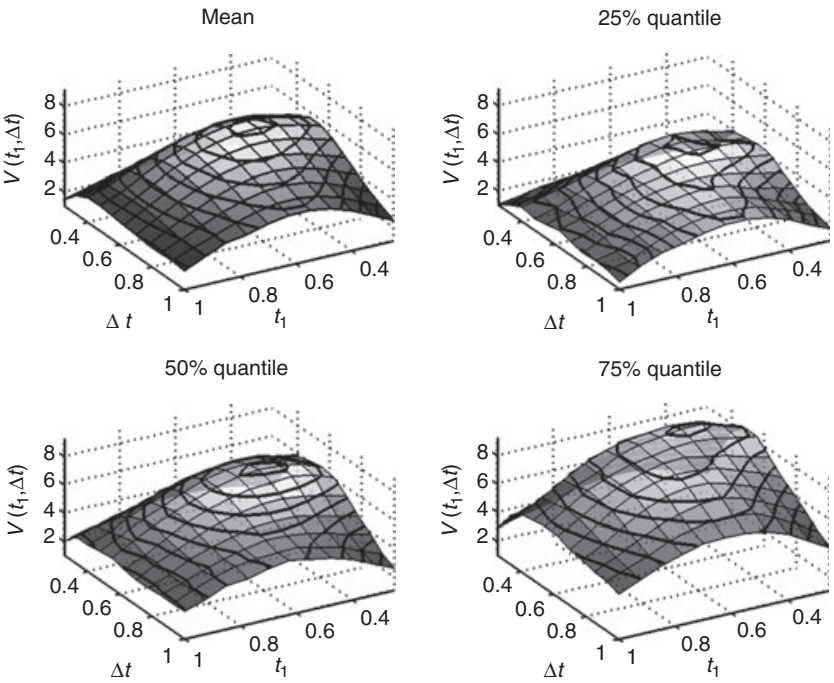

Fig. 11. For nonlinear models, the optimal design depends on the observational noise. Here, only a minor dependency of the optimal design parameters t_1 and Δt is observed between the mean, the 25% and 75% quantiles and the median performance.

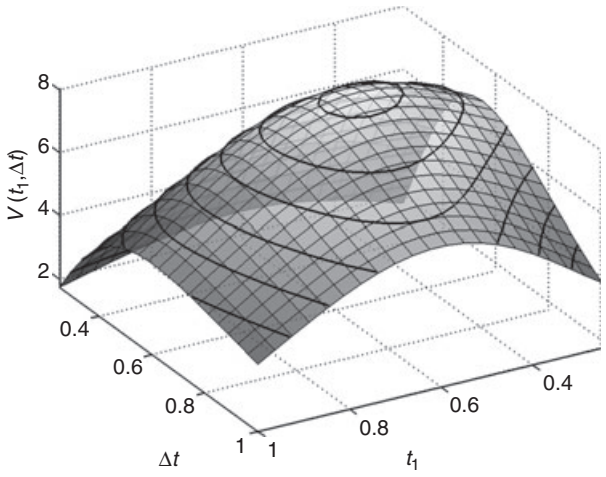

Fig. 12. The approximate performance of the design obtained for a single noise realization, i.e. for the expected measurements. The design is optimal for $t_1^* = 0.52$ and $\Delta t = 0.56$.

Figure 13 shows the dependency of $x_P(t)$ and the optimal sampling times for the initial parameter set (black curve). The protein concentration and the corresponding optimal sampling times are also displayed after changing p_1 (red), p_2 (green) and p_3 (blue) by a factor of two.

Next, design optimization for model selection is exemplified. For this purpose, we raise the question of whether the protein is degraded independently of the enzyme, i.e. model

$$\mathcal{M}_2 : \dot{x}_E(t) = p_1 - p_2 x_E(t)$$
$$\dot{x}_P(t) = p_1 - p_3 x_P(t)$$

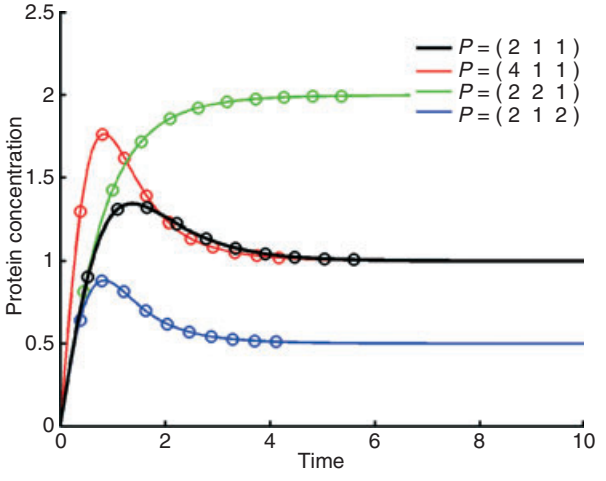

Fig. 13. The time dependency of the protein concentration for different parameter values and the optimal design for the (re)estimation of the three rates.

is compared with $\mathcal{M}_1$. In this case, the time dependency of the protein concentration yields

$$x_P(t) = p_1 t - \exp(-p_3 t) \tag{31}$$

for the case $x_P(0) = 0$. Again, the approximation $y(t) = x_P(t), t = t_1, t_1 + \Delta t, \ldots, t_1 + 9\Delta t$ is made. Because the number of parameters for both models $\mathcal{M}_1$ and $\mathcal{M}_2$ is equal, the utility functions based on the likelihood ratio (Eqn 26) and on the difference in the Akaike Information (Eqn 27) are equivalent.

Figure 14A shows the performance $V^{(\mathcal{M}_1,\mathcal{M}_2)}(t_1,\Delta t)$ if model $\mathcal{M}_1$ is assumed to be the true model. Figure 14B shows the performance if $\mathcal{M}_2$ is the correct model. If both models have equal prior probabilities $\pi(\mathcal{M}_1) = \pi(\mathcal{M}_2)$, $V^{(\mathcal{M}_1,\mathcal{M}_2)}$ and $V^{(\mathcal{M}_2,\mathcal{M}_1)}$ can be averaged to obtain an expected performance $V(t_1,\Delta t)$ according to Eqn (24) Fig. 14C. In this case, however, the average is dominated by $V^{(\mathcal{M}_1,\mathcal{M}_2)}$ because model $\mathcal{M}_1$ is hardly discriminated if model $\mathcal{M}_2$ is the truth. Therefore, depending on the purpose of the study, it could be more appropriate to optimize the worst case scenario, i.e. Eqn (25), which is plotted in panel (D) of Fig. 14.

Conclusions and outlook

In systems biology, experimental planning is becoming more and more crucial, because the establishment of mathematical models for complex biochemical networks requires huge experimental efforts. There are some studies concerning experimental design issues in the field of systems biology. However, most of them are restricted to certain applications, e.g. to microbial growth, or address only a single aspect of experimental planning.

In this minireview, an overview of experimental design aspects for systems biological applications is provided. General principles in experimental planning, i.e. replication and randomized sampling as well as the problem of confounding, are discussed. It is emphasized that clear definitions of the investigated hypotheses and the scope of the study are crucial. Also, an overview of numerical optimization of designs for the purpose of parameter estimation and for model discrimination is provided. Design optimization for parameter estimation and for model discrimination is illustrated by some examples.

In comparison to classical questions concerning design of experiments, the applications in systems biology are characterized by little prior knowledge. Therefore, experimental design considerations have to be robust against preceding assumptions. By all means, the sensitivity of a proposed experimental design with respect to the assumptions has to be considered.

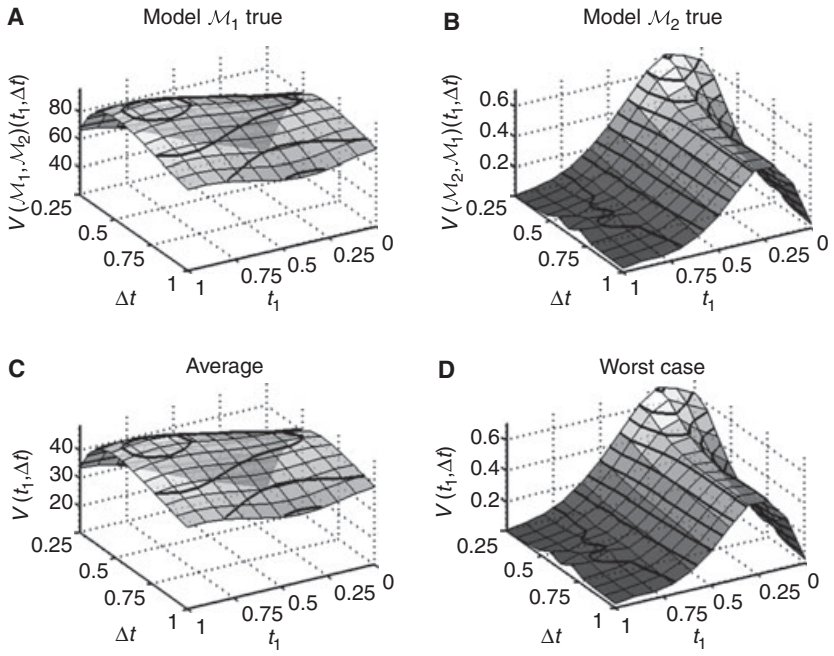

Fig. 14. The performance of model discrimination depending on the sampling times. Note the different vertical axes in the left and right panels. The performance is superior if model $\mathcal{M}_1$ is the true model (A). Therefore the average performance in panel (C) is dominated by $V^{(\mathcal{M}_1,\mathcal{M}_2)}$. The worst case scenario in panel (D) in this example is identical to the case where model $\mathcal{M}_2$ is the true one.

However, there is a general trade-off between the robustness of the designs and their efficiency for testing the hypotheses under consideration.

A related problem is that the models are often large and the number of measurements is very limited. Therefore, experiments have to be planned based on imprecise knowledge. Moreover, relative noise levels of 10% or more are standard for biochemical data. Model identification based on such noisy data is a challenging task. This situation can be improved by efficient experimental designs. However, the methods for experimental planning have to deal with the problem of non-identifiable parameters.

The models in systems biology are usually nonlinear in their parameters. Therefore, linearized models are only rough approximations and often are inadequate to show qualitatively the same behavior as the exact model. In addition, the nonlinearity hampers numerical optimization for finding globally optimal parameter estimates and their confidence intervals.

Monte Carlo approaches for experimental planning do not require any restrictive assumptions. However, an automatic and reliable optimization procedure is needed. Because the choice of an appropriate optimization technique is problem dependent, it is very difficult to implement an automatic global parameter estimation procedure without enough prior knowledge of the underlying model and the relevant part of the parameter space. Furthermore, the utility function that has to be optimized can only be estimated approximately by many realizations of the underlying model, the associated parameters and the observational noise. Therefore, approximation of the utility function is not smooth and standard optimization techniques, e.g. based on 'gradient descent', may not be applicable.

For these reasons, mathematical modeling in systems biology is a very challenging task that most likely requires the development of new methodological approaches. Proper experimental planning can decrease gaps between model based predictions, biologically motivated hypotheses and experimental validation, thus enabling the entire power of mathematical modeling to be exploited.

Acknowledgements

The authors thank Kilian Bartholome, Julia Rausenberger, Thomas Maiwald and Florian Geier for helpful discussions and for proofreading. In addition, the authors acknowledge financial support provided by the BMBF-grant 0313074D Hepatosys, FP6 EU-grant COSBICS LSHG-CT-2004-0512060 and BMBF-grant 0313921 FRISYS.

References

1 Asprey S & Macchietto S (2000) Statistical tools for optimal dynamic model building. *Comput Chem Eng* **24**, 1261–1267.

2 Montgomery DC (1991) *Design and Analysis of Experiments*, 3rd edn. John Wiley & Sons, New York, NY.
3 Mead R (1988) *The Design of Experiments: Statistical Principles for Practical Applications*. Cambridge University Press, Cambridge.
4 Black MA & Doerge RW (2002) Calculation of the minimum number of replicate spots required for detection of significant gene expression fold change in microarray experiments. *Bioinformatics* **18**, 1609–1616.
5 Churchill GA (2002) Fundamentals of experimental design for cDNA microarrays. *Nat Genet* **32**, 490–495.
6 Kerr MK (2003) Design considerations for efficient and effective microarray studies. *Biometrics* **59**, 822–828.
7 Kerr MK & Churchill GA (2001) Experimental design for gene expression microarrays. *Biostatistics* **2**, 183–201.
8 Kerr MK & Churchill GA (2001) Statistical design and the analysis of gene expression microarray data. *Genet Res* **77**, 123–128.
9 Simon RM & Dobbin K (2003) Experimental design of DNA microarray experiments. *Biotechniques*, Suppl. 16–21.
10 Eriksson J & Fenyö D (2007) Improving the success rate of proteome analysis by modeling protein-abundance distributions and experimental designs. *Nat Biotechnol* **25**, 651–655.
11 Boleda MD, Briones P, Farrs J, Tyfield L & Pi R (1996) Experimental design: a useful tool for PCR optimization. *Biotechniques* **21**, 134–140.
12 Freeman WM, Walker SJ & Vrana KE (1999) Quantitative RT-PCR: pitfalls and potential. *Biotechniques* **26**, 112–122, 124–125.
13 Ginzinger DG (2002) Gene quantification using real-time quantitative PCR: an emerging technology hits the mainstream. *Exp Hematol* **30**, 503–512.
14 Ideker TE, Thorsson V & Karp RM (2000) Discovery of regulatory interactions through perturbation: inference and experimental design. *Pacific Symposium on Biocomputing*, pp. 305–316.
15 Page D & Ong IM (2006) Experimental design of time series data for learning from dynamic Bayesian networks. *Pac Symp Biocomput* **11**, 267–278.
16 Pournara I & Wernisch L (2004) Reconstruction of gene networks using Bayesian learning and manipulation experiments. *Bioinformatics* **20**, 2934–2942.
17 Vatcheva I, Bernard O, de Jong H & Mars N (2006) Experiment selection for the discrimination of semi-quantitative models of dynamical systems. Technical report, Institute National de Recherche en Informatique et en Automatique. **170**, 472–506.
18 Yoo C & Cooper GF (2003) A computer-based microarray experiment design-system for gene-regulation pathway discovery. *AMIA Annu Symp Proc*, 733–737.
19 Atkinson A, Bogacka B & Zhigljavsky A (2000) *Optimum Design 2000*. Kluwer Publishers, Dordrecht.
20 Atkinson AC (1982) Developments in the design of experiments. *Int Statist Rev* **50**, 161–177.
21 Herzberg AM & Cox DR (1969) Recent work on the design of experiments: a bibliography and a review. *J R Statist Soc A* **132**, 29–67.
22 Chaloner K & Verdinelli I (1995) Bayesian experimental design: a review. *Stat Sci* **10**, 273–304.
23 Preece DA (1990) R. A. Fisher and experimental design: a review. *Biometrics* **46**, 925–935.
24 Dette H, Melas VB & Strigul N (2003) Design of experiments for microbiological models. Technical report, Ruhr University Bochum, Bochum.
25 Jacobsen M, Repsilber D, Gutschmidt A, Neher A, Feldmann K, Mollenkopf HJ, Kaufmann SHE & Ziegler A (2006) Deconfounding microarray analysis – independent measurements of cell type proportions used in a regression model to resolve tissue heterogeneity bias. *Methods Inf Med* **45**, 557–563.
26 Kirk R (1989) *Experimental Design: Procedures for the Behavioral Science*. Brooks/Cole Publishing Company, Belmont, CA.
27 Goulden CH (1956) *Methods of Statistical Analysis*. Wiley, New York, NY.
28 Greenland S & Morgenstern H (2001) Confounding in health research. *Annu Rev Public Health* **22**, 189–212.
29 Atkinson AC & Donev AN (1996) Experimental designs optimally balanced for trend. *Technometrics* **38**, 333–341.
30. Bailey RA, Cheng C-S & Kipnis P (1992) Construction of trend resistant factorial designs. *Stat Sin* **2**, 393–411.
31 Fisher RA (1950) *Statistical Methods for Research Workers*, 11 edn. Oliver and Boyd, Edingburgh.
32 Schilling M, Maiwald T, Bohl S, Kollmann M, Kreutz C, Timmer J & Klingmller U (2005) Computational processing and error reduction strategies for standardized quantitative data in biological networks. *FEBS J* **272**, 6400–6411.
33 Schilling M, Maiwald T, Bohl S, Kollmann M, Kreutz C, Timmer J & Klingmüller U (2005) Quantitative data generation for Systems Biology: the impact of randomization, calibrators and normalizers. *IEE Proc – Syst Biol* **152**, 193–200.
34 Kendziorski C, Irizarry RA, Chen K-S, Haag JD & Gould MN (2005) On the utility of pooling biological samples in microarray experiments. *PNAS* **102**, 4252–4257.
35 Quinn GP & Keough MJ (2002) *Experimental Design and Data Analysis for Biologists*. Cambridge University Press, Cambridge.
36 Cohen J (1988) *Statistical Power Analysis for the Behavioral Sciences*, 2nd edn. Erlbaum, Hillsdale, NJ.

37 Lee M-LT & Whitmore GA (2002) Power and sample size for DNA microarray studies. *Stat Med* **21**, 3543–3570.
38 Eng J (2003) Sample size estimation: how many individuals should be studied? *Radiology* **227,** 309–313.
39 Whitley E & Ball J (2002) Statistics review 4: sample size calculations. *Crit Care* **6**, 335–341.
40 Cilek JE & Mulrennan JA (1997) Pseudoreplication: what does it mean, and how does it relate to biological experiments? *J Am Mosq Control Assoc* **13**, 102–103.
41 Hurlbert SH (1984) Pseudoreplication and the design of ecological field experiments. *Ecol Monogr* **54**, 187–211.
42 Balsa-Canto E, Alonso AA & Banga JR (1998) *Dynamic Optimization of Bioprocesses: Deterministic and Stochastic Strategies*. Automatic Control of Food & Biological Processes.
43 Banga JR, Balsa-Canto E, Moles CG & Alonso AA (2005) Dynamic optimization of bioprocesses: efficient and robust numerical strategies. *J Biotechnol* **117**, 407–419.
44 Asprey S & Macchietto S (2002) Designing robust optimal dynamic experiments. *J Process Control* **12**, 545–556.
45 Cooney MJ & McDonald K (1995) Optimal dynamic experiments for bioreactor model discrimination. *Appl Microbiol Biotechnol* **43**, 826–837.
46 Espie D & Macchietto S (1989) The optimal design of dynamic experiments. *AIChE J* **35**, 223–229.
47 Galvanin F, Macchietto S & Bezzo F (2007) Model-based design of parallel experiments. *Ind Eng Chem Res* **46**, 871–882.
48 Maiwald T, Kreutz C, Pfeifer AC, Bohl S, Klingmüller U & Timmer J (2007) Feasibility analysis and optimal experimental design. *Ann N Y Acad Sci* **1115**, 212–220.
49 Kremling A, Fischer S, Gadkar K, Doyle FJ, Sauter T, Bullinger E, Allgoewer F & Gilles ED (2004) A benchmark for methods in reverse engineering and model discrimination: Problem formulation and solutions. *Genome Res* **14**, 1773–1765.
50 Chen BH & Asprey SP (2003) On the design of optimally informative experiments for model discrimination among dynamic crystallization process models. *Proceedings Foundations of Computer-Aided Process Operations*, pp. 455–458.
51 Baltes M, Schneider R, Sturm C & Reuss M (1994) Optimal experimental design for parameter estimation in unstructured growth models. *Biotechnol. Prog.* **10**, 480–488.
52 Kutalik Z, Cho K-H & Wolkenhauer O (2004) Optimal sampling time selection for parameter estimation in dynamic pathway modeling. *Biosystems* **75**, 43–55.
53 Cho K-H, Kolch W & Wolkenhauer O (2003) Experimental design in Systems Biology, based on parameter sensitivity analysis using a Monte Carlo method: a case study for the TNFα-mediated NF-κB signal transduction pathway. *Simulation* **79**, 726–739.
54 Dette H & Biedermann S (2003) Robust and efficient designs for the Michaelis-Menten model. *J Am Stat Assoc* **98**, 679–686.
55 Johnson PD & Besselsen DG (2002) Practical aspects of experimental design in animal research. *ILAR J* **43**, 202–206.
56 Kiefer J (1959) Optimum experimental designs. *J R Stat Soc Ser B* **21**, 272–319.
57 Swameye I, Müller T, Timmer J, Sandra O & Klingmüller U (2003) Identification of nucleocytoplasmic cycling as a remote sensor in cellular signaling by data-based modeling. *Proc Natl Acad Sci USA* **100**, 1028–1033.
58 Cho K-H & Wolkenhauer O (2003) Analysis and modeling of signal transduction pathways in systems biology. *Biochem Soc Trans* **31**, 1503–1509.
59 Mendes P & Kell D (1998) Non-linear optimization of biochemical pathways: application to metabolic engineering and parameter estimation. *Bioinformatics* **14**, 869–883.
60 Rodriguez-Fernandez M, Mendes P & Banga JR (2006) A hybrid approach for efficient and robust parameter estimation in biochemical pathways. *Biosystems* **83**, 248–265.
61 Honerkamp J (1993) *Stochastic Dynamical Systems*. VCH, New York, NY.
62 Tarantola A (2005) *Inverse Problem Theory*. SIAM, Philadelphia, PA.
63 Horbelt W (2001) *Maximum likelihood estimation in dynamical systems*. PhD thesis, University of Freiburg, Freiburg.
64 Hidalgo ME & Ayesa E (2001) Numerical and graphical description of the information matrix in calibration experiments for state-space models. *Water Res* **35**, 3206–3214.
65 Silvey SD (1970) *Statistical Inference*. Penguin Books Ltd, Harmondsworth, Middlesex, England.
66 Faller D, Klingmüller U & Timmer J (2003) Simulation methods for optimal experimental design in Systems Biology. *Simul: Trans Soc Model Comput Simul* **79**, 717–725.
67 Dette H, Melas VB & Pepelyshev A (2003) Standardized maximum E-optimal designs for the Michaelis-Menten model. *Stat Sin* **13**, 1147–1167.
68 John RCS & Draper NR (1975) D-optimality for regression designs: a review. *Technometrics* **17**, 15–23.
69 Balsa-Canto JBE & Rodriguez-Fernandez M (2007) Optimal design of dynamic experiments for improved estimation of kinetic parameters of thermal degradation. *J Food Eng* **82**, 178–188.
70 Atkinson AC & Donev AN (1992) *Optimum Experimental Designs*. Clarendon Press, Oxford.

71 Kiefer J & Wolfowitz J (1960) The equivalence of two extremum problems. *Can J Math* **12**, 363–366.
72 Kiefer J (1975) Optimal design: variation in structure and performance under change of criterion. *Biometrika* **62**, 277–288.
73 Chappell M, Godfrey K & Vajda S (1990) Global identifiability of the parameters of nonlinear systems with specified inputs: a comparison of methods. *Math Biosci* **102**, 41–73.
74 Ljung L & Glad T (1994) On global identifiability for arbitrary model parameterizations. *Automatica* **30**, 265–276.
75 Hengl S, Kreutz C, Timmer J & Maiwald T (2007) Data-based identifiability analysis of non-linear dynamical models. *Bioinformatics* **23**, 2612–2618.
76 Timmer J, Müller T & Melzer W (1998) Numerical methods to determine calcium release flux from calcium transients in muscle cells. *Biophys J* **74**, 1694–1707.
77 Titterington DM (1975) Optimal design: some geometrical aspects of D-optimality. *Biometrika* **2**, 313–320.
78 Atkinson AC (1988) Recent developments in the methods of optimum and related experimental designs. *Int Stat Rev* **56**, 99–115.
79 Studden WJ (1980) D_s-optimal designs for polynomial regression using continued fractions. *Ann Stat* **8**, 1132–1141.
80 DeFeo P & Myers RH (1992) A new look at experimental design robustness. *Biometrika* **79**, 375–380.
81 Goos P, Kobilinsky A & O'Brien TE (2005) Model-robust and model-sensitive designs. *Comput Stat Data Anal* **49**, 201–216.
82 Rojas CR, Welsh JS, Goodwin GC & Feuer A (2007) Robust optimal experiment design for system identification. *Automatica* **43**, 993–1008.
83 Sacks J & Ylvisaker D (1984) Some model robust designs in regression. *Ann Stat* **12**, 1324–1348.
84 Yue R-X & Hickernell FJ (1999) Robust designs for fitting linear models with misspecification. *Stat Sin* **9**, 1053–1069.
85 Casey FP, Baird D, Feng Q, Gutenkunst RN, Waterfall JJ, Myers CR, Brown KS, Cerione RA & Sethna JP (2006) Optimal experimental design in an EGFR signaling and down-regulation model. Technical report, Center for Applied Mathematics, Cornell University, Ithaca, NY.
86 Munack A (1989) Design of optimal dynamical experiments for parameter estimation. *Proceedings of the American Control Conference*, ACC89, Pittsburgh, PA, pp. 2011–2016.
87 Gadkar KG, Gunawan R & Doyle FJ III (2005) Iterative approach to model identification of biological networks. *BMC Bioinformatics* **6**, 1–20.
88 Steward WE, Henson TL & Box GEP (1996) Model discrimination and criticism with single-response data. *AIChE J* **42**, 3055–3062.
89 Steward WE, Shon Y & Box GEP (1998) Discrimination and goodness of fit of multiresponse mechanistic models. *AIChE J*, **66**, 1404–1412.
90 Timmer J, Müller T, Sandra O, Swameye I & Klingmüller U (2004) Modeling the non-linear dynamcis of cellular signal transduction. *Int J Bif Chaos* **14**, 2069–2079.
91 Akaike H (1974) A new look at the statistical model identification. *IEEE Trans Automat Contr* **AC-19**, 716–723.
92 Sakamoto Y, Ishiguro M & Kitagawa G (1986) *Akaike Information Criterion Statistics*. D. Reidel Publishing Company, Dordrecht.
93 Schwarz G (1978) Estimating the dimension of a model. *Ann Stat* **6**, 461–464.
94 Rissanen J (1983) A universal prior for integers and estimation by minimum description length. *Ann Stat* **11**, 416–431.
95 Cox D (1961) Tests of separate families of hypotheses. In *Proceedings of Fourth Berkeley Symposium on Mathematical Statistics and Probability*, 1, pp. 105–123. University of California Press, Berkeley, CA.
96 Honerkamp J (2002) *Statistical Physics. An Advanced Approach with Applications*. Springer-Verlag, Heidelberg.
97 Self SG & Liang KY (1987) Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions. *J Am Stat Assoc* **82**, 605–610.
98 Hunter WG & Reiner AM (1965) Designs for discriminating between two rival models. *Technometrics* **7**, 307–323.
99 Atkinson AC & Fedorov VV (1975) Optimal design: experiments for discriminating between two rival models. *Biometrika* **62**, 57–70.
100 Atkinson AC & Fedorov VV (1975) The design of experiments for discriminating between several models. *Biometrika* **62**, 289–303.
101 Box GEP & Hill WJ (1967) Discrimination among mechanistic models. *Technometrics* **9**, 57–71.
102 Buzzi Ferraris G & Forzatti P (1984) Sequential experimental design for model discrimination in the case of multiple responses. *Chem Eng Sci* **39**, 81–85.
103 Buzzi Ferraris G, Forzatti P, Emig G & Hofmann H (1983) New sequential experimental design procedure for discriminating among rival models. *Chem Eng Sci* **38**, 225–232.
104 Hsiang T & Reilly PM (1971) A practical method for discriminating among mechanistic models. *Can J Chem Eng* **38**, 225.
105 Reilly PM (1970) Statistical methods in model discrimination. *Can J Chem Eng* **48**, 168–173.

106 Atkinson AC (1981) A comparison of two criteria for the design of experiments for discriminating between models. *Technometrics* **23**, 301–305.
107 Burke AL, Duever TA & Penlidis A (1994) Model discrimination via designed experiments: Discriminating between the terminal and penultimate models on the basis of composition data. *Macromolecules* **27**, 386–399.
108 Feng X-J & Rabitz H (2004) Optimal identification of biochemical reaction networks. *Biophys J* **86**, 1270–1281.
109 Verheijen PJ (2003) Model selection: an overview of practices in chemical engineering. *Comput-Aided Chem Eng* **16**, 85–104.