Author for correspondence:
Julio R. Banga
e-mail: julio@iim.csic.es

Reverse engineering and identification in systems biology: strategies, perspectives and challenges

Alejandro F. Villaverde and Julio R. Banga

BioProcess Engineering Group, IIM-CSIC, Spanish National Research Council, Vigo 36208, Spain

The interplay of mathematical modelling with experiments is one of the central elements in systems biology. The aim of reverse engineering is to infer, analyse and understand, through this interplay, the functional and regulatory mechanisms of biological systems. Reverse engineering is not exclusive of systems biology and has been studied in different areas, such as inverse problem theory, machine learning, nonlinear physics, (bio)chemical kinetics, control theory and optimization, among others. However, it seems that many of these areas have been relatively closed to outsiders. In this contribution, we aim to compare and highlight the different perspectives and contributions from these fields, with emphasis on two key questions: (i) why are reverse engineering problems so hard to solve, and (ii) what methods are available for the particular problems arising from systems biology?

1. Introduction

In the late 1960s, Mesarović [1, p. 83] stated something that is still relevant today: 'the real advance in the application of systems theory to biology will come about only when the biologists start asking questions which are based on the system-theoretic concepts rather than using these concepts to represent in still another way the phenomena which are already explained in terms of biophysical or biochemical principles'.

Four decades later, Csete & Doyle [2], considering the reverse engineering of biological complexity, argued that, although biological entities and engineered advanced technologies have very different physical implementations, they are quite similar in their systems-level organization. Furthermore, they also noted that the level of complexity in engineering design was approaching that of living systems. When viewed as networks, biological systems share some important structural features with engineered systems, such as modularity, robustness and use of recurring circuit elements [3]. Frequently, important aspects of the functionality of a network can be derived solely from its structure [4]. It seems therefore natural that systems engineering and related disciplines can play a major role in modern systems biology [5–9].

Today, a decade after the reverse engineering paper of Csete & Doyle, recent research [10] clearly shows the feasibility of comprehensive large-scale whole-cell computational modelling. This class of models includes the necessary detail to provide mechanistic explanations and allows for the investigation of how changes at the molecular level influence behaviour at the cellular level [11]. Multi-scale modelling, which considers the interactions between metabolism, signalling and gene regulation at different scales both in time and space, is key to the study of complex behaviour and opens opportunities to facilitate biological discovery [12,13]. The interplay between experiments and computational modelling has led to models with improved predictive capabilities [14]. In the case of evolutionary and developmental biology, reverse engineering of gene regulatory networks (GRNs) and numerical (*in silico*) evolutionary simulations have been used [15,16] to explain observed phenomena

and, more importantly, to suggest new hypotheses and future experimental work. Finally, model-based approaches are already in place for the next step, namely, synthetic biology [17].

Most reverse engineering studies of biological systems have considered microbial cells. In this context, a wide range of modelling approaches have been adopted, which can be classified according to different taxonomies. Stelling [18] distinguished between three large groups: interaction-based (no dynamics, no parameters [19,20]), constraint-based (no dynamics, only stoichiometry parameters [21,22]) and mechanism-based models (dynamic, with both stoichiometry and kinetic parameters). Other classifications can be found in more recent literature, such as those based on modelling formalisms [23,24], which include Boolean networks, Bayesian networks (BNs), Petri nets, process algebras, constraint-based models, differential equations, rule-based models, interacting state machines, cellular automata and agent-based models.

Regardless of the type of representation chosen, the importance of taking into account the system dynamics has to be acknowledged [25,26]. It has been stated that the central dogma of systems biology is that the functioning of cells is a consequence of system dynamics [5]. In particular, regulation—usually achieved by feedback—plays a key role in biological processes [27]. Hence, the study of the rich behaviour exhibited by biological systems requires the use of engineering tools, namely from the systems and control areas [7]. Furthermore, it has been argued that even more interesting than the application of systems engineering ideas to biological problems is the inspiration that these problems provide in the development of new theories [8]. Systems engineering aims to design systems, while biology aims to understand (reverse engineering) them; it is natural then that these two communities have traditionally specialized in solving different problems. However, the interplay between both disciplines can be mutually beneficial [6]; in this sense, systems biology can be seen 'not as the application of engineering principles to biology but as a merger of systems and control theory with molecular and cell biology' [5].

This work reviews different perspectives for the reverse engineering problem in biological systems. The first step in the identification of a dynamic model is to establish its components and connectivity, a task for which either prior knowledge or data-driven statistical methods is required [28]. We begin by discussing these methods in §2, where we address the reduced problem of recovering interaction structures. We classify the methods proposed for this task in three main strategies: correlation-based, information-theoretic and Bayesian. Then in §3, we discuss the different perspectives for the reverse engineering of complete dynamic models,[1] grouping them in eight areas: inverse problems, optimization, systems and control theory, chemical reaction network theory, Bayesian statistics, physics, information theory and machine learning. We finish this review with some conclusions about the convergence of these perspectives in §4.

2. Interaction networks: three main strategies

We address now the question of reverse engineering systems modelled as interaction networks. This problem can be formulated as follows: given a list of nodes (variables), infer the connections (dependencies) among them using the information contained in the available datasets. The goal is the determination of the existing interactions, not the detailed characterization of these interactions. Thus, the recovered models do not include differential equations, and there is no need to estimate parameter values such as kinetic constants. This problem can be considered as a reduced version of the general reverse engineering problem, which will be considered in the following sections. However, this does not mean that it is easy to solve; on the contrary, it is still a very active area of research. The key task is to estimate the strength of the dependence among variables using the available data.

Most of the methods used to infer interactions are ultimately related to statistics. In this context, it is worth mentioning that there are several schools of thought in statistics: Bayesian, frequentist, information-theoretic and likelihood (the latter being a common element in all of them). However, roughly speaking, the Bayesian and frequentist approaches are usually considered the main paradigms [33,34].

The history of statistics reveals that the Bayesian approach was initially developed in the eighteenth century. Bayes himself only considered a special case of the theorem that receives his name, which was actually rediscovered independently and further developed in its modern form by Laplace years later. At that time, the theory received the name of inverse probability. Frequentist statistics was developed during the first decades of the twentieth century by Pearson, Neyman and Fisher, among others. The frequentist theory rapidly displaced the inverse probability (Bayesian) approach and became the dominant school in statistics. Bayesian ideas barely survived, mostly outside statistics departments (a detailed history is given in [35]). The use of Bayes prior information was regarded by many frequentists as the introduction of subjectivity, and therefore, a biased approach, something not acceptable in the scientific method. Although refinements, e.g. empirical Bayes methods [36] (prior distribution based on existing data, not assumptions), tried to surmount this, Bayesian approaches still had another major problem: the computations needed were extremely demanding.

In the 1980s, the application of Markov chain Monte Carlo (MCMC) methods [37,38] changed everything. MCMC and related techniques [39,40] made feasible many of the complex computations necessary in Bayesian methods and the theory resurfaced and started to be applied in many areas [41], including bioinformatics and computational systems biology [42–44].
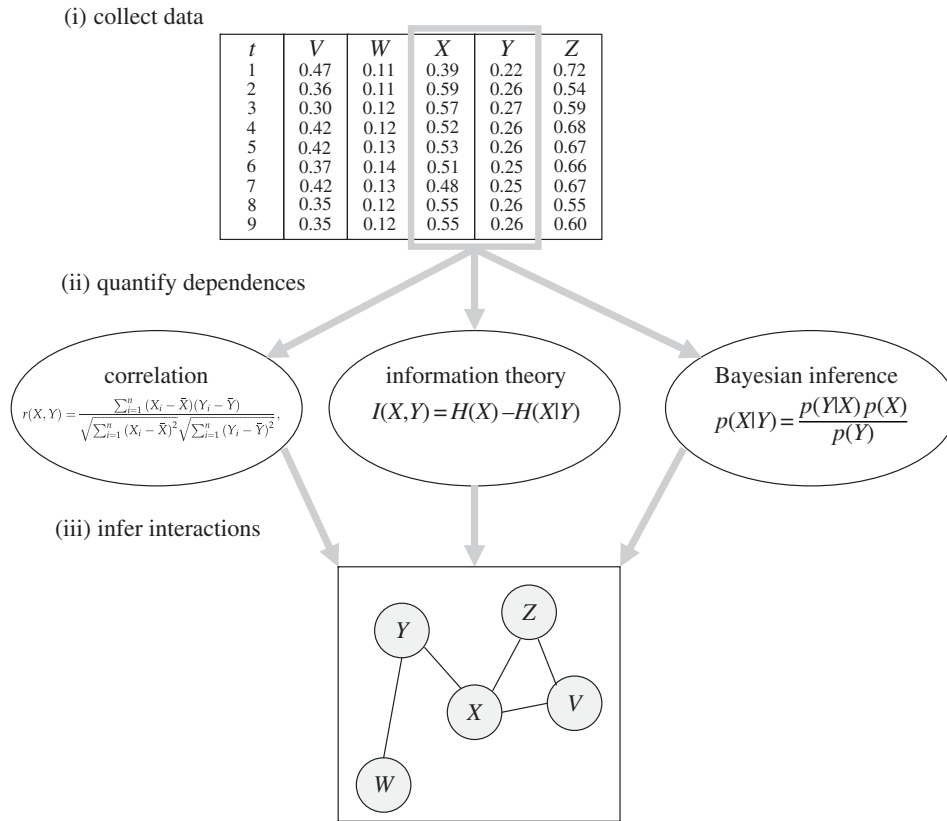
Depending on the statistic used to measure the interaction strength, the most common reverse engineering approaches can be classified into three classes: correlation, mutual information and Bayesian (see figure 1). Their main characteristics are discussed in the following subsections; more detailed surveys can be found in [45–49]. With a more specific focus, Bayesian methods were covered in [50,51] and information-theoretic approaches in [52] (figure 1).

2.1. A classical tool: correlation

The correlation coefficient r, commonly referred to as the Pearson correlation coefficient, quantifies the dependence between two random variables X and Y as

$$r(X,Y) = \frac{\sum_{i=1}^{n} (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^{n} (X_i - \bar{X})^2}\sqrt{\sum_{i=1}^{n} (Y_i - \bar{Y})^2}}, \qquad (2.1)$$

where X_i, Y_i are the n data points and $\bar{X}$, $\bar{Y}$ are their averages. If both variables are linearly independent, $r(X,Y) = 0$; in the opposite situation, where one variable is completely determined by the other, $r(X,Y) = \pm 1$.

(i) collect data

t	V	W	X	Y	Z
1	0.47	0.11	0.39	0.22	0.72
2	0.36	0.11	0.59	0.26	0.54
3	0.30	0.12	0.57	0.27	0.59
4	0.42	0.12	0.52	0.26	0.68
5	0.42	0.13	0.53	0.26	0.67
6	0.37	0.14	0.51	0.25	0.66
7	0.42	0.13	0.48	0.25	0.67
8	0.35	0.12	0.55	0.26	0.55
9	0.35	0.12	0.55	0.26	0.60

(ii) quantify dependences

correlation
$$r(X,Y) = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^{n}(X_i - \bar{X})^2}\sqrt{\sum_{i=1}^{n}(Y_i - \bar{Y})^2}},$$

information theory
$$I(X,Y) = H(X) - H(X|Y)$$

Bayesian inference
$$p(X|Y) = \frac{p(Y|X)\,p(X)}{p(Y)}$$

(iii) infer interactions

Figure 1. Approaches for inferring interaction networks. Schematic of the process of inferring a network structure from data, showing three approaches for measuring dependence among variables: correlation-based, information theoretic and Bayesian.

Correlation-based methods can be used for unsupervised learning from data and have been widely used to discover biological relationships. While most applications have been developed for genetic networks [53,54], there are also examples in reverse engineering metabolic networks. One such method is correlation metric construction [55], which takes into account time lags among species and was successfully tested on the glycolytic pathway [56]. A more sophisticated measure of association between variables is the distance correlation method [57,58], which has theoretical advantages over Pearson's coefficient and has been recently used in biological applications [59,60].

2.2. Perspective from information theory

While the Pearson coefficient is appropriate for measuring linear correlations, its accuracy decreases for strongly non-linear interactions. A more general measure is mutual information, a fundamental concept of information theory defined by Shannon [61]. It is based on the concept of entropy, which is the uncertainty of a single random variable: let X be a discrete random vector with alphabet χ and probability mass function $p(x)$. The entropy is

$$H(X) = -\sum_{x \in \chi} p(x) \log p(x) \qquad (2.2)$$

and the conditional entropy $H(Y|X)$ is the entropy of a variable Y conditional on the knowledge of another variable X

$$H(Y|X) = \sum_{x} p(x)H(Y|X=x)$$
$$= -\sum_{x}\sum_{y} p(x,y)\log p(y|x). \qquad (2.3)$$

The mutual information I of two variables measures the amount of information that one variable contains about another; equivalently, it is the reduction in the uncertainty of one variable owing to the knowledge of another. It can be defined in terms of entropies as [62]

$$I(X,Y) = H(X) - H(X|Y). \qquad (2.4)$$

As mutual information is a general measure of dependencies between variables, it may be used for inferring interaction networks: if two components have strong interactions, their mutual information will be large; if they are not related, it will be theoretically zero. Mutual information has been applied for reverse engineering biological networks since the 1990s. In early applications [63–67], genetic interactions were hypothesized from high values of pairwise mutual information between genes. The success of this approach encouraged further research and increasingly sophisticated techniques were developed during the following decade. One of the most popular methods for GRN inference is ARACNE [68], which exploits the data processing inequality (DPI, [62]) to discard indirect interactions. The DPI states that if $X \to Y \to Z$ is a Markov chain, then $I(X,Y) \geq I(X,Z)$. ARACNE examines the gene triplets (X,Y,Z) that have a significant value of mutual information and removes the edge with the smallest value, thus reducing the number of false positives. A time-delay version of ARACNE, which is especially suited for time-course data, is also available [69].

In reverse engineering applications, the probability mass functions $p(x)$, $p(y)$ are generally unknown; however, they can be estimated from experimental data using several methods. The simplest one is to partition the data into bins of a fixed width, and approximate the probabilities by the

frequencies of occurrence. This naive solution has the drawback that the mutual information is systematically overestimated [70]. To avoid this problem, one can either make the bin-size dependent on the density of data points (adaptive partitioning, [71]), or use kernel density estimation [72]. The influence of the choice of estimators on the network inference problem has been studied in [73].

Information-theoretic methods have a rigorous theoretical foundation on concepts that allow for an intuitive interpretation. This facilitates the development of new methods that are aimed at specific purposes. An example is the distinction between direct and indirect interactions, which has motivated the design of methods, such as minimum redundancy networks [74], three-way mutual information [75], entropy metric construction and entropy reduction technique [76], among others. Another example is the modification of the calculation of mutual information by taking into account the background distribution for all possible interactions, as done by the context likelihood of relatedness technique (CLR) [77]. The combination of CLR with another method, the Inferelator [78], became one of the top performers at the DREAM4 100-gene *in silico* network inference challenge [79]. In yet another example, a recently presented statistic called maximal information coefficient [80] aims to enforce equitability, a property that consists of assigning similar values to equally noisy relationships, independently of the type of association.

Cantone *et al.* [81] argued in 2009 that information-theoretic methods were not appropriate for reconstruction of small networks, because they could not infer the direction of regulations. However, in the years following that statement some progress was made, and some information-theoretic methods capable of recovering directions are already available [69,82].

2.3. Incorporating prior knowledge: the Bayesian inference perspective

Prior knowledge can be incorporated into the inference procedure using a Bayesian framework. The Bayes rule for two variables X and Y is

$$p(X|Y) = \frac{p(Y|X)p(X)}{p(Y)}, \qquad (2.5)$$

where $p(Y)$ and $p(X|Y)$ are called prior and posterior probabilities, respectively. In a typical scenario, X may be the value of a parameter and Y the available data. The Bayes rule allows the belief in a prediction to be updated given new observations. In practice, complications may arise owing to the fact that typically neither $p(Y)$ nor $p(Y|X)$ is known. Estimation of these quantities usually involves computationally costly calculations, which do not scale up well for large networks. It may be necessary to decompose the full problem according to the underlying conditional independence structure of the model [42]; graphical models appear in this context. Probabilistic graphical models represent joint probability distributions as a product of local distributions that involve only a few variables [83]. BNs are probabilistic graphical models in which the variables are discrete; their graphical representation is given by a directed acyclic graph (DAG).

BNs can be automatically inferred from data, a problem known as Bayesian inference. Reverse engineering a BN consists of finding the DAG that 'best' describes the data. The goodness of fit to the data is given by a score calculated from

the Bayes rule. It should be noted that the search for the best BN is an NP-hard problem [84], and therefore heuristic methods are used for solving it. Additionally, it is possible to look for approximations that help to decrease the computational complexity: approximate Bayesian computation (ABC) methods estimate posterior distributions without explicitly calculating likelihoods, using instead simulation-based procedures [85,86]. A Bayesian method for constructing a probabilistic network from a database was first presented in [87].

Genetic networks can be represented as probabilistic graphical models, by associating each gene with a random variable. The expression level of the gene gives the value of this random variable. Bayesian approaches were first used for reverse engineering genetic networks from expression data in [88]. An important limitation of BNs is that they are acyclic, while in reality most biological networks contain loops. An extension of BNs called dynamic Bayesian networks (DBNs) can be used to overcome this issue. Unlike BNs, DBNs can include cycles and may be constructed when time-course data are available [89–92].

## 3. Dynamic models: perspectives from different areas

Here, we focus on dynamic (kinetic) models of biological systems. These models typically consist of systems of differential equations. From the identification point of view, one can distinguish among three main problem classes (in decreasing order of generality):

(1) Full network inference (reverse engineering or reconstruction): given (high-throughput) dynamic data (i.e. time-series of measured concentrations and other properties), one seeks to find the full network (kinetic model structure and kinetic parameters) that fits (explains) the data.
(2) Network selection (network refinement, retrofitting): given dynamic data and an existing dynamic model with possible structural modifications (or a set of alternative kinetic model structures), the objective is to find the structural modifications and the kinetic parameters that fit the data.
(3) Kinetic parameter estimation (model calibration, parametric identification): given dynamic data and a fixed kinetic model structure, the objective is to find the kinetic parameters that fit the data.

Problem (1) above is the most general, while problem (2) is somewhere in the middle between the general inference problem and the more focused parameter estimation problem. Although fitting existing data is usually the first objective sought, one should also perform cross-validation studies with a different set of existing data. Ultimately, one should also seek to use the inferred model to allow for high-quality predictions under different conditions. Problem (1) has been usually solved using a bilevel approach, first determining the interaction network (as discussed in §2), and then identifying the kinetic details.

It has been widely recognized that all of the above problems are hard. Many approaches have been proposed for solving them, using different theoretical foundations. Several authors have carried out comparisons among methods using simulated or experimental data; early examples can be found
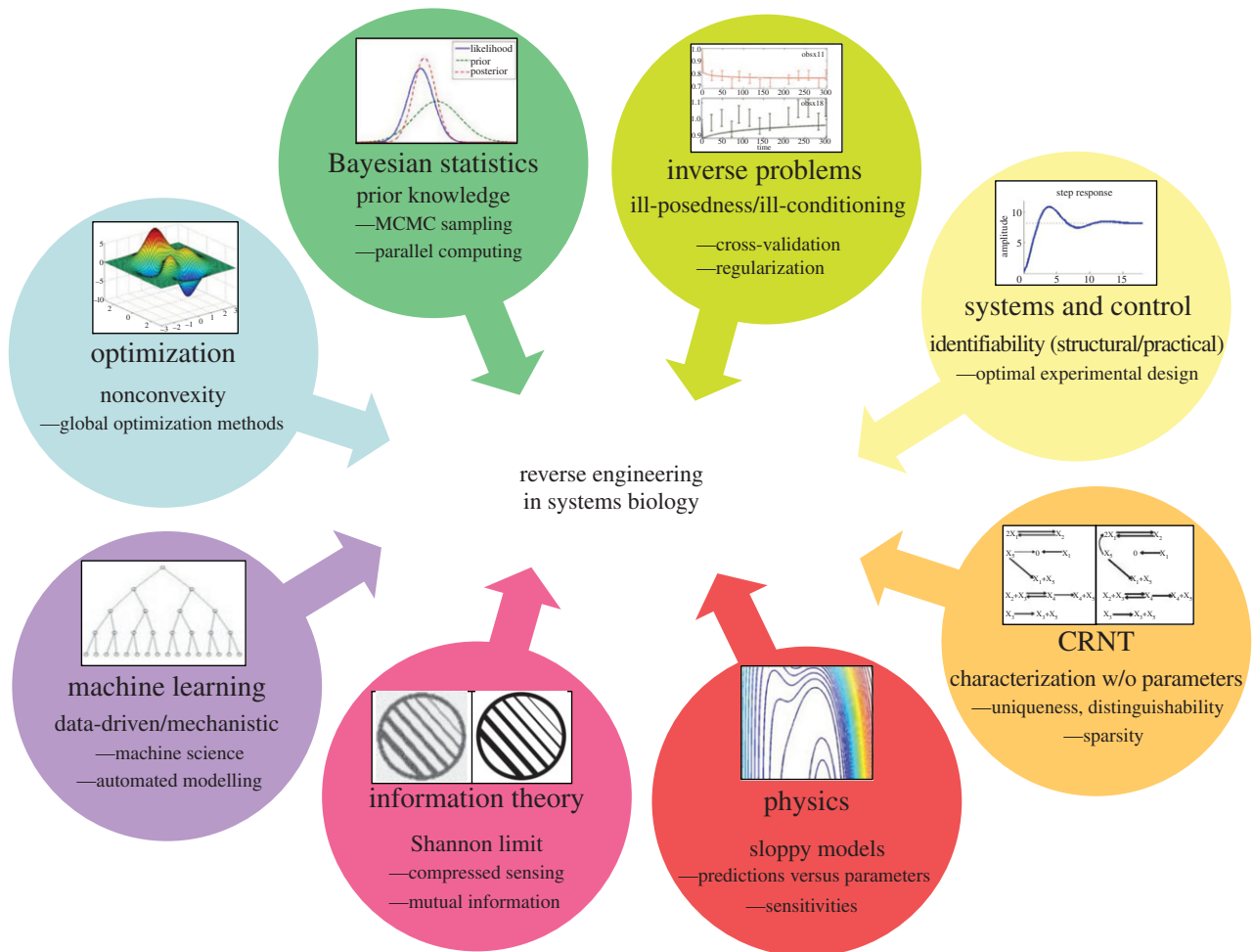
Figure 2. Perspectives on reverse engineering. An overview of the different perspectives that converge in the area of systems biology, showing some of their key concepts and tools. (Online version in colour.)

in [46,93]. It is particularly interesting to explore the conclusions of organizers of the DREAM (Dialogue for Reverse Engineering Assessment and Methods) challenge, which is probably the best current source of comparisons of different methods. The DREAM challenges take place annually and they seek to promote the interactions between theoretical and experimental methods in the area of cellular network inference and model building. In Prill *et al.* [94], the organizers state: 'The vast majority of the teams' predictions were statistically equivalent to random guesses. Moreover, even for particular problem instances like gene regulation network inference, there was no one-size-fits-all algorithm'. In other words, reliable network inference remains an unsolved problem. The organizers identify two major hurdles to be surmounted: lack of data and deficiencies in the inference algorithms. We agree with this diagnostic but, as we show below, we also think that there are other hurdles that are as important and that have been mostly ignored until recently.

The above problems are obviously not exclusive of systems biology and have been (and continue to be) studied in different areas, such as statistics, machine learning, artificial intelligence, nonlinear physics, (bio)chemical kinetics, systems and control theory, optimization (local and global), inverse problems theory, etc. (figure 2). This is a rather *ad hoc* list, because there is significant overlap between these disciplines, and some people might claim that some are simply subareas of others. However, our intention here is not to come up with a consensus classification but rather to highlight that these different areas

(or, maybe better, communities) have looked in depth at the reverse engineering problem during the last decades, arriving at several powerful principles. However, despite the interdisciplinary nature of systems biology, these different perspectives have apparently not exchanged notes to the degree that one might expect for such a general problem.

In the following, we intend to give the reader the principal components of these different perspectives. With the aim of facilitating the readability of the associated literature, the presentation of the different perspectives is ordered according to the timeline of their key developments. In particular, we want to consider the different answers to two main questions:

— Why are the problems (1–3) above so challenging?
— Which methods are available to solve them?

3.1. Perspective from inverse problems

Inverse problem theory [95,96] is a discipline that aims to find the best model to explain (at least in an approximate way) a certain set of observed data. The name comes from the fact that it is the reverse of the direct (or forward) problem, i.e. given a model and its parameters, generate predictions by solving the model. Hadamard [97] was already aware of the difficulties associated with such an exercise, and defined well-posed problems as those with the following properties:

— existence: a solution exists;
— uniqueness: the solution is unique; and

— stability: the solution's behaviour hardly changes when there is a slight change in the initial condition or parameters (the solution depends continuously on the data).

Inverse problems are often ill-posed in the sense of Hadamard. Furthermore, many problems are well-posed but ill-conditioned, meaning that the solution of the inverse problem is very sensitive to errors and noise in the data. In these situations, solving the original problem can result in overfitting, i.e. the fitted model will describe the noise instead of the underlying relationship. An overfitted model might be able to describe the data well but will have poor predictive value. This situation can be avoided by using cross-validation and/or regularization methods.

Cross-validation [98,99] tries to estimate the performance of a predictive model in practice. In its simplest form, the available data are partitioned into two subsets, using the first to solve the inverse problem, and then evaluating its predictive performance with the second subset.

Regularization tries to reduce the ill-conditioning by introducing additional information via a penalty function in the cost term to be minimized. For linear systems, Tikhonov regularization [100] is the most popular approach. For nonlinear dynamical systems, it remains an open question, although successful applications of Tikhonov-inspired schemes have been reported. Engl *et al.* [101] review these topics in the context of systems biology and present results supporting the use of sparsity-enforcing regularization. We will revisit the sparsity-enforcing concept and its consequences below.

3.2. Perspective from optimization

Identification problems are usually formulated using an optimization framework, seeking to minimize a cost function which is a metric of the distance between the predicted values and the real data. Convex optimization [102] problems have nice properties: the minimum is unique and algorithms for solving them scale up well with problem size. However, the identification of nonlinear dynamic models results in non-convex problems, which exhibit a wide range of possible pitfalls and difficulties [103] when one attempts to solve them with standard local optimization methods: convergence to local solutions, badly scaled and non-differentiable model functions, flat objective functions in the vicinity of solutions, etc. Therefore, the use of popular local methods, such as Levenberg–Marquardt or Gauss–Newton, will result in different solutions depending on the guess for the starting point in the parameter space.

It is sometimes argued that these difficulties can be avoided by using a local method in a multi-start fashion (i.e. repeated solutions of the problem starting with different guesses of the parameters). However, this folklore approach [104] is neither robust (it fails with even small problems) nor efficient (the same local optima are found repeatedly since many of the initial guesses are inside the same basins of attraction of local minima).

As a consequence, there is a need for proper non-convex (global) optimization methods [105,106]. Deterministic approaches for global optimization in dynamic systems [107,108] can guarantee the global optimality of the solution, but the associated computational effort increases very rapidly with problem size. This is a consequence of the NP-hard nature of these problems. In fact, global optimization problems are undecidable in unbounded domains [109], and NP-hard on bounded domains [110]. Therefore, based on the current status of the NP issue [111], approximate methods (such as stochastic algorithms and metaheuristics) are a more attractive alternative for problems of realistic size [112–114]. The price to pay is the lack of guarantees regarding the global optimality of the solution found. However, as the objective function to be minimized has a lower bound, which can be estimated from *a priori* considerations, obtaining a value close to that bound gives us enough indirect confidence of the near-global nature of a solution. These methods have been successfully applied to different benchmark problems with excellent results [115]. Moreover, they can be parallelized, so their application to large-scale kinetic models is feasible [116]. Further computational efficiency can be gained by following divide and conquer strategies [117].

A common question in this context is to identify the best performing method to solve a particular global optimization problem. Wolpert & Macready [118] caused quite a stir with the publication of the NFL (no free lunch) theorem. Basically, the theorem shows that if method A outperforms method B in solving a certain set of problems, then B will outperform A in a different set. Thus, considering the space of all possible optimization problems, all methods are equally efficient (so there is no free lunch in optimization). A number of misconceptions from this theorem were derived by others, including (i) the claim that there is no point in comparing metaheuristics for global optimization, as there can be no winner owing to NFL and (ii) the whole enterprise of designing global optimization methods is pointless owing to the NFL nature of optimization. What is fundamentally wrong in these claims is that the NFL theorem considers ALL possible problems in optimization, which is certainly not the case in practical applications such as parameter estimation. Furthermore, the theorem considers methods without resampling, an assumption not met by most modern metaheuristics. Finally, many modern metaheuristics exploit the problem structure to increase efficiency. For example, scatter search has proved to be a very efficient method when the local search phase is performed by a specialized local method [114,116]. Again, in these conditions, the NFL theorem does not apply.

The above does not mean in any way that global optimization problems cannot be extremely hard. It is quite easy to build a needle-in-a-haystack type of problem, which will be pathologically difficult for any algorithm, because it has no structure, and therefore requires full exploration of the search space (or a lot of luck). For this type of problem, it becomes obvious that on average, no method will perform better than pure random search, and therefore we might be tempted to assume that the NFL theorem is right after all. Fortunately, needle-in-a-haystack problems do not appear in practice, and if they do, they will very likely be the consequence of extremely poor modelling.

In summary, the NFL theorem can be regarded as one of those impossibility theorems which, although true for the general assumptions considered, do not really have major implications in a real practice framework, and therefore it offers a pessimistic view which is the consequence of its universality ('all possible problems'). This is similar to Godel's incompleteness theorems, which have not stopped advances in mathematics [119]. As we will see below with yet another impossibility theorem, the fact that our practical problems have a structure that can be exploited allows us to escape from such a pessimistic trap.

3.3. Perspective from systems and control theory

System identification theory [120,121] was developed and applied in the control engineering field with the purpose of building dynamic models of systems from measured data. This theory is well developed for linear systems, but remains as a very active research area for the nonlinear dynamic case [122].

Although the systems and control area has been primarily focused on engineered systems (mechanical, electrical and chemical), it also has a long record of applications in biology. For example, back in 1978 Bekey & Beneken [123] published a review paper on the identification of biological systems. In fact, we could also consider the pioneer contributions of Wiener [124] and Ludwig von Bertalanffy [125] as seminal examples of the interactions between biology and systems and control theory. It has been increasingly noted that these interactions can be instrumental in solving relevant problems in areas, such as medicine and biotechnology [126].

A key concept in system identification is the property of identifiability: roughly speaking, a system is identifiable if the parameters can be uniquely determined from the given input/output information (data). One can distinguish between structural [127] and practical identifiability [128]. In the structural case, identifiability is a property of the model structure (its dynamics), and the observation and stimuli (control inputs) functions (perfect measurements are assumed). In the case of practical identifiability, the property is related to the experimental data available (and their information content). Despite its importance, most modelling studies in systems biology have overlooked identifiability. Fortunately, recent literature is correcting this (e.g. [129–140]). Despite the frequent problems of lack of full identifiability, models can still be useful to predict variables of interest [141,142]. To address the issues of sparse and noisy data, Lillacci & Khammash [143] propose a combination of an extended Kalman filter (a recursive estimator well known in control engineering) with *a posteriori* identifiability tests and moment-matching optimization. The resulting approach may be used for obtaining more accurate estimates of the parameters as well as for model selection.

A closely related topic is that of optimal experimental design (OED), i.e. how we should design experiments that would result in the maximum amount of information so as to identify a model with the best possible statistical properties (which are user defined and can be related to precision, decorrelation, etc.). The advantages for efficient planning of biological experiments are obvious and have been demonstrated in real practice. For example, Bandara *et al.* [144] showed how two cycles of optimization and experimentation were enough to increase parameter identifiability very significantly. The topic of optimal design of dynamic experiments in biological systems is receiving increased attention [144–152]. Balsa *et al.* [145] presented computational procedures for OED, which was formulated as a dynamic optimization problem and solved using control vector parametrization. He *et al.* [148] compared two robust design strategies, maximin (worst-case) and Bayesian, finding a trade-off between them: while the Bayesian design led to less conservative results than the maximin, it also had a higher computational cost.

Improving the quality of parameter estimates is not the only purpose of OED; it can also be used for inferring the network topology. Tegnér *et al.* [153] proposed a reconstruction scheme where genes in the network were iteratively perturbed, selecting at each iteration the perturbation that maximized the amount of information of the experiment.

Another common application of OED is discrimination among competing models [147]. With this aim, Apgar *et al.* [129] proposed a control-based formulation, where the stimulus is designed for each candidate model so that its outputs follow a target trajectory; the quality of a model is then judged by its tracking performance. In [149], three different approaches were considered, each of which optimized initial conditions, input profiles or parameter values corresponding to structural changes in the system. Other methods have exploited sigma-point approaches [151] or Kullback–Leibler optimality [150,152].

OED with dynamic stimuli is therefore a powerful strategy to maximize the informative value of experiments while minimizing their number and associated cost. Ingolia & Weissman [154] highlight the importance of choosing the way to perturb biological systems, because it determines what characteristics of those systems can be observed and analysed, as illustrated in [155,156]. In summary, there is a need for technologies that permit a wide range of perturbations and for OED methods which can make the most out of them.

A topic that deserves special attention is the analysis of kinetic models under uncertainty. Kaltenbach *et al.* [157] offer an interesting study focused on epistemic uncertainty (lack of knowledge about the cellular networks) owing to practical limitations. These authors support the idea that the structure of these networks is more important than the fine tuning of their rate laws or parameters. As a result, methods that are based on structural properties are able to extract useful information even from partially observed and noisy systems. Kaltenbach *et al.* [157] also offer an excellent overview of methods from different areas, noting the 'cultural' differences that need to be addressed in systems biology. Vanlier *et al.* [140] provide an introduction to various methods for uncertainty analysis (focusing on parametric uncertainty). In addition to giving an overview of current methods (including frequentist and Bayesian approaches), these authors highlight how the applicability of each type of method is linked to the properties of the system considered and the assumptions made by the modeller. This type of study is of great interest as it provides system biologists with a balanced view of the requirements and results that are expected in each method. Ensemble modelling is a particularly interesting type of Monte Carlo methodology that has been used to account for uncertainty in many areas, from weather forecasting to machine learning. Applications in systems biology have already appeared [158,159]. Another related successful approach for robust inference is the wisdom of crowds [160].

Finally, advances in the identification of biological systems ultimately lead to their control [8], and here the possibilities are enormous, especially in synthetic biology [161–167].

3.4. Perspective from chemical reaction network theory

The fundamentals of chemical reaction network theory (CRNT) were established back in the 1970s by Horn, Jackson and Feinberg [168–170]. The theory remained rather dormant until authors like Bailey [171] highlighted its potential for the

analysis of biological networks. The basic idea is that, using CRNT, we can characterize kinetic models (multi-stability, oscillations, etc.) without knowing the precise values of the kinetic parameters. During the last decade, research based on CRNT has gained momentum [172–178] leading to major contributions [179].

Regarding the identification of biological systems, CRNT offers several results of considerable importance. Craciun & Pantea [180] make use of CRNT to show that, given a (mass action) reaction network and its dynamic equations (ODEs), it might be impossible to identify its rate constants uniquely (even with perfect measurements of all species). Furthermore, they also show that, given the dynamics, it might be impossible to identify the reaction network uniquely.

Szederkenyi et al. [181] make use of CRNT principles to explore inherent limitations in the inference of biological networks. Their results show that, in addition to the obstacles identified by Prill et al. [94] (lack of data and deficiencies in the inference algorithms), we must be also aware of fundamental problems related to the uniqueness and distinguishability of these networks (even for the utopian case of fully observed networks with no noise). More importantly, uniqueness and distinguishability of models can be guaranteed by carefully adding extra constraints and/or prior knowledge. A topic that deserves further investigation is the effect of imposing a sparse network topology. Data from cellular networks suggest such a sparse topology, so it is a common prior enforced in many inference methods [182,183]. However, Szederkenyi et al. [181] show that the sparsity assumption alone is not enough to ensure uniqueness. Moreover, in the case of linear dynamic genetic network models, too sparse structures can be harmful.

3.5. Perspective from Bayesian statistics

As previously mentioned, the origin of the Bayesian approach goes back to the eighteenth century, and the statistical methods used during the nineteenth century were mostly Bayesian as well. However, during the twentieth century the frequentist paradigm clearly dominated statistics [184]. Frequentism was the default approach used for estimation and inference of kinetic (dynamic) models, where most studies (cited in the previous subsections) considered maximum-likelihood and related metrics as the cost functions to optimize. However, fuelled by important developments in MCMC methods in the 1990s, the beginning of the twenty-first century witnessed a Bayes revival, and studies on Bayesian methods for dynamic models started to appear as a result of theoretical and computational advances and the greater availability of more powerful computers. In parallel, systems biology was taking off with the new century, requiring methods that were able to handle the biological complexity. Bayesian methods, which are especially useful to extract information from uncertain and noisy data (the most common scenario in bioinformatics and computational systems biology), started to receive greater attention [42,44,185]. Bayesian estimation in stochastic kinetic models was considered in several seminal works regarding diffusion models [186,187]. Similarly, in the case of deterministic kinetic models, the last decade has seen a rapidly growing Bayesian literature. Pioneering work using Monte Carlo methods were presented by Battogtokh et al. [188] and Brown & Sethna [189]. Sanguinetti et al. [190], considering a discrete time state space model, presented a Bayesian method for genome-wide

quantitative reconstruction of transcriptional regulation. Girolami [191] illustrated the use of the Bayesian framework to systematically characterize uncertainty in models based on ordinary differential equations. Vyshemirsky & Girolami [192] compared four methods for estimating marginal likelihoods, investigating how they affect the Bayes factor estimates, which are used for kinetic model ranking and selection.

When the formulation of a likelihood function is difficult or impossible, ABC-like approaches can be adopted [85]. ABC schemes replace the evaluation of the likelihood function with a measure of the distance between the observed and simulated data. Briefly, ABC algorithms sample a parameter vector from the distribution and use it for generating a simulated dataset. Then, they calculate the distance between this dataset and the experimental data, and if it is below a certain threshold they accept the candidate parameter vector. The weakness of this approach, at least in its simplest form, is that it can have a low acceptance rate when the prior and posterior are very different. To overcome this problem, Marjoram et al. [193] presented a MCMC algorithm (ABC MCMC) that accepts observations more frequently and does not require the computation of likelihoods. The price to pay is the generation of dependent outcomes, and the risk of getting stuck in regions of low probability of the state space for long periods of time. An alternative is to use sequential Monte Carlo (SMC) techniques, which use an ensemble of particles to represent the posterior density, with each sample having a weight that represents its probability of being sampled. SMC particles are uncorrelated, and the approach avoids being stuck in low probability regions. Sisson et al. [194] proposed a likelihood-free ABC sampler based on SMC simulation (ABC SMC) and a related formulation was proposed by Toni et al. [195,196], who applied it for parameter estimation and model selection in several biological systems.

ABC schemes can also be used to improve computational efficiency, which is an important issue in Bayesian approaches. Using the full probability distribution of parameters instead of single estimates of parameter values entails calculating the likelihood across the whole parameter space, a step that can be very costly.

The availability of these theoretical and computational advances has led to their successful application in combination with biological experimentation. For example, Xu et al. [197] considered the ERK cell signalling pathway and found unexpected new results of biological significance, demonstrating the capability of Bayesian approaches to infer pathway topologies in practical applications, even when measurements are noisy and limited. In another recent application, Eydgahi et al. [198] used Bayes factor analysis to discriminate between two alternative kinetic models of apoptosis. It is interesting to note that the approach allowed these authors to assign a much greater plausibility to one of the models even though both presented equally good fits to data. Moreover, it is also remarkable that, despite non-identifiability of the models, the Bayesian approach resulted in predictions with small confidence intervals. Regarding experimental design, Liepe et al. [199] illustrated the combination of Bayesian inference with information theory to design experiments with maximum information content and applied it to three different problems.

Recently, Raue et al. [200] presented an interesting study combining the frequentist and the Bayesian approaches. These authors note that for kinetic models with lack of identifiability (structural and/or practical), the Markov chain in

MCMC-based Bayesian methods cannot ensure convergence and will result in inaccurate results. To surmount this, they suggest a two-step procedure. In the first step, a frequentist profile-likelihood approach is used in iterative combination with experimental design until the identifiability problems are solved. Then, in the second step, the MCMC approach can be used reliably.

Another important question concerns the scalability of Bayesian approaches, i.e. can they handle large-scale kinetic models? In a recent contribution, Hug *et al.* [201] discuss the conceptual and computational issues of Bayesian estimation in high-dimensional parameter spaces and present a multi-chain sampling method to address them. The feasibility and efficiency of the method is illustrated with a signal transduction model with more than 100 parameters. The study is a significant proof of principle and also a good example of the care that must be taken regarding the verification of results.

The existing literature indicates the importance of adequate selection of priors. Gaussian processes (GPs) can be used for specifying a prior directly over the function space, which is often simpler than over the parameter space. A GP [202] is a stochastic process for which any set of variables have a joint multi-variate Gaussian distribution. Gaussian processes are generalizations of Gaussian probability distributions: they describe the properties of functions rather than of scalars or vectors. They have also been applied in the development of efficient and reliable sampling schemes. Here, Calderhead *et al.* [203] illustrated how GPs can be used to greatly accelerate Bayesian inference in nonlinear dynamic models. Other notable recent advances in sampling methods have been presented by Girolami & Calderhead [204,205] and Schmidl *et al.* [206].

3.6. Perspective from physics

Physics has made numerous and highly relevant contributions to inference and mathematical modelling in general. In fact, the origins of many of the ideas classified in the sections above can be traced back to developments in physics. Therefore, our intention here is not to present any type of overview of such vast history.

Rather, we will focus on recent research that has spurred a broad discussion about whether there exist fundamental limitations regarding dynamic modelling of biological systems. Gutenkunst *et al.* [207] discuss the concept of sloppy models (introduced by Brown & Sethna [189]), i.e. multi-parametric models whose behaviour (and predictions) depends only on a few combinations of parameters, with many other sloppy parameter directions which are basically unimportant. These authors tested a collection of 17 systems biology models and concluded that (i) sloppiness is universal in systems biology models and (ii) sloppy parameter sensitivities help to explain the difficulty of extracting precise parameter estimates from collective fits, even from comprehensive data. The previous study by Brown & Sethna [189] presents a sound theoretical analysis based on statistical thermodynamics and Bayesian inference.

This work has received great attention from the systems biology community. Here, we would like to highlight some open questions to be addressed, and comment on possible misconceptions surrounding it. Some of our remarks below can also be found in the correspondence by Apgar *et al.* [208] and related comments [209,210].

Although the work of Gutenkunst *et al.* [207] is a valuable contribution which nicely illustrates the difficulties that plague parameter estimation problems in dynamic models, we believe that:

(i) links between sloppiness and previous works on identifiability (which are not cited) should have been established. Our own biased opinion is that identifiability is probably a better framework to analyse the above-mentioned challenges. To begin with, sloppiness seems to lump together lack of structural and practical identifiability. However, structural problems can be addressed by model reformulation or reduction. Practical identifiability problems can be surmounted by more informative data and, ideally, by OED (see related comments by Apgar *et al.* [208]). Consequently, identifiablity seems to be a more powerful concept in the sense that it also provides us with guidelines on how to improve it;

(ii) as model sloppiness can be reduced by the above strategies, it is not a universal property in systems biology. See also Apgar *et al.* [208] for more on this. To be fair, Gutenkunst *et al.* [207] clearly state that 'universal' has a technical meaning from statistical physics (a shared property with a deep underlying cause), so universality in this sense does not imply that all models must necessarily share the property. But from conversations with many colleagues, it seems that the latter incorrect meaning has been often assumed; and

(iii) related with (i), the study by Brown & Sethna [189] concludes that sloppiness is not a result of lack of data. Does this imply that it is only related to lack of structural identifiability? We do not get this impression, because e.g. in the study by Apgar *et al.* [208] and related comments [209,210], the issues discussed seem to be only related to practical identifiability.

In a way, sloppiness has created a somewhat pessimistic view towards parameter estimation in dynamic modelling, similar to the one created by the NFL theorem in optimization as described above. But, in this case, there is no theorem, and there are ways to surmount structural and practical identifiability problems (model reformulaton and/or better experiments can indeed lead to good parameter estimations). We believe that an integrative study of sloppiness and identifiability would be very valuable. We also believe that, in many situations, the lack of informative data is the source of such lack of identifiability, because most biological systems of interest are only partially observed and current measurement technologies often result in large errors. However, advances in such technologies, coupled with new ways of introducing perturbations and the use of OED methods, should lead to identifiable dynamic models [208,211], so we should be optimistic about the calibration of these models.

3.7. Perspective from information theory

Information theory was initiated by the work of Shannon [61], who was interested in finding fundamental limits to signal-processing processes in communication and compression of data. The so-called sampling theorem (frequently attributed to Shannon & Nyquist [212]) is one of such

fundamental results: a signal's information is preserved if it is uniformly sampled at a rate at least two times faster than its Fourier bandwidth (higher frequency). Or, in other words, a time-varying signal with no frequencies higher than N hertz can be perfectly reconstructed by sampling the signal at regular intervals of $1/(2N)$ seconds. Therefore, if we do not have a sampling above this threshold, we cannot recover exactly the original signal.

We find again a theorem that establishes a fundamental constraint on what we can infer from data. Once more, this seems to be another case of a pessimistic view that can be avoided if we exploit other information about the system. Indeed, recent work [213–215] showed how sparsity patterns can be used to perfectly reconstruct signals with sampling rates below the Shannon limit. These works have created the burgeoning new field of compressed (or compressive) sensing, which has already seen the publication of a large number of works, not only regarding the methodology and its extensions but also applications. In the case of biological data, they have been successfully applied in bioinformatics [216]. Very recently, Pan *et al.* [217] presented a very interesting compressive sensing approach for the reverse engineering of biochemical networks, assuming fully observed networks. A key question remains open: can we apply this framework to partially observed networks?

3.8. Perspective from machine learning

Machine learning, generally considered a subfield of artificial intelligence, aims to build systems (usually programs running on computers) that can learn from data and act according to requirements. In other words, it is based on data-driven approaches where the systems learn from experience (data). Machine learning methods have been widely used in bioinformatics [218,219] and computational and systems biology [220,221]. Traditionally, the data-driven models used in machine learning have been considered on the other side of the spectrum from mechanistic models. However, recent machine learning advances have made it possible to somehow link both and consider the automatic generation of mechanistic models via data-driven methods. In this line, Kell & Oliver [222] argue that data-driven approaches should be regarded as complementary to the more traditional hypothesis-driven programmes.

During the last decade, several studies have been presented which examine the full automation of reverse engineering, from hypothesis generation to experiments and back, in what has been termed machine science [223]. A prominent example is the robot scientist developed by King *et al.* [224–226] and its applications in functional genomics, illustrating how a machine can discover novel scientific knowledge in a fully automatic manner.

An automated process for reverse engineering of nonlinear dynamical systems was presented by Bongard & Lipson [227], illustrating how the method could be used for automated modelling in systems biology, including the automatic generation of testable hypotheses. More recently, Schmidt & Lipson [228] presented an approach to automatically generate free-form natural laws from experimental data. Despite these success stories, the large-scale and partially observed nature of most biological systems will undoubtedly pose major challenges for the widespread application of these procedures in the laboratory.

4. Conclusion: lessons from converging perspectives

Reverse engineering can help us to infer, understand and analyse the mechanisms of biological systems. In this sense, modelling is a systematic way to efficiently encapsulate our current knowledge of these systems. However, the value of models can (and should) go beyond their explanatory value: they can be used to make predictions, and also to suggest new questions and hypotheses that can be tested experimentally. Systems biology will succeed if the practical value of theory is realized [5].

The above perspectives from different areas clearly show overlaps and convergent ideas. For example, the ill-posed nature of many identification problems, as described in inverse problem theory, has obvious parallels in optimization (multi-modality, flatness of cost functions), systems identification (lack of identifiability) or CRNT (non-uniqueness). Similarly, some regularization techniques can be regarded as Bayesian approaches where certain prior distributions are enforced. Other overlaps and synergies are not so obvious (e.g. the role of sparsity in inference) and will require careful study.

Several basic lessons can be extracted from the different perspectives that we have briefly reviewed. The first lesson is that modelling should start with questions associated with the intended use. These questions will also help us to choose the level of description that must be selected [229]. We should focus on making the right questions, even if we can only give approximate answers to them (an exact answer to the wrong question is of little use [230]).

The second lesson is that these reverse engineering problems are extremely challenging, so pessimistic views are understandable (e.g. Brenner [231] thinks that they are not solvable). But, as nicely argued by Noble [232], the history of science contains many incorrect claims to impossibility. In fact, we have seen in previous sections that the existence of several pessimistic theorems has not precluded advances in related areas. Brenner [231] cites an article [233] on inverse problems to justify his skepticism. In that work, Tarantola [233] comments on the difficulties that plague inverse problems in geophysics, concluding that observations should not be used to deduce a particular solution but to falsify possible solutions. In our opinion, even if this holds for inverse problems in systems biology (which is questionable), it does not mean that we are doomed to fail (Popper considers science as falsification, and Tarantola's view builds on that). Besides, fortunately, optimistic views are also present in the community, and modern statistical methods are here to help (e.g., see the excellent preface in the book by Stumpf *et al.* [185]).

In this context, it is also worth mentioning that, as described by Silver [41], modelling and simulation have been very successful in some areas (notably, short-term weather forecasts), but have failed dramatically in others (e.g. earthquake predictions). Many decades of research have been invested in both topics. However, in the case of weather, we have better data and a deeper knowledge of the physico-chemical mechanisms involved. The atmosphere and its boundaries are much easier to explore than the tensions and displacements underground. The optimistic system biologist will rejoice in the description of systems biology as cellular weather forecasting [5]. But we should also bear in mind that it took many years to develop

the theoretical and computational methods behind current weather models.

The third lesson is that approximate methods can give us rather good solutions to many of these hard problems. Notably, we have seen how randomized algorithms of several types (e.g. stochastic methods for global optimization, or MCMC sampling methods in Bayesian inference) can produce good results in reasonable computation times. Needless to say, this does not mean that deterministic algorithms should be abandoned (e.g. in global optimization, they are making good progress). Rather, it will be very interesting to see how hybrids between deterministic and stochastic methods result in techniques that scale up well with problem size.

The fourth lesson is that, although the Bayesian versus frequentism controversy continues [234], Bayesian methods are probably better suited for many of the inference problems in systems biology. Stumpf *et al.* [185] mention the difficulties of classical statistics with an area that is data-rich but also hypotheses-rich. Incidentally, it is interesting to note that Lindley [235] predicted a Bayesian twenty-first century. However, as it has happened in other areas of science, joining forces might be a strategy worth exploring, as recently illustrated by Raue *et al.* [200].

The fifth lesson is that we need to establish links between identifiability, as developed in the systems and control area, and related concepts developed in other fields, such as sloppiness [207]. The Bayesian view will also help in establishing the practical limits for reverse engineering of kinetic models [236].

The sixth lesson is that we need to exploit the structure of dynamic models. In addition to CRNT, Kaltenbach *et al.* [157] also mention the theory of monotone systems [237] as a promising avenue, highlighting the need of further research

to be able to apply these theories to biological networks of realistic complexity.

A final seventh lesson is that, although systems biology is a truly interdisciplinary area, we need to coordinate more efforts and exchange more notes. Different communities have developed theories and tools that have major implications for the identification and reverse engineering of biological systems, but in many cases they have been doing so in isolation from each other. There are several notable examples where collaborations have been very successful, such as SBML [238], BioModels Database [239] or the DREAM challenges [94]. As indicated by Kitano [240], international alliances for quantitative modelling in systems biology might be needed. Whole-cell models will require robust and scalable inference and estimation methods. Much reverse engineering work lies ahead.

Endnote

[1]This distinction between the reduced problem of recovering interactions and the general reverse engineering problem is commonly made. For example, the challenges designed in the DREAM initiative belong to one of these two classes [29]. This entails that constraint-based approaches, which are intermediate between the dynamic and interaction-based models, will not be explicitly considered here. We refer the interested reader to the existing reviews on the subject [22,30–32].

References

1. Mesarović MD. 1968 Systems theory and biology—view of a theoretician. In *Systems Theory and Biology: Proc., III Systems Symp. at Case Institute of Technology*, pp. 59–87. New York, NY: Springer. (doi:10.1007/978-3-642-88343-9_3)

2. Csete ME, Doyle JC. 2002 Reverse engineering of biological complexity. *Science* **295**, 1664–1669. (doi:10.1126/science.1069981)

3. Alon U. 2003 Biological networks: the tinkerer as an engineer. *Science* **301**, 1866–1867. (doi:10.1126/science.1089072)

4. Stelling J, Klamt S, Bettenbrock K, Schuster S, Gilles ED. 2002 Metabolic network structure determines key aspects of functionality and regulation. *Nature* **420**, 190–193. (doi:10.1038/nature01166)

5. Wolkenhauer O, Mesarovic' M. 2005 Feedback dynamics and cell function: why systems biology is called systems biology. *Mol. Biosyst.* **1**, 14–16. (doi:10.1039/b502088n)

6. Doyle FJ, Stelling J. 2006 Systems interface biology. *J. R. Soc. Interface* **3**, 603–616. (doi:10.1098/rsif.2006.0143)

7. Kremling A, Saez-Rodriguez J. 2007 Systems biology—an engineering perspective. *J. Biotechnol.* **129**, 329–351. (doi:10.1016/j.jbiotec.2007.02.009)

8. Sontag ED. 2004 Some new directions in control theory inspired by systems biology. *Syst. Biol.* **1**, 9–18. (doi:10.1049/sb:20045006)

9. Khammash M, El-Samad H. 2004 Systems biology: from physiology to gene regulation. *IEEE Contr. Syst.* **24**, 62–76. (doi:10.1109/MCS.2004.1316654)

10. Karr JR, Sanghvi JC, Macklin DN, Gutschow MV, Jacobs JM, Bolival B, Assad-Garcia N, Glass JI, Covert MW. 2012 A whole-cell computational model predicts phenotype from genotype. *Cell* **150**, 389–401. (doi:10.1016/j.cell.2012.05.044)

11. Coveney PV, Fowler PW. 2005 Modelling biological complexity: a physical scientist's perspective. *J. R. Soc. Interface* **2**, 267–280. (doi:10.1098/rsif.2005.0045)

12. Noble D. 2002 Modeling the heart—from genes to cells to the whole organ. *Science* **295**, 1678–1682. (doi:10.1126/science.1069881)

13. Dada JO, Mendes P. 2011 Multi-scale modelling and simulation in systems biology. *Integr. Biol.* **3**, 86–96. (doi:10.1039/c0ib00075b)

14. Di Ventura B, Lemerle C, Michalodimitrakis K, Serrano L. 2006 From *in vivo* to *in silico* biology and back. *Nature* **443**, 527–533. (doi:10.1038/nature05127)

15. Jaeger J, Crombach A. 2012 Life's attractors. In (ed. OS Soyer) *Evolutionary systems biology*, pp. 93–119. Berlin, Germany: Springer.

16. Crombach A, Wotton KR, Cicin-Sain D, Ashyraliyev M, Jaeger J. 2012 Efficient reverse-engineering of a developmental gene regulatory network. *PLoS Comput. Biol.* **8**, e1002589. (doi:10.1371/journal.pcbi.1002589)

17. Marchisio MA, Stelling J. 2011 Automatic design of digital synthetic gene circuits. *PLoS Comput. Biol.* **7**, e1001083. (doi:10.1371/journal.pcbi.1001083)

18. Stelling J. 2004 Mathematical models in microbial systems biology. *Curr. Opin. Microbiol.* **7**, 513–518. (doi:10.1016/j.mib.2004.08.004)

19. Bader JS, Chaudhuri A, Rothberg JM, Chant J. 2003 Gaining confidence in high-throughput protein interaction networks. *Nat. Biotechnol.* **22**, 78–85. (doi:10.1038/nbt924)

20. Sharan R, Ideker T. 2006 Modeling cellular machinery through biological network comparison. *Nat. Biotechnol.* **24**, 427–433. (doi:10.1038/nbt1196)

21. Förster J, Famili I, Fu P, Palsson BØ, Nielsen J. 2003 Genome-scale reconstruction of the *Saccharomyces cerevisiae* metabolic network. *Genome Res.* **13**, 244–253. (doi:10.1101/gr.234503)

22. Llaneras F, Picó J. 2008 Stoichiometric modelling of cell metabolism. *J. Biosci. Bioeng.* **105**, 1–11. (doi:10.1263/jbb.105.1)

23. Machado D, Costa RS, Rocha M, Ferreira EC, Tidor B, Rocha I. 2011 Modeling formalisms in systems biology. *AMB Expr.* **1**, 1–14. (doi:10.1186/2191-0855-1-45)

24. Tenazinha N, Vinga S. 2011 A survey on methods for modeling and analyzing integrated biological networks. *IEEE-ACM Trans. Comput. Biol. Inform.* **8**, 943–958. (doi:10.1109/TCBB.2010.117)

25. Kholodenko BN. 2006 Cell-signalling dynamics in time and space. *Nat. Rev. Mol. Cell Biol.* **7**, 165–176. (doi:10.1038/nrm1838)

26. Novák B, Tyson JJ. 2008 Design principles of biochemical oscillators. *Nat. Rev. Mol. Cell Biol.* **9**, 981–991. (doi:10.1038/nrm2530)

27. Tyson JJ, Chen KC, Novak B. 2003 Sniffers, buzzers, toggles and blinkers: dynamics of regulatory and signaling pathways in the cell. *Curr. Opin. Cell Biol.* **15**, 221–231. (doi:10.1016/S0955-0674(03)00017-6)

28. Aldridge BB, Burke JM, Lauffenburger DA, Sorger PK. 2006 Physicochemical modelling of cell signalling pathways. *Nat. Cell Biol.* **8**, 1195–1203. (doi:10.1038/ncb1497)

29. Marbach D, Prill R, Schaffter T, Mattiussi C, Floreano D, Stolovitzky G. 2010 Revealing strengths and weaknesses of methods for gene network inference. *Proc. Natl Acad. Sci. USA* **107**, 6286–6291. (doi:10.1073/pnas.0913357107)

30. Price ND, Papin JA, Schilling CH, Palsson BO. 2003 Genome-scale microbial *in silico* models: the constraints-based approach. *Trends Biotechnol.* **21**, 162–169. (doi:10.1016/S0167-7799(03)00030-1)

31. Orth JD, Thiele I, Palsson BØ. 2010 What is flux balance analysis? *Nat. Biotechnol.* **28**, 245–248. (doi:10.1038/nbt.1614)

32. Hyduke DR, Lewis NE, Palsson BØ. 2013 Analysis of omics data with genome-scale models of metabolism. *Mol. BioSyst.* **9**, 167–174. (doi:10.1039/c2mb25453k)

33. Bayarri MJ, Berger JO. 2004 The interplay of Bayesian and frequentist analysis. *Stat. Sci.* **19**, 58–80. (doi:10.1214/088342304000000116)

34. Cox DR. 2006 *Principles of statistical inference.* Cambridge, UK: Cambridge University Press.

35. McGrayne SB. 2011 *The theory that would not die: how Bayes rule cracked the enigma code, hunted down Russian submarines, and emerged triumphant from two centuries of controversy.* New Haven, CT: Yale University Press.

36. Carlin BP, Louis TA. 1997 Bayes and empirical Bayes methods for data analysis. *Stat. Comput.* **7**, 153–154. (doi:10.1023/A:1018577817064)

37. Geman S, Geman D. 1984 Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Trans. Pattern Anal. Mach. Intell.* **6**, 721–741. (doi:10.1109/TPAMI.1984.4767596)

38. Gelfand AE, Smith AF. 1990 Sampling-based approaches to calculating marginal densities. *J. Amer. Stat. Assoc.* **85**, 398–409. (doi:10.1080/01621459.1990.10476213)

39. Gilks WR, Richardson S, Spiegelhalter DJ. 1996 *Markov chain Monte Carlo in practice*, vol. 2. Boca Raton, FL: CRC press.

40. Smith AF, Roberts GO. 1993 Bayesian computation via the Gibbs sampler and related Markov chain Monte Carlo methods. *J. R. Stat. Soc. B* **55**, 3–23.

41. Silver N. 2012 *The signal and the noise: why so many predictions fail—but some don't.* Baltimore, MD: Penguin Press.

42. Wilkinson DJ. 2007 Bayesian methods in bioinformatics and computational systems biology. *Brief. Bioinform.* **8**, 109–116. (doi:10.1093/bib/bbm007)

43. Needham CJ, Bradford JR, Bulpitt AJ, Westhead DR. 2007 A primer on learning in Bayesian networks for computational biology. *PLoS Comp. Biol.* **3**, e129. (doi:10.1371/journal.pcbi.0030129)

44. Lawrence ND, Girolami M, Rattray M, Sanguinetti G. 2010 *Learning and inference in computational systems biology.* Cambridge, UK: MIT Press.

45. D'Haeseleer P, Liang S, Somogyi R. 2000 Genetic network inference: from co-expression clustering to reverse engineering. *Bioinformatics* **16**, 707–726. (doi:10.1093/bioinformatics/16.8.707)

46. Bansal M, Belcastro V, Ambesi-Impiombato A, Di Bernardo D. 2007 How to infer gene networks from expression profiles. *Mol. Syst. Biol.* **3**, 122.

47. De Smet R, Marchal K. 2010 Advantages and limitations of current network inference methods. *Nat. Rev. Microbiol.* **8**, 717–729.

48. Penfold CA, Wild DL. 2011 How to infer gene networks from expression profiles, revisited. *Interface Focus* **1**, 857–870. (doi:10.1098/rsfs.2011.0053)

49. López-Kleine L, Leal L, López C. 2013 Biostatistical approaches for the reconstruction of gene co-expression networks based on transcriptomic data. *Brief. Funct. Genomics* **12**, 457–467. (doi:10.1093/bfgp/elt003)

50. Friedman N. 2004 Inferring cellular networks using probabilistic graphical models. *Science* **303**, 799–805. (doi:10.1126/science.1094068)

51. Markowetz F, Spang R. 2007 Inferring cellular networks—a review. *BMC Bioinform.* **8**(Suppl. 6), S5. (doi:10.1186/1471-2105-8-S6-S5)

52. Villaverde AF, Ross J, Banga JR. 2013 Reverse engineering cellular networks with information theoretic methods. *Cells* **2**, 306–329. (doi:10.3390/cells2020306)

53. Butte A, Tamayo P, Slonim D, Golub T, Kohane I. 2000 Discovering functional relationships between RNA expression and chemotherapeutic susceptibility using relevance networks. *Proc. Natl Acad. Sci. USA* **97**, 12 182–12 186. (doi:10.1073/pnas.220392197)

54. Stuart J, Segal E, Koller D, Kim S. 2003 A gene-coexpression network for global discovery of conserved genetic modules. *Science* **302**, 249–255. (doi:10.1126/science.1087447)

55. Arkin A, Ross J. 1995 Statistical construction of chemical reaction mechanisms from measured time-series. *J. Phys. Chem.* **99**, 970–979. (doi:10.1021/j100003a020)

56. Arkin A, Shen P, Ross J. 1997 A test case of correlation metric construction of a reaction

57. pathway from measurements. *Science* **277**, 1275–1279. (doi:10.1126/science.277.5330.1275)

57. Székely G, Rizzo M, Bakirov N. 2007 Measuring and testing dependence by correlation of distances. *Ann. Stat.* **35**, 2769–2794. (doi:10.1214/009053607000000505)

58. Szekely G, Rizzo M. 2009 Brownian distance correlation. *Ann. Appl. Stat.* **3**, 1236–1265. (doi:10.1214/09-AOAS312)

59. Roy A, Post C. 2012 Detection of long-range concerted motions in protein by a distance covariance. *J. Chem. Theory Comput.* **8**, 3009–3014. (doi:10.1021/ct300565f)

60. Kong J, Klein B, Klein R, Lee K, Wahba G. 2012 Using distance correlation and SS-ANOVA to assess associations of familial relationships, lifestyle factors, diseases, and mortality. *Proc. Natl Acad. Sci. USA* **109**, 20 352–20 357. (doi:10.1073/pnas.1217269109)

61. Shannon C. 1948 A mathematical theory of communication. *Bell Syst. Tech. J.* **27**, 379–423. (doi:10.1002/j.1538-7305.1948.tb01338.x)

62. Cover T, Thomas J. 1991 *Elements of information theory.* New York, NY: Wiley.

63. Farber R, Lapedes A, Sirotkin K. 1992 Determination of eukaryotic protein coding regions using neural networks and information theory. *J. Mol. Biol.* **226**, 471–479. (doi:10.1016/0022-2836(92)90961-I)

64. Korber B, Farber R, Wolpert D, Lapedes A. 1993 Covariation of mutations in the v3 loop of human immunodeficiency virus type 1 envelope protein: an information theoretic analysis. *Proc. Natl Acad. Sci. USA* **90**, 7176–7180. (doi:10.1073/pnas.90.15.7176)

65. Liang S, Fuhrman S, Somogyi R. 1998 Reveal, a general reverse engineering algorithm for inference of genetic network architectures. *Pac. Symp. Biocomput.* **3**, 18–29.

66. Michaels G, Carr D, Askenazi M, Fuhrman S, Wen X, Somogyi R. 1998 Cluster analysis and data visualization of large scale gene expression data. *Pac. Symp. Biocomput.* **3**, 42–53.

67. Butte A, Kohane I. 2000 Mutual information relevance networks: functional genomic clustering using pairwise entropy measurements. *Pac. Symp. Biocomput.* **5**, 418–429.

68. Margolin A, Wang K, Lim W, Kustagi M, Nemenman I, Califano A. 2006 Reverse engineering cellular networks. *Nat. Protoc.* **1**, 662–671. (doi:10.1038/nprot.2006.106)

69. Zoppoli P, Morganella S, Ceccarelli M. 2010 TimeDelay-ARACNE: reverse engineering of gene networks from time-course data by an information theoretic approach. *BMC Bioinform.* **11**, 154. (doi:10.1186/1471-2105-11-154)

70. Steuer R, Kurths J, Daub C, Weise J, Selbig J. 2002 The mutual information: detecting and evaluating dependencies between variables. *Bioinformatics* **18**(Suppl. 2), S231–S240. (doi:10.1093/bioinformatics/18.suppl_2.S231)

71. Cellucci C, Albano A, Rapp P. 2005 Statistical validation of mutual information calculations: comparison of alternative numerical algorithms.

Phys. Rev. E **71**, 066208. (doi:10.1103/PhysRevE. 71.066208)

72. Moon Y, Rajagopalan B, Lall U. 1995 Estimation of mutual information using kernel density estimators. *Phys. Rev. E* **52**, 2318–2321. (doi:10.1103/ PhysRevE.52.2318)

73. de Matos Simoes R, Emmert-Streib F. 2011 Influence of statistical estimators of mutual information and data heterogeneity on the inference of gene regulatory networks. *PLoS ONE* **6**, e29279. (doi:10.1371/journal.pone.0029279)

74. Meyer PE, Kontos K, Lafitte F, Bontempi G. 2007 Information-theoretic inference of large transcriptional regulatory networks. *EURASIP J. Bioinform. Syst. Biol.* **2007**, 79879. (doi:10.1155/ 2007/79879)

75. Luo W, Hankenson K, Woolf P. 2008 Learning transcriptional regulatory networks from high throughput gene expression data using continuous three-way mutual information. *BMC Bioinform.* **9**, 467. (doi:10.1186/1471-2105-9-467)

76. Samoilov M. 1997 Reconstruction and functional analysis of general chemical reactions and reaction networks. PhD thesis, Stanford University, CA, USA.

77. Faith J, Hayete B, Thaden J, Mogno I, Wierzbowski J, Cottarel G, Kasif S, Collins J, Gardner T. 2007 Large-scale mapping and validation of *Escherichia coli* transcriptional regulation from a compendium of expression profiles. *PLoS Biol.* **5**, e8. (doi:10.1371/ journal.pbio.0050008)

78. Bonneau R, Reiss D, Shannon P, Facciotti M, Hood L, Baliga N, Thorsson V. 2006 The inferelator: an algorithm for learning parsimonious regulatory networks from systems-biology data sets de novo. *Genome Biol.* **7**, R36. (doi:10.1186/ gb-2006-7-5-r36)

79. Greenfield A, Madar A, Ostrer H, Bonneau R. 2010 DREAM4: combining genetic and dynamic information to identify biological networks and dynamical models. *PLoS ONE* **5**, e13397. (doi:10.1371/journal.pone.0013397)

80. Reshef D, Reshef Y, Finucane H, Grossman S, McVean G, Turnbaugh P, Lander E, Mitzenmacher M, Sabeti P. 2011 Detecting novel associations in large data sets. *Science* **334**, 1518–1524. (doi:10. 1126/science.1205438)

81. Cantone I *et al*. 2009 A yeast synthetic network for *in vivo* assessment of reverse-engineering and modeling approaches. *Cell* **137**, 172–181. (doi:10. 1016/j.cell.2009.01.055)

82. Villaverde AF, Ross J, Morán F, Banga JR. 2013 Mider: network inference with mutual information distance and entropy reduction. See http://www. iim.csic.es/gingproc/mider.html.

83. Larranaga P, Inza I, Flores JL. 2005 A guide to the literature on inferring genetic networks by probabilistic graphical models. In *Data analysis and visualization in genomics and proteomics* (eds F Azuaje, J Dopazo), pp. 215. New York, NY: Wiley.

84. Chickering D. 1996 Learning Bayesian networks is NP-complete. In *Learning from data: artificial intelligence and statistics* (eds D Fisher, H-J Lenz), pp. 121–130. New York, NY: Springer.

85. Beaumont MA, Zhang W, Balding DJ. 2002 Approximate Bayesian computation in population genetics. *Genetics* **162**, 2025–2035.

86. Toni T, Stumpf MP. 2010 Simulation-based model selection for dynamical systems in systems and population biology. *Bioinformatics* **26**, 104–110. (doi:10.1093/bioinformatics/btp619)

87. Cooper G, Herskovits E. 1992 A Bayesian method for the induction of probabilistic networks from data. *Mach. Learn.* **9**, 309–347. (doi:10.1007/BF00994110)

88. Friedman N, Linial M, Nachman I, Pe'er D. 2000 Using Bayesian networks to analyze expression data. *J. Comput. Biol.* **7**, 601–620. (doi:10.1089/ 106652700750050961)

89. Yu J, Smith V, Wang P, Hartemink A, Jarvis E. 2004 Advances to Bayesian network inference for generating causal networks from observational biological data. *Bioinformatics* **20**, 3594–3603. (doi:10.1093/bioinformatics/bth448)

90. Kim S, Imoto S, Miyano S. 2003 Inferring gene networks from time series microarray data using dynamic Bayesian networks. *Brief. Bioinform.* **4**, 228–235. (doi:10.1093/bib/4.3.228)

91. Kim S, Imoto S, Miyano S. 2004 Dynamic Bayesian network and nonparametric regression for nonlinear modeling of gene networks from time series gene expression data. *Biosystems* **75**, 57–65. (doi:10. 1016/j.biosystems.2004.03.004)

92. Dondelinger F, Husmeier D, Lèbre S. 2012 Dynamic Bayesian networks in molecular plant science: inferring gene regulatory networks from multiple gene expression time series. *Euphytica* **183**, 361–377. (doi:10.1007/s10681-011-0538-3)

93. Camacho D, Vera Licona P, Mendes P, Laubenbacher R. 2007 Comparison of reverse-engineering methods using an *in silico* network. *Ann. NY Acad. Sci.* **1115**, 73–89. (doi:10.1196/annals.1407.006)

94. Prill RJ, Marbach D, Saez-Rodriguez J, Sorger PK, Alexopoulos LG, Xue X, Clarke ND, Altan-Bonnet G, Stolovitzky G. 2010 Towards a rigorous assessment of systems biology models: the DREAM3 challenges. *PLoS ONE* **5**, e9202. (doi:10.1371/journal.pone.0009202)

95. Kirsch A. 2011 *An introduction to the mathematical theory of inverse problems*, vol. 120. New York, NY: Springer Science+Business Media.

96. Tarantola A. 2005 *Inverse problem theory and methods for model parameter estimation*. Philadelphia, PA: Society for Industrial and Applied Mathematics.

97. Hadamard J. 1902 Sur les problèmes aux dérivées partielles et leur signification physique. *Princet. Univ. Bull.* **13**, 28.

98. Kohavi R *et al*. 1995 A study of cross-validation and bootstrap for accuracy estimation and model selection. In *Int. Joint Conf. on Artificial Intelligence*, vol. 14, pp. 1137–1145. London, UK: Lawrence Erlbaum Associates Ltd.

99. Picard RR, Cook RD. 1984 Cross-validation of regression models. *J. Am. Stat. Assoc.* **79**, 575–583. (doi:10.1080/01621459.1984.10478083)

100. Tikhonov AN, Goncharsky A, Stepanov V, Yagola A. 1995 *Numerical methods for the solution of ill-posed problems*. Dordrecht, The Netherlands: Kluwer Academic Publishers.

101. Engl HW, Flamm C, Kügler P, Lu J, Müller S, Schuster P. 2009 Inverse problems in systems biology. *Inverse Probl.* **25**, 123014. (doi:10.1088/ 0266-5611/25/12/123014)

102. Boyd S, Vandenberghe L. 2004 *Convex optimization*. Cambridge, UK: Cambridge University Press.

103. Schittkowski K. 2002 *Numerical data fitting in dynamical systems: a practical introduction with applications and software*, vol. 77. Dordrecht, The Netherlands: Kluwer Academic Publishers.

104. Kan A, Timmer G. 1985 A stochastic approach to global optimization. *Numer. Optim.* **84**, 245–262.

105. Horst R, Romeijn HE. 2002 *Handbook of global optimization*, vol. 2. Dordrecht, The Netherlands: Kluwer Academic Publishers.

106. Floudas C, Gounaris C. 2009 A review of recent advances in global optimization. *J. Glob. Optim.* **45**, 3–38. (doi:10.1007/s10898-008-9332-8)

107. Esposito WR, Floudas CA. 2000 Global optimization for the parameter estimation of differential-algebraic systems. *Ind. Eng. Chem. Res.* **39**, 1291–1310. (doi:10.1021/ie990486w)

108. Miró A, Pozo C, Guillén-Gosálbez G, Egea JA, Jiménez L. 2012 Deterministic global optimization algorithm based on outer approximation for the parameter estimation of nonlinear dynamic biological systems. *BMC Bioinform.* **13**, 90. (doi:10.1186/1471-2105-13-90)

109. Zhu W. 2006 Unsolvability of some optimization problems. *Appl. Math. Comput.* **174**, 921–926. (doi:10.1016/j.amc.2005.05.025)

110. Murty KG, Kabadi SN. 1987 Some NP-complete problems in quadratic and nonlinear programming. *Math. Program* **39**, 117–129. (doi:10.1007/ BF02592948)

111. Fortnow L. 2009 The status of the P versus NP problem. *Commun. ACM* **52**, 78–86. (doi:10.1145/ 1562164.1562186)

112. Moles CG, Mendes P, Banga JR. 2003 Parameter estimation in biochemical pathways: a comparison of global optimization methods. *Genome Res.* **13**, 2467–2474. (doi:10.1101/gr.1262503)

113. Egea JA, Rodrguez-Fernández M, Banga JR, Mart R. 2007 Scatter search for chemical and bio-process optimization. *J. Glob. Optim.* **37**, 481–503. (doi:10. 1007/s10898-006-9075-3)

114. Rodriguez-Fernandez M, Egea JA, Banga JR. 2006 Novel metaheuristic for parameter estimation in nonlinear dynamic biological systems. *BMC Bioinform.* **7**, 483. (doi:10.1186/1471-2105-7-483)

115. Banga JR. 2008 Optimization in computational systems biology. *BMC Syst. Biol.* **2**, 47. (doi:10.1186/ 1752-0509-2-47)

116. Villaverde A, Egea J, Banga J. 2012 A cooperative strategy for parameter estimation in large scale systems biology models. *BMC Syst. Biol.* **6**, 75. (doi:10.1186/1752-0509-6-75)

117. Kotte O, Heinemann M. 2009 A divide-and-conquer approach to analyze underdetermined biochemical models. *Bioinformatics* **25**, 519–525. (doi:10.1093/ bioinformatics/btp004)

118. Wolpert DH, Macready WG. 1997 No free lunch theorems for optimization. *IEEE Trans. Evol. Comput.* **1**, 67–82. (doi:10.1109/4235.585893)

119. Ho Y-C. 1999 The no free lunch theorem and the human–machine interface. *IEEE Contr. Syst.* **19**, 8–10. (doi:10.1109/37.768535)

120. Ljung L. 1999 *System identification*. New York, NY: Wiley Online Library.

121. Walter E, Pronzato L. 1997 *Identification of parametric models from experimental data. Communications and Control Engineering Series*. London, UK: Springer.

122. Ljung L. 2010 Perspectives on system identification. *Annu. Rev. Control* **34**, 1–12. (doi:10.1016/j.arcontrol.2009.12.001)

123. Bekey GA, Beneken JE. 1978 Identification of biological systems: a survey. *Automatica* **14**, 41–47. (doi:10.1016/0005-1098(78)90075-4)

124. Wiener N. 1948 *Cybernetics: or, control and communication in the animal and the machine*. New York, NY: Wiley.

125. Von Bertalanffy L. 1968 *General system theory: foundations, development, applications*. New York, NY: George Braziller.

126. Cury JE, Baldissera FL. 2013 Systems biology, synthetic biology and control theory: a promising golden braid. *Ann. Rev. Control* **37**, 57–67. (doi:10.1016/j.arcontrol.2013.03.006)

127. Bellman R, Åström K. 1970 On structural identifiability. *Math. Biosci.* **7**, 329–339. (doi:10.1016/0025-5564(70)90132-X)

128. Cobelli C, Distefano 3rd JJ. 1980 Parameter and structural identifiability concepts and ambiguities: a critical review and analysis. *Am. J. Physiol.* **239**, R7–R24.

129. Kremling A, Fischer S, Gadkar K, Doyle F, Sauter T, Bullinger E, Allgower F, Gilles E. 2004 A benchmark for methods in reverse engineering and model discrimination: problem formulation and solutions. *Genome Res.* **14**, 1773–1785. (doi:10.1101/gr.1226004)

130. Jaqaman K, Danuser G. 2006 Linking data to models: data regression. *Nat. Rev. Mol. Cell Biol.* **7**, 813–819. (doi:10.1038/nrm2030)

131. Nikerel IE, van Winden WA, Verheijen PJ, Heijnen JJ. 2009 Model reduction and *a priori* kinetic parameter identifiability analysis using metabolome time series for metabolic reaction networks with linlog kinetics. *Metab. Eng.* **11**, 20–30. (doi:10.1016/j.ymben.2008.07.004)

132. Quaiser T, Mönnigmann M. 2009 Systematic identifiability testing for unambiguous mechanistic modeling—application to JAK-STAT, map kinase, and NF-κb signaling pathway models. *BMC Syst. Biol.* **3**, 50. (doi:10.1186/1752-0509-3-50)

133. Raue A, Kreutz C, Maiwald T, Bachmann J, Schilling M, Klingmüller U, Timmer J. 2009 Structural and practical identifiability analysis of partially observed dynamical models by exploiting the profile likelihood. *Bioinformatics* **25**, 1923–1929. (doi:10.1093/bioinformatics/btp358)

134. Ashyraliyev M, Fomekong-Nanfack Y, Kaandorp JA, Blom JG. 2009 Systems biology: parameter estimation for biochemical models. *FEBS J.* **276**, 886–902. (doi:10.1111/j.1742-4658.2008.06844.x)

135. Balsa-Canto E, Alonso AA, Banga JR. 2010 An iterative identification procedure for dynamic modeling of biochemical networks. *BMC Syst. Biol.* **4**, 11. (doi:10.1186/1752-0509-4-11)

136. Srinath S, Gunawan R. 2010 Parameter identifiability of power-law biochemical system models. *J. Biotechnol.* **149**, 132–140. (doi:10.1016/j.jbiotec.2010.02.019)

137. Nemcová J. 2010 Structural identifiability of polynomial and rational systems. *Math. Biosci.* **223**, 83–96. (doi:10.1016/j.mbs.2009.11.002)

138. Chis O-T, Banga JR, Balsa-Canto E. 2011 Structural identifiability of systems biology models: a critical comparison of methods. *PLoS ONE* **6**, e27755. (doi:10.1371/journal.pone.0027755)

139. Berthoumieux S, Brilli M, Kahn D, De Jong H, Cinquemani E. 2012 On the identifiability of metabolic network models. *J. Math. Biol.* **67**, 1795–1832. (doi:10.1007/s00285-012-0614-x)

140. Vanlier J, Tiemann C, Hilbers P, van Riel N. In press. Parameter uncertainty in biochemical models described by ordinary differential equations. *Math. Biosci.* (doi:10.1016/j.mbs.2013.03.006)

141. Cedersund G. 2012 Conclusions via unique predictions obtained despite unidentifiability—new definitions and a general method. *FEBS J.* **279**, 3513–3527. (doi:10.1111/j.1742-4658.2012.08725.x)

142. Schaber J, Klipp E. 2011 Model-based inference of biochemical parameters and dynamic properties of microbial signal transduction networks. *Curr. Opin. Biotechnol.* **22**, 109–116. (doi:10.1016/j.copbio.2010.09.014)

143. Lillacci G, Khammash M. 2010 Parameter estimation and model selection in computational biology. *PLoS Comput. Biol.* **6**, e1000696. (doi:10.1371/journal.pcbi.1000696)

144. Bandara S, Schlöder JP, Eils R, Bock HG, Meyer T. 2009 Optimal experimental design for parameter estimation of a cell signaling model. *PLoS Comput. Biol.* **5**, e1000558. (doi:10.1371/journal.pcbi.1000558)

145. Balsa-Canto E, Alonso AA, Banga JR. 2008 Computational procedures for optimal experimental design in biological systems. *IET Syst. Biol.* **2**, 163–172. (doi:10.1049/iet-syb:20070069)

146. Banga J, Balsa-Canto E. 2008 Parameter estimation and optimal experimental design. *Essays Biochem.* **45**, 195–210. (doi:10.1042/BSE0450195)

147. Apgar JF, Toettcher JE, Endy D, White FM, Tidor B. 2008 Stimulus design for model selection and validation in cell signaling. *PLoS Comput. Biol.* **4**, e30. (doi:10.1371/journal.pcbi.0040030)

148. He F, Brown M, Yue H. 2010 Maximin and Bayesian robust experimental design for measurement set selection in modelling biochemical regulatory systems. *Int. J. Robust Nonlinear Control* **20**, 1059–1078. (doi:10.1002/rnc.1558)

149. Mélykúti B, August E, Papachristodoulou A, El-Samad H. 2010 Discriminating between rival biochemical network models: three approaches to optimal experiment design. *BMC Syst. Biol.* **4**, 38. (doi:10.1186/1752-0509-4-38)

150. Skanda D, Lebiedz D. 2010 An optimal experimental design approach to model discrimination in dynamic biochemical systems. *Bioinformatics* **26**, 939–945. (doi:10.1093/bioinformatics/btq074)

151. Flassig R, Sundmacher K. 2012 Optimal design of stimulus experiments for robust discrimination of biochemical reaction networks. *Bioinformatics* **28**, 3089–3096. (doi:10.1093/bioinformatics/bts585)

152. Stegmaier J, Skanda D, Lebiedz D. 2013 Robust optimal design of experiments for model discrimination using an interactive software tool. *PLoS ONE* **8**, e55723. (doi:10.1371/journal.pone.0055723)

153. Tegner J, Yeung MS, Hasty J, Collins JJ. 2003 Reverse engineering gene networks: integrating genetic perturbations with dynamical modeling. *Proc. Natl Acad. Sci. USA* **100**, 5944–5949. (doi:10.1073/pnas.0933416100)

154. Ingolia NT, Weissman JS. 2008 Systems biology: reverse engineering the cell. *Nature* **454**, 1059–1062. (doi:10.1038/4541059a)

155. Mettetal JT, Muzzey D, Gómez-Uribe C, van Oudenaarden A. 2008 The frequency dependence of osmo-adaptation in *Saccharomyces cerevisiae*. *Sci. Signal* **319**, 482.

156. Bennett MR, Pang WL, Ostroff NA, Baumgartner BL, Nayak S, Tsimring LS, Hasty J. 2008 Metabolic gene regulation in a dynamically changing environment. *Nature* **454**, 1119–1122. (doi:10.1038/nature07211)

157. Kaltenbach H-M, Dimopoulos S, Stelling J. 2009 Systems analysis of cellular networks under uncertainty. *FEBS Lett.* **583**, 3923–3930. (doi:10.1016/j.febslet.2009.10.074)

158. Kuepfer L, Peter M, Sauer U, Stelling J. 2007 Ensemble modeling for analysis of cell signaling dynamics. *Nat. Biotechnol.* **25**, 1001–1006. (doi:10.1038/nbt1330)

159. Tan Y, Lafontaine Rivera JG, Contador CA, Asenjo JA, Liao JC. 2011 Reducing the allowable kinetic space by constructing ensemble of dynamic models with the same steady-state flux. *Metab. Eng.* **13**, 60–75. (doi:10.1016/j.ymben.2010.11.001)

160. Marbach D *et al.* 2012 Wisdom of crowds for robust gene network inference. *Nat. Methods* **9**, 796–804. (doi:10.1038/nmeth.2016)

161. Tigges M, Marquez-Lago TT, Stelling J, Fussenegger M. 2009 A tunable synthetic mammalian oscillator. *Nature* **457**, 309–312. (doi:10.1038/nature07616)

162. Holtz WJ, Keasling JD. 2010 Engineering static and dynamic control of synthetic pathways. *Cell* **140**, 19–23. (doi:10.1016/j.cell.2009.12.029)

163. Menolascina F, Di Bernardo M, Di Bernardo D. 2011 Analysis, design and implementation of a novel scheme for *in-vivo* control of synthetic gene regulatory networks. *Automatica* **47**, 1265–1270. (doi:10.1016/j.automatica.2011.01.073)

164. Menolascina F, Siciliano V, di Bernardo D. 2012 Engineering and control of biological systems: a new way to tackle complex diseases. *FEBS Lett.* **586**, 2122–2128. (doi:10.1016/j.febslet.2012.04.050)

165. Milias-Argeitis A, Summers S, Stewart-Ornstein J, Zuleta I, Pincus D, El-Samad H, Khammash M, Lygeros J. 2011 *In silico* feedback for *in vivo* regulation of a gene expression circuit. *Nat. Biotechnol.* **29**, 1114–1116. (doi:10.1038/nbt.2018)

166. Sootla A, Strelkowa N, Ernst D, Barahona M, Stan G-B. 2013 Toggling a genetic switch using reinforcement learning. (http://arxiv.org/abs/1303.3183).

167. Uhlendorf J, Miermont A, Delaveau T, Charvin G, Fages F, Bottani S, Batt G, Hersen P. 2012 Long-term model predictive control of gene expression at the population and single-cell levels. *Proc. Natl Acad. Sci. USA* **109**, 14 271 – 14 276. (doi:10.1073/pnas.1206810109)

168. Horn F, Jackson R. 1972 General mass action kinetics. *Arch. Rat. Mech. Anal.* **47**, 81 – 116. (doi:10.1007/BF00251225)

169. Feinberg M. 1972 Complex balancing in general kinetic systems. *Arch. Rat. Mech. Anal.* **49**, 187 – 194. (doi:10.1007/BF00255665)

170. Feinberg M, Horn FJ. 1974 Dynamics of open chemical systems and the algebraic structure of the underlying reaction network. *Chem. Eng. Sci.* **29**, 775 – 787. (doi:10.1016/0009-2509(74)80195-8)

171. Bailey JE. 2001 Complex biology with no parameters. *Nat. Biotechnol.* **19**, 503 – 504. (doi:10.1038/89204)

172. Craciun G, Tang Y, Feinberg M. 2006 Understanding bistability in complex enzyme-driven reaction networks. *Proc. Natl Acad. Sci. USA* **103**, 8697 – 8702. (doi:10.1073/pnas.0602767103)

173. Wang L, Sontag ED. 2008 On the number of steady states in a multiple futile cycle. *J. Math. Biol.* **57**, 29 – 52. (doi:10.1007/s00285-007-0145-z)

174. Otero-Muras I, Banga JR, Alonso AA. 2009 Exploring multiplicity conditions in enzymatic reaction networks. *Biotechnol. Prog.* **25**, 619 – 631. (doi:10.1002/btpr.112)

175. Craciun G, Feinberg M. 2010 Multiple equilibria in complex chemical reaction networks: semiopen mass action systems. *SIAM J. Appl. Math.* **70**, 1859 – 1877. (doi:10.1137/090756387)

176. Szederkényi G. 2010 Computing sparse and dense realizations of reaction kinetic systems. *J. Math. Chem.* **47**, 551 – 568. (doi:10.1007/s10910-009-9525-5)

177. Szederkényi G, Hangos KM. 2011 Finding complex balanced and detailed balanced realizations of chemical reaction networks. *J. Math. Chem.* **49**, 1163 – 1179. (doi:10.1007/s10910-011-9804-9)

178. Shinar G, Feinberg M. 2011 Design principles for robust biochemical reaction networks: what works, what cannot work, and what might almost work. *Math. Biosci.* **231**, 39 – 48. (doi:10.1016/j.mbs.2011.02.012)

179. Shinar G, Feinberg M. 2010 Structural sources of robustness in biochemical reaction networks. *Science* **327**, 1389 – 1391. (doi:10.1126/science.1183372)

180. Craciun G, Pantea C. 2008 Identifiability of chemical reaction networks. *J. Math. Chem.* **44**, 244 – 259. (doi:10.1007/s10910-007-9307-x)

181. Szederkényi G, Banga J, Alonso A. 2011 Inference of complex biological networks: distinguishability issues and optimization-based solutions. *BMC Syst. Biol.* **5**, 177. (doi:10.1186/1752-0509-5-177)

182. Yeung M, Tegnér J, Collins J. 2002 Reverse engineering gene networks using singular value decomposition and robust regression. *Proc. Natl Acad. Sci. USA* **99**, 6163 – 6168. (doi:10.1073/pnas.092576199)

183. Hayete B, Gardner TS, Collins JJ. 2007 Size matters: network inference tackles the genome scale. *Mol. Syst. Biol.* **3**, 77. (doi:10.1038/msb4100118)

184. Efron B. 2013 A 250-year argument: belief, behavior, and the bootstrap. *Bull. Am. Math. Soc.* **50**, 129 – 146. (doi:10.1090/S0273-0979-2012-01374-5)

185. Stumpf M, Balding DJ, Girolami M. 2011 *Handbook of statistical systems biology*. New York, NY: Wiley.

186. Golightly A, Wilkinson DJ. 2008 Bayesian inference for nonlinear multivariate diffusion models observed with error. *Comput. Stat. Data Anal.* **52**, 1674 – 1693. (doi:10.1016/j.csda.2007.05.019)

187. Wilkinson DJ. 2012 *Stochastic modelling for systems biology*, vol. 44. Boca Raton, FL: CRC press.

188. Battogtokh D, Asch D, Case M, Arnold J, Schüttler H-B. 2002 An ensemble method for identifying regulatory circuits with special reference to the *qa* gene cluster of *Neurospora crassa*. *Proc. Natl Acad. Sci. USA* **99**, 16 904 – 16 909. (doi:10.1073/pnas.262658899)

189. Brown KS, Sethna JP. 2003 Statistical mechanical approaches to models with many poorly known parameters. *Phys. Rev. E* **68**, 021904. (doi:10.1103/PhysRevE.68.021904)

190. Sanguinetti G, Lawrence ND, Rattray M. 2006 Probabilistic inference of transcription factor concentrations and gene-specific regulatory activities. *Bioinformatics* **22**, 2775 – 2781. (doi:10.1093/bioinformatics/btl473)

191. Girolami M. 2008 Bayesian inference for differential equations. *Theor. Comput. Sci.* **408**, 4 – 16. (doi:10.1016/j.tcs.2008.07.005)

192. Vyshemirsky V, Girolami MA. 2008 Bayesian ranking of biochemical system models. *Bioinformatics* **24**, 833 – 839. (doi:10.1093/bioinformatics/btm607)

193. Marjoram P, Molitor J, Plagnol V, Tavaré S. 2003 Markov chain Monte Carlo without likelihoods. *Proc. Natl Acad. Sci. USA* **100**, 15 324 – 15 328. (doi:10.1073/pnas.0306899100)

194. Sisson S, Fan Y, Tanaka MM. 2007 Sequential Monte Carlo without likelihoods. *Proc. Natl Acad. Sci. USA* **104**, 1760 – 1765. (doi:10.1073/pnas.0607208104)

195. Secrier M, Toni T, Stumpf MP. 2009 The ABC of reverse engineering biological signalling systems. *Mol. Biosyst.* **5**, 1925 – 1935. (doi:10.1039/b908951a)

196. Toni T, Welch D, Strelkowa N, Ipsen A, Stumpf MP. 2009 Approximate Bayesian computation scheme for parameter inference and model selection in dynamical systems. *J. R. Soc. Interface* **6**, 187 – 202. (doi:10.1098/rsif.2008.0172)

197. Xu T-R *et al*. 2010 Inferring signaling pathway topologies from multiple perturbation measurements of specific biochemical species. *Sci. Signal.* **3**, ra20. (doi:10.1126/scisignal.2000517)

198. Eydgahi H, Chen WW, Muhlich JL, Vitkup D, Tsitsiklis JN, Sorger PK. 2013 Properties of cell death models calibrated and compared using Bayesian approaches. *Mol. Syst. Biol.* **9**, 644. (doi:10.1038/msb.2012.69)

199. Liepe J, Filippi S, Komorowski M, Stumpf MP. 2013 Maximizing the information content of experiments in systems biology. *PLoS Comput. Biol.* **9**, e1002888. (doi:10.1371/journal.pcbi.1002888)

200. Raue A, Kreutz C, Theis FJ, Timmer J. 2013 Joining forces of Bayesian and frequentist methodology: a study for inference in the presence of non-identifiability. *Phil. Trans. R. Soc. A* **371**, 20110544. (doi:10.1098/rsta.2011.0544)

201. Hug S, Raue A, Hasenauer J, Bachmann J, Klingmüller U, Timmer J, Theis F. In press. High-dimensional Bayesian parameter estimation: case study for a model of JAK2/STAT5 signaling. *Math. Biosci.* (doi:10.1016/j.mbs.2013.04.002)()

202. Rasmussen C, Williams C. 2006 *Gaussian processes for machine learning*. Cambridge, MA: MIT Press.

203. Calderhead B, Girolami M, Lawrence ND. 2008 Accelerating Bayesian inference over nonlinear differential equations with Gaussian processes. In *Advances in Neural Information Processing Systems* (eds D Koller, D Schuurmans, Y Bengio, L Bottou), pp. 217 – 224. Neural Information Processing Systems Foundation, http://books.nips.cc/nips21.html.

204. Girolami M, Calderhead B. 2011 Riemann manifold Langevin and Hamiltonian Monte Carlo methods. *J. R. Stat. Soc. B* **73**, 123 – 214. (doi:10.1111/j.1467-9868.2010.00765.x)

205. Calderhead B, Girolami M. 2011 Statistical analysis of nonlinear dynamical systems using differential geometric sampling methods. *Interface Focus* **1**, 821 – 835. (doi:10.1098/rsfs.2011.0051)

206. Schmidl D, Czado C, Hug S, Theis FJ. 2013 A vine-copula based adaptive MCMC sampler for efficient inference of dynamical systems. *Bayesian Anal.* **8**, 1 – 22. (doi:10.1214/13-BA801)

207. Gutenkunst RN, Waterfall JJ, Casey FP, Brown KS, Myers CR, Sethna JP. 2007 Universally sloppy parameter sensitivities in systems biology models. *PLoS Comput. Biol.* **3**, e189. (doi:10.1371/journal.pcbi.0030189)

208. Apgar JF, Witmer DK, White FM, Tidor B. 2010 Sloppy models, parameter uncertainty, and the role of experimental design. *Mol. Biosyst.* **6**, 1890 – 1900. (doi:10.1039/b918098b)

209. Chachra R, Transtrum MK, Sethna JP. 2011 Comment on 'sloppy models, parameter uncertainty, and the role of experimental design'. *Mol. Biosyst.* **7**, 2522 – 2522. (doi:10.1039/c1mb05046j)

210. Hagen DR, Apgar JF, Witmer DK, White FM, Tidor B. 2011 Reply to comment on 'sloppy models, parameter uncertainty, and the role of experimental design'. *Mol. Biosyst.* **7**, 2523 – 2524. (doi:10.1039/c1mb05200d)

211. Hagen DR, White JK, Tidor B. 2013 Convergence in parameters and predictions using computational experimental design. *Interface Focus* **3**, 20130008. (doi:10.1098/rsfs.2013.0008)

212. Shannon CE. 1949 Communication in the presence of noise. *Proc. IRE* **37**, 10 – 21. (doi:10.1109/JRPROC.1949.232969)

213. Candes EJ, Romberg JK, Tao T. 2006 Stable signal recovery from incomplete and inaccurate

measurements. *Commun. Pure Appl. Math.* **59**, 1207–1223. (doi:10.1002/cpa.20124)

214. Donoho DL. 2006 Compressed sensing. *IEEE Trans. Inform. Theory* **52**, 1289–1306. (doi:10.1109/TIT.2006.871582)

215. Candès EJ, Romberg J, Tao T. 2006 Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information. *IEEE Trans. Inform. Theory* **52**, 489–509. (doi:10.1109/TIT.2005.862083)

216. Sheikh MA, Milenkovic O, Baraniuk RG. 2007 Designing compressive sensing DNA microarrays. In *Computational Advances in Multi-Sensor Adaptive Processing, 2007. CAMPSAP 2007. 2nd IEEE International Workshop on*, pp. 141–144. Piscataway Township, NJ: IEEE.

217. Pan W, Yuan Y, Stan G. 2012 Reconstruction of arbitrary biochemical reaction networks: a compressive sensing approach. In *Proc. 51st IEEE Conf. Dec. Contr. 10–13 December 2012, Maui, HI*, pp. 2334-2339, Piscataway Township, NJ: Institute of Electrical and Electronics Engineers (IEEE). (doi:10.1109/CDC.2012.6426216)

218. Baldi P, Brunak S. 2001 *Bioinformatics: the machine learning approach*. Bradford Book. Cambridge, MA: MIT Press.

219. Larrañaga P *et al.* 2006 Machine learning in bioinformatics. *Brief. Bioinform.* **7**, 86–112. (doi:10.1093/bib/bbk007)

220. Kell DB. 2006 Metabolomics, modelling and machine learning in systems biology—towards an understanding of the languages of cells. *FEBS J.* **273**, 873–894. (doi:10.1111/j.1742-4658.2006.05136.x)

221. Tarca AL, Carey VJ, Chen X-w, Romero R, Drăghici S. 2007 Machine learning and its applications to biology. *PLoS Comput. Biol.* **3**, e116. (doi:10.1371/journal.pcbi.0030116)

222. Kell DB, Oliver SG. 2004 Here is the evidence, now what is the hypothesis? The complementary roles of inductive and hypothesis-driven science in the post-genomic era. *Bioessays* **26**, 99–105. (doi:10.1002/bies.10385)

223. Evans J, Rzhetsky A. 2010 Machine science. *Science* **329**, 399–400. (doi:10.1126/science.1189416)

224. King RD, Whelan KE, Jones FM, Reiser PG, Bryant CH, Muggleton SH, Kell DB, Oliver SG. 2004 Functional genomic hypothesis generation and experimentation by a robot scientist. *Nature* **427**, 247–252. (doi:10.1038/nature02236)

225. Soldatova LN, Clare A, Sparkes A, King RD. 2006 An ontology for a robot scientist. *Bioinformatics* **22**, e464–e471. (doi:10.1093/bioinformatics/btl207)

226. King RD *et al.* 2009 The automation of science. *Science* **324**, 85–89. (doi:10.1126/science.1165620)

227. Bongard J, Lipson H. 2007 Automated reverse engineering of nonlinear dynamical systems. *Proc. Natl Acad. Sci. USA* **104**, 9943–9948. (doi:10.1073/pnas.0609476104)

228. Schmidt M, Lipson H. 2009 Distilling free-form natural laws from experimental data. *Science* **324**, 81–85. (doi:10.1126/science.1165893)

229. Goldenfeld N, Kadanoff LP. 1999 Simple lessons from complexity. *Science* **284**, 87–89. (doi:10.1126/science.284.5411.87)

230. Tukey JW. 1962 The future of data analysis. *Ann. Math. Stat.* **33**, 1–67. (doi:10.1214/aoms/1177704711)

231. Brenner S. 2010 Sequences and consequences. *Phil. Trans. R. Soc. B* **365**, 207–212. (doi:10.1098/rstb.2009.0221)

232. Noble D. 2011 Systems: what's in a name? *Physiology* **26**, 126–128. (doi:10.1152/physiol.00014.2011)

233. Tarantola A. 2006 Popper, Bayes and the inverse problem. *Nat. Phys.* **2**, 492–494. (doi:10.1038/nphys375)

234. Efron B. 2013 Bayes' theorem in the 21st century. *Science* **340**, 1177–1178. (doi:10.1126/science.1236536)

235. Lindley D. 1975 The future of statistics: a Bayesian 21st century. *Adv. Appl. Probab.* **7**, 106–115. (doi:10.2307/1426315)

236. Erguler K, Stumpf M. 2011 Practical limits for reverse engineering of dynamical systems: a statistical analysis of sensitivity and parameter inferability in systems biology models. *Mol. Biosyst.* **7**, 1593–1602. (doi:10.1039/c0mb00107d)

237. Sontag ED. 2007 Monotone and near-monotone biochemical networks. *Syst. Synth. Biol.* **1**, 59–87. (doi:10.1007/s11693-007-9005-9)

238. Hucka M *et al.* 2003 The systems biology markup language (SBML): a medium for representation and exchange of biochemical network models. *Bioinformatics* **19**, 524–531. (doi:10.1093/bioinformatics/btg015)

239. Le Novere N *et al.* 2006 Biomodels database: a free, centralized database of curated, published, quantitative kinetic models of biochemical and cellular systems. *Nucleic Acids Res.* **34**(Suppl. 1), D689–D691. (doi:10.1093/nar/gkj092)

240. Kitano H. 2005 International alliances for quantitative modeling in systems biology. *Mol. Syst. Biol.* 1:2005.0007. (doi:10.1038/msb4100011)