Zen and the Art of Parameter Estimation in Systems Biology

Christopher R. Myers

Department of Physics and Center for Advanced Computing

Cornell University

1 Introduction

A mathematical model describes a space of possibilities. As this volume illustrates, models come in many shapes and sizes, and discerning an appropriate form of a model for a given problem is in many ways as much art as science, suggested by an intuitive feel for a problem and a drive to distill the important degrees-of-freedom needed to capture some phenomenon of interest. Alongside the poetry of identifying an ideal model form lies the more prosaic work of estimating the values of parameters that provide reality to that form.

Broadly speaking, mechanistic mathematical models typically consist of state variables, interaction rules among those variables, and parameters that quantify aspects of those state variables and interaction rules. Parameters dictate the space of possible model outputs, given a specified model structure. Thus, parameters represent a class of model inputs that impact what possible predictions a model can make. Parameter estimation is the process by which a modeler, having identified a plausible model structure, endeavors to determine the numerical values of parameters within that model in order to be able to assess model outcomes. As such, parameter estimation is an aspect of inference, and typically refers to a process of fitting parameters to data through their collective effects in a model; if one were able to measure parameters directly and with sufficient certainty,

one would simply use those measured values as parameter inputs. But for many biological systems of interest, carrying out such measurements is not easy or may not be possible, and one is left instead with the process of reverse engineering plausible parameter values from measurements of state variables (or functions thereof) rather than forward simulation based on experimentally determined parametric inputs.

Many models of interest in the field of systems biology contain many unknown parameters, resulting in high-dimensional parameter spaces that must be characterized, with complex structure that is not well understood. Understanding such structure, and how it impacts the predictivity of models and the potential for the construction of alternative models, is an active area of research. I will begin by providing an overview of some of the mechanics of parameter estimation, although "mechanics" more perhaps in the style of *Zen and the Art of Motorcycle Maintenance* [1] than in that of the Chilton manuals for auto repair. Just as importantly, however, I will also endeavor to consider parameter estimation within the broader context of modeling, to describe how it relates to model construction, inference, selection, reduction, and analysis. I will close with some thoughts about the somewhat fractured and multifaceted field of systems biology, highlighting how issues of parameterization and parameter estimation lie at the crossroads of different schools of thought.

Even though I will mostly address here mechanistic models of cellular processes, many of the concepts and techniques introduced here are broadly applicable to a wide range of models relevant not just to immunology and systems biology, but to other fields as well [2]. In the field of immunology, this might include statistical models, or descriptions at other levels of biological resolution, such as models of the population dynamics of pathogens replicating within hosts, or spreading among hosts [3]. Where possible, I will endeavor to point out generalities and abstractions that are useful across different classes of models, while also noting some of the particular aspects that arise in analyzing complex cellular networks.

2 The mechanics of parameter estimation

2.1 Estimating parameters from data

Let us assume that we have a set of M state variables $\mathbf{x}(t) = x_1(t), ..., x_M(t)$. Since we are focused here primarily on dynamical models, we assume there are a set of initial conditions $\mathbf{x}^0(t = 0)$, and a prescription for solving for the dynamics of the system at later times. If our model is deterministic, this prescription might involve formulating and integrating a set of coupled ordinary differential equations (ODEs) describing the time rate of change of chemical concentrations; if the model is stochastic, we might instead use a stochastic simulation algorithm such as Gillespie's method [4] to step the system forward in time. The particular trajectory that the system traces out in its state space will also depend on the choice of model parameters. We will denote this set of N parameters $\theta = \theta_1, ..., \theta_N$ and denote the trajectory's dependence on parameters as $\mathbf{x}(t; \theta)$. The goal of parameter estimation is to infer the numerical values of the parameters θ based on available data. For the case of deterministic dynamics described by ODEs, the state variables will unfold in time according to the dynamical equation $d\mathbf{x}/dt = \mathbf{f}(\mathbf{x}, t; \theta)$, where $\mathbf{f}$ reflects the sum of all the reaction fluxes in and out of the states.

Experimental data might reflect individual state variables at specific time points $x_i(t_j; \theta)$, or they might reflect functions of multiple state variables. Often, one is able to measure the abundance of some chemical entity, but is unable to distinguish its participation in different molecular compounds or states; in such a case, the relevant observed quantity would reflect some weighted sum of individual state variables. In other cases, data might be available on reaction fluxes, which are (potentially nonlinear) functions of state variables. We might have measurements of the system under different experimental conditions, such as with different initial amounts of chemical species, or in a mutant where we have knocked out or overexpressed some particular component. Let $y^*_{o,c}(t_j)$ represent the value of observable o in condition c at time t_j; denote the uncertainty of that measurement as $\sigma_{o,c}(t_j)$; and let $y_{o,c}(t; \theta)$ represent the corresponding predictions of the model for a given set of parameters θ. We are interested in finding a choice of parameters θ for

which the model trajectories $\mathbf{y}(t;\theta)$ best approximate the measured values $y^*_{o,c}(t_j)$. We can define a cost function $C(\theta)$ that represents the squared deviation of a given set of model predictions from the data:

$$C(\theta) = \frac{1}{2}\sum_{o,c,t_j}\left(\frac{y_{o,c}(t_j;\theta) - y^*_{o,c}(t_j)}{\sigma_{o,c}(t_j)}\right)^2$$

where t_j refers to the timepoints at which the data are measured. The best fitting parameter set θ^* is that which minimizes the cost: $\theta^* = argmin[C(\theta)]$.

In statistics, one is often interested in computing the *likelihood* of a particular model, that is, the probability of observing the particular data measured given a set of parameters θ. If the model residuals are independent and normally distributed, then the cost function $C(\theta)$ corresponds to the negative logarithm of the likelihood $P(Data|\theta)$, where "Data" refers here to the full set of measured observables [5]:

$$P(Data|\theta) = \prod_{o,c,t_j}\frac{1}{\sqrt{2\pi}\sigma_{o,c,t_j}}\exp\left(-\frac{1}{2}\frac{y_{o,c}(t_j;\theta) - y^*_{o,c}(t_j)}{\sigma_{o,c}(t)}\right)^2$$

Minimizing the cost $C(\theta)$ corresponds to maximizing the likelihood function. Different statistical frameworks emphasize different aspects of this data-fitting problem: frequentist statistics typically focuses on estimating the parameters that maximize the likelihood, while Bayesian statistics uses the likelihood to estimate the posterior distribution, reflecting the probability of estimating different sets of parameters given the data. A nice discussion that emphasizes distinctions between frequentist and Bayesian statistical treatments of parameter estimation for models in systems biology can be found in [6].

Our goal is not just to fit the data, however, since the whole point of building a model is to be able to make predictions about situations for which we do not have data. One common pitfall in parameter estimation is overfitting, typically from having an overly complicated model with many parameters that can fit the existing data very well, but which has poor predictive performance on unseen data. In such a case, the data are often "overfit" in the sense that the power of the model

is used to fit random fluctuations in the data rather than the underlying trend of the data. Having laid this groundwork, we will now consider parameter estimation in the context of some specific models of biological networks.

2.2 Examples: JAK-STAT signaling and Epo receptor trafficking

Information processing lies at the core of many important biological functions, implemented by molecular networks supporting signal perception and integration. A central function of the immune system is to process signals in the environment and to decide whether those signals are associated with self or non-self. An important class of perception and signaling networks, relevant both to immune system function and other biological processes, are the JAK-STAT pathways [7, 8]. JAK-STAT signaling involves the perception of extracellular ligands (cytokines, growth factors, etc.) by membrane-bound receptors (so-called Janus kinases, or JAKs) which trigger the activitation of an intracellular molecular complex that subsequently translocates to the nucleus and affects gene transcription (so-called signal transducers and activators of transcription, or STATS). A schematic figure of this pathway is shown in Fig. 1(a). This basic structural theme is played out again and again throughout systems biology, although this relatively simple signaling architecture is more reminiscent of bacterial two-component signaling systems [9] than the more deep and elaborate signaling cascades that have evolved in animals. In mammals, there are 7 different JAK proteins and 4 different STATS. Therefore, this basic architecture can in principle be instantiated in various forms by combining different JAKs and STATs, although this molecular flexibility can also introduce the possibility of crosstalk among different components [10]. Whereas engineered communication systems can be constructed to support efficient and nonambiguous codes [11], living systems that communicate through the interaction of many paralogous components evolving through duplication and divergence face the nontrivial challenge of communicating reliably in the face of constraints posed by crosstalk [12, 13].

The development and analysis of mathematical models of JAK-STAT signaling has a rich history in the field of systems biology, with particular relevance to the problem of parameter estima-

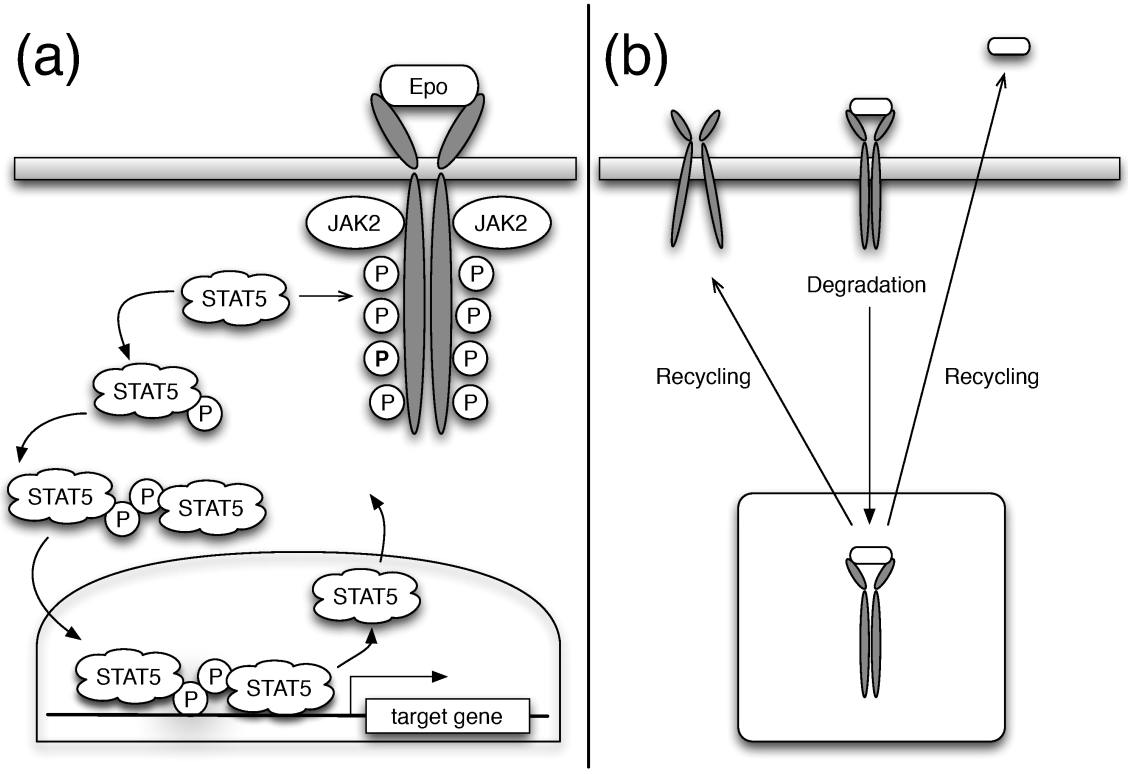

Figure 1: (a) Schematic of the JAK2-STAT5 signaling pathway (adapted from [14]), in which binding to extracellular Epo induces recruitment and phosphorylation of intracellular STAT5 via JAK2. Phosphorylated STAT5 dimerizes and translocates to the nucleus, where it regulates expression of target genes, followed by recycling of monomeric STAT5 back to the cytoplasm. (b) Schematic of receptor trafficking, involving the degradation and recycling of receptor proteins and ligands; an example is EpoR trafficking (adapted from [15]), in which bound Epo-EpoR complexes are internalized, and either degraded or recycled back to the membrane as free EpoR.

tion. Owing largely to the work of Timmer and colleagues, various models of JAK-STAT pathways have been constructed, fit to experimental data, and used to assess new theoretical and computational tools for parameter estimation and sensitivity analysis. Much of this work has centered on the JAK2-STAT5 pathway, which is involved in detecting the growth factor erythropoietin to stimulate the production of red blood cells*. The original JAK-STAT signaling model developed by Timmer and collegues ([14, 16]) emphasized the importance of inference through "databased modeling", and worked through a series of related models to demonstrate the importance of STAT recycling from the nucleus back to the cytoplasm and time delays associated with that process. That model has been used as an application example on parameter estimation in various software packages, such as the SBML-PET Parameter Estimation Toolkit [17], SloppyCell [18, 19], and Data2Dynamics [20]. Raue *et al.* examined this same model in their use of the profile likelihood to analyze structural and practical parameter identifiability [21]. Vanlier *et al.* used the model and underlying data to probe the relationship between parameter estimation and experimental design [22]. And Toni and Stumpf revisited JAK-STAT signaling as an application example to highlight their use of Approximate Bayesian Computation [6]. Readers interested in both the details of JAK-STAT signaling and the intricacies of parameter estimation are encouraged to dig through that rich history of research.

I will consider briefly a related model, involving not the downstream signaling through the JAK-STAT pathway, but the regulation and trafficking of membrane-bound receptors involved in Epo recognition at the gateway of JAK-STAT. Signaling networks are not static scaffolds along which information is communicated (as is often the case with our hard-wired engineered systems), but are instead dynamic entities themselves subject to regulation and control, as indicated schematically in Fig. 1(b). The regulation and organization of membrane-bound receptors is important in a number of problems in systems biology [23, 15, 24] and their misregulation is sometimes implicated in diseases such as cancer, where an excess of growth factor receptors can lead to enhanced rates of cell growth. The model in question here was developed by Becker *et al.* as part of a

*Erythropoietin is also known as Epo, probably best known to the world through its surreptitious abuse by the cyclist Lance Armstrong and others interested in enhancing aerobic performance.

larger study considering Epo receptor (EpoR) trafficking [15]. I will address here only the "core" model presented in [15]. The dynamical equations for the core model describe the concentrations of various molecular states and complexes:

$$\frac{d}{dt}[EpoR] = k_t B_{max} - k_t[EpoR] - k_{on}[Epo][EpoR] + k_{off}[Epo\text{-}EpoR] + k_{ex}[Epo\text{-}EpoR_i]$$
$$\frac{d}{dt}[Epo] = -k_{on}[Epo][EpoR] + k_{off}[Epo\text{-}EpoR] + k_{ex}[Epo\text{-}EpoR_i]$$
$$\frac{d}{dt}[Epo\text{-}EpoR] = k_{on}[Epo][EpoR] - k_{off}[Epo\text{-}EpoR] - k_e[Epo\text{-}EpoR]$$
$$\frac{d}{dt}[Epo\text{-}EpoR_i] = k_e[Epo\text{-}EpoR] - k_{ex}[Epo\text{-}EpoR_i] - k_{di}[Epo\text{-}EpoR_i] - k_{de}[Epo\text{-}EpoR_i]$$
$$\frac{d}{dt}[dEpo_i] = k_{di}[Epo\text{-}EpoR_i]$$
$$\frac{d}{dt}[dEpo_e] = k_{de}[Epo\text{-}EpoR_i]$$

While the model contains these six state variables, experimentally the authors were only able to measure pools of Epo in various compartments: Epo in the extracellular medium (Epo_medium = [Epo] + [dEpoe]), Epo at the cellular membrane (Epo_membrane = [Epo-EpoR]), and Epo in the interior of the cell (Epo_cells = [Epo-EpoR$_i$] + [dEpo$_i$]).

2.3 Data fitting

The actual process used by the authors for parameter estimation for this Epo receptor trafficking model is much more complicated than the simplified analysis that I will present here [15]. A variety of different assays were carried out to characterize and estimate subprocesses within the model, such as the binding of Epo to EpoR. In addition, multiple model variants were developed (the "core" model, the "core model + k_{mob}", and the "auxiliary" model), and parameter estimation was performed simultaneously for both the core and auxiliary model. I will not delve into all the complexity of the estimation process presented in [15], but will simply use the model and some of the data to illustrate some basic points.

As noted, experimental data are available characterizing the levels of Epo in various pools in three different locations: extracellular, on the membrane, and internal to the cell [25]. Experiments provide these levels at six time points (ranging from approximately 1 to 300 minutes after introduction of Epo) for three replicated versions of the the experiment. From the replicated data, we can estimate the average levels of Epo in each location and at each time point; if we have no intrinsic information about the uncertainties inherent in the experimental data, we can estimate uncertainties in these mean quantities by computing the standard error of the triplicate points. From these 18 data points and their uncertainties, we can estimate the best-fitting set of model parameters. In [15], the parameter B_{max} was fixed based upon estimation from other data, as was the ratio of $K_D = k_{off}/k_{on}$; in this example, I will leave k_{off} and k_{on} as separate parameters and fit them to the time-course data. Thus the set of parameters to fit includes: $k_t, k_{on}, k_{off}, k_e, k_{ex}, k_{di}, k_{de}$. Using SloppyCell, a Python package we have developed to support the simulation and analysis of reaction networks, I can import the SBML version of the core model deposited by the authors at BioModels.net (BIOMD0000000271.xml, which encodes the reaction network and associated kinetic laws in the dynamical equations above), define a model which links together the experimental data and the model, add some priors to keep the parameters within broad ranges identified by the authors, and optimize the parameters to fit the data. Details on these sorts of operations can be found in the SloppyCell user's guide (http://sloppycell.sourceforge.net/user.pdf). The best-fit time-courses for the three observables are shown in Fig. 2(a).

Much of the power that one gains in dissecting complex biological systems arises from the ability to perturb such systems in a sufficiently diverse set of conditions so as to tease apart the contributions of different components and subsystems. If every such condition required estimation of an entirely new and disjoint set of parameters, nothing would be gained by combining different experiments. For example, suppose one could alter the pH within a cell, and conduct a series of experiments at different pH levels, but without having a model that incorporates the effects of pH directly, and with no idea how model parameters should vary as a function of pH. Every experiment would effectively be independent of all others, involving a different and unrelated parameter for

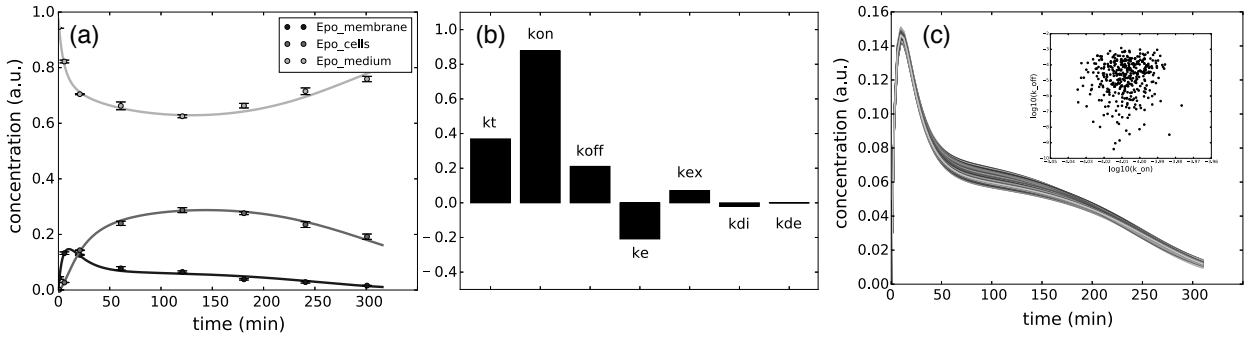

Figure 2: (a) Best-fit time courses in the Epo-EpoR model for the observed data. (b) Stiffest eigenvector of the approximate Hessian $J^T J$ at the best-fit set of parameters. (c) Main panel: Ensemble of trajectories for [Epo_membrane] over a sampled parameter ensemble. Inset: projection of parameter ensemble in the (k_{on}, k_{off}) plane.

each pH, even if all models shared the same underlying mathematical structure.[†] Fortunately, many experimental perturbations are often local in nature, affecting a node or an edge in a biochemical reaction network. In such a situation, one can fit data to different variants of a reaction network, differing only in these local perturbations. Unless there is reason to think otherwise, one can fit these multiple variants of a network – each of which makes different predictions about the specific experimental conditions relevant to that variant – to multiple datasets. More concretely, if we knock out some component of a network, we expect to be able to estimate the same numerical values for parameters elsewhere in the network that are not involved in that knockout. Different model variants that share parameters can typically have those parameters fit to system-specific data across all relevant experimental conditions.

In order to fit model parameters across sets of different conditions, one needs to coordinate the integration of experimental data with the appropriate model variants, running each model variant separately to fit the relevant data. One advantage to using existing tools targeted specifically for parameter estimation in reaction networks is that there is often support for that sort of data integration. The SloppyCell system, for example, stitches together objects representing collec-

[†]In some cases, however, one can relate parametric variation to global environmental parameters, such as temperature. In ref. [26], we modeled temperature compensation in bacterial circadian oscillation networks, modulating chemical activation barriers exponentially with temperature according to the Arrhenius activation law. Even though the model was manifestly dependent on temperature, there exist extended regions of parameter space where the system dynamics is effectively independent of temperature, resulting in a circadian clock unaffected by temperature variation.

tions of Experiments with objects representing collections of Calculations in order to carry out the combined inference. The Data2Dynamics system provides additional intelligence in this regard, automatically creating model variants for different experimental conditions and organizing output data accordingly [27, 25]. These sorts of multi-experiment parameter estimation problems can be complicated in their structure, however: witness the different sets of experiments and estimation processes involved in Epo receptor trafficking discussed in [15]. The field of systems biology has been accelerated through the development of a standard format for the specification of reaction network models (the Systems Biology Markup Language, or SBML (sbml.org)), and there have been related efforts to develop additional ontologies and formats to standardize other aspects of the modeling process (e.g., characterization of simulation time courses, http://sed-ml.org). Given the importance of parameter estimation to the modeling process and the potential complexities inherent in carrying it out in practice, it might be worthwhile for the community to focus on the development of standard formats and data structures to organize parameter estimation activities, to enable the reuse and exchange of data, metadata and models for parameter estimation in the same way that SBML facilitates the reuse of the underlying model specifications themselves.

2.4 Optimization

Abstracted appropriately, the specific form of a mathematical model is irrelevant insofar as the numerical optimization of the cost function is concerned, as long as it can evaluate the least-squares deviation of a model from data for a given set of parameters θ. Optimizing an arbitrary nonlinear function of a set of variables is a widespread problem throughout all of science, and accordingly, much algorithmic and development work has been devoted to producing numerical tools capable of carrying out this essential computational task. Numerical optimization is something of an art: there is a vast set of different algorithms that one might possibly make use of, and determining which is most appropriate for a given problem can require a bit of experimentation. Perhaps the most relevant distinguishing feature among different algorithms are those that are capable of identifying global optima and those that make do with finding local optima. In some cases, there can be

multiple distinct local minima of the cost function $C(\theta)$ separated by barriers of higher cost. This can complicate the process of finding the global minimum, especially since most numerical minimization routines move locally downhill. Some optimization algorithms are in fact designed to avoid getting stuck in local minima, and one can often use local methods in conjunction with multiple independent restarts in order to converge to different minima. Several of those methods are addressed specifically in the context of biological parameter estimation in [28] and [27]. A more subtle complication in finding a global minimum, even in cases where there may be no other competing local minima, arises from a near degeneracy in the cost function near the minimum. This degeneracy is often associated with long, thin canyons in parameter space, resulting in multiple different combinations of parameters that are able to fit the data within the underlying uncertainty, along which the numerical values of individual parameters might vary wildly. I will return to this issue in more detail below in describing parametric sloppiness, but note here that in some cases a detailed understanding of the structure of cost surfaces under evaluation can lead to the development of novel optimization algorithms that exploit that structure, such as those involving geodesic flow along model manifolds [29].

2.5 Sensitivity analysis

Parameter sensitivity analysis quantifies how model trajectories depend upon variations in parameters. Formally, one can differentiate the dynamical equation for a system of interest with respect to parameters, and then reorder derivatives to arrive at a set of sensitivity equations for the parametric derivatives that can be solved alongside the underlying dynamical equations:

$$\frac{\partial}{\partial t}\frac{\partial x}{\partial \theta} = \frac{\partial f}{\partial x}\frac{\partial x}{\partial \theta} + \frac{\partial f}{\partial \theta}$$

Carrying out this sensitivity integration with analytically calculated derivatives is generally preferable to numerically estimating such sensitivities using finite-difference approximations to derivatives by imposing small variations on parameter values. Finite-difference approaches to sen-

sitivity analysis generally do a poorer job of characterizing complicated cost surfaces, especially those plagued with long, thin, shallow canyons. Tools such as SloppyCell and Data2Dynamics calculate these analytic derivatives from the underlying kinetic laws and reaction network topology, supporting sensitivity integration.

One is not limited to characterizing sensitivities with respect to single parameters, and deeper insights come from considering the collective effects of interactions among parameters. This can be done by examining the structure of the Hessian matrix of parametric second derivatives $H_{ij} = \partial^2 C / \partial \theta_i \partial \theta_j$ or its approximation $J^T J$, where J represents the Jacobian matrix of first derivatives. Local sensitivities to combined parameter variation can be assessed by computing the eigenvalues and eigenvectors of H, as described in detail elsewhere [5, 30]. For the Epo-EpoR trafficking model, sensitivity integration can be used to compute $J^T J$, and subsequently its eigenvalues and eigenvectors. Fig. 2(b) shows the "stiffest" parameter combination associated with the fit of the Epo-EpoR model to the experimental data, that is, the eigenvector associated with the largest eigenvalue of $J^T J$. Nonlocal exploration of parameter space can be undertaken using parameter sampling approaches, as described below.

2.6 Parameter sampling, posterior distributions, and identifiability

Identifying a single set of parameters which best fits the available data does not acknowledge that other parameter sets might be almost as effective in describing observations. Within the framework of Bayesian statistics, one aims to ascertain posterior distributions on parameters, that is, the probability of parameters (and hence a model) given the available data. Bayes' theorem, along with the ability to compute the likelihood of the probability of the data given the parameters, allows us to compute the posterior distribution $P(\theta|Data)$:

$$P(\theta|Data) \propto P(Data)P(Data|\theta)$$

Markov Chain Monte Carlo (MCMC) sampling can be used to explore parameter space, sam-

pling from the posterior distribution to identify an ensemble of parameter sets that fit the data within the experimental uncertainty [30, 5]. For the Epo-EpoR model, Fig. 2(b, inset) shows a projection of such an ensemble in the k_{on}, k_{off} plane. While the distribution of k_{on} values is well-localized about a central value, the values of k_{off} vary by several orders of magnitude. In the original Epo-EpoR paper, however, additional binding kinetics data were incorporated into the estimation process that enabled the authors to estimate and fix a value for the dissociation constant $K_D = k_{off}/k_{on}$.

The goal of parameter estimation is to identify the values of model parameters, but this is often confounded for a variety of reasons. The subject of "parameter identifiability" (or, of greater concern, nonidentifiability) addresses such issues [31, 32]. Typically a distinction is made between *structural nonidentifiability* and *practical nonidentifiability*. Structural identifiability refers to the way that two or more parameters enter into the mathematical structure of a model; for example, if two parameters only ever enter a model in terms of their ratio or their product, then the separate values of those parameters cannot be determined. Mathematically, relationships associated with structural nonidentifiabilities correspond to zero modes of the Hessian, that is, eigenvectors of the Hessian matrix of second derivatives of the cost function $H_{ij} = \partial^2 C/\partial\theta_i\partial\theta_j$ with zero eigenvalue. In such cases, it is best to first mathematically reformulate the problem in order to replace nonidentifiable parameters.

Practical nonidentifiability is a more subtle problem, involving parameters that are not strictly degenerate but which are difficult to disentangle in practice. A canonical example of this arises in problems with a separation of time scales, such as in the case of binding-unbinding kinetics that is fast compared to other reactions. I will return to the issue of practical nonidentifiability below in the context of sloppy parameter sensitivities.

2.7 Quantifying prediction uncertainties

One constructs a model in order to make predictions, and thus for those predictions to be testable through further experimentation, it is necessary to indicate what it means for the model to be

wrong. (Of course, "all models are wrong" [33], so it may seem odd that I would need to delineate under what conditions any particular model is wrong.) But to test predictions, we need to quantify the uncertainty of those predictions, given the information they are built upon. Just in the same way that uncertainties and errors can be propagated through simple arithmetic calculations, we can propagate parameter uncertainties (derived from the sorts of sampling procedures described above) through the action of mathematical models, in order to specify model prediction uncertainties.

Either the local analysis of sensitivities, or the nonlocal sampling of parameter space, can be used to estimate prediction uncertainties. Fig. 2(c), for example, shows a set of trajectories for [Epo_membrane] for the parameter sets contained in the sampled ensemble. Even though some parameters in the ensemble vary considerably, the ensemble of trajectories shows much less variation.

It should be noted that much of the machinery described above assumes that there are experimental data available to fit to. Often there are, and clearly one will be on uneasy ground making predictions from models that are not built on a foundation of experimental verifiability. Nonetheless, the reality is that many mathematical models are published with parameters that are not systematically fit to data, but which are instead estimated based on values in the literature or assumed to be of the right order of magnitude based on data. One is often nonetheless interested in understanding the sensitivity of model predictions to parameters (e.g., to suggest possible perturbations of interest), and can use many of the same methods described above. In this case, we are interested not in fitting to experimental data, but rather to synthetic data generated by the model itself.

2.8 Resources for parameter estimation in systems biology

Several excellent treatments and reviews of many of the issues touched upon here are available in the literature, with specific application to mechanistic systems biology models [5, 6, 34, 28]. A summary of lessons learned in computational approaches to parameter estimation – with specific application to the JAK-STAT and EpoR models discussed above – can be found in [27].

In addition, while parameter estimation is a generic problem in many different areas of mathematical modeling, software tools specifically engineered to support parameter estimation and related analyses in the area of systems biology are available in several packages, including Copasi, Data2Dynamics, ABC-SysBio, the MATLAB SimBiology toolbox, and SloppyCell. Many such systems make use of the SBML standard for encoding models of reaction networks, which facilitates not only model interchange but also model meta-analyses such as those contained in [35] and [36]. The Data2Dynamics website contains numerous examples that weave together computational models and associated experimental data for fitting, and the ABC-SysBio system is described in considerable detail in a protocol article describing its use [37].

2.9 Stochastic models

Most of the discussion here has focused on deterministic descriptions of biological processes, whereby a given choice of parameters and initial conditions always leads to the same set of model predictions. Models of this sort, typically in the form of coupled ODEs describing the dynamics of molecular concentrations, are most appropriate when fluctuations are insignificant, although they are sometimes used even if that is not the case due to the relative ease with which they can be analyzed. Investigation of a stochastic model is complicated by the fact that (a) system trajectories will vary from realization to realization [4], and (b) solving the master equation for the joint probability distribution characterizing those variable trajectories becomes prohibitively expensive as the system size grows due to the combinatorial explosion of the configuration space [38]. The process of parameter estimation for stochastic models is further complicated by the fact that estimating the likelihood is not as tractable as for the case of deterministic models with normally distributed noise. A discussion of inference methods for stochastic models is beyond the scope of this article, but several useful theoretical descriptions and computational tools supporting these types of analyses are available [39, 40, 41, 42, 37, 43].

3 Parameter estimation and the process of modeling

Having laid out some of the basic procedures involved in parameter estimation (and pointed to references where much more detailed and expert descriptions are available), I will now step back to examine how the process of parameter estimation relates to other aspects of modeling. As noted above, parameter estimation is not an end in itself, but a means to an end of making predictions about new behaviors or analyzing the structure of existing behaviors. As such, parameter estimation functions as an inner loop in service of these other goals.

3.1 Sloppiness and the geometry of parameter sensitivities

I alluded above to long, thin canyons in high-dimensional parameter spaces, and collective interactions among parameters in the sensitivity of model predictions to parameter variation. These features are related to a phenomenon that my colleagues and I have termed "sloppiness", whereby system dynamics is very insensitive to some particular combinations of parameter changes while being very sensitive to others; see Fig. 3(a). This wide range of sensitivities, involving a few stiff modes that determine system dynamics and a large space of sloppy modes that do not impact system behavior, is seen in a number of complex, multiparameter, nonlinear models [30, 35, 2]. This anisotropy in parameter sensitivities can be quantitatively characterized by the eigenvalue spectrum of the Hessian matrix as described above, which has a characteristic form where the eigenvalues decay by a roughly constant factor, implying that only a few stiff modes contribute significantly [44]. Parameter ensembles generated by MCMC sampling can exhibit large variation in individual parameters, some fluctuating by many orders of magnitude over an ensemble [45]. These parameters are practically nonidentifiable, but it is possible for a model with a set of widely divergent parameter estimates to nonetheless show well-constrained model predictions. This is counterintuitive, since one usually expects that if one puts junk into a model (in the form of uncharacterized parameter values) one would expect junk out (in the form of useless predictions). But sloppiness reveals that due to internal redundancies and correlations in parameter sensitivities,

constrained and therefore testable predictions can often be made from sloppy models. More recent work by Transtrum, Machta and Sethna has extended this theory by combining insights from sloppy models with techniques from differential geometry and information geometry to characterize the underlying manifolds associated with how models map from parameter space to data space [29, 46].

There are a variety of implications arising from this characteristic geometric structure of sensitivities. Unless all model parameters can be well-constrained through other measurements, model predictions can vary substantially if even one parameter is not so constrained [35]. Efficient sampling of parameter space in MCMC requires taking step sizes in different directions consistent with the underlying anisotropy [48]. As described in some more detail below, simplified models with better-identified parameters can be constructed using model reduction procedures that leverage the hierarchical structure of model manifolds. And we have hypothesized that sloppiness in complex biological networks might provide a mechanism for evolution to explore extended parametric neutral spaces, allowing for robustness to some types of parameter variation while enabling evolvability of new phenotypes [26].

3.2 Model inference, refinement, selection and reduction

Traditionally, parameter estimation has served as a separate inner loop within the broader processes of model construction, inference, refinement, selection and reduction. One posits a model structure, attempts to fit parameters in that model, and refines the model structure if it is unable to describe the data. A model successfully fit to data can be used to make predictions about unseen scenarios, although whether those predictions are falsifiable depends on how constrained they are by the available data. (New scenarios might involve, for example, different initial conditions, or the knockout, inhibition or overexpression of particular model components.) Upon making a falsifiable prediction about a new condition and testing it experimentally, one will find either that the prediction is validated, or that it is not. Those experiments might suggest possible changes to the model structure, as well as contribute additional data to feed into the parameter estimation

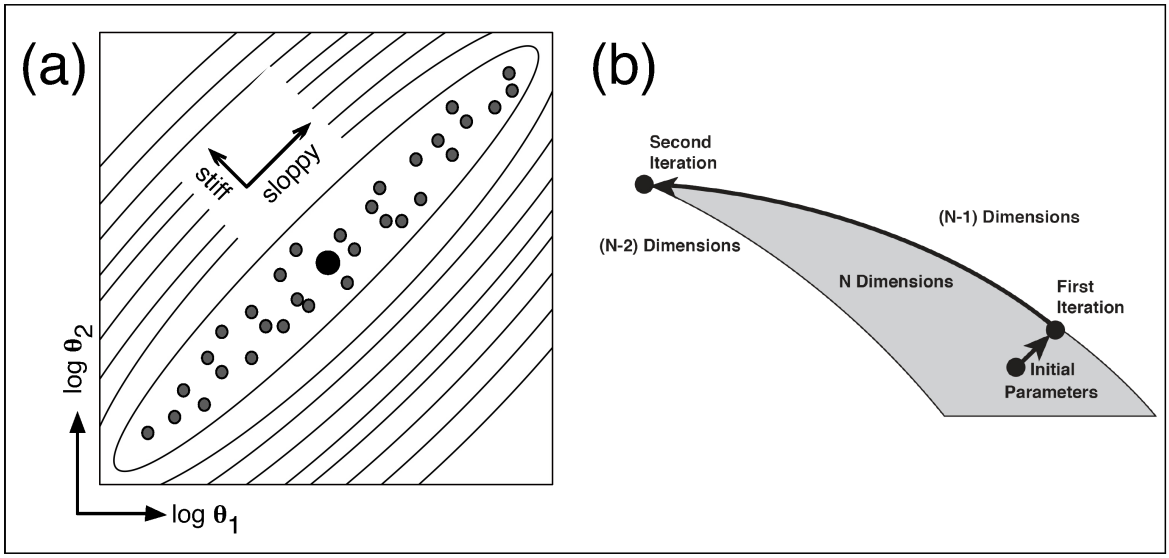

Figure 3: (a) Schematic of the cost surface of a sloppy model (adapted from [35]), projected on a two-dimensional parameter subspace (θ_1, θ_2). Curved lines/ellipses represent contours in the cost surface $C(\theta)$ in the vicinity of the best-fit set (large black dot). Stiff directions are directions in parameter space where system behavior changes rapidly with parameters; in sloppy directions, behavior is largely insensitive to parameter variations. The anistropy of the cost contours is proportional to the square root of the eigenvalues of the Hessian matrix. Gray dots represent other parameter sets near the cost minimum, generated by MCMC sampling. (b) Schematic depicting model reduction via the manifold boundary approximation method (adapted with permission from Ref. [47] – copyrighted by the American Physical Society). The manifold associated with N-dimensional parameterization of the full model is bounded by lower-dimensional facets, edges and corners. Model reduction proceeds by repeatedly flowing to boundaries and lumping parameters together in accordance with the zero mode that emerges at the boundary, resulting at each step in a new model with one less parameter.

pipeline.

A key aspect of model inference, as a first step, and model refinement subsequently is experimental design [49, 50]. In some cases, separate experiments can be carried out in order to estimate particular parameter values and then fix those values in subsequent efforts to estimate other parameters from different types of data; the multistage estimation efforts carried out in [15] and [51] are representative of this approach. A significant part of the experimental design literature aims to address issues of parameter non-identifiability. Within the context of the sorts of sloppy parameter sensitivities described above, papers by Tonsing *et al.* [52], Apgar *et al.* [53], and Hagen *et al.* [54] have demonstrated that appropriate design of experiments can reduce most of the uncertainty in parameter estimates observed in sloppy models, albeit by introducing more complex sorts of experiments [5]. That is, despite the ubiquity of sloppiness arising in parameter estimation for many different models, it is not intrinsic, as both model reparameterizations and targeted experiments can result in identifiable parameters. Motivated by the sense that constraining prediction uncertainties is a more fruitful endeavor than constraining parameter uncertainties *per se*, Casey *et al.* used experimental design techniques to propose and test new measurements needed to optimally reduce predicition uncertainties for a molecular complex identified as relevant to the regulation of EGF receptor trafficking [24].

Model selection refers broadly to the process of deciding from among a set of competing models based on their ability to describe available data and to make testable predictions about unseen data. Likelihood ratio tests are one common method for formally characterizing the goodness of fit of different competing models; the early papers on modeling the JAK-STAT pathway, for example, show this practice in great detail [14]. In situations where a large amount of training data are available, cross-validation is a common approach to model comparison: some portion of the available data are included in a training set for parameter estimation, while the remaining data are left out for validation. Unfortunately, while technological advances have placed some facets of biology squarely in the land of Big Data, parameterization of many complex dynamical models of biological networks is still usually on more barren terrain, often with barely more data points than

parameters to be estimated.

The term "model selection" also refers somewhat more specifically to methods for deciding among different structural models that contain different numbers of parameters. Within this latter context, one acknowledges that structurally different models are inherently unequal, in that one might expect a more complex model to be more able to fit existing data simply because it contains more fitting degrees-of-freedom. (Although the caveats of having complicated models that overfit the data must be heeded.) In this sense, model selection echoes Occam's razor, a principle encouraging the development of parsimonious descriptions, in which "Entities must not be multiplied beyond necessity" [55]. Additional model complexity in the form of parameters should be pursued only if those additional degrees-of-freedom provide sufficient extra predictive capacity to warrant their inclusion, and various methods have been developed which effectively penalize larger models. Some mathematical approaches are based either on information theory (leading, for example, to the Akaike Information Criterion (AIC) for model selection) or Bayesian statistics (resulting in the Bayesian Information Criterion (BIC)). An excellent recent overview of methods for model selection, with specific application to the sorts of dynamical models of interest in systems biology, can be found in [34].

While the inner loop of parameter estimation is relatively automated and robust (the challenges described above notwithstanding), the process of model refinement can be rather slow, laborious and requiring human intervention. This has led some researchers to develop creative new procedures that aim to couple the processes of structural inference, parametric inference, model selection and model reduction into more synergistic wholes. The automated discovery of model forms highlighted by Francois and colleagues [56, 57] integrates parameter estimation directly into the broader mode of structural inference. Kuepfer *et al.* advocated the role of "ensemble modeling" that examines not just sets of parameters within a single model structure, but more broadly ensembles of model structures [58]. Sunnaker *et al.* have combined exploration of parameter space with simplication of models, recognizing that quantitative changes in some parameters that act as edge weights connecting different components can lead to qualitative and topological changes in

reaction networks with fewer edges [59].

A promising new model reduction procedure (the manifold boundary approximation method), recently introduced by Transtrum and Qiu [47], makes use of insights into the hierarchical hyperribbon structure of model manifolds in complex models. These high-dimensional manifolds, which map from parameter space to data space, are typically bounded by lower-dimensional surfaces (facets, edges, corners, etc.) that represent simpler models with fewer parameters. The interior of the manifold is characterized by a number of sloppy modes with small but not strictly zero eigenvalues. The manifold boundaries, on the other hand, are associated with zero modes in the full model. The model reduction procedure involves flowing along geodesics on the manifold until a boundary is reached, and inspecting the resulting zero mode that separates out from the spectrum. The associated eigenvector describes a relationship among a subset of parameters that can be used for parameter removal. For example, the binding-unbinding kinetics in the Epo-EpoR model described above can be simplified – in the limit that the rates k_{on} and k_{off} are much faster than other rates – by a single parameter, the dissociation constant $K_D = k_{off}/k_{on}$. This simplification of the model corresponds to a bounding facet of the manifold with one fewer parametric degree-of-freedom than the original model. This procedure can be repeated, leading to smaller and smaller models, as depicted schematically in Fig. 3(b); the process is terminated when all effective parameters in the model can be identified from the available data and any further reductions lead to a loss of predictive power. Not only can complex models be considerably simplified, but the effective parameters that result through such a procedure are themselves functions of the underlying bare parameters that were introduced in the original model [60].

3.3 The many faces of systems biology

Systems biology is a field of study that is broadly interested in understanding how the vast array of genomic and molecular components that make up living organisms are organized to produce the bewildering variety of phenotypic behaviors exhibited by those systems, in short, how genotype is mapped to phenotype. Perhaps not surprisingly for a field with such lofty goals, the manner in

which different people seek to reach those goals vary considerably. For some, the ability to gather large amounts of data through the use of various high-throughput experimental techniques is the dominant theme, with the hope that insights can come from mining patterns in those data. Although in many cases, such analyses do little more than apply relatively simple statistical techniques to identify or confirm interesting trends in the data, without digging down into the mechanistic layers that are responsible for producing phenotype from genotype. For others, systems biology suggests the ability to move beyond molecular "parts lists" to construct cellular "wiring diagrams", complex dynamical descriptions that can be investigated to characterize the emergent phenomena that drive living systems. Others motivated to understand such emergence, however, are deeply skeptical of such models, in large part because so many parameters are unknown; for them, characterizing the phenomenology of living systems is paramount, even if the resulting models might be more difficult to link directly to the sorts of data generated by high-throughput assays and the sorts of microscopic manipulations that are in the molecular biologist's toolbox. Still others aim to develop mechanistic models within other mathematical frameworks that do not rely on the proliferation of unknown parameters characteristic of chemical kinetic networks. These include boolean models that discretize state spaces and transition rules [61], and in the arena of metabolic modeling, constraint-based models that compute flux distributions consistent with a set of stoichiometric and flux-bound constraints [62].

Model parameterization and parameter estimation are on the front lines of these sorts of debates. Complex wiring diagrams are easy to construct yet difficult to parameterize. For some, a model with nonidentifiable parameters is a nonstarter, although as noted above, it is still possible to have predictive models with poorly characterized parameters. Rather than simply rejecting overly complicated models in favor of parsimonious phenomenological models, insights into sloppiness in complex biological systems allow us to migrate along this spectrum, identifying model structures with lower effective dimensionality that can help point the way toward phenomenological models. Model reduction procedures that retain the full underlying parameter space in effective parameters allow bridges to be built between mechanism and phenomenology [60].

In addition, top-down statistical inference from large data sets can potentially make use of machine learning techniques focused on dimensionality reduction: Big Data are not always as big as they seem, but often reside on lower-dimensional subspaces. Developing better methods for incorporating prior information about mechanistic processes into statistical inferences from system-wide datasets remains a big challenge for the field, but perhaps bridges could be built from mechanism to inference by integrating our insights about low-dimensional subspaces in mechanistic models with discovery of reduced-dimensional statistical descriptions of large datasets.

As prosaic as the parameter estimation process is, it sits in many ways on a knife edge that cuts across many important themes in the modern world of quantitative biology. Appreciating both the mechanics of parameter estimation and its place in the larger realm of modeling, experimental design, and interpretation of data are keys to making further progress in this area.

Acknowledgements

This review would not be possible were it not for the work I have been fortunate to carry out over several years along with a number of collaborators: Kevin Brown, Josh Waterfall, Ryan Gutenkunst, Fergal Casey, Bryan Daniels, Yan-Jiun Chen, Ben Machta, Mark Transtrum, Rick Cerione and Jim Sethna. In addition, I acknowledge support from NSF grant IOS-1127017.

References

[1] R M Pirsig. *Zen and the Art of Motorcycle Maintenance: An Inquiry into Values*. William Morrow, 1974.

[2] M K Transtrum, B B Machta, K S Brown, B C Daniels, C R Myers, and J P Sethna. Perspective: Sloppiness and emergent theories in physics, biology, and beyond. *The Journal of Chemical Physics*, 143(1):010901, 2015.

[3] S Singh, D J Schneider, and C R Myers. Using Multitype Branching Processes to Quantify Statistics of Disease Outbreaks in Zoonotic Epidemics. *Physical Review E*, 89(3):032702, 2014.

[4] D T Gillespie. Exact stochastic simulation of coupled chemical reactions. *The Journal of Physical Chemistry*, 1977.

[5] B K Mannakee, A P Ragsdale, M K Transtrum, and R N Gutenkunst. Sloppiness and the geometry of parameter space. In L Geris and D Gomez-Cabrero, editors, *Uncertainty in Biology: A Computational Modeling Approach*, pages 271–299. Springer International Publishing, Cham, 2016.

[6] T Toni and M P H Stumpf. Parameter inference and model selection in signaling pathway models. *Methods in Molecular Biology*, 673:283–295, 2010.

[7] J S Rawlings. The JAK/STAT signaling pathway. *Journal of Cell Science*, 117(8):1281–1283, 2004.

[8] D S Aaronson. A Road Map for Those Who Don't Know JAK-STAT. *Science*, 296(5573):1653–1655, 2002.

[9] MT Laub and M Goulian. Specificity in two-component signal transduction pathways. *Annual Review of Genetics*, 41(1):121–145, 2007.

[10] Y F Qi, Y X Huang, H Y Wang, Y Zhang, Y L Bao, L G Sun, Y Wu, C L Yu, Z B Song, L H Zheng, Y Sun, G N Wang, and Y X Li. Elucidating the crosstalk mechanism between IFN-gamma and IL-6 via mathematical modelling. *BMC Bioinformatics*, 14:41–41, 2013.

[11] C E Shannon. A mathematical theory of communication. *Bell System Technical Journal*, 27:379–656, 1948.

[12] S Itzkovitz, T Tlusty, and U Alon. Coding Limits on the Number of Transcription Factors. *BMC Genomics*, 7(1471-2164):239, 2006.

[13] C R Myers. Satisfiability, sequence niches and molecular codes in cellular signalling. *IET Systems Biology*, 2(5):304–312, 2008.

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

[15] V Becker, M Schilling, J Bachmann, U Baumann, A Raue, T Maiwald, J Timmer, and U Klingmüller. Covering a Broad Dynamic Range: Information Processing at the Erythropoietin Receptor. *Science*, 328(5984):1404–1408, 2010.

[16] J Timmer, T G Müller, I Swameye, and O Sandra. Modeling the nonlinear dynamics of cellular signal transduction. *International Journal of Bifurcation and Chaos in Applied Sciences and Engineering*, 2004.

[17] Z Zi and E Klipp. SBML-PET: a Systems Biology Markup Language-based parameter estimation tool. *Bioinformatics*, 22(21):2704–2705, 2006.

[18] R N Gutenkunst, J C Atlas, F P Casey, B C Daniels, R S Kuczenski, J J Waterfall, C R Myers, and J P Sethna. SloppyCell, 2007. http://sloppycell.sourceforge.net.

[19] C R Myers, R N Gutenkunst, and J P Sethna. Python unleashed on systems biology. *Computing in Science and Engineering*, 9(3):34–37, 2007.

[20] A Raue, B Steiert, M Schelker, C Kreutz, T Maiwald, H Hass, J Vanlier, C Tönsing, L Adlung, R Engesser, W Mader, T Heinemann, J Hasenauer, M Schilling, T Höfer, E Klipp, F Theis, U Klingmuller, B Schöberl, and J Timmer. Data2Dynamics: a modeling environment tailored to parameter estimation in dynamical systems. *Bioinformatics*, 31(21):3558–3560, 2015.

[21] A Raue, C Kreutz, T Maiwald, J Bachmann, M Schilling, U Klingmuller, and J Timmer. Structural and practical identifiability analysis of partially observed dynamical models by exploiting the profile likelihood. *Bioinformatics*, 25(15):1923–1929, 2009.

[22] J Vanlier, C A Tiemann, P A J Hilbers, and N A W van Riel. A Bayesian approach to targeted experiment design. *Bioinformatics*, 28(8):1136–1142, 2012.

[23] M L Skoge, R G Endres, and N S Wingreen. Receptor-receptor coupling in bacterial chemotaxis: evidence for strongly coupled clusters. *Biophysical Journal*, 90(12):4317–4326, 2006.

[24] F P Casey, D Baird, Q Feng, R N Gutenkunst, J J Waterfall, C R Myers, K S Brown, R A Cerione, and J P Sethna. Optimal experimental design in an epidermal growth factor receptor signalling and down-regulation model. *IET Systems Biology*, 1(3):190–202, 2007.

[25] Data2Dynamics. http://data2dynamics.org.

[26] B Daniels, Y Chen, J P Sethna, R N Gutenkunst, and C R Myers. Sloppiness, Robustness, and Evolvability in Systems Biology. *Current Opinion in Biotechnology*, 19(4):389–395, 2008.

[27] A Raue, M Schilling, J Bachmann, A Matteson, M Schelke, D Kaschek, S Hug, C Kreutz, B D Harms, F J Theis, U Klingmüller, and J Timmer. Lessons learned from quantitative dynamical modeling in systems biology. *PLoS ONE*, 8(9):e74335, 2013.

[28] G Cedersund, O Samuelsson, G Ball, J Tegnér, and D Gomez-Cabrero. Optimization in biology parameter estimation and the associated optimization problem. In Liesbet Geris and David Gomez-Cabrero, editors, *Uncertainty in Biology: A Computational Modeling Approach*, pages 177–197. Springer International Publishing, Cham, 2016.

[29] M K Transtrum, B B Machta, and J P Sethna. Geometry of Nonlinear Least Squares with Applications to Sloppy Models and Optimization. *Physical Review E*, 83(3 Pt 2):036701, 2011.

[30] K S Brown and J P Sethna. Statistical mechanical approaches to models with many poorly known parameters. *Physical Review E*, 68(2 Pt 1):021904, 2003.

[31] O-T Chis, J R Banga, and E Balsa-Canto. Structural Identifiability of Systems Biology Models: A Critical Comparison of Methods. *PLoS ONE*, 6(11):e27755, 2011.

[32] A Raue, V Becker, U Klingmuller, and J Timmer. Identifiability and observability analysis for experimental design in nonlinear dynamical models. *Chaos*, 20(4):045105, 2010.

[33] G E P Box. Science and statistics. *Journal of the American Statistical Association*, 71(356):791–799, 1976.

[34] M Sunnåker and J Stelling. Model extension and model selection. In L Geris and D Gomez-Cabrero, editors, *Uncertainty in Biology: A Computational Modeling Approach*, pages 213–241. Springer International Publishing, Cham, 2016.

[35] R N Gutenkunst, J J Waterfall, F P Casey, K S Brown, C R Myers, and J P Sethna. Universally Sloppy Parameter Sensitivities in Systems Biology Models. *PLoS Computational Biology*, 3(10):e189 EP –, 2007.

[36] K Erguler and M P H Stumpf. Practical limits for reverse engineering of dynamical systems: a statistical analysis of sensitivity and parameter inferability in systems biology models. *Molecular BioSystems*, 7(5):1593–1602, 2011.

[37] J Liepe, P Kirk, S Filippi, T Toni, C P Barnes, and M P H Stumpf. A framework for parameter estimation and model selection from experimental data in systems biology using approximate Bayesian computation. *Nature Protocols*, 9(2):439–456, 2014.

[38] M J Keeling and J V Ross. On Methods for Studying Stochastic Disease Dynamics. *Journal of The Royal Society Interface*, 5(19):171–181, 2008.

[39] D J Wilkinson. *Stochastic Modelling for Systems Biology, Second Edition (Chapman & Hall/CRC Mathematical and Computational Biology)*. CRC Press, 2011.

[40] E L Ionides, C Bretó, and A A King. Inference for nonlinear dynamical systems. *Proceedings of the National Academy of Sciences*, 103(49):18438–18443, 2006.

[41] T Toni, D Welch, N Strelkowa, A Ipsen, and M P H Stumpf. Approximate Bayesian computation scheme for parameter inference and model selection in dynamical systems. *Journal of the Royal Society, Interface / the Royal Society*, 6(31):187–202, 2009.

[42] A A King, E L Ionides, C M Bretó, S Ellner, and B Kendall. *pomp: Statistical inference for partially observed Markov processes (R package)*. URL http://pomp. r-forge. r- . . . , 2010.

[43] A Golightly and D J Wilkinson. Bayesian parameter inference for stochastic biochemical network models using particle Markov chain Monte Carlo. *Interface focus*, 1(6):807–820, 2011.

[44] JJ Waterfall, FP Casey, RN Gutenkunst, KS Brown, CR Myers, PW Brouwer, V Elser, and JP Sethna. Sloppy-model universality class and the Vandermonde matrix. *Physical Review Letters*, 97(15):150601, 2006.

[45] K S Brown, C C Hill, G A Calero, C R Myers, K H Lee, J P Sethna, and R A Cerione. The statistical mechanics of complex signaling networks: nerve growth factor signaling. *Physical Biology*, 1(3):184, 2004.

[46] B B Machta, R Chachra, M K Transtrum, and J P Sethna. Parameter Space Compression Underlies Emergent Theories and Predictive Models. *Science (New York, NY)*, 342(6158):604–607, 2013.

[47] M K Transtrum and P Qiu. Model reduction by manifold boundaries. *Physical Review Letters*, 113(9):098701, 2014.

[48] R N Gutenkunst, F P Casey, J J Waterfall, C R Myers, and J P Sethna. Extracting falsifiable predictions from sloppy models. *Annals of the New York Academy of Sciences*, 1115(1):203–211, 2007.

[49] D Silk, P D W Kirk, C P Barnes, T Toni, and M P H Stumpf. Model selection in systems biology depends on experimental design. *PLoS Computational Biology*, 10(6):e1003650, 2014.

[50] M K Transtrum and P Qiu. Optimal experiment selection for parameter estimation in biological differential equation models. *BMC Bioinformatics*, 13:181, 2012.

[51] E Lee, A Salic, R Krüger, R Heinrich, and M W Kirschner. The roles of APC and axin derived from experimental and theoretical analysis of the Wnt pathway. *PLoS Biology*, 1(1):e10, 2003.

[52] C Tönsing, J Timmer, and C Kreutz. Cause and cure of sloppiness in ordinary differential equation models. *Phys. Rev. E*, 90:023303, Aug 2014.

[53] J F Apgar, D K Witmer, F M White, and B Tidor. Sloppy models, parameter uncertainty, and the role of experimental design. *Molecular BioSystems*, 6(10):1890, 2010.

[54] D R Hagen, J K White, and B Tidor. Convergence in parameters and predictions using computational experimental design. *Interface Focus*, 3(4):20130008, 2013.

[55] Occam's razor. https://en.wikipedia.org/wiki/Occam%27s_razor.

[56] P Francois and E D Siggia. Predicting embryonic patterning using mutual entropy fitness and in silico evolution. *Development*, 137(14):2385–2395, June 2010.

[57] P François, V Hakim, and E D Siggia. Deriving Structure from Evolution: Metazoan Segmentation. *Mol Syst Biol*, 3, 2007.

[58] L Kuepfer, M Peter, U Sauer, and J Stelling. Ensemble modeling for analysis of cell signaling dynamics. *Nature Biotechnology*, 25(9):1001–1006, 2007.

[59] M Sunnåker, E Zamora-Sillero, R Dechant, C Ludwig, A G Busetto, A Wagner, and J Stelling. Automatic generation of predictive dynamic models reveals nuclear phosphorylation as the key msn2 control mechanism. *Science Signaling*, 6(277):ra41, 2013.

[60] M K Transtrum and P Qiu. Bridging mechanistic and phenomenological models of complex biological systems. *PLOS Computational Biology*, 12(5):1–34, 05 2016.

[61] J Thakar and R Albert. Boolean Models of Within-host Immune Interactions. *Current Opinion in Microbiology*, 2010.

[62] E Bogart and C R Myers. Multiscale metabolic modeling of C4 plants: connecting nonlinear genome-scale models to leaf-scale metabolism in developing maize leaves. *PLoS ONE*, 11(3):e0151722, 2016.