The inverse problem in mathematical biology

Gilles Clermont[1] **Sven Zenker**[2]

[1]Crisma Center, Departments of Critical Care Medicine, Mathematics, and Chemical Engineering, University of Pittsburgh, 200 Lothrop St, Pittsburgh, PA 16123

[2]Department of Anesthesiology and Intensive Care Medicine, University of Bonn Medical Center, Sigmund-Freud-Str. 25, 53105 Bonn, Germany

Abstract

Biological systems present particular challengers to model for the purposes of formulating predictions of generating biological insight. These systems are typically multi-scale, complex, and empirical observations are often sparse and subject to variability and uncertainty. This manuscript will review some of these specific challenges and introduce current methods used by modelers to construct meaningful solutions, in the context of preserving biological relevance. Opportunities to expand these methods are also discussed.

Introduction

Biological systems at all levels of organizations are modeled for the purposes of gaining insight, testing hypotheses, or formulating predictions on future states of these systems following their normal evolution, or consequent to some outside perturbation or experiment. Depending on the intent of the modeler, models are formulated with varying levels of complexity. Simulation refers to applying a model in order to extract predictions. Depending on the nature of the model, these predictions are associated with varying levels of certainty, which may be quantifiable with appropriate methodology. This is the forward problem [1]. Before a model can be used for offering predictions, it must be appropriately parameterized. The activity of parameterizing a model from empirical observations, which may involve various tasks depending on the precise mathematical formulation of this model, constitutes the inverse problem [2,3]. Depending on the particular biological system and its precise model representation, solving the inverse problem ranges from manageable, to impractical, to impossible. Solutions to the inverse problems present significant challenges for all but the most simple of biological systems. While most of the effort in systems biology has focused on the interface of complex biological networks and high-throughput data [4,5], biological problems at other scales of description and triangulated by much sparser empirical datasets have received relatively little attention [1].

Posing the problem

Mathematical formulation

This manuscript will discuss solutions of biological problems that are mathematically represented by systems of difference or differential equations [6], or even more generally by sets of rules defining interactions such as present in rule-based or agent-based models [7,8]. With little loss of generality, let's assume that a biological system is modeled using a system of ordinary differential equations:

$$\dot{\underline{X}} = f(\underline{X}, \underline{p}, t) + \underline{U}(\underline{X}, t) + \underline{N}(t)$$
$$\underline{O} = g(\underline{X})$$

Where $\dot{\underline{X}}$ is the first time derivative a vector of model variables, $\underline{O}$ is a vector of observables on the biological system being modeled, $\underline{U}(\underline{X},t)$ a vector of external controls or influences on the system, $\underline{p}$ is a vector of parameters, $\underline{N}(t)$ are noise terms, the function f embodies the biology governing the interaction between the different variables on the system, and g is a map between system observables and model variables. For the model to be usable for prediction, additional data, typically in the form of a vector of initial conditions $\underline{X}_{t=0}$, turning the forward problem into an initial value problem, need to be specified. In addition, a number of observations are made on the system across several experimental conditions E which systematically or randomly vary initial conditions, boundary conditions, or manipulate the temporal evolution of the system in some other way. In its simplest form, solving the inverse problem therefore involves the definition of a function $\underline{\Psi}(\underline{O}, \underline{p}, f, \underline{U})$ quantifying the error between model predictions and observations across all experiments, and identifying $\underline{p}$, but occasionally also f or $\underline{U}$ that will minimize this function, resulting in a model that best matches all available data across experiments. The careful construction of $\underline{\Psi}$ and how its various components are combined and weighted when multiple objectives are pursued is of paramount importance in biological systems, involving layers of considerations beyond identifying local minima of the error.

Sources of uncertainty

Any quantitative scientist collaborating with biological or social scientists is awed by the character of the empirical data derived from apparently perfectly reproducible experiments. Biological systems are complex and noisy. Noise may appear to govern a lot of the observed dynamics, either because of actual random dynamics of the data generating processes or because a significant portion of what drives the dynamical evolution of the system is either not represented or implicit in the model. In addition, there is generally an observability problem, that is, the knowledge obtained from the available observation insufficiently constrains the inverse problem of identifying states and/or parameters. Sources of variability in the data can be grouped under three major categories: (1) the experiment is difficult to reproduce because a number of known factors cannot be easily controlled, (2) the measurement is difficult to reproduce because instruments and tests have intrinsic variability, and (3) there is residual variability stemming from a large number of unknown and possibly uncontrollable factors, including the highly variable makeup of biological organisms, which can oftentimes turn out to be the dominant source of variability overall. There are also

additional sources of variability that could be present in longitudinal data. Small animal experiments often require animal sacrifice at preset time points since the measurements cannot be performed while leaving the animal intact, so a given animal only provides data at a single time point. Many experiment in live animals may also result in death, caused by an experimental manipulation or not. Data from late time points are therefore populated by animals that survived to that time point, and data may therefore be biased in favor of animals naturally more resistant. A separate source of uncertainty resides in lack of a full understanding of the exact relationships between observables and model variables. In biological systems, an identity relationship is often assumed, thereby further simplifying biology. Data sparsity is also a common source of uncertainty for the modelers. It is therefore centrally important that experimental data be obtained over the dynamic range of the biological response. This problem is particularly pernicious when dealing with observational studies or clinical data. An example of this is related to efforts to model the molecular response to severe infections in humans: most humans with severe infections are first encountered well into the disease process, and at different, and unknown, times along that process [9].

Some of these shortcomings in data can be alleviated with careful experimental design or more sophisticated statistical treatment. Yet most cannot be addressed effectively, especially in situations where the experimental and modeling groups do not work in close and bilateral collaboration, particularly when data has already been collected. Any attempt at solving inverse problems must address these shortcoming in such a way as to map data uncertainty into model uncertainty with minimal bias.

A satisfactory solution to the inverse problem

A more precise formulation of the inverse problem in the context of biological systems may thus be more appropriately stated along the following lines: given the data at hand and its limitation, and given prior knowledge of the system one is trying to model, what is the least biased model representation of this system that can be offered. Whether this representation is useful in the broader sense remains to be investigated and will generally depend on the nature of the mechanistic insight or predictions being sought. A vast class of models offer a linear representation of biological systems. Given sufficient data, the optimization problems arising from the corresponding inverse problems are convex, that is admit a single, global, optimal parametrization. One must realize that such representations are in fact a choice made by the modeler and that the actual system might only be approximately linear under very restricted biological operating conditions, a specific instance of the general fact that any mathematical representation of reality will involve simplifying assumptions. Such local linearization, performed explicitly or implicitly in the modeling process, is a useful tool as long as it is recognized that as soon as the system is perturbed away from the validity of this approximation, the model may lose any potential validity and thus usefulness, which will almost always be the case when studying biological systems under stress caused by environmental alterations, disease, etc..

More broadly, the modeler may also mandate or make the implicit assumption that, despite the existence of a large number of local optima, there exist among them a best one that

nature has somehow adopted. An example of such a situation is instantiated in the protein folding problem where generally, the best local optimum (lowest energy) is promoted to global optimum within a range of conformations of interest and predicted to be the three-dimensional conformation of a sequence of amino acids. In this particular situation, folding is a process which is itself highly optimized requiring the presence of chaperon proteins at different steps of a growing primary sequence of amino-acids and thus the final product is "guided" to a local energy minimum. Therefore, approaching folding as a local optimization problem is sensible. Despite the knowledge of the existence of a large number of other minima, they are not favored by nature where function dictates structure. Such a conformation, although stable upon small perturbations representative of routine biological function, is not upon larger, biologically irrelevant perturbations such as heating beyond temperatures compatible with life.

Therefore, although methods for estimating convex or "almost" convex systems are plentiful, well understood and described and implemented in major software packages, they are of limited use and interest in estimating complex biological systems. Rather, we are interested in a class of problems where there are many, potentially infinitely many, solutions to the inverse problem, that is ill-posed problems, and where there is no clear biological indication as to which of those solutions are biologically implausible. Techniques have been developed to construct ensemble models, where each member of the ensemble is a parameter vector for a given a model structure [1,10–13]. Ensemble modeling, although computationally demanding, represents the most satisfactory solution to ill-posed inverse problems to date [14]. This suite of methods is undergoing active development in the context of complex biological systems and presents a number of open problems still to be addressed effectively (see below).

Parameter uncertainty and model identification

There exists good reviews on the topic of model identifiability [15]. Model identifiability is generally defined along an axis ranging from formal, or a priori global identifiability, to practical identifiability. A priori global identifiability or structural identifiability stipulates whether, given a model structure and a set of parameters and initial conditions, model parameters can always be uniquely identified assuming sufficient data. Several formal methods exist to resolve global and local structural identifiability, each having strength and weaknesses depending on the mathematical structure of the problem [16]. Generally, the more complex the system (e.g. presence of Hill or Michaelis-Menten kinetic terms) and the higher the number of parameters compared to observables, the more difficult it will be to resolve structural identifiability of a system. Formal methods based on differential algebra [17](e.g.) have become popular recently. Software implementing differential algebra [18,19] (http://www.dei.unipd.it/~pia/), generating series [20], or a combination of approaches [21] have been developed to investigate a priori identifiability in systems with dynamics described by systems of equations and can be implemented successfully by non-experts, although software output typically requires critical interpretation [16].

Indeed, when a system is not identifiable, such methods can often identify algebraic relationships between parameters which are identifiable even when the parameters

themselves are not. A trivial and intuitive illustration would be: one attempts to model a cloud of points in a plane with a model of the form $y = (a \cdot b)x + b$. The algebraic expression $a \cdot b$ is identifiable, but obviously not the individual parameters comprising this expression if the only empirical data available to estimate the system parameters are (x, y) pairs. Such methods have limited applicability to models of biological systems. More generally, there has been much literature on practical identifiability, the notion that even a structurally identifiable system cannot be practically identified if the number or quality of empirical observations on the system is insufficient. In the simple example above, if the uncertainty on y is such that one cannot say with certainty if the slope is positive or negative, then the model is of limited usefulness if the ultimate purpose on the model is to predict y given x.

Unfortunately, the situation with biological systems is slightly different. Observations are limited and uncertain, combined with the fact that even simplified models of these systems are typically *a priori* unidentifiable. Yet, even under these adverse conditions there might be sufficient residual identifiability to achieve meaningful predictions from even grossly over-parameterized and poorly specified models. The focus shifts from the ability to robustly identify model parameters and their variance to the ability of the model to generate predictions, even in the absence on robust parametric estimates. This class of problems, referred to as sloppy problems, are encountered regularly in mathematical biology [4] and probably constitute the rule rather than the exception. As we will see below, there is a very real possibility that, depending on their intended use, sloppy models are good enough and useful for critical objectives. Indeed, poorly specified model may still be perfectly adequate for some practical applications, but not for others. Some examples come to mind. At a fundamental level, there still no definitive theory of gravitation. Yet, incrementally as theories have evolved and become more sophisticated, and observational data has become extremely accurate, the predictions of contemporary macroscopic gravitational physics, in spite of the fact that they completely lack any thus far accessible connection to the underlying microscopic quantum reality, have achieved outstanding precision, with no practical limitations in terms of its application in everyday human endeavors. Similarly, considering the human body as a complex system, contemporary physicians must estimate the macroscopic state of this system (diagnosis) such that proper corrective action is undertaken (therapy) given the estimated evolution of this state (prediction). Despite more accurate diagnostic methods, knowledge of the current microscopic state remains very partial under the best of circumstances, and complete knowledge may be utterly unattainable. Yet, arguably, corrective actions appropriately improve patients trajectories in a, surprisingly maybe, large proportion of instances, even if inference and prediction in clinical medicine today is, in most instances, not performed quantitatively using mathematical tools, as has been common practice in physics for centuries. This extreme case points out a valid, yet largely unresolved question with respect to modeling biological systems: how accurate must a model be to be effectively useful?

Model validation and selection

The concept of model validation can intuitively be stated as the process of verifying the ability of a model to provide sufficiently accurate predictions and is to be distinguished from the ability of the model to reproduce the data used to estimate this model, which is termed

goodness-of-fit. For statistical models, the broadest concept of validation is referred to as external validity, or generalizability: can the model accurately predict an outcome, or a time course in data not used to estimate the model? It is assumed that a valid statistical model has effectively learned the structure of a multi-dimensional joint distribution of outcomes and predictor variables, such that, if presented with a vector of predictions, outcomes are robustly inferred. For types of models that describe the evolution of interacting variables over time using mechanistic rules rather than statistical inference, the approach pioneered in physics, the concept of model validation differs essentially from statistical model validation. A valid model will in fact "understand" the system. A first level of validation resides in the *a priori* construct validity of the dynamical model: are the interactions included in the formulation of the model plausible representations of what is believed to be known about the biological system being modeled? Validity in the statistical sense is a consequence of a dynamical model "understanding" the biological system, while both the inability to fit the data and statistical invalidity, which can be present in spite of perfect ability to fit the data [22], falsify the combination of assumptions made in the formulation of a mechanistic model and may thus provide useful mechanistic insight. However, a well-fitting model that is valid in the statistical sense may or may not be valid under the more stringent requirement made of mechanistic models of dynamical systems: can a dynamical system predict the outcome of a new experiment, the data from which was unavailable when estimating the model? This step, along with the ability to incorporate pre-existing knowledge of a system's structure in a principled way, is probably the single most important reason why mechanistic models have attracted the attention of biologists and clinicians. If one could use a valid mathematical model, one could predict the effect of a new drug, or of a new experiment, provided the system is modeled in sufficient detail to represent the biological effect of this new drug [23,24], or the proposed experimental manipulation.

Validation of a dynamical system should be conceived as an iterative process: erroneous or inaccurate predictions, provided general construct validity and good fit on prior experiments, begs the question as to why the model failed to "understand" this new experiment. This generally indicates an opportunity to reexamine finer details of construct validity: the experiment draws from mechanisms that the model fails to include or fails to understand, indicating that the model must be corrected. The specifics of the failure may well guide how models need to be adjusted.

Model selection comes into play when the task is to choose among a set of structurally distinct models, given some observational data along with methodology to solve the corresponding forward, and, if necessary, inverse problems. A plethora of quantitative statistical approaches has been proposed for this problem type [25–28], including the Bayes Information Criterion and the Akaike Information Criterion, and the Bayes factor. The fundamental idea behind the majority of these approaches is that, when choosing a model, a balance needs to be found between model complexity, typically quantified by the number of free parameters, and goodness-of-fit, since the goodness-of-fit will usually increase monotonically with the number of free parameters, at the expense of parsimony and usually model identifiability. In a sense, model selection thus aims to quantitatively implement Occam's razor. While the available formulaic model selection criteria may certainly contribute to informing the model development process, we suggest that in the

context of mechanistic models of biology, where the solution of the inverse problem is one of the major challenges on the path to obtaining useful predictions and also a major source of prediction uncertainty, model selection should preferentially be informed directly by quantifying the model's predictive performance on real data given the available solution methodology.

Individual models and personalized medicine

In statistical regression model theory, there generally exists a single model structure which explains the relationship between predictors and outcome. Personalization is conveyed through the use of mixed-effect models. An individual, or a group of individuals, characterizes a random effect. This approach has been extended to dynamical models such as ordinary differential equations, where individual rate parameters k_i are estimated. Depending on the estimation method, the individual rates can be estimated directly from individual data, or evaluated in several stages, where a mean population rate k_p is first computed, to serve as prior guess for k_i, with the advantage that individual with poor data quality can still benefit from a robust prior. Alternatively, and more in line with statistical mixed-effect modeling, individual deviations $k_i \equiv k_p + \Delta k_i$ or $k_i \equiv k_p^{\Delta k_i}$ can be computed, assuming k_p has already been computed using data from a population of individuals. Much of these approaches have been developed and studied in the context of population pharmacokinetic models [29], where modelers can typically benefit from dense time series from individual patients and structurally simple models.

However, these methods can generate similar approaches, within the framework of ensemble models, to personalization. In this context, an ensemble of population parameters are computed from a naïve pool of data from the entire population, and in analogy with multistage Bayesian estimation of mixed effect models, patient or subgroup-specific ensembles are computed as a posterior distribution, using data from an individual, or pooled data from a homogeneous subgroup of individuals, and the population ensemble as prior distribution. Models for individuals (subgroups) with sparse data would resemble the population ensemble, while ensemble models for individuals (subgroups) with richer data could vary substantially from the population ensemble model.

A frequent misconception is that, in the process of computing a population ensemble from pooled data, a particular instance of a parameter set can be interpreted as representing a potential individual member of the population. This is an incorrect interpretation of the ensemble modeling methodology, unless it can ascertained that the ensemble is a representative sample of the population distribution, which, in general, is extremely challenging to achieve, see below. On the other hand, mining subgroup ensembles could provide an alternative way to statistical data mining in identifying mechanistically-based [30], meaningful phenotypes such as possible drug responders, or identifying groups of, as individuals that could potentially be harmed by a potential intervention [23].

Unsolved problems in inverse problem theory

Evaluating fits of models to empirical data is a routine task of model development. The error between a model's prediction and data is typically quantified as the sum of squared errors between prediction and data, where the sum ranges over all observables, all time points, and all experimental conditions. This is the error function we previously discussed. This deviation, termed the objective function, is then mapped to a likelihood of the model, given the data. This map is typically assumed to take the form of a Gibbs distribution, as this effectively implements the assumption that uncertainty in observations are normally distributed. Although, such a formulation allows easy computation of the relative likelihood of two parameter sets, this assumption if often not met in biological systems. . Thus, it remains unclear whether there are preferred means of constructing the objective function, given the biological system to be modeled, or the structure of the model representation of this system? In the absence of data on specific experiments of interest, there may still be knowledge of the overall behavior of a system, given some initial conditions or control conditions of interest. For example, biological constraints might limit the magnitude or dynamic rate of growth/decline of certain unobserved states, thus imposing restrictions on parameter values. How do modelers best include this general knowledge in the estimation process? Some have adopted the approach of mitigating against unlikely model behavior by imposing heuristic penalties within the objective function $\underline{\Psi}$ [31]. These may take a variety of forms, including Jacobian evaluation to ensure that a particular parameter set will provide asymptotically stable solutions, or the insertion of suitable weighed fictional data points implementing such heuristics. The construction of a meaningful objective function $\underline{\Psi}$, and a likelihood based on $\underline{\Psi}$, especially in the context of multiple objectives, or in the context of multiple experiments with differential reliability, remains an art more than a science at this point in time, although quantitative information about, e.g., measurement error distributions, if available, can and should of course be incorporated into $\underline{\Psi}$ in a principled way. In the process of pooling data from several individuals, there is typically a cross-correlation structure across the observed times series expressing the join distribution of observations in the population. Therefore, naïve pooling of the data, with the implicit assumption of uncorrelated errors between time-point of a single time-series and across time-series inflates actual variance in the data. Yet, methods that leverage the existing structure on this join distribution in the construction of $\underline{\Psi}$ are not widely implemented.

In Bayesian parameter estimation, prior knowledge about parameters and states, if expressible in the form of probability distributions, can naturally be incorporated into the inference process. However, the generic expression of the absence of prior knowledge is an open question. This is the problem of choosing a non-informative prior distribution. Although the common practice of setting wide bounds de facto constitutes a weak prior, more can be done depending on the structure of uncertainty on the data. There exists a broad literature on the choice of un-informative priors for Bayesian inference [32], providing the user with a choice of approaches supported by, among others, reparametrization invariance arguments, as is the case for Jeffrey's prior. Given the relative sparsity of data characterizing many biological systems, the choice of prior is likely to weigh heavily on the posterior distribution and should be performed with care. Yet, without providing details here, it would

seem very important to explore this question theoretically on the basis of how the model structure maps volume elements of parameter space to data space and vice versa, given that these spaces are often of very different dimensionality [33].

Conclusion

Modeling complex biological systems is a difficult exercise, particularly when one intends to parameterize such models for generating actual predictions, suggesting experimental design, or for providing mechanistic insight. In this manuscript, we outline key obstacles to generating such models, some of which lie in the nature of biological data, and some of which lie in the theoretical and computational nature of the ill-posed estimation process. We also outline the importance of collaborative groups with strong data acquisition, modeling and computation expertise, where such models can be iterated and validated, e.g. [34].

References

[1]. Zenker S, Rubin J, Clermont G, From inverse problems in mathematical physiology to quantitative differential diagnoses, PLoS Comput Biol. 3 (2007) e204–e204. [PubMed: 17997590]

[2]. Tarantola A, Inverse Problem Theory and Methods for Model Parameter Estimation, SIAM, 2005.

[3]. Kirsch A, An Introduction to the Mathematical Theory of Inverse Problems, Springer Science & Business Media, 2011.

[4]. Gutenkunst RN, Waterfall JJ, Casey FP, Brown KS, Myers CR, Sethna JP, Universally sloppy parameter sensitivities in systems biology models, Plos Comput. Biol. 3 (2007) 1871–1878. [PubMed: 17922568]

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

[6]. Daun S, Rubin J, Vodovotz Y, Clermont G, Equation-based models of dynamic biological systems., J. Crit. Care. 23 (2008) 585–594. doi:10.1016/j.jcrc.2008.02.003. [PubMed: 19056027]

[7]. An G, Mi Q, Dutta-Moscato J, Vodovotz Y, Agent-based models in translational systems biology, Wiley.Interdiscip.Rev.Syst.Biol.Med. 1 (2009) 159–171. [PubMed: 20835989]

[8]. Faeder JR, Blinov ML, Hlavacek WS, Rule-Based modeling of Biochemical Systems with BioNetGen, Methods Mol.Biol. 500 (2009) 113–167. [PubMed: 19399430]

[9]. Kellum JA, Kong L, Fink MP, Weissfeld LA, Yealy DM, Pinsky MR, et al., Understanding the inflammatory cytokine response in pneumonia and sepsis: results of the Genetic and Inflammatory Markers of Sepsis (GenIMS) Study, Arch.Intern.Med. 167 (2007) 1655–1663. [PubMed: 17698689]

[10]. Daun S, Rubin J, Vodovotz Y, Roy A, Parker R, Clermont G, An ensemble of models of the acute inflammatory response to bacterial lipopolysaccharide in rats: results from parameter space reduction, J Theor Biol. 253 (2008) 843–853. [PubMed: 18550083]

[11]. Mochan E, Swigon D, Ermentrout GB, Lukens S, Clermont G, A mathematical model of intrahost pneumococcal pneumonia infection dynamics in murine strains., J. Theor. Biol. 353 (2014) 44–54. doi:10.1016/j.jtbi.2014.02.021. [PubMed: 24594373]

[12]. Song SO, Hogg JS, Parker RS, Peng Z, Kellum JA, Clermont G, Ensemble models of neutrophil trafficking in severe sepsis, PLoS Comput Biol. 8 (2012) e1002422–e1002422. [PubMed: 22412365]

[13]. Toni T, Stumpf MP, Parameter inference and model selection in signaling pathway models, Methods Mol.Biol. 673 (2010) 283–295. [PubMed: 20835806]

[14]. Swigon D, Ensemble modeling of biological systems, in: Antoniouk AV, Melnik RVN (Eds.), Math. Life Sci, The Publishing House, Berlin, 2012: p. 316.

[15]. Audoly S, Belu G, Angio L, Saccomani MP, Cobelli C, Global Identifiability of Nonlinear Models of Biological Systems, IEEE Trans. Biomed. Eng. 48 (2001) 55–65. [PubMed: 11235592]

[16]. Chis O-T, Banga JR, Balsa-Canto E, Structural identifiability of systems biology models: a critical comparison of methods., PLoS One. 6 (2011) e27755. doi:10.1371/journal.pone.0027755. [PubMed: 22132135]

[17]. Moles CG, Mendes P, Banga JR, Parameter estimation in biochemical pathways: a comparison of global optimization methods., Genome Res. 13 (2003) 2467–74. doi:10.1101/gr.1262503. [PubMed: 14559783]

[18]. Bellu G, Saccomani MP, Audoly S, Angio L, DAISY: A New Software Tool to Test Global Identifiability of Biological and Physiological Systems, Comp.Meth.Prog.Biomed. 88 (2007) 52–61.

[19]. Saccomani MP, Audoly S, Bellu G, D'Angiò L, Examples of testing global identifiability of biological and biomedical models with the DAISY software., Comput. Biol. Med. 40 (2010) 402–407. doi:10.1016/j.compbiomed.2010.02.004. [PubMed: 20185123]

[20]. Chi O, Banga JR, Balsa-Canto E, GenSSI: a software toolbox for structural identifiability analysis of biological models., Bioinformatics. 27 (2011) 2610–1. doi:10.1093/bioinformatics/btr431. [PubMed: 21784792]

[21]. Balsa-Canto E, Banga JR, AMIGO, a toolbox for advanced model identification in systems biology using global optimization., Bioinformatics. 27 (2011) 2311–3. doi:10.1093/bioinformatics/btr370. [PubMed: 21685047]

[22]. Chow CC, Clermont G, Kumar R, Lagoa C, Tawadrous Z, Gallo D, et al., The acute inflammatory response in diverse shock states, Shock. 24 (2005) 74–84. [PubMed: 15988324]

[23]. Clermont G, Bartels J, Kumar R, Constantine G, Vodovotz Y, Chow C, In silico design of clinical trials: a method coming of age, Crit Care Med. 32 (2004) 2061–2070. [PubMed: 15483415]

[24]. An G, Bartels J, Vodovotz Y, In Silico Augmentation of the Drug Development Pipeline: Examples from the study of Acute Inflammation, Drug Dev.Res. 72 (2011) 187–200. [PubMed: 21552346]

[25]. Moses T, Holland PW, A comparison of statistical selection strategies for univariate and bivariate log-linear models, Br J Math Stat.Psychol. (2009).

[26]. Akaike H, A New Look at the Statistical Model Identification, IEEE Trans.Aut.Control. AC-19 (1974) 716–723.

[27]. Berger JO, Pericchi LR, Objective bayesian methods for model selection: introduction and comparisons, (2001) 137–193.

[28]. Toni T, Stumpf MPH, Parameter inference and model selection in signaling pathway models., Methods Mol. Biol. 673 (2010) 283–95. doi:10.1007/978-1-60761-842-3_18. [PubMed: 20835806]

[29]. Pillai GC, Mentré F, Steimer J-L, Non-linear mixed effects modeling - from methodology and software development to driving implementation in drug development science, J. Pharmacokinet. Pharmacodyn. 32 (2005) 161–83. doi:10.1007/s10928-005-0062-y. [PubMed: 16283536]

[30]. Prince JM, Levy RM, Bartels J, Baratt A, Kane JM III, Lagoa C, et al., In silico and in vivo approach to elucidate the inflammatory complexity of CD14-deficient mice, Mol Med. 12 (2006) 88–96. [PubMed: 16953560]

[31]. Hancioglu B, Swigon D, Clermont G, A dynamic model of human immune response to influenza A virus infection, J. Theor. Biol. 246 (2007) 70–86. [PubMed: 17266989]

[32]. Jeffreys H, Theory of probability, Clarendon Press, Oxford Oxfordshire, 1983.

[33]. Zenker S, Rubin J, Clermont G, Unbiased inference of parameter distributions for nonlinear models: The underdetermined case, J. Crit. Care. 23 (2008).

[34]. Vodovotz Y, Clermont G, Hunt CA, Lefering R, Bartels J, Seydel R, et al., Evidence-based modeling of critical illness: an initial consensus from the Society for Complexity in Acute Illness, J.Crit Care. 22 (2007) 77–84. [PubMed: 17371750]

Highlights

- Estimation of computational models of biological systems poses formidable challenges

- Mechanistic inference usually require practical parameter identifiability

- Prediction of system evolution may not require parameter identifiability

- A satisfactory solution to the inverse problem awaits further theoretical progress