

Druckfreigabe/approval for printing	
Without corrections/ ohne Korrekturen	<input type="checkbox"/>
After corrections/ nach Ausführung der Korrekturen	<input type="checkbox"/>
Date/Datum: .....	
Signature/Zeichen: .....	

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45

**Part Three**  
**Analysis of Gene Networks**

UNCORRECTED PROOF

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45

UNCORRECTED PROOF

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45

## 9 What if the Fit is Unfit? Criteria for Biological Systems Estimation Beyond Residual Errors

*Eberhard O. Voit*

### 9.1 Introduction

The analysis of biological networks has made enormous strides in recent years. In the context of static networks, which do not change over short periods of time, new biological techniques have begun to permit the characterization of very large interaction maps (e.g., among proteins), and computational graph theory has been the tool of choice for analyzing and interpreting what these maps entail. Complementing these activities has been the exploration of dynamically changing, regulated biological systems. On the experimental side, these efforts have enormously benefitted from the astounding advances in high-throughput biology at the genomic, proteomic, metabolic, and physiological levels. On the analytical side, the procedures and results of this new field of experimental systems biology have first been supported by a rapidly expanding repertoire of bioinformatics tools that permit the storage, retrieval, and analysis of very large datasets. More recently, the bioinformatics tools have become tightly interwoven with analytical and simulation techniques that are at the heart of the emerging field of computational systems biology.

At this point in time, it is no longer a real challenge to simulate large linear or even nonlinear systems in the form of algebraic or differential equations. It has also become feasible to simulate hybrid systems that contain continuous and discrete events, stochastic effects, and delays (e.g., [1, 2]). The simplicity with which we can perform large-scale simulations is in stark contrast to the overwhelming challenges we face much earlier in any biological systems analysis, namely when the biological phenomenon of interest is to be translated into a mathematical or computational model. This translation task may be subdivided into three aspects. The first concerns the acquisition of data. While biology is producing high-quality data in large quantities, these data are not always of the type and completeness that elucidate all aspects of the biological phenomenon from sufficiently many angles to construct a mathematical model. For instance, models of dynamic processes in the brain are hampered by the extremely difficult access to specific, restricted neuronal areas in living organisms [3]. This aspect of data availability will without doubt continue to

1 improve throughout the foreseeable future and we will not discuss issues of data  
2 generation in this chapter. The other two components of the translation from biology  
3 into a computational construct are distinct, but closely related to each other. The first  
4 is the determination of suitable mathematical descriptions for all relevant details of  
5 the biological phenomenon, while the second is the identification of numerical values  
6 for the parameters in these descriptions [4]. These two fundamental tasks of  
7 computational systems biology are the focus of this chapter.

8 Before we discuss details, challenges, methods, and pitfalls associated with the  
9 construction of biological systems models, we should ask why such an effort appears  
10 to be worth our while. One might begin by pre-empting a widespread critique of  
11 modeling, namely that models merely recreate, often in a much abstracted and  
12 simplified fashion, what “real” biologists had known all along and in greater detail. So  
13 what, if a model produces results similar to those observed? Modelers are sometimes  
14 stunned by this critique, because it is certainly not a trivial matter to write computer  
15 code that fits a large collection of biological data well. The truth behind this (mis-)  
16 conception is that a well-fitting model is a necessary but not sufficient condition for  
17 greater things to come. Indeed, without further analysis, exploration, explanation, or  
18 prediction, an accurate fit by itself does not earn the modeler much more than  
19 bragging rights.

20 In generic terms, the construction of a model is worthwhile if the model is able to  
21 answer specific questions or helps decide between acceptance and rejection of a  
22 hypothesis. Such a hypothesis may take many different forms. It may be qualitative in  
23 a sense that one is primarily interested in whether some key variable in the system  
24 responds to a specific input with an increase or a decrease. It may be semiquantitative  
25 if one is interested in the rough extent of the response and it is quantitative or  
26 numerical if the model is supposed to show the correct value of the affected key  
27 variable. What level of accuracy is needed in a model result depends on the questions  
28 asked, on the effort one is willing to invest in the modeling effort, and on many issues  
29 associated with the biological phenomenon itself and with the model.

## 32 9.2 33 Model Design

34  
35 The challenge of converting a biological system into a mathematical structure  
36 requires the specification of functions that describe all pertinent processes, as well  
37 as the identification of suitable parameter values. The selection of process descrip-  
38 tions is by no means trivial. Granted, there are situations where a function can be  
39 inferred from the type and mechanism of the process. For example, there is good  
40 reason to choose an exponential function for the description of the growth of a small  
41 bacterial population, because the process is biologically driven by repeated cell  
42 doubling. However, such cases of mechanism-based model selection are actually  
43 rare and even traditional choices like a Michaelis–Menten rate function for an  
44 enzyme-catalyzed reactions are not without troubling questions, because deep  
45 underlying assumptions like homogeneity of the medium and free movement of

1 enzymes and substrates are usually not satisfied *in vivo* [5]. Physics affords us with a  
 2 rich repertoire of proven formulations for fundamental characteristics like forces and  
 3 energy, but in biology these fundamental aspects are convoluted and often mixed  
 4 together in a complicated manner. As an example, consider the process of gene  
 5 expression, which involves the opening of the DNA strands, the right spatial and  
 6 temporal availability and action of transcription factors, and the complex process of  
 7 transcription into RNA. It is simply impossible to reduce this collective event into  
 8 mechanistic pieces that permit elementary, physics-based representations.

9 Faced with similar challenges, engineers typically resort to linear approximations.  
 10 These are very convenient, because there are stringent rules for their design, as well  
 11 as for their analysis. Indeed, the repertoire of analytical and computational methods  
 12 for linear systems is huge. The problem with linear approaches in biology is that most  
 13 phenomena are genuinely nonlinear. They saturate or oscillate in a stable fashion,  
 14 show switches, and sometimes appear to be chaotic. Reducing their dynamics to  
 15 linear functions would not permit a proper analysis of these features. At the same  
 16 time, the number of nonlinear functions is infinite and there are no guidelines as to  
 17 which of these might be optimal or even appropriate descriptions of biological  
 18 processes. A useful alternative is a nonlinear approximation. The first idea presumably  
 19 coming to mind might be a second-order (quadratic) approximation, but this  
 20 choice actually turns out to be rather inconvenient for later analyses [6]. Instead, it has  
 21 proven beneficial to approximate biological processes with linear functions in  
 22 logarithmic coordinates. This procedure is mathematically sound, as it directly  
 23 adheres to the tenets of Taylor's theory and leads to nonlinear descriptions that  
 24 can capture all types of responses, including different types of oscillations and  
 25 chaos [7, 8]. Besides, these representations have desirable properties for mathematical  
 26 and computational analysis. The concept of linearization in logarithmic coordinates  
 27 is the core of biochemical systems theory (BST) [9, 10], which has been  
 28 documented in several hundred articles and book chapters; book-length descriptions  
 29 include [11–14].

30 BST comes in two main variants. In the generalized mass action (GMA) formulation,  
 31 every process is represented with one product of power-law functions. For  
 32 instance, in the simple branched pathway with two feedback signals that is shown in  
 33 Figure 9.1, the equation for  $X_2$  can be formulated directly as:

$$34 \quad \dot{X}_2 = \gamma_{21} X_1^{f_{21}} - \gamma_{22} X_2^{f_{22}} - \gamma_{23} X_2^{f_{22}} X_4^{f_{24}} \quad (9.1)$$

35 where the  $\gamma$  parameters denote rate constants, which can take any non-negative  
 36 values, while  $f_{21}$ ,  $f_{22}$ ,  $f_{22}$ , and  $f_{24}$  are kinetic orders that may take any real values. As  $X_1$   
 37 is the substrate of the production reaction,  $f_{21}$  is positive. By contrast,  $f_{24}$  is negative,  
 38 because it represents the inhibitory signal exerted by  $X_4$ . In general, a GMA system  
 39 always has the format:  
 40

$$41 \quad \dot{X}_i = \gamma_{i1} \prod_{j=1}^{n+m} X_j^{f_{j1}} \pm \gamma_{i2} \prod_{j=1}^{n+m} X_j^{f_{j2}} \pm \dots \pm \gamma_{ik} \prod_{j=1}^{n+m} X_j^{f_{jk}} \pm \dots \quad i = 1, \dots, n \quad (9.2)$$

42 In addition to the dynamically changing variables, the system may also contain  
 43 independent variables,  $X_{n+1}, \dots, X_{n+m}$ , which affect the system, but are not  
 44  
 45

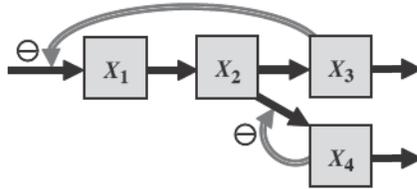


Figure 9.1 Generic pathway with one branch and two feedback signals.

affected by the system. In many cases, these variables are constant during a given mathematical experiment.

In the alternative S-system formulation, all processes entering a variable or pool are collectively represented with a single product of power-law functions that contains all variables affecting the collection of fluxes. Similarly, all processes leaving a variable or pool are collectively represented with a single product of power-law functions that contains all variables affecting the collection of fluxes. Revisiting the pathway in Figure 9.1, the only difference to the GMA formulation occurs for the degradation of  $X_2$ , which is now represented with only one term that contains both  $X_2$  and  $X_4$ . Using the conventional parameter names for S-systems, the equation for  $X_2$  is therefore:

$$\dot{X}_2 = \alpha_2 X_1^{g_{21}} - \beta_2 X_2^{h_{22}} X_4^{h_{24}} \quad (9.3)$$

All other equations are the same as before, with the minor deviation of traditionally different names for the parameters. Accounting again for independent variables, the generic S-system format is:

$$\dot{X}_i = \alpha_i \prod_{j=1}^{n+m} X_j^{g_{ij}} - \beta_i \prod_{j=1}^{n+m} X_j^{h_{ij}} \quad i = 1, \dots, n \quad (9.4)$$

BST models have a number of advantages over *ad hoc* formulations. First, in order to formulate the system equations one does not have to know the true mechanisms governing the phenomenon of interest. As long as it is clear how the system is “connected” (i.e., which component affects which other components), either as a source of material or as the source of a regulatory signal, it is a straightforward procedure to set up a symbolic model. Here, “symbolic” means that we formulate the equations of the model, but that we do not know what the values of their parameter are; examples are Equations 9.1–9.4. The second advantage of BST models is that each parameter, even if it does not yet have a specific value, has a clearly defined meaning. An example is  $f_{24}$ , which exclusively represents the inhibition by  $X_4$  of the conversion between  $X_2$  and  $X_4$ . Similarly, it is immediately clear into which parameter a particular feature of a biological system has to be translated. This one-to-one relationship between biological aspects and parameters is very helpful both during the model design and the interpretation of results from the model analysis. The third advantage of the BST formulation is that the particular format has convenient mathematical properties. The main disadvantage of BST is that there is no guarantee that the model will capture all relevant features of the biological system of interest with sufficient accuracy. While the Taylor approximation is mathematically guaranteed to match any

1 data close to a chosen operating point, this guarantee extends only over a very small  
2 range. Outside this range, which is of unknown size, the model may or may not fit  
3 well. BST shares this feature with all other models in biology.

4 BST models are just one option for nonlinear representations, but they are  
5 especially powerful as initial default models when not much is known about  
6 the details of the biological system. Specifically, since the construction of the  
7 symbolic equations is essentially automatic, the model design challenge is reduced  
8 to determining optimal parameter values with which the model matches the  
9 observed data well. This determination, while still difficult, is to be seen in  
10 comparison with the task of setting up a model with unspecified functions that are  
11 supposed to capture the dynamics of complex and ill-characterized processes. Many  
12 methods are available for parameter estimation purposes (for a recent review, see [4]),  
13 but none of them works perfectly or even satisfactorily in cases of moderately large  
14 biological systems.

### 17 9.3 18 Concepts and Challenges of Parameter Estimation

19  
20 Methods of parameter estimation for biochemical systems fall into two broad  
21 categories, which are directly tied to the available types of data (e.g., see chapter 5  
22 in [14]). In the past, the data almost always consisted of kinetic features associated  
23 with a particular step in the biochemical pathway system. Such features included the  
24  $K_M$  of the enzyme, sometimes a flux rate or  $V_{max}$ , a dissociation constant, or some  
25 other kinetic characteristic. Also, once in a while steady-state values for the variables  
26 or fluxes of the system were available. Given such data, the “bottom-up” strategy of  
27 parameter estimation consisted of formulating each step symbolically and optimiz-  
28 ing the parameter values such that they matched the alleged shape of the reaction  
29 step. Subsequently, all “local” descriptions (of individual reaction steps) were merged  
30 into a system of differential equations that described the entire pathway, the  
31 equations were integrated, and the numerical solutions of the system were compared  
32 against additional observation data. As the comparison typically led to numerical  
33 inconsistencies, one had to go back to the individual process representations many  
34 times and adjust functions or parameter values. This iterated reformulation and data  
35 matching could easily take months if not years. The vast majority of all existing  
36 biochemical models have been estimated with methods of this type; a very detailed  
37 example is [15].

38 Recent advances in molecular biology have begun to offer an alternative “top-  
39 down” approach to parameter estimation. The data here consist of measurements of  
40 metabolite concentrations at many subsequent time points, usually following some  
41 stimulus. The estimation now occurs in the “opposite” direction. Namely, one  
42 attempts to determine parameter values such that the solutions of the differential  
43 equations match the observed time-series data. The local parameters, which earlier  
44 were the starting point, are now the result of the top-down estimation. This type of an  
45 estimation task is known in mathematics and computing as an inverse problem, for

1 which many algorithms are available. However, these algorithms tend to fail for  
 2 moderately complicated biological systems [4].

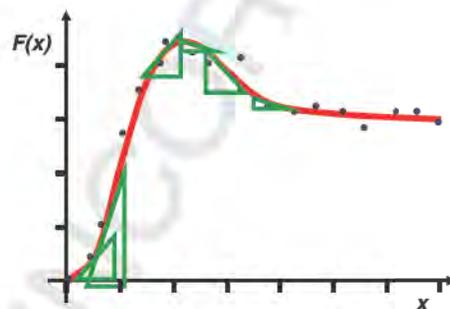
3 As the identification of parameter values is a severe bottleneck of systems biology,  
 4 enormous efforts have been dedicated to the development of general and specific  
 5 techniques. One crucial issue is the fact that essentially all estimation algorithms are  
 6 iterative. They use some method to determine a candidate set of parameter values,  
 7 solve the differential equations with these parameter values, compare the solution  
 8 with the observed data, and evoke some method for improving the parameter set for  
 9 the next iteration. As an immediate consequence, the set of differential equations  
 10 must be solved thousands of times and while each solution may be relatively fast,  
 11 their collection can become prohibitively long. To circumvent this particular issue,  
 12 two independent groups [16, 17] proposed almost 30 years ago the smoothing of the  
 13 raw data and the interpretation of the slopes of the time course for each metabolite as  
 14 estimates for the differentials on the left-hand sides of the differential equations  
 15 (Figure 9.2). Thus, according to this method, each differential equation is replaced  
 16 with a set of  $K$  algebraic equations, where  $K$  is the number of time points where the  
 17 metabolite concentrations and slopes are measured or estimated. Each of the  
 18 algebraic equations contains on the left-hand side the estimated slope value  $S(t_k)$   
 19 at a given time point  $t_k$  and on the right-hand side the expression given by the  
 20 differential equation and evaluated at  $t_k$ . Pursuing this strategy, estimating the  
 21 parameters  $p_1, \dots, p_M$  in the differential equation:  
 22

$$23 \quad \dot{X}_i = f_i(X_1, X_2, \dots, X_n; p_{i1}, \dots, p_{iM_i}) \quad (9.5)$$

24 becomes a task involving a system of  $K$  algebraic equations of the form:

$$25 \quad S_i(t_k) \approx f_i(X_1(t_k), X_2(t_k), \dots, X_n(t_k); p_{i1}, \dots, p_{iM_i}), \quad i=1, \dots, n, \quad k=1, \dots, K \quad (9.6)$$

26 As the estimation is based on observed time courses, all quantities  $X_i(t_k)$  and  $S_i(t_k)$   
 27 are known or estimated, and the only unknowns are the parameter values of  
 28  $p_1, \dots, p_M$ .  
 29  
 30  
 31

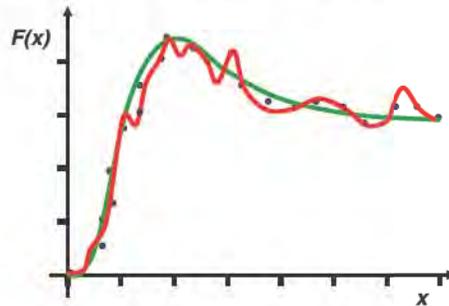


43  
 44  
 45  
**Figure 9.2** Generic univariate dataset (blue dots), a smoothing function (red), and examples of estimated slopes (indicated by green triangles) along the time course. The slopes thus estimated are used as surrogates of the differentials in the differential equations of the systems model.

1 This slope estimation strategy has two significant advantages. First, the need  
2 for solving differential equations is eliminated and the estimation is done instead  
3 with a purely algebraic system of equations. This conversion is very consequential,  
4 because in excess of 95% of the estimation time for a system of differential  
5 equations is spent on numerically integrating the equations, and this percentage  
6 can approach 100% [18]. Second, each time course is “decoupled” from the  
7 others, because it can be addressed independently of all other time courses.  
8 Intriguingly, if this method of slope estimation and decoupling is applied to  
9 Lotka–Volterra models, the initially nonlinear parameter estimation task becomes  
10 a simple linear regression [19]. Similarly, applied to S-systems within BST, the  
11 nonlinear estimation can be converted into an iteration of linear regression tasks [20]  
12 (see also [21]).

13 Clearly, the computational speed-up of the slope estimation and decoupling  
14 strategy is very appealing, but one might wonder about its statistical rigor. For a long  
15 time, the method was seen as such a convenient shortcut that it would outweigh  
16 possible concerns of statistical bias. The argument was that the solution obtained  
17 with this method could at least be used as a starting point for more conventional and  
18 possibly less biased estimations. However, Brunel recently showed that the procedure  
19 is asymptotically normal, consistent, and indeed statistically sound [22].

20 As the optimization of parameters in algebraic equations is much easier than in  
21 differential equations, one might be tempted to assume that the estimation task is  
22 essentially solved. However, this is not always the case. The slope estimation and  
23 decoupling strategy consists of two key steps – the smoothing of the data, which is  
24 necessary for the estimation of slopes, and the parameter estimation of the systems of  
25 nonlinear algebraic equations. The smoothing step is typically achieved with  
26 splines [23], although more sophisticated methods have also been proposed for this  
27 purpose [24, 25]. The choice of a spline or another smoother necessarily requires a  
28 decision regarding the degree of desired smoothness, and this decision cannot be  
29 made with total objectivity, because the smoothness of the data – or the lack thereof –  
30 depends on the processes governing the system from which the data were obtained,  
31 including experimental noise. At the same time it is quite evident that the degree of  
32 smoothing will affect the second step of slope estimation. No matter which method is  
33 chosen, smoothing incurs an approximation error, which is generally larger for  
34 smoother splines that are of lower order and consist of fewer pieces. If more spline  
35 pieces or higher-order splines are used, the approximation error is generally lower,  
36 but the appearance of the smoothing function is “bumpier” (Figure 9.3). This  
37 bumpiness may be due to time courses with many true ups and downs or to  
38 experimental noise in the data. If the latter is the case, a bumpy smoother simply  
39 tries to mimic the noise, the slopes along the smoothed time course increase and  
40 decrease to an unreasonable extent, and the subsequent parameter estimation results  
41 in a larger residual error. Thus, if the data contain even moderate noise, the  
42 smoothing error and the parameter estimation error are inversely related to each  
43 other, and the two must be weighed against each other. Addressing this issue with  
44 statistical rigor, Ramsay *et al.* developed algorithms that optimize weights associated  
45 with the two types of error [26]. A remaining question in this context is to what degree



**Figure 9.3** Different degrees of smoothing incur different residual errors. A smooth interpolation (green) generally results in a larger residual error with respect to the data (blue dots), whereas the more detailed

smoother (red) is associated with a smaller error but a “bumpier” appearance, which may cause problems with the slope estimation and decoupling technique.

different extents of smoothing may lead to a loss of biological information, especially if the true “nature” of the data is unknown.

Overall, the slope estimation and decoupling strategy makes the estimation task simpler and much faster, but it does not solve all problems. We discuss some remaining issues in the following.

### 9.3.1

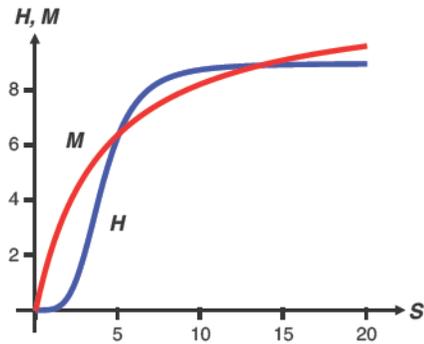
#### Typical Parameter Estimation Problems

The problems typically encountered in the estimation of parameter values fall into two classes. In the first class, the algorithm simply does not find a suitable solution and, as a consequence, the residual error is unacceptably high. In the second class, an algorithm does find a solution, but something is not quite right with this solution. We discuss the different cases one by one.

##### 9.3.1.1 Data Fit is Unacceptable

If no satisfactory fit can be found, the reasons may be manifold. In a relatively clear-cut case, the algorithm does not converge at all, reaches the maximally permitted number of iterations, or produces a fit that is obviously very different from the observed data. In such cases, the foremost (although not necessarily only) reasons are likely of a technical nature. It might be that the computer- or user-suggested initial guesses for all parameters are simply so bad that the algorithm does not reach a basin of attraction surrounding the optimal solution. It is also possible that the algorithm is attracted to an unacceptable local minimum.

A distinctly different cause for not yielding a good solution may be that the alleged functions in the model are so far from the truth that the algorithm cannot determine a satisfactory solution. In contrast to purely technical issues, the result in this case is often “some” fit that, however, is clearly not optimal. For instance, if one attempts to fit a Michaelis–Menten function to a sigmoidal time course, it is clear that the initial



**Figure 9.4** An algorithm tries to fit a sigmoidal Hill function ( $H$ ; blue) with a Michaelis–Menten function ( $M$ ; red) that simply does not have sufficient shape flexibility. As a consequence, no adequate fit can be reached.

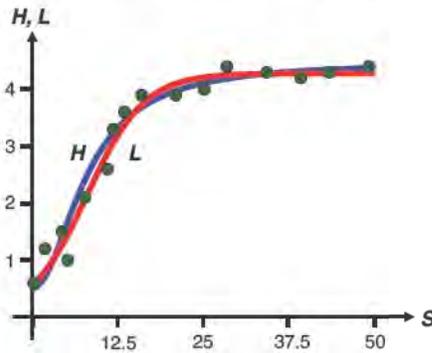
phase cannot be matched appropriately, because the structure of the Michaelis–Menten model is not equipped to capture S-shaped datasets (Figure 9.4). In simple cases like the one described, the problem is easily detected and diagnosed. However, this analysis is not so readily accomplished in a high-dimensional parameter space. Of course, all combinations of the above causes may be encountered. Many studies have focused on these and other technical issues.

#### 9.3.1.2 Differently Structured Candidate Models are Difficult to Compare

Within the realm of linear regression, methods have been developed for assessing the relative worth of an additional parameter. Specifically, objective criteria exist, based on residual errors and numbers of parameters and data points, for deciding the superiority of one of two candidate fits where one involves  $M$  and the other one  $M + 1$  parameters [27]. For nonlinear estimation tasks, such comparisons are much more difficult, especially if different model structures are involved. Surprisingly, there is even ambiguity in the number of parameters. As an example, consider a Hill function of the type:

$$V(S) = \frac{V_{\max} S^2}{K_M^2 + S^2} \quad (9.7)$$

which has a sigmoidal shape as shown in Figure 9.4. One will easily recognize  $K_M$  and  $V_{\max}$  as parameters, but should the Hill coefficient (i.e., the power associated with  $S$ ) be counted, even if it is *a priori* set equal to 2? After all, it clearly affects the shape of the function and if 2 did not fit, we could easily change it to a different value. It is difficult to find an objective criterion accounting for this issue. Thus, it is in general not a trivial matter to compare two fits, such as one with a Michaelis–Menten model (which is a special case with the Hill coefficient equaling 1) and one with a Hill model with a fixed or tunable Hill coefficient. Similarly, it is difficult to compare fits with a Hill model and a logistic model, which can both capture sigmoidal processes equally well but have different mathematical structures (Figure 9.5).



**Figure 9.5** Data (green symbols) following a sigmoidal trend are equally well represented by a Hill function ( $H$ ; blue;  $H(S) = 4S^2/(8^2 + S^2) + 0.5$ ) or a logistic function ( $L$ ; red;  $L(S) = 4.3/[1 + \exp(-0.24 \cdot (S-8))]$ ).

### 9.3.1.3 Fit is Acceptable, But...

In spite of all technical challenges, one often obtains a fit that is good, as judged by visual inspection and an acceptable sum of squared residual errors ( $SSE$ ), and it might appear that the problem is solved. In some cases this is the case, but a good  $SSE$  should not be taken as the sole criterion.

The best-known situation is convergence to a local minimum. Even if an optimization algorithm returns an acceptable fit, there is no guarantee that the fitted model is truly the best option. It is easily possible that the solution corresponds to a local minimum and that other solutions, maybe far away, are even better. However, the algorithm may not necessarily find these superior solutions because they are separated from the current solution by domains of parameter sets that correspond to drastically inferior model fits. Thus, any time the algorithm attempts to move toward the global optimum, it hits the separating areas of high errors, deems the direction futile, and searches in other parts of the parameter space. A partial, but not always effective solution to the problem is repeated optimization with different initial guesses for some or all parameters.

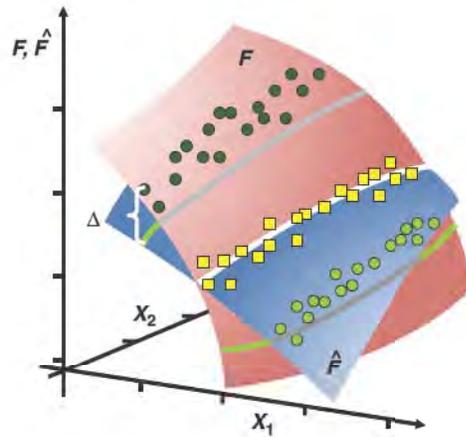
A slightly different situation occurs if the identified optimum is surrounded by a large area of solutions with very similar errors. In the simplest cases, these “almost optimal” solution sets form slightly distorted ellipses in a higher-dimensional space, but this is not necessarily so. Recent years have seen quite a bit of attention dedicated to what is now called “sloppy” solutions (e.g., [28, 29]). One could argue that the optimized solution is still a tiny bit better than its neighbors in the sloppy set and that it should therefore be preferred over all other candidates. However, in some cases the residual errors within the sloppy set are so similar that a slight change in just one data point, which could easily correspond to experimental noise, would identify a different solution as optimal. Thus, one should not discard the range of parameters surrounding the optimized solution. In a way, sloppy solution sets are not necessarily a cause for concern, because the corresponding models are quite similar in their fits to the data. In fact, it might be possible to simplify the optimized solution by setting certain

1 parameters equal to zero, if the sloppy set includes this setting as a legitimate  
2 possibility. The *SSE* would very slightly increase, but arguments of simplicity might  
3 favor this increase in exchange for a simpler model structure. Intriguingly, if the  
4 sloppy set permits positive, zero, as well as negative values for a certain parameter, the  
5 interpretation of the corresponding models differs. For instance, within the context  
6 of BST, a negative kinetic order is to be interpreted as an inhibitory signal, while a  
7 positive kinetic order suggests an activating or augmenting influence and a value of  
8 zero corresponds to an unimportant role of the parameter. This result poses the  
9 interesting question of whether sloppy solutions of this type are computational  
10 artifacts or whether inter-individual variation could go so far as to allow activation in  
11 one organism and inhibition in another [30]. At present, this question cannot be  
12 answered with confidence. Finally, while Ockham's razor might suggest that the  
13 simpler solution is to be preferred, biology has presented us in many cases with  
14 solutions that initially seem more complex than necessary. Later we may find that the  
15 more complex solutions are preferable because of improved robustness or other  
16 higher-order features.

17 Somewhat related to sloppy solutions is the situation where the data are not  
18 plentiful of informative enough to allow a precise determination of all parameter  
19 values. A very simple example arises if two parameters  $p_1$  and  $p_2$  always appear in the  
20 model in the same constellation, such as the ratio  $p_1/p_2$ . Clearly, by multiplying the  
21 two parameters with a nonzero factor, the ratio is unchanged. Turning the argument  
22 around, a search algorithm cannot find a unique solution in this situation, but only  
23 one where the ratio fits well. In these cases, the solution is "structurally non-  
24 identifiable," which entails that infinitely many combinations of parameter values  
25 can yield solutions with the same *SSE*. It is recommended to remove these  
26 identifiability issues, for instance, with methods of model reduction [31].

27 Quite a different concern with an apparently good fit is the occasional identification  
28 of parameter values that are impossible or unreasonable from a biological point of  
29 view. The most obvious case is an optimized value that turns out to be negative,  
30 although the parameter must be positive. Examples include rate constants and  $K_M$   
31 values. If this situation occurs, the optimization should be redone with correspond-  
32 ing constraints on the parameters. A more subtle variation of this issue is a resulting  
33 model that is unstable or extremely sensitive. It may happen in this case that a very  
34 small percent change in a parameter value would lead to dramatic changes in an  
35 important system feature such as the steady state. It is rare that such a high sensitivity  
36 is realistic in biology. Whether the situation is caused by a faulty parameter or by the  
37 misidentification or omission of some process or regulatory signal cannot be said  
38 without further analysis.

39 Another common and generic issue with an apparently good fit is often overlooked  
40 at first and becomes an inconvenient surprise later. This issue is the frequent inability  
41 of the parameterized model to predict responses to new stimuli in an adequate  
42 fashion [32]. Thus, the original data are matched quite nicely and the model traverses  
43 the cloud of data points seemingly fine. It may also be able to predict responses of the  
44 model throughout a modest time period beyond the measured time points. However,  
45 if new observations are to be modeled, the model may woefully fail. How is that



**Figure 9.6** Inability of a model to represent new data. In this generic example, the true function of a flux  $F(X_1, X_2)$  (red surface) is modeled with the function  $\hat{F}(X_1, X_2)$  (blue surface), based on one dataset (yellow squares). The fit (white line) is quite good.

However, since  $F$  and  $\hat{F}$  only intersect in the (white) model fit and otherwise diverge, extrapolated fits to other data (light and dark green circles) are no longer accurate and lead to significant residual errors (exemplified with  $\Delta$ ).

possible? One explanation is that the fitted curve constitutes the intersection of two (usually curved) surfaces – one describing the model function with optimized parameter values and the other representing the true dynamics of the biological system (see Figure 9.6). If the data points are close to this intersection, the fit not only appears to be good, but actually is so. However, analyzing new data means moving along the true surface and away from the intersection, and it becomes significant that the two surfaces diverge. As a result, if the extrapolation differs substantially from the original data, the model is no longer able to capture it. Again, several root causes for this situation are possible. One is the compensation of error among different terms within a system of equations. As an example, consider an equation with two terms, of which one is modeled quite badly. It is not difficult to imagine that, at least in certain situations, the other term could compensate for the error. In fact by also being modeled badly, the overall results can be surprisingly good. At first glance, it appears that two wrongs could indeed make a right. However, the error compensation usually no longer holds for new situations where the system variables in the two terms of the equation change to different degrees. As consequence, the fit to the extrapolated dataset can be unacceptably bad. To some degree, the method of dynamic flux estimation can remedy the extrapolation problem [32]. However, this method requires ideal conditions that are seldom satisfied and therefore requires additional information, which is not always easy to obtain [33, 34].

Finally, a genuine challenge with the otherwise appealing slope estimation and decoupling technique is that dependencies among equations are *a priori* ignored. For instance, if the same parameter appears in two equations, the decoupling causes

1 independent estimations and unless countermeasures are put into place, the same  
2 parameter might be identified with two different values [21].  
3

#### 4 9.3.1.4 Needed: A Better Fit! Or Not?

5 The previous sections have demonstrated that even an apparently good fit may not be  
6 the ultimate solution, because, for instance, it may not hold up in extrapolations. If so,  
7 what should be done? This is a complex question, and as in so many cases the answer  
8 is: it depends. The need to search for other solutions depends on the purpose of the  
9 model and the availability of appropriate data.

10 In some cases, even symbolic models, in which parameters are not at all  
11 numerically defined, can yield valuable insights. In fact, if insights can be gained  
12 for arbitrary parameter values, they are usually more general than results obtained for  
13 certain numerical parameter combinations. A beautiful example is the exploration of  
14 design principles [35]. In this line of research one asks what the role of a particular  
15 feature of a biological system is. For instance, one might observe a feedback  
16 inhibition signal in a pathway system and ask why it is there. According to the  
17 method of controlled mathematical comparisons (MCMC), one sets up two essentially  
18 identical systems models in parallel; however, one represents the observed  
19 signal and the other one does not [36]. The responses of these two models are  
20 compared with respect to performance criteria, such as robustness and response  
21 time. Studying a whole roster of such features and results, one system design is  
22 ultimately declared superior, either in general or within a certain environment. Many  
23 such comparisons have been performed without the specification of numerical  
24 parameter values, while other comparisons required the definition of relevant ranges  
25 of parameters. Using this MCMC strategy, Savageau proposed general rules for the  
26 regulation of bacterial gene circuits that have held up in all cases tested so far [37, 38].  
27 These rules were independent of specific parameter values and identified the  
28 superior circuit designs primarily based on demands exerted by the environmental  
29 of the bacteria. Other examples included more complex gene circuit designs and a  
30 variety of other system structures [39–44]. In the context of signaling cascades, the  
31 structural design and performance demands even helped determine the ranges of  
32 effective parameter values [45].

33 Trusting in the observation that accurate parameter values are often not as critical  
34 as the correct model structure, Alves *et al.* exhaustively evaluated likely parameter  
35 ranges on discrete grid points, thereby yielding sufficiently good, although maybe not  
36 optimal solutions [46]. Parameter ranges can also be restricted by biological and  
37 clinical constraints, and coarse solutions can be refined by means of simulations. In  
38 spite of the rather uncertain nature of this process, the model results can be  
39 surprisingly strong. For example, in an effort to construct a model of dopamine  
40 metabolism in the human brain, Qi *et al.* collected semiquantitative input from  
41 clinicians, biochemists, and toxicologists regarding the relative concentrations of  
42 relevant metabolites and the flux split ratios at diverging branch points in the pathway  
43 system [3]. This information turned out to be sufficient for setting up complex  
44 pathway models with coarse parameter values. Even though these models did not  
45 capture the precise numerical features of the biological system, they helped explain

1 the inner workings of the system. For instance, Qi *et al.* used these models to perform  
2 exhaustive analyses of root causes leading to Parkinson's disease and schizophre-  
3 nia [3, 47, 48]. Models of this type can also indicate whether the responses to targeted  
4 alterations occur in the right direction. Thus, even if the parameter values are coarse,  
5 some types of valuable insights may be gained.

6 Alas, there are situations where the parameterization should be as good and  
7 reliable as possible. For instance, it is becoming possible to develop health and  
8 disease models, which could be used for designing specific, and maybe even  
9 personalized, treatment strategies [49]. Clearly such strategies need to be quantitative  
10 and accurate enough to allow at least modest extrapolations from the normal state.  
11 These models should also permit reliable predictions of what might happen to an  
12 individual if s/he is or is not treated for some abnormality. One might be tempted to  
13 discard such predictions into the future, because our current models are simply too  
14 inexact. However, even short-term predictions can be extremely beneficial. For  
15 instance, a 5-min warning that a critical care patient is diverging from the normal  
16 trajectory might be sufficient to initiate efficient countermeasures [50]. Such short  
17 time horizon seems to be within reach of computational models, even if they are  
18 based on relatively coarse approximations.

19 If the fit is good with respect to the *SSE* of one data set, but fails in extrapolations,  
20 the situation is dire, because there is no general diagnostic tool for identifying where  
21 the problem lies or whether the reason of failure is a combination of several  
22 problems. To some degree, the modeling process needs to be restarted in such a  
23 case. Of course, the process does not really start at the very beginning, because a  
24 reasonable model for at least one situation is now available and this model may serve  
25 as a starting point or as a constraint for further model development. One might also  
26 expect both, greater reliability of the model and insight into the true nature of the  
27 biological processes if more and ideally diverse datasets are available that cover a  
28 greater portion of the space of variation in the system variables. Nevertheless, the  
29 difficult question at this point is whether the chosen functions in the model are  
30 appropriate or not. If they are, renewed parameter estimation with additional data  
31 might lead to a better numerical model implementation, but if they are not, new  
32 functions need to be determined. This task can be very difficult, because the structure  
33 of these functions is *a priori* unclear and because one usually does not even have data  
34 regarding the individual functions, but only regarding the entire dynamics of the  
35 system, which is governed by numerous processes simultaneously [34].

36 One strategy is to use biological insights to find functions that might be appropriate  
37 representations of individual processes. Accordingly, one designs a model with more  
38 complex, biologically relevant functions, and fits it to the original data and to the new  
39 data at the same time. The original, simpler model may be evoked as a constraint in a  
40 sense that the new function should coincide with the original model function (and  
41 possibly its slopes) with respect to the original data. In principle, this strategy seems  
42 to make sense, but in reality it often requires considerable effort [34]. Ultimately, this  
43 strategy leads to the question of when enough is enough. Is it really desirable to  
44 develop a very complicated process description that now captures two datasets? It is  
45 known that even slight overparameterization tends to lead to extrapolation problems

1 and the more parameters are involved in a functional description, the more severe the  
2 potential problem is.

3 An alternate strategy is an extension of the concept of generic approximations.  
4 Experience has shown that extensions toward higher-order Taylor terms become  
5 mathematically cumbersome and while they naturally fit better, the facility of their  
6 analysis is compromised [6]. Instead, one may employ a piecewise approximation [51].  
7 This strategy is straightforward in principle, but realistically requires the determi-  
8 nation of breakpoints, at which one approximation is substituted with another. For  
9 univariate functions, the breakpoints can almost be determined by visual inspection  
10 or with a simple algorithm, but the determination is much harder for multivariate  
11 functions. For linear and power-law functions, algorithms have been developed that  
12 automate this process [52, 53]. The result is an approximation, consisting of a certain  
13 number of pieces, throughout which the overall *SSE* is within desired bounds. The  
14 actual estimation of parameter values for these piecewise approximations requires  
15 considerably more data and of course the number of parameter values grows with  
16 each added piece. Nevertheless, this strategy is relatively unbiased and therefore may  
17 offer a first default in situations where biology does not suggest candidate functions  
18 or where such candidates are so complicated [54] that there is hardly a chance that  
19 parameters could be estimated without serious overfitting.  
20

## 21

### 22 9.4

### 23 Conclusions

### 24

25 The goal of parameter estimation is clear – find numerical values that render a model  
26 optimal for the representation of a biological system. While clear in principle, the task  
27 is often convoluted in practice. In the past, most data were coarse and scarce, and  
28 simple model fits had to be considered adequate. With recent advances in molecular  
29 biology, the situation has changed and some data are so good now that it is difficult to  
30 excuse bad fits.

31 As parameter estimation is so important, many groups have devoted substantial  
32 effort to it. In most cases, these investigations focused on the substantial technical  
33 challenges associated with the task and on algorithmic improvements. In addition to  
34 distinctly different optimization methods, which include regression, simulated  
35 annealing, and numerous variants of genetic, colony, and swarm algorithms, it has  
36 turned out that the estimation of slopes and the subsequent conversion of differential  
37 into algebraic equations is very beneficial. While these methodological and algorithmic  
38 improvements have made parameter estimation manageable in principle, they  
39 have not solved all problems. There are still issues with slow convergence or the  
40 trapping of algorithms in local minima.

41 Beyond the well-known technical issues, even apparently good fits should be  
42 subjected to additional muster. Criteria for such additional tests should be the  
43 reasonableness of the numerical values of all parameters, model stability, sensitivity,  
44 and robustness, and the ability to provide good fits in extrapolations. Possible causes  
45 for models to fail these criteria are plentiful. Particularly hideous among them is the

1 compensation of fitting errors within and among the system equations, because such  
 2 compensation is not always easy to diagnose and remediate. The recent method of  
 3 dynamic flux estimation [32] is able to help, but only under ideal conditions, and it will  
 4 be necessary to expand such methods toward more realistic conditions.

5 If a fit with a good *SSE* is obtained, but extrapolations cause problems, the entire  
 6 model structure may have to be revisited, including the choice of functions, the  
 7 numbers and types of parameters, and the availability of data. Clearly, several  
 8 replicates of time-series data and data obtained under different conditions will allow  
 9 better estimations, along with their statistical analyses and interpretations. Ulti-  
 10 mately, one must consider the purpose of the model and judge the real need for  
 11 accurate parameter values. Indeed, a simple model with fewer parameters is often  
 12 more robust and less sensitive to overparameterization, yet may yield as much, if not  
 13 more insight.

#### 16 Acknowledgments

17  
 18 The author expresses his gratitude to Yun Lee for help with one of the figures. This  
 19 work was supported in part by a National Heart, Lung, and Blood Institute Proteomics  
 20 Initiative (contract N01-HV-28181; D. Knapp, PI), a Molecular and Cellular Biosciences  
 21 Grant (MCB-0517135; E.O. Voit, PI) from the National Science Foundation,  
 22 and a Biological Energy Science Center grant from the Department of Energy  
 23 (M. Keller, PI). Any opinions, findings, and conclusions or recommendations  
 24 expressed in this material are those of the authors, and do not necessarily reflect  
 25 the views of the sponsoring institutions.

#### 28 References

- 29  
 30  
 31  
 32  
 33  
 34  
 35  
 36  
 37  
 38  
 39  
 40  
 41  
 42  
 43  
 44  
 45
- 1 Chen, L., Qi-Wei, G., Nakata, M., Matsuno, H., and Miyano, S. (2007) Modelling and simulation of signal transduction in an apoptosis pathway by using times Petri nets. *J. Biosci.*, **32**, 113–127.
  - 2 Wu, J. and Voit, E. (2009) Hybrid modeling in biochemical systems theory by means of functional Petri nets. *J. Bioinform. Comput. Biol.*, **7**, 107–134.
  - 3 Qi, Z., Miller, G.W., and Voit, E.O. (2008) Computational systems analysis of dopamine metabolism. *PLoS ONE*, **3**, e2444.
  - 4 Chou, I.-C. and Voit, E.O. (2009) Recent developments in parameter estimation and structure identification of biochemical and genomic systems. *Math. Biosci.*, **219**, 57–83.
  - 5 Savageau, M.A. (1995) Enzyme kinetics *in vitro* and *in vivo*: Michaelis–Menten revisited, in *Principles of Medical Biology* (ed. E.E. Bittar), JAI Press, Greenwich, CT, pp. 93–146.
  - 6 Cascante, M., Sorribas, A., Franco, R., and Canela, E.I. (1991) Biochemical systems theory: increasing predictive power by using second-order derivatives measurements. *J. Theor. Biol.*, **1991**, 521–535.
  - 7 Savageau, M.A. and Voit, E.O. (1987) Recasting nonlinear differential equations as S-systems: a canonical nonlinear form. *Math. Biosci.*, **87**, 83–115.
  - 8 Voit, E.O. (1993) S-system modeling of complex systems with chaotic input. *Environmetrics*, **4**, 153–186.

- 1 9 Savageau, M.A. (1969) Biochemical  
 2 systems analysis. I. Some mathematical  
 3 properties of the rate law for the  
 4 component enzymatic reactions.  
 5 *J. Theor. Biol.*, **25**, 365–369.
- 6 10 Savageau, M.A. (1969) Biochemical  
 7 systems analysis. II. The steady-state  
 8 solutions for an  $n$ -pool system using a  
 9 power-law approximation. *J. Theor. Biol.*,  
 10 **25**, 370–379.
- 11 11 Savageau, M.A. (1976) *Biochemical Systems*  
 12 *Analysis: A Study of Function and Design in*  
 13 *Molecular Biology*, Addison-Wesley,  
 14 Reading, MA, p. 379.
- 15 12 Voit, E.O. (ed.) (1991) *Canonical Nonlinear*  
 16 *Modeling. S-System Approach to*  
 17 *Understanding Complexity*, Van Nostrand  
 18 Reinhold, New York, p. 365.
- 19 13 Torres, N.V. and Voit, E.O. (2002) *Pathway*  
 20 *Analysis and Optimization in Metabolic*  
 21 *Engineering*, Cambridge University Press,  
 22 Cambridge.
- 23 14 Voit, E.O. (2000) *Computational Analysis of*  
 24 *Biochemical Systems: A Practical Guide for*  
 25 *Biochemists and Molecular Biologists*,  
 26 Cambridge University Press, Cambridge,  
 27 p. 530.
- 28 15 Alvarez-Vasquez, F., Sims, K.J.,  
 29 Hannun, Y.A., and Voit, E.O. (2004)  
 30 Integration of kinetic information on yeast  
 31 sphingolipid metabolism in dynamical  
 32 pathway models. *J. Theor. Biol.*, **226**,  
 33 265–291.
- 34 16 Varah, J.M. (1982) A spline least squares  
 35 method for numerical parameter  
 36 estimation in differential equations.  
 37 *SIAM J. Sci. Stat. Comput.*, **3**, 28–46.
- 38 17 Voit, E.O. and Savageau, M.A. (1982)  
 39 Power-law approach to modeling  
 40 biological systems; III. Methods of  
 41 analysis. *J. Ferment. Technol.*, **60**,  
 42 233–241.
- 43 18 Voit, E.O. and Almeida, J. (2004)  
 44 Decoupling dynamical systems for  
 45 pathway identification from  
 46 metabolic profiles. *Bioinformatics*, **20**,  
 47 1670–1681.
- 48 Q1 19 Voit, E.O. and Chou, I.C. (2010)  
 49 Parameter estimation in canonical  
 50 biological systems models. *Int. J. Syst.*  
 51 *Synth. Biol.*, in press.
- 52 20 Chou, I.C., Martens, H., Voit, E.O. (2006)  
 53 Parameter estimation in biochemical  
 54 systems models with alternating  
 55 regression. *Theor. Biol. Med. Model.*, **3**, 25.
- 56 21 Vilela, M., Chou, I.C., Vinga, S. *et al.*  
 57 (2008) Parameter optimization in S-  
 58 system models. *BMC Syst. Biol.*, **2**, 35.
- 59 22 Brunel, N.J.-B. (2008) Parameter  
 60 estimation of ODE's via nonparametric  
 61 estimators. *Electr. J. Stat.*, **2**, 1242–1267.
- 62 23 Seatzu, C. (2000) A fitting based method  
 63 for parameter estimation in S-systems.  
 64 *Dyn. Syst. Appl.*, **9**, 77–98.
- 65 24 Eilers, P.H. (2003) A perfect smoother.  
 66 *Anal. Chem.*, **75**, 3631–3636.
- 67 25 Vilela, M., Borges, C.C., Vinga, S. *et al.*  
 68 (2007) Automated smoother for the  
 69 numerical decoupling of dynamics  
 70 models. *BMC Bioinformatics*, **8**, 305.
- 71 26 Ramsay, J.O., Hooker, G., Cao, J., and  
 72 Campbell, D. (2007) Parameter estimation  
 73 for differential equations: a generalized  
 74 smoothing approach. *J. Roy. Stat. Soc. B*,  
 75 **69**, 741–796.
- 76 27 Neter, J. and Wasserman, W. (1974)  
 77 *Applied Linear Statistical Models*, Irwin,  
 78 Homewood, IL.
- 79 28 Gutenkunst, R.N., Casey, F.P.,  
 80 Waterfall, J.J., Myers, C.R., and  
 81 Sethna, J.P. (2007) Extracting falsifiable  
 82 predictions from sloppy models. *Ann. NY*  
 83 *Acad. Sci.*, **1115**, 203–211.
- 84 29 Gutenkunst, R.N., Waterfall, J.J.,  
 85 Casey, F.P. *et al.* (2007) Universally  
 86 sloppy parameter sensitivities in systems  
 87 biology models. *PLoS Comput. Biol.*, **3**,  
 88 1871–1878.
- 89 30 Vilela, M., Vinga, S.,  
 90 Mattoso Maia, M.A.G., Voit, E.O., and  
 91 Almeida, J. (2009) Identification of neutral  
 92 sets of biochemical systems models from  
 93 time series data. *BMC Syst. Biol.*, **3**, 47.
- 94 31 Raue, A., Kreutz, C., Maiwald, T. *et al.*  
 95 (2009) Structural and practical  
 96 identifiability analysis of partially  
 97 observed dynamical models by  
 98 exploiting the profile likelihood.  
 99 *Bioinformatics*, **25**, 1923–1929.
- 100 32 Goel, G., Chou, I.-C., and Voit, E.O. (2008)  
 101 System estimation from metabolic time  
 102 series data. *Bioinformatics*, **24**, 2505–2511.
- 103 33 Voit, E.O., Goel, G., Chou, I.-C., and  
 104 da Fonseca, L. (2009) Estimation of  
 105 metabolic pathway systems from different  
 106 data sources. *IET Syst. Biol.*, **3**, 513–522.

- 1 34 Goel, G. (2009) Dynamic Flux Estimation:  
2 a novel framework for metabolic pathway  
3 analysis. Dissertation, Department of  
4 Biomedical Engineering, Georgia  
5 Institute of Technology, Atlanta, GA.
- 6 35 Savageau, M.A. (2001) Design principles  
7 for elementary gene circuits: elements,  
8 methods, and examples. *Chaos*, **11**,  
9 142–159.
- 10 36 Savageau, M.A. (1985) A theory of  
11 alternative designs for biochemical  
12 control systems. *Biomed. Biochim. Acta*,  
13 **44**, 875–880.
- 14 37 Savageau, M.A. (1998) Demand theory  
15 for gene regulation: quantitative  
16 development of the theory. *Genetics*, **149**,  
17 1665–1676.
- 18 38 Savageau, M.A. (1998) Demand theory for  
19 gene regulation: quantitative application  
20 to the lactose and maltose operons of  
21 *Escherichia coli*. *Genetics*, **149**, 1677–1691.
- 22 39 Coelho, P.M., Salvador, A., and Savageau,  
23 M.A. (2009) Quantifying global tolerance  
24 of biochemical systems: design  
25 implications for moiety-transfer cycles.  
26 *PLoS Comput. Biol.*, **5**, e1000319.
- 27 40 Igoshin, O.A., Price, C.W., and  
28 Savageau, M.A. (2006) Signalling network  
29 with a bistable hysteretic switch controls  
30 developmental activation of the sigma  
31 transcription factor in *Bacillus subtilis*. *Mol.*  
32 *Microbiol.*, **61**, 165–184.
- 33 41 Alves, R. and Savageau, M.A. (2000) Effect  
34 of overall feedback inhibition in  
35 unbranched biosynthetic pathways.  
36 *Biophys. J.*, **79**, 2290–2304.
- 37 42 Hlavacek, W.S. and Savageau, M.A. (1996)  
38 Rules for coupled expression of regulator  
39 and effector genes in inducible circuits.  
40 *J. Mol. Biol.*, **255**, 121–139.
- 41 43 Irvine, D.H. and Savageau, M.A. (1985)  
42 Network regulation of the immune  
43 response: alternative control points for  
44 suppressor modulation of effector  
45 lymphocytes. *J. Immunol.*, **134**,  
2100–2116.
- 46 44 Schwacke, J.H. and Voit, E.O. (2004)  
Improved methods for the mathematically  
controlled comparison of biochemical  
systems. *Theor. Biol. Med. Model.*, **1**, 1.
- 47 45 Schwacke, J.H. and Voit, E.O. (2007)  
Concentration-dependent effects on the  
rapid and efficient activation of MAPK.  
*Proteomics*, **7**, 890–899.
- 48 46 Alves, R., Herrero, E., and Sorribas, A.  
(2004) Predictive reconstruction of the  
mitochondrial iron–sulfur cluster  
assembly metabolism. II. Role of  
glutaredoxin Grx5. *Proteins Struct. Funct.*  
*Bioinform.*, **57**, 481–492.
- 49 47 Qi, Z., Miller, G.W., and Voit, E.O. (2008)  
A mathematical model of presynaptic  
dopamine homeostasis: implications for  
schizophrenia. *Pharmacopsychiatry*, **41**,  
S89–S98.
- 50 48 Qi, Z., Miller, G.W., and Voit, E.O. (2009)  
Computational analysis of determinants  
of dopamine dysfunction. *Synapse*, **63**,  
1133–1142.
- 51 49 Voit, E.O. and Brigham, K.L. (2008)  
The role of systems biology in predictive  
health and personalized medicine.  
*Open Path J.*, **2**, 68–70.
- 52 50 Lu, Y., Burykin, A., Deem, M.W., and  
Buchman, T.G. (2009) Predicting clinical  
physiology: a Markov chain model of heart  
rate recovery after spontaneous breathing  
trials in mechanically ventilated patients.  
*J. Crit. Care*, **24**, 347–361.
- 53 51 Savageau, M.A. (2001) Design  
principles for elementary gene circuits:  
elements, methods, and examples.  
*Chaos*, **11**, 142.
- 54 52 Ferrari-Trecate, G. and Muselli, M. (2002)  
A new learning method for piecewise  
linear regression. *Lecture Notes Comput.*  
*Sci.*, **2415**, 444–449.
- 55 53 Machina, A., Ponosov, A., and Voit, E.O.  
(2010) Automated piecewise power-law  
modeling of biological systems.  
*J. Biotechnol.*, **149**, 154–165.
- 56 54 Peskov, K., Goryanin, I., and Demin, O.  
(2008) Kinetic model of  
phosphofructokinase-1 from *Escherichia*  
*coli*. *J. Bioinform. Comput. Biol.*, **6**,  
843–867.

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45

**Abstract**

The identification of functions and their parameter values for biological systems models is currently among the most severe bottlenecks in computational systems biology. While many papers and books have discussed the advantages and pitfalls of estimation algorithms of various types, including nonlinear regression, genetic algorithms, ant colony and swarm optimization, and simulated annealing, issues beyond the technical challenges of parameter estimation have seldom received much attention. This chapter discusses several situations where the computationally determined model fit to experimental data is satisfactory with respect to the sum of squared residual errors, but unsatisfactory or undesirable for other reasons. Examples include unrealistic parameter values, nonunique or sloppy solutions, and the inability of a fully parameterized model to make reliable predictions regarding so far untested scenarios.

**Keywords:** biochemical systems theory; extrapolation; inverse problem; model identification; parameter estimation; sloppiness.

UNCORRECTED PROOF

**Author Query**

1. Update [19]?

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45

UNCORRECTED PROOF