Decoupling dynamical systems for pathway identification from metabolic profiles

Eberhard O. Voit¹,²,* and Jonas Almeida¹

¹Department of Biometry and Epidemiology and ²Department of Biochemistry and Molecular Biology, Medical University of South Carolina, Charleston, SC 29425, USA

Received on April 2, 2003; revised on November 19, 2003; accepted on December 22, 2003

Advance Access publication February 26, 2004

ABSTRACT

Rationale: Modern molecular biology is generating data of unprecedented quantity and quality. Particularly exciting for biochemical pathway modeling and proteomics are comprehensive, time-dense profiles of metabolites or proteins that are measurable, for instance, with mass spectrometry, nuclear magnetic resonance or protein kinase phosphorylation. These profiles contain a wealth of information about the structure and dynamics of the pathway or network from which the data were obtained. The retrieval of this information requires a combination of computational methods and mathematical models, which are typically represented as systems of ordinary differential equations.

Results: We show that, for the purpose of structure identification, the substitution of differentials with estimated slopes in non-linear network models reduces the coupled system of differential equations to several sets of decoupled algebraic equations, which can be processed efficiently in parallel or sequentially. The estimation of slopes for each time series of the metabolic or proteomic profile is accomplished with a ‘universal function’ that is computed directly from the data by cross-validated training of an artificial neural network (ANN).

Conclusions: Without preprocessing, the inverse problem of determining structure from metabolic or proteomic profile data is challenging and computationally expensive. The combination of system decoupling and data fitting with universal functions simplifies this inverse problem very significantly. Examples show successful estimations and current limitations of the method.

Availability: A preliminary Web-based application for ANN smoothing is accessible at http://bioinformatics.musc.edu/webmetabol/. S-systems can be interactively analyzed with the user-friendly freeware PLAS© (http://correio.cc.fc.ul.pt/~aent/plas.html) or with the MATLAB module BSTLab (http://bioinformatics.musc.edu/bstlab/), which is currently being beta-tested.

Contact: Voiteo@musc.edu

*To whom correspondence should be addressed.

INTRODUCTION

A serious bottleneck of metabolic modeling in the past has often been the lack of sufficient data for parameter estimation. Even for systems with only five or six metabolites and some modulators, the number of parameters is typically in the dozens and grows much faster than linearly with the number of variables involved. As a specific example, the number of parameters in the systems we use here is 2n(n + 1), where n is the number of species. Metabolic modelers have been responding to this challenge in two ways. Either they have limited the amount of necessary information to the stoichiometry of fluxes (e.g. Stephanopoulos et al., 1998; Reed and Palsson, 2003), thereby avoiding the need for kinetic details like regulatory feedback signals, or fully kinetic models have been constructed with in vivo or in vitro information from different sources and, quite often, different organisms (e.g. Mulquiney and Kuchel, 2003; Voit, 2000). It has been clear to all involved that this mixing and matching carries risks.

The advent of comprehensive metabolic profiles provides new means of parameter estimation. Metabolic profiles consist of simultaneous measurements of metabolites within the same cell, cell system or organism. At present, metabolic profiles are typically obtained as snapshots of many metabolites at one time point or as time series measurements of a limited number of metabolites (Goodacre and Harrigan, 2003). An example of the former is the simultaneous analysis of about 6000 metabolites in ripening strawberries, using modern variants of mass spectrometry (Goodenowe, 2001, 2003), Examples of the latter are in vivo NMR measurements of a few metabolites in rather dense-time sequences of a few seconds or minutes (e.g. Szymerski, 1998; Gerner et al., 2002; Neves et al., 2002), ‘rapid sampling’ methods with subsequent high-performance liquid chromatography (HPLC) (Theobald et al., 1997; Ostergaard et al., 2001) and protein kinase phosphorylation (Mckenzie and Strauss, 2003). It seems quite obvious that it is only a matter of time before mass spectrometric methods will be extended to establishing detailed time courses of simultaneously measured metabolites, for instance by combination with gas chromatography (Fiehn et al., 2000) or capillary electrophoresis (Soga et al., 2002) and that the current nuclear
magnetic resonance (NMR) time series of a few metabolites will be extended to larger numbers of simultaneous measurements. With these types of trace data immanent, the modeling community must be prepared to offer methods that make optimal use of them.

The premier strategies for analyzing time traces have been massive computer power, subdivision into smaller subtasks and parallelization (e.g. Kikuchi et al., 2001; Maki et al., 2001; Sakamoto and Iba, 2001). As a case in point, Kikuchi et al. (2003) estimated parameters in a differential equation model with five variables, using a genetic algorithm. Even though they used a cluster of 1040 processors Pentium III (933 MHz) on this rather small system with noise-free data, the time for each algorithmic loop was \( \sim 10 \) h.

While intensive—and most likely parallelized—computational effort will be unavoidable for pathways of realistic proportions, we submit that much time and effort can be saved if one preprocesses the data in a convenient fashion. Specifically, it is useful to smooth the data traces, estimate their slopes and substitute these for the differentials in the dynamic system model. This process is accomplished with ‘universal algebraic functions’ that are computed with artificial neural networks (ANNs). Even though a universal function usually fits the data quite well, the data fit alone does not provide structural or mechanistic insight. A good fit may be used for interpolations between data points and possibly for limited extrapolations, but all further explanations and predictions require a fully structured and parameterized mathematical model. We use for this purpose the S-system form within Biochemical Systems Theory (BST; Savageau, 1969a,b), which has a proven track record as a modeling framework for biological systems (cf. Voit, 2000) and is defined through parameters that can be interpreted directly as structural features of the pathway.

**METHODS**

**Choice of modeling framework**

Any modeling framework for complex metabolic networks must necessarily be a compromise that on one hand is general enough not to exclude observed dynamical trends but on the other hand does not require so much a priori information that it becomes unattainable for practical identification tasks. For instance, a polynomial model or neural network may be the least biased, but it does not help with the identification of structural details because the optimized parameters have no biological meaning. In contrast, a traditional kinetic model, such as a Michaelis–Menten rate law with inhibition, explains to some degree the mechanism of the reaction but requires a priori knowledge that may not be available. Competitive inhibition calls for one structure, non-competitive inhibition for another (e.g. Savageau, 1976); the rate law for an ordered, bi–bi, sequential reaction differs from that of a ping–pong reaction (e.g. Schulz, 1994).

As an alternative, the S-system form within BST (Savageau, 1969a,b) is ideally suited for the identification of pathway structures. BST is based on linearization in a logarithmic coordinate system and leads to a non-linear power-law approximation in Cartesian space that involves two types of parameters: kinetic orders, usually represented as \( g_{ij} \) and \( h_{ij} \), and rate constants, which are typically called \( \alpha_i \) and \( \beta_i \). The greatest benefit of BST in the present context is that these parameters map one-to-one onto the structure of the metabolic network modeled. As a consequence, correct estimation of parameters is in principle sufficient to deduce the structure of the underlying pathway. It is furthermore known that BST is rich enough to allow an unlimited range of non-linearities (Savageau and Voit, 1987), yet can be formulated symbolically without much prior knowledge about the processes involved. Due to these and other unique features, BST has been the subject of numerous articles and several books (Savageau, 1976; Voit, 1991, 2000; Torres and Voit, 2002) and therefore needs no detailed review here.

**From coupled differential to uncoupled algebraic equations**

Consider the differential equation

\[
\dot{X} = f(X) \quad X(t_0) = X_0, \tag{1}
\]

where \( X \) is a variable or a vector of variables, the notation \( \dot{X} = dX/dt \) indicates differentiation with respect to time \( t \), \( f(X) \) is some kind of kinetic rate function, which we choose as the difference between two power-law terms in accordance with the S-system form within BST, and \( X_0 \) represents the initial condition(s) at the start time, \( t_0 \). Equation (1) states that the instantaneous change (slope) in \( X \) at any given time point \( t_k \) is equal to \( f(X) \) at \( t_k \), i.e. \( \dot{X}(t_k) = f[X(t_k)] \). Given \( f(X) \) and \( X_0 \), the time courses of \( X(t) \) are usually computable, either analytically or with a numerical integrator.

The estimation of parameter values of \( f(X) \) from experimental trace data requires computation in the opposite direction. This ‘inverse problem’ is not trivial because \( f(X) \) is non-linear in all relevant cases, thus requiring iterative optimization in a larger parameter space, where the most computationally expensive component is the numerical integration of the differential equations, which may require in excess of 95% of the total search time (see Appendix 1). This time can be drastically reduced if one interprets the derivatives \( \dot{X} \) at all measured time points \( t_k \) as slopes, estimates these slopes from the data as \( S(t_k) \), and approximates the differential equations as

\[
S(t_k) \approx \dot{X} \mid_{t_k} = f[X(t_k)] \tag{2}
\]

(cf. Voit and Savageau, 1982b). If the metabolic profile consists of metabolites \( X_1, \ldots, X_i, \ldots, X_n \), which are measured at \( N \) time points \( t_1, \ldots, t_k, \ldots, t_N \), one estimates \( n \times N \) slopes \( S_i(t_k) \) that correspond to \( n \times N \) measured metabolite levels \( X_i(t_k) \). The inverse problem is thus reformulated from one
involving \( n \) differential equations to a larger system of \( n \times N \) algebraic equations of the form

\[
S_1(t_1) \approx f_1[X_1(t_1), X_2(t_1), \ldots, X_n(t_1); p_{11}, \ldots, p_{1M}], \\
S_1(t_2) \approx f_1[X_1(t_2), X_2(t_2), \ldots, X_n(t_2); p_{11}, \ldots, p_{1M}], \\
\vdots \\
S_1(t_N) \approx f_1[X_1(t_N), X_2(t_N), \ldots, X_n(t_N); p_{11}, \ldots, p_{1M}], \\
S_2(t_1) \approx f_2[X_1(t_1), X_2(t_1), \ldots, X_n(t_1); p_1, \ldots, p_M], \\
S_2(t_2) \approx f_2[X_1(t_2), X_2(t_2), \ldots, X_n(t_2); p_1, \ldots, p_M], \\
\vdots \\
S_2(t_N) \approx f_n[X_1(t_N), X_2(t_N), \ldots, X_n(t_N); p_{n1}, \ldots, p_{nM}], \\
\]

(3)

The functions \( f_i \) are the same as in Equation (1), but the parameters to be optimized, \( p_{i1}, \ldots, p_{iM} \), have been made explicit in Equation (3) for clarity of discussion. The parameters may differ in number among the \( n \) equations and are the only unknowns in these equations once the slopes have been determined from the data.

It may be surprising at first that a tightly coupled system of non-linear differential equations can be validly decoupled and analyzed as sets of algebraic equations. Indeed, for solving the system in the traditional sense of integration, one would only be able to use the algebraic equations in an approximate, step-by-step fashion, evaluating all equations simultaneously, which is how the Euler algorithm numerically integrates differential equations. The situation is different for the purpose of parameter estimation. In the solution strategy proposed here, the decoupled algebraic equations simply provide numerical values of variables (metabolites) and slopes (in typical jar-kinetic order \( g_{ij} \) or \( h_{ij} \) with the smaller magnitude is sub-

Evaluation and interpretation

If the alleged S-system model is adequate, one should expect the regression to identify a numerical solution with sufficiently good data fit. The parameter values of this solution provide direct insight into the underlying pathway structure. In particular, kinetic orders of zero suggest that the corresponding connection between two variables is not existent. In practice, it is advisable to analyze these initial regression results further.

The conclusion that several pathway designs fit the data is not entirely bad because it excludes a myriad of other designs and suggests very targeted hypotheses for additional, decisive experiments.

In addition to presenting a roster of solutions, one may execute two further types of analyses. First, one may use each solution as a set of high-quality starting values for a non-linear regression based directly on the differential equations. Since the proposed estimation is possibly biased by the fact that it is executed in a space of \( dX \) versus \( X \), the direct estimation of \( X \) versus \( t \) may give further insight as to which solution may be most likely. Second, it is useful to explore close-by solutions that have a simpler network structure. For example, one may force ‘small’ kinetic orders to be zero in a follow-up model identification, which corresponds to saying that very weak predicted interactions are in truth artifacts. For the illustrative example discussed below, we will select 0.1 as the threshold for this ‘pruning’ of kinetic orders. Furthermore, experience with metabolic systems suggests that it is relatively rare that the same variable is present in both the production and degradation terms of another variable. Consequently, one may explore solutions where the kinetic order \( g_{ij} \) or \( h_{ij} \) with the smaller magnitude is subtracted from both \( g_{ij} \) and \( h_{ij} \). The mathematical rationale is that this simplification corresponds to factoring out \( X_t^{\delta} \) or \( X_t^{\gamma} \) from both terms in the equation of \( X_j \). If \( X_j \) is more or less constant throughout the measured time interval, this term, \( X_t^{\delta} \) or \( X_t^{\gamma} \), is more or less constant as well and could be merged with the two rate constants. If it is not constant, the term in effect constitutes a scaling of the time variable, which can often be absorbed by other system variables and the rate constant. Any such simplification in structure must be tested with respect to the quality of fit.
Pathway identification profiles

Scheme 1. Logistic flow of a typical identification analysis. The data and model are prepared in parallel. The results of these steps are smoothed traces and their slopes, as well as a minimal S-system model that is still general enough to represent the desired pathway structure. This model and data from the smoothed traces and their slopes are used in a regression step that reveals optimal parameter values. These present a possible solution that is to be interpreted in terms of pathway features. For additional candidate solutions, it is useful to explore simpler solutions that result from pruning, as discussed in the text.

and if it is comparable with that of the original fit, the solution presents an additional possible network design. Apart from the pragmatism of Ockham’s razor, only further data or additional information can declare this design better or worse than others with similar residual error. The suggested simplifications may also be executed one step (kinetic order) at a time.

Summary of identification strategy

Because the proposed strategy for pathway identification consists of a combination of diverse methods, it may be useful to summarize the flow of a typical analysis, which is illustrated in Scheme 1. The typical raw data consist of a metabolic profile, i.e. of time traces of \( n \) variables, measured at \( N \) time points. The first step after assuring that the data are in an appropriate format is the smoothing of each trace with an ANN. The smoothed data henceforth replace the measured raw data. Each smoothed trace is furthermore used to compute slopes. Parallel to this data preprocessing, an S-system model is designed in symbolic form, i.e. without numerical specification of parameter values. If partial information on the underlying pathway is available, it is taken into consideration at this point (cf. Almeida and Voit, 2003); otherwise the S-system model corresponds to a fully connected graph. For numerical purposes, the equations may be log-transformed (Savageau, 1976), but this is not mandatory. The differentials on the left-hand sides of the S-system equations are replaced by ANN-estimated slopes, a process that converts the differential with algebraic equations. These algebraic equations are subjected to a parameter estimation, which may be based on genetic algorithms, non-linear regression or both. If kinetic orders are estimated as close to zero, one may consider fixing them at zero and rerunning the regression with the pruned system. It may also be useful to resample the data and to redo the analysis in order to explore alternative solutions. If different solutions with similar residual errors exist, the final result of the method is a roster of possible solutions; further data or other additional pieces of information are needed for definite structure identification.

RESULTS

As an illustration of the combination of methods, consider the dynamics of the pathway in Figure 1. The production of \( X_1 \) depends on the source variable, \( X_0 \), and is also affected by the inhibition exerted by \( X_2 \), which is generated from \( X_1 \) via the intermediate \( X_2 \). \( X_1 \) can also be used for the production of \( X_4 \), and this product inhibits the degradation of \( X_3 \). Although only containing four dependent variables, the system is already difficult to predict. For instance, it is not intuitive what would happen to \( X_3 \) if the input to the system were increased for some while. By including those and only those variables in each term that have a direct effect on this term, one obtains the describing S-system model in symbolic form. A numerical implementation with typical parameter values is

\[
\begin{align*}
\dot{X}_1 &= 20X_0X_3^{-0.8} - 10X_1^{0.5} & X_1(t_0) &= 1.4, \\
\dot{X}_2 &= 8X_1^{0.5} - 3X_2^{0.75} & X_2(t_0) &= 2.7, \\
\dot{X}_3 &= 3X_2^{0.75} - 5X_3^{0.5}X_4^{0.2} & X_3(t_0) &= 1.2, \\
\dot{X}_4 &= 2X_1^{0.5} - 6X_4^{0.8} & X_4(t_0) &= 0.4, \\
X_0 &= 0.6.
\end{align*}
\]

We used this model to create artificial datasets, which resemble experimentally measured metabolic profiles (Fig. 2) in structure. It may be seen as an unfair advantage that we use an S-system model for the generation of test data. However, this illustration is a proof of concept that should not be burdened by questions of how well the S-system model captures the dynamics of metabolic systems. It is noted that these
Questions have been addressed quite extensively in the literature (e.g. Voit and Savageau, 1987; Sorribas and Savageau, 1989).

It must be stressed that we pretend in this analysis that none of the structural information in Figure 1 is available. Thus, as far as the ANN and the following regression are concerned, the four dependent variables could be connected in any imaginable fashion, and the task is to deduce the most likely network pattern(s) compatible with the traces.

We analyzed various measurement frequencies and different levels of noise in the data but only show the scenario without noise; a smoothing of data with noise is shown in Appendix 2. The data sets consisted of profiles at 100 time points, which we fitted with the ANN in five overlapping pieces. The result was a set of four universal function representations, one for each trace, which subsequently replaced the data points and allowed the computation of slopes. An important, genuine feature of this approach is that the quality of fit in one trace does not affect the fit in other traces. Because of the decoupling and the fact that data points are replaced with corresponding trace values, the fits are entirely independent of each other.

The piecewise ANN-fitted traces and their slopes were used to estimate S-system parameters. While greatly simplified in comparison with the original problem of fitting simultaneous differential equations, the algebraic non-linear regression task is still far from trivial. In many realistic situations, some biological information is available that excludes certain network connections. For instance, if it is a priori known that $X_j$ does not directly affect $X_i$, the corresponding kinetic orders $g_{ij}$ and $h_{ij}$ are zero and this reduces the parameter space that needs to be searched by a full dimension (cf. Almeida and Voit, 2003). No such information was used in the regression examples shown here, however. The only ‘help’ provided to the regression was a restriction of the ranges of possible kinetic orders, which was based on general knowledge in the field (cf. Voit, 2000: Chapter 5). As an illustration, we show results where we restricted the search space by allowing kinetic orders to fall only in the range $[-1, 1]$ and by penalizing high kinetic orders, as proposed elsewhere (Kikuchi et al., 2003; Voit and Almeida, 2003). Also based on experience, we set the kinetic orders $g_{ii} = 0$, which precluded a direct effect of a variable on its own production and required the kinetic orders $h_{ii}$ to be greater than 0, indicating that the degradation of a metabolite almost always depends on the metabolite concentration itself. If these assumptions happened to be violated in a specific case, we would expect unsatisfactory fits and solutions with parameters at the admissible boundaries, which would subsequently be relaxed.

The first result of our analysis involved single datasets, for which we obtained perfectly fitting solutions, whose parameter values, however, were strikingly different from those of the true system (data not shown). Indeed, any single data set allowed multiple distinctly different numerical solutions, especially if constraints on kinetic orders were set loosely. This was not unexpected because even one-variable S-systems are flexible enough to permit different parameter sets generating very similar graphs (e.g. Voit et al., 1995; Sands and Voit, 1996; Sorribas et al., 2000). Without further information, each such ‘faulty’ solution is a priori as valid as the original solution because it fits the data essentially equally well.

Instead of supplying structural information [e.g. where this was successfully done, see Almeida and Voit (2003)], we generated from the same system six data sets, which consisted of concentrations at 100 time points and differed only in their initial values. This situation is comparable with experimentally testing the same pathway under different conditions, such as different substrate availability, and has recently been employed by other investigators with similar goals (e.g. Kikuchi et al., 2003; Torralba et al., 2003; Veflingstad et al., 2003). Using several data sets clearly constrains the flexibility of the underlying model considerably, and in fact we obtained a solution whose graphs for all variables and all test sets were very close to the true solution. Examples are shown in Figures 3 and 4.

The parameter values of the original system [Equation (4)] and the numerical values from one fitting process are given in Table 1 (Panels A and B). For reasons given in the Discussion section, we fitted the data over the overlapping time intervals [0, 0.5], [0.3, 1], [0.8, 2], [1.5, 3] and [2.5, 5] and
subsequently omitted the first six and last three data points from each set. The solution provides good fits to all traces (e.g. Figs 3 and 4) and could be considered the final result. The correlation coefficients, combined for the six traces of each of the four variables, are [0.987, 0.995, 0.991, 0.998]. While the quality of the fit is certainly sufficient for most purposes, we explored the quality of structurally simpler models, as described in the Methods section, by subtracting the smaller kinetic order, $g_{ij}$ or $h_{ij}$, in Table 1 (Panel B) from each side of the corresponding S-system equation and setting all kinetic orders with a magnitude below 0.1 equal to 0. The data were then refitted with the constrained set of parameters. The result is given in Table 1 (Panel C). It is characterized by fewer non-zero parameters (i.e. a more parsimonious network topology) and correlation coefficients [0.986, 0.994, 0.990, 0.998] that are only very slightly lower than those obtained in the less constrained first fit. Using the same type of simplification again, we obtained the results in Table 1 (Panel D) with correlation coefficients [0.986, 0.994, 0.989, 0.997]. This solution is essentially identical to the true solution in its parameter values. Thus, the solutions in panels B, C and D of Table 1 are equivalent in their quality of fit but differ in the network complexity they imply.

**DISCUSSION**

Metabolic profiles contain a wealth of information about the structure and regulation of pathways. This information is not immediately explicit in the profiles themselves but requires mathematical methods of retrieval. We propose for this purpose the combination of three components: S-system models, decoupling of equations and smoothing and slope estimation with ANNs.

S-systems have been shown to provide valid representations in a large number of theoretical and applied studies (e.g. Voit, 2000). Their key benefit in the present context is that their parameters have a well-defined role and biological meaning, which allows interpretation in terms of pathway structure.

The reformulation of the system of differential equations as a system of algebraic equations has the obvious advantage of making numerical integration unnecessary, but it has other very beneficial consequences. First and foremost, it constitutes an efficient partitioning of the estimation task. If all $X$s and $S$s at $N$ time points are known, the system may be split up into $n$ groups, each corresponding to one of the $n$ differential equations of the system, each of which is evaluated at all $N$ available time points [Equation (3)]. Each of these sets can be analyzed independent of all other sets because all equations within the same set contain the same parameters, $p_{ij}$, whose values are to be determined such that they fit the $N$ 'data sets' consisting of the values of the metabolites and slopes of the given equation at the $N$ available data points. This possibility of separate analysis for each set immediately suggests the option of a parallel implementation. Second, the $n$ sets do not have to be analyzed at the same time but may be evaluated sequentially. Third, the decoupled equations are amenable to the same estimation technique if data are missing or the different state variables are measured asynchronously. Fourth, if traces and their slopes are estimated with universal functions, the parameter estimation is not limited to measured time points. Other points on these traces can be added or used to substitute for measured data points, which may be necessary to balance the sampling distribution. Fifth, poor estimations in some equations do not affect the quality of estimations in other equations, and if some metabolite traces are modeled...
simultaneously, as opposed to fitting the signal piecewise single algebraic representation that captures the entire signal the computationally simpler options because it generates a degree dependent on the data and the system in question.

The parameter estimation itself is based on smoothing of traces and estimation of their slopes. Voit and Savageau (1982a,b) accomplished this with curves that were to be drawn by hand through all data points of a given trace. This method, while successful for the authors’ simple system and nearly error-free data, is cumbersome, possibly subjective and inaccurate, and infeasible for large systems. The three-point method, popular in biochemical engineering, is a possible alternative that uses the values immediately before and after a missing data. An example of a more advanced moving filter technique would be the use of dual-pass Kalman filters. Each of these methods has advantages and drawbacks in terms of simplicity, accuracy, stability and its feasibility for automation, and the choice of the most efficient method is to some degree dependent on the data and the system in question.

The ANN procedure proposed here has advantages over the computationally simpler options because it generates a single algebraic representation that captures the entire signal simultaneously, as opposed to fitting the signal piecewise within a forward and/or reverse moving window, works similarly well for data with no, modest or substantial noise and is still computationally efficient: the ANNs we use produce traces within a few minutes on a now commonplace dual 2.4 GHz CPU 4 GB RAM. Other advantages of the ANN approach are that (a) theoretically, any time course can be modeled with a smooth ANN (Funahashi, 1989; Hornik et al., 1989); (b) the user or an automated cross-validation algorithm may decide how smooth the outcome should be and what deviation in the data should be considered as noise; (c) the output from the ANN, formally an algebraic equation, allows the direct computation of slopes, which one may use for the next step of parameter estimation, and of sensitivities that indicate how robust the output model is and (d) the ANN-estimated universal functions do not exhibit spurious maxima and minima as polynomials are known to do. The most relevant details of the particular ANN used here are presented in Appendix 2. General background information on ANNs as well as on matters of implementation and stopping rules for the regression and validation steps, combined with the selection of optimal network topologies, were recently discussed in Almeida (2002).

Because ANNs are based on logistic transfer functions, they sometimes yield fits that are subtly sigmoidal, especially if the number of hidden nodes is high or the number of data points is proportionally low. This is often hardly noticeable in the data fits themselves, but it does introduce bias in the slope

<table>
<thead>
<tr>
<th>Panel A</th>
<th>(a_1)</th>
<th>(g_{11})</th>
<th>(g_{12})</th>
<th>(g_{13})</th>
<th>(g_{14})</th>
<th>(b_1)</th>
<th>(h_{11})</th>
<th>(h_{12})</th>
<th>(h_{13})</th>
<th>(h_{14})</th>
</tr>
</thead>
<tbody>
<tr>
<td>(X_1)</td>
<td>12</td>
<td>0</td>
<td>0</td>
<td>-0.8</td>
<td>0</td>
<td>10</td>
<td>0.5</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>(X_2)</td>
<td>8</td>
<td>0.5</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>0</td>
<td>0.75</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>(X_3)</td>
<td>3</td>
<td>0</td>
<td>0.75</td>
<td>0</td>
<td>0</td>
<td>5</td>
<td>0</td>
<td>0</td>
<td>0.5</td>
<td>0.2</td>
</tr>
<tr>
<td>(X_4)</td>
<td>2</td>
<td>0.5</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>6</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.8</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Panel B</th>
<th>(a_1)</th>
<th>(g_{11})</th>
<th>(g_{12})</th>
<th>(g_{13})</th>
<th>(g_{14})</th>
<th>(b_1)</th>
<th>(h_{11})</th>
<th>(h_{12})</th>
<th>(h_{13})</th>
<th>(h_{14})</th>
</tr>
</thead>
<tbody>
<tr>
<td>(X_1)</td>
<td>20</td>
<td>0.01277</td>
<td>-0.92739</td>
<td>-0.03803</td>
<td>17.2129</td>
<td>0.3738</td>
<td>0.001631</td>
<td>-0.33314</td>
<td>-0.05072</td>
<td></td>
</tr>
<tr>
<td>(X_2)</td>
<td>14.5085</td>
<td>0.30756</td>
<td>0</td>
<td>-0.06147</td>
<td>0.1574</td>
<td>7.5914</td>
<td>-0.03354</td>
<td>0.49879</td>
<td>-0.04841</td>
<td>0.1694</td>
</tr>
<tr>
<td>(X_3)</td>
<td>1.0146</td>
<td>0.31548</td>
<td>1</td>
<td>0</td>
<td>-0.45731</td>
<td>1.9404</td>
<td>0.29759</td>
<td>-0.01078</td>
<td>0.63904</td>
<td>-0.21437</td>
</tr>
<tr>
<td>(X_4)</td>
<td>11.1751</td>
<td>0.87564</td>
<td>-0.30983</td>
<td>0.54805</td>
<td>15.3778</td>
<td>0.72458</td>
<td>-0.32224</td>
<td>0.52265</td>
<td>0.22029</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Panel C</th>
<th>(a_1)</th>
<th>(g_{11})</th>
<th>(g_{12})</th>
<th>(g_{13})</th>
<th>(g_{14})</th>
<th>(b_1)</th>
<th>(h_{11})</th>
<th>(h_{12})</th>
<th>(h_{13})</th>
<th>(h_{14})</th>
</tr>
</thead>
<tbody>
<tr>
<td>(X_1)</td>
<td>17.9593</td>
<td>0</td>
<td>0</td>
<td>-0.88287</td>
<td>0</td>
<td>15.5005</td>
<td>0.38389</td>
<td>0</td>
<td>-0.25944</td>
<td>0</td>
</tr>
<tr>
<td>(X_2)</td>
<td>7.9673</td>
<td>0.52322</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2.8616</td>
<td>0</td>
<td>0.78023</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>(X_3)</td>
<td>2.1971</td>
<td>0</td>
<td>0.83445</td>
<td>0</td>
<td>0</td>
<td>3.7405</td>
<td>0</td>
<td>0</td>
<td>0.55261</td>
<td>0.14447</td>
</tr>
<tr>
<td>(X_4)</td>
<td>2.5685</td>
<td>0.59129</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>6.5295</td>
<td>0</td>
<td>0.15187</td>
<td>0</td>
<td>0.68664</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Panel D</th>
<th>(a_1)</th>
<th>(g_{11})</th>
<th>(g_{12})</th>
<th>(g_{13})</th>
<th>(g_{14})</th>
<th>(b_1)</th>
<th>(h_{11})</th>
<th>(h_{12})</th>
<th>(h_{13})</th>
<th>(h_{14})</th>
</tr>
</thead>
<tbody>
<tr>
<td>(X_1)</td>
<td>11.5926</td>
<td>0</td>
<td>0</td>
<td>-0.87442</td>
<td>0</td>
<td>9.4683</td>
<td>0.54275</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>(X_2)</td>
<td>7.9673</td>
<td>0.52321</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2.8616</td>
<td>0</td>
<td>0.78023</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>(X_3)</td>
<td>3.2301</td>
<td>0</td>
<td>0.7014</td>
<td>0</td>
<td>0</td>
<td>5.0974</td>
<td>0</td>
<td>0</td>
<td>0.45625</td>
<td>0.17084</td>
</tr>
<tr>
<td>(X_4)</td>
<td>1.8317</td>
<td>0.49001</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>5.8583</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.82772</td>
</tr>
</tbody>
</table>

*Set of true parameter values of the S-system [Equation (4)] from which six data sets were generated. First solution obtained from piecewise ANN-smoothing and regression (see Figs 3 and 4). The correlation coefficients for each variable, combined for the six data sets, are [0.987, 0.995, 0.991, 0.998]. Results of first simplification (see text) with correlation coefficients [0.986, 0.994, 0.990, 0.996]. Results of second simplification with correlation coefficients [0.986, 0.994, 0.989, 0.997]. The graphs in all solutions, and for a variety of other initial conditions, are essentially indistinguishable from those of the true system.
computation, which may, in particular, render the first and last few points of a fit unreliable. This bias can be ameliorated to some degree by overlapping—piecewise fitting, which reduces the necessary number of hidden nodes in each fit thereby yielding smoother traces. Once fitted, the first and last few trace points and slopes are ignored. The result is a set of universal function representations, one for each trace, which in the further analyses replace the data points and allow the computation of slopes. This strategy is conceptually similar to techniques used in adaptive filtering (Diniz, 1997).

Although the ANN-derived universal function is obtained directly from the metabolic profile data and without a predefined mathematical structure, it may not reflect the ‘true’ underlying function. This problem is intrinsic to any structure identification experiment, and its resolution will in most cases require additional subject area information and/or multiple data sets that represent responses to different types of perturbations. As typical problems to be expected, the data could be sparse enough to miss subtle details, or it could be that noise masks properties that otherwise would give crucial clues about the structure and regulation of the pathway.

Even in ideal, low-dimensional cases, a single data set—or even a series of several data sets—may not always suffice to identify uniquely the structure of an underlying system completely [for a discussion of concepts and typical challenges, see Bellman and Åström (1970) and Godfrey et al. (1994)]. For S-system models, power-law relationships among the dependent variables may lead to different yet mathematically equivalent model descriptions and thus preclude unique determinations (e.g. Voit, 1992a; Sands and Voit, 1996). Two approaches can be brought to bear upon this issue. The first, theoretical, diagnostic could be a Lie analysis that identifies transformation groups under whose actions the solutions of the differential equations remain unchanged. Algebraic methods for such an analysis have been developed for S-systems (Voit, 1991: Chapter 15, 1992b). Second, and maybe more important from a practical point of view, any identification algorithm probably will—and should—generate a roster of alternative, data-consistent solution ‘proposals’, from which a subject area scientist has to decide which solutions would make sense and which not and how to design critical experiments distinguishing between these alternatives.

A somewhat related, subtler problem is that any smoothing may bias the parameter estimation in some form, especially if slopes are estimated as a first, auxiliary step. Possible undesired consequences of this bias may be remedied if the solution is merely used as the starting point of a regression that is based on the metabolic profile itself and uses the alleged model directly in its differential form.

The combination of methods proposed here speeds up the inverse problem tremendously. The illustrative four-species modeling problem analyzed here took ∼5 min of automatic ANN smoothing and under 10 min for the regression, using the Matlab™ coded application implemented on the machine described above (see Availability note for public access). This should be seen in contrast to uncoupled approaches that use heavy parallelization on clusters of several hundred processors and still require many hours to obtain solutions even for systems of small size. Combining the decoupling with optimized, parallel implementations of the regression step (e.g. Kikuchi et al., 2003) may make it possible to scale up structure identification algorithms to larger sizes with existing technology. However, even these improvements will not suffice for identifications of very large structures as are emerging in transcriptomics and proteomics profiling. These will require not only further fine-tuning of the techniques shown here but also, more importantly, novel methods for deducing additional information directly from the measured traces. Initial steps have been taken in this direction (Vance et al., 2002; Torralba et al., 2003; Almeida and Voit, 2003; Veflingstad et al., 2003), but much further work remains to be done.

Metabolic networks, which were the focus here, have the advantage that they must obey the law of mass conservation, which imposes significant stoichiometric constraints that can be utilized to restrict the space of possible parameter values. This advantage is lost in proteomic and transcriptomic systems, which can in principle be analyzed with the same methods. Unless at least some constraints can be established, time series analysis of these high-throughput systems must therefore be expected to pose additional and possibly very significant challenges. Even within the realm of metabolic systems, challenging problems remain to be addressed. Clearly, there are numerous technical issues, but there are also conceptual modeling questions that arise from the spatial heterogeneity within cells, which may require partial differential equations and the consideration of delays, and issues of small numbers of molecules, which invalidate many explicit and implicit assumptions underlying kinetic models and probably require stochastic methods.

ACKNOWLEDGEMENTS
Any opinions, findings and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the sponsoring institutions. This work was supported in part by a Quantitative Systems Biotechnology grant (BES-0120288; E.O.V, PI) from the National Science Foundation, a Cancer Center grant from the Department of Energy (Carolyn E. Reed, PI), a National Heart, Lung and Blood Institute Proteomics Initiative through contract N01-HV-28181 (Daniel Knapp, PI) and an Interdisciplinary USC/MUSC grant (Edward P. Gatzke, PI).

REFERENCES


Downloaded from https://academic.oup.com/bioinformatics/article-abstract/20/11/1670/300113 by guest on 23 April 2019
needed to execute all integrations and found that integration required in excess of 95% of the total search time.

If the target model is stiff, the time for solution increases further. As an example, consider a test problem that was introduced by Robertson (1966) and is now frequently used as a benchmark for the efficiency of stiff solvers (cf. Hairer and Wanner, 1997). The system consists of a two-step biochemical reaction in which the first step is reversible with fast forward and slow reverse reaction rates and the second step is very fast. Coding the three metabolites and their concentrations as $C_1$, $C_2$ and $C_3$, Robertson’s example may be formulated as

\[
\begin{align*}
    \dot{C}_1 &= k_1 C_2 C_3 - k_2 C_1 & C_1(0) &= 1, \\
    \dot{C}_2 &= k_2 C_1 - k_3 C_2 C_3 - k_1 C_2^2 & C_2(0) &= 0, \\
    \dot{C}_3 &= k_3 C_2^2 & C_3(0) &= 0,
\end{align*}
\]

where the rate constants have the values $k_1 = 10^4$ s$^{-1}$, $k_2 = 0.04$ s$^{-1}$ and $k_3 = 3 \times 10^7$ s$^{-1}$. High-resolution integration shows that the dynamics starts to be interesting as early as $t = 10^{-4}$ s, when mass begins to flow from $C_1$ to $C_3$, via $C_2$. Most mass is exchanged between $t = 1$ s and $t = 10^2$ s, but the steady state is only reached between $t = 10^6$ s and $t = 10^3$ s, depending on requirements on precision (Burchard et al., 2003, http://www.io-warnemuende.de/homepages/burchard/pdf/papers/Patankar.pdf). To estimate parameters in this example, one would have to solve this system thousands of times, and because each solution requires small integration steps, at least at the beginning, and is relatively slow, the time to adjust parameters in the gradient search becomes negligible in comparison with the total integration time, which may exceed 99%, depending on the time interval for the integration and other settings.

Noteworthy is that even if the true model is not particularly stiff, it may happen that some transient models that are evaluated on the path toward the optimal solution have unfortunate combinations of parameter values that cause stiffness and slow down the search or even fail. The frequency of these events must be expected to increase in systems with larger numbers of components, which constitutes precisely the scenario that is likely to be the norm in the future.

As an example of system failure in a non-stiff parent model, consider the simple pathway with S-system representation

\[
\begin{align*}
    \dot{X}_1 &= \alpha_1 X_2^{0.4} - \beta_1 X_1^{0.5} & X_1(0) &= 1.2, \\
    \dot{X}_2 &= \beta_1 X_1^{0.5} - \beta_2 X_1^{1.5} X_2^{0.2} & X_2(0) &= 1,
\end{align*}
\]

(Voit, 1993), where the symbolic parameters of the ‘true’ model have the values $\alpha_1 = \beta_1 = \beta_2 = 1$, $h_{22} = -0.4$. With these values, the system exhibits damped, stable oscillations and is very robust to various initial conditions. However, if during the search for these parameter values the algorithm explores close-by values like $\alpha_1 = 0.9, \beta_1 = 1.1, \beta_2 = 1.2$ or $h_{22} = -0.5$, one of the variables becomes negative, its non-integer power is undefined and the solution terminates.
transfer function that involves connections to nodes and their weights. [cf. Equations (A1) and (A2)]. Each layer evaluates information from matrices that are associated with each connection between layers h\(j\) (as shown in Figure 5. The value at each node \(h_j\) in the hidden layer is computed from the linear combination of \(m\) input values \(x_i\) with weights \(w_{ij}^{<1>}\), and bias (baseline value) \(b_{xj}\), according to the multivariate sigmoidal function

\[
h_j = \tanh \left( b_{xj} \sum_{i=1}^{m} w_{ij}^{<1>} x_i \right). \tag{A1}\]

Similarly, the values \(y_k\) of the output layer are determined as linear combinations of the \(l\) hidden node values \(h_j\) with weights \(w_{jk}^{<2>}\), and bias (baseline value) \(b_{hk}\), according to the multivariate logistic function

\[
y_k = \frac{1}{1 + \exp(b_{hk} + \sum_{j=1}^{l} w_{jk}^{<2>} h_j)}. \tag{A2}\]

While each individual function \(h_j\) and \(y_k\) is a simple sigmoid, the weighted summation and nesting of multiples of these sigmoids is able to produce functions of arbitrary flexibility if the numbers of hidden nodes is high enough (Funahashi, 1989). The resulting output is therefore referred to as a ‘universal approximator’ (Hornik et al., 1989) or ‘universal function’. Details are presented in Mendes and Kell (1996) and Almeida (2002).

Smoothing the trace of metabolite \(X_i\) requires finding a universal function that approximates the true solution \(X_i(t)\) of the differential equation \(X_i = f_i(X_1, \ldots, X_n)\). Thus, the input data for the ANN are the measurement times, and the outputs are the corresponding values of \(X_i\). These output values are ultimately generated from the optimally weighted summation and nesting of the multiple sigmoids shown above, and the training of the ANN thus consists of optimizing the number of hidden nodes and the weights for both layers. As an illustration of Figure 6 the result of such a training process, consider the trace of \(X_2\) (Fig. 6a) of the example in the text (Fig. 1), as was determined by training an ANN with seven nodes in the hidden layer. For noise-free data, the ANN trace follows the data very precisely, and even if uniformly distributed \(\pm 15\%\) random errors are superimposed on the data, the ANN is able to determine a reasonable trace (Fig. 6b).

The number of hidden nodes, which is a priori unknown and depends on the desired degree of smoothness, can be determined by automated cross-validation (Almeida, 2002) or manually. Large numbers of nodes in the hidden layer(s) tend to follow the time points, including noise, more closely, while fewer nodes sometimes lead to output functions that have fewer maxima and minima, identify outliers by not modeling them precisely and often correspond to more appealing non-linear ‘regression’ lines that distinguish signal from noise in a fashion similar to how human intelligence would do it.

The identification of a well-fitting ANN requires cross-validation for terminating the regression process as soon as validation and test errors diverge, and this process is compounded by the need to optimize the ANN topology, which is determined by the numbers of nodes and their weights. These and other issues associated with ANN models for complex data were reviewed recently (Almeida, 2002) and led us to develop a Matlab library of functions that includes optimal strategies and remedies for all typical problems. It is particularly noteworthy that all configuration parameters for optimal ANN parameterizations are fully automated, which facilitates high-throughput processing of arbitrary experimental time series.

Because the universal function that results from training the ANN is an algebraic function, it is a cumbersome, yet straightforward task to compute its slopes at all time points of interest. This differentiation task is best accomplished symbolically with computer algebra software such as available in Matlab, Maple and Mathematica. The result is convoluted, spanning several lines, and it is easiest to display it in pseudo-code. Consider a system with \(n\) equations, \(x\) time points and \(nh\) hidden nodes. As before, \(wx\) and \(wy\) are the weight matrices in Equations (A1) and (A2), respectively, \(sx\) and \(sy\) are scaling vectors and \(y\) is the concentration of a target metabolic species. Each derivative is then computed in the following fashion:

\[
A(i,:) = \tanh((wx(:,1') \cdot (sx(1,1) - sx(1,1))/((sx(2,1) - sx(1,1)) + wx(:,2')));
B(i,1) = \exp(-sum(wy(1,1 : end - 1)) \cdot A(i,:)) - wy(1, nh + 1);
C(i,1) = sum(wy(1,1 : end - 1) \cdot (1 - A(i,:)'\cdot (wx(:,1'))) / (sx(2,1) - sx(1,1)));
\]
Fig. 6. The ANN method has the advantage that traces are determined very efficiently even if the data contain considerable noise. (a) Noise-free data from the example in the text, along with the ANN-derived trace. (b) Data and trace from the same pathway, but with $\pm 15\%$ uniformly distributed random noise.

\[
dy(i, 1) = \left(\frac{1}{1 + B(i, 1)^2}\right) \ast (sy(2, 1) - sy(1, 1)) \ast C(i, 1) \ast B(i, 1);
\]

A generic problem with these slopes is that they are rather sensitive to the ANN’s preference for sigmoidal fits. Partial solutions to this problem are to ignore the first and last data points and/or to fit the data in several overlapping pieces, which tend to average out the bias to some degree. A full solution will require further research in numerical methods and algorithm identification.