Nonparametric Dynamic Modeling

Mojdeh Faraji and Eberhard O. Voit*

Department of Biomedical Engineering
Georgia Institute of Technology
950 Atlantic Drive, Suite 2115
Atlanta, GA, 30332-2000, USA

* Corresponding author

E-mail: mojdeh@gatech.edu, eberhard.voit@bme.gatech.edu

Key Words

Dynamic Flux Estimation (DFE); Metabolic Pathway Analysis; Nonlinear Compartment Model, Pathway Structure Identification, Systems Biology

Highlights

- It is possible to design dynamic models from time series data without estimating parameters
- The proposed method is a variant of the method of Dynamic Flux Estimation
- The nonparametric model allows typical simulation studies with altered model settings
- Steady states of the nonparametric model may be obtained from non-steady-state data
- Sensitivity and stability analyses are feasible

Abstract

Challenging as it typically is, the estimation of parameter values seems to be an unavoidable step in the design and implementation of any dynamic model. Here, we demonstrate that it is possible to set up, diagnose, and simulate dynamic models without the need to estimate parameter values, if the situation is favorable. Specifically, it is possible to establish nonparametric models for nonlinear compartment models, including metabolic pathway models, if sufficiently many high-quality time series data are available that describe the biological phenomenon under investigation in an appropriate and representative manner. The proposed nonparametric strategy is a variant of the method of Dynamic Flux Estimation (DFE), which permits the estimation of numerical flux profiles from metabolic time series data. However, instead of attempting to formulate these numerical profiles as explicit functions and to optimize their parameter values, as it is done in DFE, the metabolite and flux profiles are used here directly as a scaffold for a library from which values are interpolated and retrieved for the simulation of the differential equations describing the model. Beyond simulations, the proposed methods render it possible to determine steady states from non-steady state data, perform sensitivity analyses, and estimate the Jacobian of the system at a steady state.

Introduction

Ever since the digital revolution drove analog computing to the brink of extinction, the design of computational models for complex systems has become an effort in choosing optimal mathematical representations and their parameter values. For most physical and engineering systems, the choice of model functions is directly guided by our rather solid understanding of basic physical concepts, such as mechanical or electrical forces, dilution and dispersion phenomena, optical processes, and the features of electric circuits. Biological systems are, of course, objects of the physical world and must therefore obey the laws of physics, but most processes that govern even moderately sized biological systems are so convoluted that they cannot be dissected into elementary physical representations [1]. As an example, the transmission of a neuronal signal at a dopamine synapse requires electrical activation, the prior biochemical production of dopamine and its packaging into membrane vesicles, the move of these vesicles through the crowded cytoplasm toward the synapse, the merging of vesicle and cell membranes, the opening of this membrane toward the synapse, the release of dopamine out of the vesicle and through the synaptic cleft to a receptor on the postsynaptic neuron, possible interactions with other neurotransmitters, and binding to the receptor. This binding in turn triggers a slew of additional mechanisms inside the signal receiving cell, including the complex process of signal interpretation which in the case of dopamine is often accomplished through multiple phosphorylation of the specific protein DARPP32, and the possible long-term adaptation to repeated stimuli [2; 3; 4; 5]. Thus, a very coarse model could easily capture the fact that a signal moved from one neuron to another, but a detailed mechanistic model becomes quickly bogged down in the minutiae of the numerous intertwined biophysical processes that are involved in signal transduction.

Because elementary physical descriptions are often infeasible, the biological systems modeler is forced to resort to “higher-order” process representations, *ad hoc* models, suitable approximations, or combinations thereof. A good example is the Michaelis-Menten function of enzyme kinetics [6]. Its underlying concept is a process that postulates the reversible formation of a biochemical complex between an enzyme and its substrate and the subsequent release of the product of the reaction and of the enzyme, which is used over and over again. Under idealistic conditions *in vitro*, this concept is believed to be quite realistic. However, within living cells, the prerequisites for the involved mass-action functions are clearly not satisfied, and the so-called quasi-steady-state assumption, which is needed to formulate the process with a simple, explicit function, only holds under certain conditions. Thus, an idealized concept, formulated with the help of somewhat doubtful approximations, becomes a higher-order process

representation for enzyme catalyzed reactions. Indeed, the Michaelis-Menten function performs well *in vitro* and, in an approximate sense, presumably *in vivo*, although this is not really known. For simulations of large pathway systems, this function is often used as well, but its mathematical features become rather cumbersome, even for standard model assessments such as sensitivity analyses [7].

Notwithstanding these mathematical issues, it is common for the biological modeling community to base simulation studies in a variety of fields on a rather small set of functions, which are used time and again and prominently include mass-action, Michaelis-Menten and Hill functions, which often include regulatory terms [8]. The users of these functions rely on the argument that these functions suit their purposes—quasi as black boxes—and are sufficiently accurate if one considers the typical noise encountered in biological data. Furthermore, these particular functions at least have some foundation and rationale in biology, whereas the use of a function like a shifted arctangent has very little justification, except that its graph is s-shaped and therefore might resemble some saturation processes in biology.

True alternatives to these *ad hoc* approaches are generic approximations. Linearization, the simplest of these, has been enjoying enormous successes in engineering applications for many decades. For the representation of biological phenomena, by contrast, linear models tend to run into conflicts with the genuine nonlinearities that characterize living systems. For instance, common features like saturation, stable oscillations, threshold phenomena, synergisms, or chaos cannot directly be modeled with linear equations. A logical solution might seem to be the expansion of linear models to second-order Taylor approximations, but these become so awkward for larger systems [9] that very few modelers have resorted to this option. Instead, many biological modeling groups have been using power-law approximations, which are nonlinear, but have linear characteristics in logarithmic space. Biochemical Systems Theory (BST; [10; 11; 12; 13]) and Metabolic Control Analysis (MCA; [14; 15; 16; 17]), which directly or indirectly utilize power-law representations [18; 19], respectively, have had success with analyses of a wide variety of complex biological systems (for a review, see [20]). Notwithstanding their successes, power-law representations are local approximations and therefore genuinely limited in their accuracy of capturing phenomena over large ranges of variation in the involved variables. As a case in point, univariate power-law functions in BST do not saturate for large substrate concentrations, and lin-log models, which are associated with MCA, become negative for small substrate concentrations and tend toward $-\infty$ for substrate concentrations approaching 0 [21; 22; 23]. As an alternative to these canonical power-law

models, one could use sigmoidal basis functions, but for realistic models this option requires correspondingly larger numbers of parameters that need to be estimated [24; 25].

Even if reasonable guideposts could be found to justify the choice of appropriate model representations, the second step of model identification is still to be performed, namely the estimation of parameter values. For moderately sized or large models, this estimation is always challenging [26; 27; 28], due to noise in the data, non-convergence of the search algorithm and other problems, or because the wrong model was chosen after all. To make matters worse, even an excellent fit is not necessarily optimal, and the parameterized model may perform poorly in extrapolations, because the original fit was obscuring the compensation of errors among some terms within the model (e.g., see [29; 30]). Furthermore, an excellent fit may be the result of overfitting with a model containing too many parameters.

These challenges and compromises lead to the obvious question of whether it might be possible to glean appropriate functions directly from experimental biological data, without presupposing potentially unjustified mathematical formats. The method of Dynamic Flux Estimation (DFE), which permits a relatively unbiased estimation of fluxes within a system and which will be reviewed later, took a first step toward answering this question affirmatively, at least for metabolic systems under ideal conditions [31]. Still, DFE requires some choices of model frameworks when the task is setting up a model from scratch.

In this article, we describe a novel variant of DFE that makes such choices unnecessary, at least under favorable conditions. Given such conditions, the overall result of the proposed strategy is that it is possible to develop dynamic models in a nonparametric manner. Intriguingly, the resulting nonparametric models, which make no assumptions regarding parameter values or even mathematical formats, beyond the topology of the system, permit most of the typical diagnoses and analyses that are possible with a fully parametric model, which may be considered the gold standard in the field. As a consequence, simulations and other analyses can be performed without the complicated and often biased step of choosing models and parameterizing them, if suitable data are available. The data needed for this purpose consist of sets of time series that representatively capture the dynamics of a system under relevant inputs.

Both DFE and the nonparametric variant proposed here are particularly well suited for nonlinear, dynamic, regulated compartment models, because these possess the property of mass conservation, which imposes strong, unbiased constraints that greatly aid the formulation of appropriate models. As an illustration, and for ease of discussion, we will focus here on metabolic pathway systems, but it appears that other nonlinear compartment systems, such as SIR models of epidemiology and pharmacokinetic models, can be treated in the same manner.

Methods

Dynamic Flux Estimation (DFE)

The stoichiometric equation

$$\dot{\mathbf{X}} = \mathbf{S} \cdot \mathbf{V} \quad (1)$$

provides a generic description of the dynamics of a metabolic pathway system. This well-known equation collectively formulates dynamic changes in each metabolite of the system, $\frac{dX}{dt} = \dot{X}$, as a product between the stoichiometric matrix $\mathbf{S}$ and a vector of reactions or fluxes, $\mathbf{V}$. This product formulation is remarkable, as it naturally separates the linear aspects of the system from its nonlinear features. Specifically, consider the situation where the slopes of all metabolites on the left-hand sides are known for some given time point. If so, Eq. (1) is a system of linear algebraic equations, where each variable V_j represents the state of a flux at this time point, rather than a metabolite. The nonlinear features enter the system secondarily, by virtue of the fact that each component of the flux vector is a possibly complicated function of metabolites and regulators, and therefore of time. Dynamic Flux Estimation (DFE) makes maximal use of this separation of the model into linear and nonlinear components.

In typical analyses, such as Flux Balance Analysis, the stoichiometric equation (1) is studied at a steady state of the system [32; 33; 34], where the vector on the left-hand side contains zeros. DFE reaches beyond the steady state, by addressing the system at many time points of a system's trajectory, where the vector of derivatives is different from zero. In its first phase, DFE uses time series measurements of metabolite concentrations, $X_1, \dots, X_n$, along with estimates of the slopes of these time courses. Thus, DFE evaluates equations of the type

$$\frac{dX_i}{dt} \text{ at } t_k = \text{Slope of } X_i \text{ at } t_k = \sum_{j=1}^m s_{i,j} V_j(t_k) \quad (2)$$

where the slopes are numerical values, estimated from the data. Since data are typically noisy or incomplete, it is advisable to apply one of various available preprocessing, smoothing and data substitution techniques (e.g., [35; 36; 37; 38]). We do not discuss these further in this article, because data smoothing techniques and the methods proposed here constitute clearly distinct steps within the model design procedure. Thus, we will assume in the following that the data had been successfully smoothed.

The slope substitution is performed for m time points, with the result that each differential equation in (1) is replaced with a set of m linear algebraic equations at these time points. Collectively, these equations may be formulated as a matrix equation, where the variables are the fluxes V_j , rather than the metabolites [39; 40; 41; 42]. If this matrix has full rank, the solution is unique, and if the equations are overdetermined, the best-fitting solution is computed via linear regression. In the most common case, the system is underdetermined. We will briefly skip this case here, but return to it later. The result of solving the linear equations is a set of numerical flux values at each time point. Collecting these sets for all time points it is straightforward to create a plot of each flux as a function of time. It is furthermore possible to plot each flux against the system variables upon which it depends. In the case of a single substrate and no regulation, the plot is a simple line graph. By the same token, if the flux depends on two or more variables, the result is a line on the manifold that is given by the unknown flux function in three or more dimensions. This first phase of DFE typically requires knowledge of the connectivity of the system. However, it is to some degree possible to infer formerly unknown reaction steps and regulatory signals [43; 44].

Figure 1. Dynamic Flux Estimation (DFE). The method consists of two phases. Phase 1 (top) is model free in a sense that only the stoichiometry is assumed to be known, whereas functional forms of the process representations are not. The procedure in this phase is based on raw experimental time series data in the form of metabolite concentrations. It is beneficial to smooth these data with some numerical algorithm, such as a spline. Next, the rate of change in each metabolite is obtained from the smoothed time courses. These numerical slopes correspond to the values on the left-hand side of Eq. (2), so that numerical evaluation of the slopes at m time points converts each differential equation of the model into a set of m linear algebraic equations, in which the flux states are the driving variables. The system is solved, potentially with the aid of additional constraints, and the result is a time dependent, numerical profile of all fluxes. In Phase 2 (bottom), the dynamic flux profiles are fitted with appropriate functions or rate laws, and the result is a fully parameterized dynamic model.

In Phase 2 of DFE, the numerical flux representations are converted into mathematical functions. For this purpose, assumptions must be made regarding the functional format of each flux. In some cases, the shape of the flux profile or independent biological considerations may suggest a mathematical format, but this is not guaranteed (Figure 2). Once a format has been chosen, the parameters of each flux representation are to be estimated from the available data, as it is typical for any other modeling effort. This estimation is much simpler than for the entire system, because it is performed for a single explicit function at a time, rather than simultaneously for all fluxes in the system of ODEs. The result of the two phases combined is a fully parameterized model of the pathway system.

It is not directly possible to solve the linear system when the stoichiometric matrix is underdetermined. Unfortunately, this is actually the most common case for metabolic systems, which typically contain more reactions than metabolite pools. In the case of an underdetermined system, the stoichiometric equation admits infinitely many solutions, and these can differ tremendously, even if the system is small, with some having monotonic shapes, while others may overshoot or exhibit oscillations [45]. To study these sets of solutions, it is advisable to reduce the system mathematically to a system whose dimension equals the degrees of freedom [45]; for instance, in a system with six metabolites and eight reactions, the dimension will typically be two.

A solution to the challenge of under-determination is the Moore-Penrose pseudoinverse [46; 47; 48]. While effective, the pseudoinverse usually contains negative values, which are not consistent with biological fluxes. Several other approaches have been proposed. First, characterizability analysis, based on the pseudoinverse, shows directly where additional information is needed about the system, or which metabolite pools could be merged, to make the equations uniquely solvable [49]. Second, it is sometimes possible to measure influxes or effluxes experimentally or to infer them from the data [43; 44]. Third, it might be feasible to obtain additional biological information regarding some of the internal fluxes. For instance, one might be able to deduce internal fluxes from biochemical features of the involved metabolites [50]. Fourth, one may use the dynamic concentration profile associated with one metabolite to estimate flux functions for its influx and efflux [51]. This procedure requires assumptions regarding the functional forms of these fluxes, but if the variable does not change much in magnitude, an approximate representation is likely to be sufficient. Fifth, sufficiently many datasets may allow the inference of a few flux profiles directly from the data [52]. Finally, it might be possible to utilize biological constraints, such as energy minimization, to reduce the degrees of freedom within the system [45].

Figure 2. Typical results of Phase 1 of DFE. The flux in panel A may be representable with a Michaelis-Menten function. By contrast, it might be difficult to choose appropriate formats for the fluxes in panels B and C.

If independent information regarding a flux can be found, the system of equations becomes simpler. For instance, suppose that flux v_3 is known in sufficient detail. Then, Eq. (2) becomes

$$\frac{dX_i}{dt} - s_{i,3} \cdot V_3 = \sum_{j=1, j \neq 3}^m s_{i,j} V_j, \quad (3)$$

where the left-hand side is numerically known and the number of free degrees is typically reduced by 1.

Concepts of Nonparametric Dynamic Modeling

Overview. The core idea of nonparametric modeling is to forgo Phase 2 of DFE and instead to replace the functional representations of the processes in the system with a library of numerical results directly obtained from Phase 1 of DFE. Of course, nothing comes for free: The proposed substitution requires good, comprehensive data. While such datasets may currently be scarce, the decreasing cost of generating rich datasets renders the proposed method increasingly more appealing.

Thus, let us suppose that rich data are available, where “richness” refers to more or less complete datasets and enough time series to represent the phenomenon under investigation appropriately. For instance, one could imagine sets of time series experiments with many combinations of different input and inhibitor concentrations. Collectively these datasets form the scaffold for the proposed modeling strategy. One should note that even a rich dataset rarely produces complete coverage of the imaginable metabolic profile space. One reason is that most metabolic systems have a single non-trivial steady state and that trajectories even from a variety

of initial values tend to approach this steady state, so that many regions of the mathematically imaginable solution space are only scarcely covered or not at all. While this situation may seem to be undesirable, it actually reflects those portions of the metabolic profile space that are most relevant and reliably represented, while ignoring situations that may be biologically less relevant. Figure 3 shows a generic example where the trajectories approach a common steady state.

For metabolic pathway systems and many other compartment models, the topology of the underlying network of components is typically known, or assumed to be known. It is also often, although not always, known which metabolites affect each flux as substrates or modulators. Nonparametric modeling takes advantage of this situation. It directly uses Eq. (2), where the stoichiometric coefficients are assumed to be given and where the flux states are the dependent variables on the right-hand sides. In contrast to parametric modeling, the explicit functions on the right-hand sides of the differential equations are here replaced with numerical flux descriptions. Thus, no mathematical formalism is used to represent the fluxes and hence no parameterization is needed. Instead, the numerical flux-substrate profiles are generated directly from the smoothed experimental data processed in Phase 1 of DFE and recorded in a library that is subsequently used for looking up values required by the ODE solver. As a result, the library, combined with an interpolation algorithm, constitutes the primary simulation tool for analyses with the nonparametric model (Figure 4). Expressed differently, the processes of making functional assumptions and of parameterizing these functions are replaced by the use of a scaffolding library and geometric interpolations, once dynamic flux profiles have been obtained. The details of setting up the library and of the interpolation algorithm are discussed in the next sections.

Figure 3. Illustrative example for scaffolding the library with different datasets. The green surface is the graph of the unknown flux function $V(x, y)$. Each white line represents the result of a time series experiment. All time series approach the same steady state (yellow) so that many regions of the surface, such as the top right, are not supported by measurements and may be biologically irrelevant. Thus, the set of time series results may be sufficient to interpolate between the data (white lines) but does not allow reliable extrapolations far beyond the data.

Figure 4. Nonparametric modeling framework. Functional assumptions and the parameterization of fluxes are circumvented in nonparametric modeling. Instead, the flux values needed at each step of a system simulation are provided by look-up tables that are generated from the measured dynamic metabolite profiles and the inferred flux profiles. They are stored in a library for call-up.

Library Construction. Phase 1 of DFE results in dynamic flux profiles that can be displayed against time or against the systems variables that affect the flux. Combining the smoothed metabolic time series and flux profiles, flux-substrate relationships are generated and recorded as arrays in which the columns contain values of substrates and regulatory agents that affect each flux, as well as the corresponding flux value itself. The rows represent snapshots of the state of the system at various time points and possibly from different experiments (Figure 5). To ensure sufficient scaffolding of the flux-substrate subspace, each array should ideally include the dynamic profile of a flux for several experimental settings, for instance, with different initial conditions.

Of course, the quality of further analyses and simulations depends on the quality of the library, which in turn depends on the quality of the available data. It is difficult to quantify how many data have to be available, as this quantity depends on the complexity of the flux functions. For example, oscillatory data will often require more data points than smooth monotonic data for a good representation of the true time trend. Thus, one can only pose a vague criterion that must be satisfied: the data must be representative of the trends they quantify.

In theory, every metabolite could be involved in every flux of the system, thereby causing an unmanageable combinatorial explosion for larger systems. In practice, this situation does not arise, and most metabolites are only affected by a few other metabolites.

Interpolation Algorithm. To be useful, the nonparametric model must permit simulations of the evolution of the modeled system over time. The model is formally specified as a set of ODEs whose right-hand sides are defined by fluxes whose values are stored in the libraries described before. By its nature, each ODE solver needs to have flux values available for essentially arbitrary substrate and modulator amounts within reasonable ranges, which are, for instance, bounded by the experimental datasets. In the case of nonparametric modeling, the library consists of a discrete set of values, which covers only a finite subset of all values that are possibly needed. To overcome this issue, a method is needed that allows the ODE solver quickly to estimate each required flux value from the data in the library. Obviously, the quality of this type of gap-filling is directly correlated with the density and quality of the data in the library.

	Substrate	Regulator	Flux
Experiment 1	0.506	0.234	9.168
	0.580	0.529	10.35
	0.647	0.779	11.33
	...	...	...
Experiment 2	1.012	0.766	15.82
	1.026	0.719	15.96
	1.038	0.679	16.08
	...	...	...
Experiment 3	1.963	0.428	25.58
	1.857	0.435	24.55
	1.762	0.490	23.67
	...	...	...

Figure 5. Generic format of the flux library. The data are arranged in arrays where the columns represent a specific flux along with all substrate(s) and regulator(s) that affect this flux directly. The rows represent snapshots of the state of the system for different time points and experiments. The specific numerical values shown here refer to the later example of a fermentation pathway but are not of import here.

A convenient tool for this task is the function *scatteredInterpolant* in MATLAB (version R2014a, The MathWorks, Natick, MA), which is applied to the smoothed data obtained from

DFE. This function generates 2D or 3D interpolations from the dataset in the library in an unbiased manner. Specifically, the interpolant passes through the original data and uses adjustable methods such as linear or nearest-neighbor interpolation, which enable the algorithm to estimate flux values for arbitrary substrate and modulator values during the ODE simulation.

For our demonstration of the concepts of nonparametric modeling with fluxes of two or three arguments, we used the linear interpolation method. For fluxes with a single substrate and no regulators, the function *interp1* was preferred, as it is more efficient for one-dimensional interpolations. Figure 6 presents some output from this interpolation algorithm. Panel (a) exhibits data points for a flux V that depends on two substrates S_1 and S_2 . Panel (b) visualizes the interpolated surface over a grid of inputs.

Preliminary studies indicate that the interpolation, under favorable conditions, can generate sufficiently accurate libraries even for data that had not been smoothed. Indeed, skipping the smoothing step may lead to means of assessing the effects of intrinsic noise in the data. For instance, single data points or subsets could be removed from the library in Monte-Carlo simulations, thereby ultimately yielding distributions of slightly different trajectories and steady states. This option will be addressed elsewhere. It also remains to be more formally investigated under what conditions the chosen interpolation algorithm is optimal for the tasks discussed here or whether different, specifically customized interpolation methods might perform better.

Figure 6. Interpolation of a flux-versus-substrate surface for two substrates. The algorithm uses an interpolant function that efficiently connects the data points. Panel (a) shows the raw data points, while panel (b) shows the interpolated surface.

Expanded Data Coverage by Moderate Extrapolation. Parametric models may typically be extrapolated without bounds. However, one must ask to what degree such extrapolations beyond the domain of experimental data are really justified. In our situation of nonparametric modeling, the interpolating function in MATLAB is able to extrapolate datasets to some degree. However, the accuracy of extrapolation is only guaranteed for relatively small ranges in the vicinity of the scaffolding data subspace. While the lack of far-reaching extrapolations may be seen as a disadvantage over parametric models, it also provides a healthy warning against extrapolations that are not supported by data and might even be biologically infeasible. We used the nearest-neighbor variant of the interpolation technique in MATLAB for extrapolations. An example will be shown later.

Library-based Simulations. The flow of a typical simulation with a nonparametric model is illustrated in Figure 7. Starting from an arbitrary initial state, within or sufficiently near the recorded flux-substrate subspace, the library, together with the interpolation algorithm, provides for each set of metabolites the appropriate flux values V_i . The flux values are then used to compute changes in the substrates, $\frac{d\mathbf{X}}{dt}$, using Eq. (1). These changes are utilized to compute the state of the system at the next time point; the base concept of this procedure is shown in Eq. (4).

$$\mathbf{X}_{\tau+\Delta t} \approx \mathbf{X}_{\tau} + \frac{d\mathbf{X}_{\tau}}{dt} \cdot \Delta t, \quad (4)$$

where $\frac{d\mathbf{X}_{\tau}}{dt}$ represents the collection of appropriate flux values at time τ . While equation 4 demonstrates the solution procedure with Euler's forward method, modern ODE solvers employ more sophisticated methods, some of which are expansions of Euler's method. Specifically, the standard ODE solver in MATLAB (version R2014a, The MathWorks, Natick, MA) uses the Runge-Kutta method with variable step size. In contrast to supplying the ODE solver with the parametric functions on the right-hand sides of the ODEs, as it is commonly done, we specify the appropriate interpolants from the library. This substitution does not affect the computation speed much.

Figure 7. Flowchart for a typical nonparametric simulation. Once initial conditions are specified, the algorithm extracts from the library the flux values that correspond to the metabolite profile at the initial time point. From these values, the algorithm computes the slopes of all variables and moves the simulation to the next time step, according to Eq. (4), until the desired final time point T_F is reached.

Typical Model Analyses

Analyses with the nonparametric model are primarily based on simulations. The simplest of these are changes in initial values of one or more of the dependent variables. Similarly easy to assess are changes in independent variables, which are often used to model constant quantities like an enzyme activity or a fixed input. As an example, suppose that some experimental technique raises the activity of an enzyme, which results in a concomitant 20% increase in the corresponding flux by. An analogous parametric situation would be a 20% raise in the V_{max} of a Michaelis-Menten or Hill rate law. In the nonparametric case, such an alteration is simply implemented by increasing all pertinent flux values in the library by 20%. The analogous strategy holds for an external inhibitor, as long as it decreases one or more fluxes in a multiplicative manner. A change in K_m corresponds to a different scale for the substrate concentration. For example, if K_m is to be doubled, an appropriate solution is to double the substrate concentrations in the library without changing the corresponding flux values. To see the rationale for this strategy, consider the standard Michaelis-Menten rate law, $V = \frac{V_{max}S}{K_m + S}$. If K_m is doubled, the same functional values of V are achieved if each value of S is doubled.

To some degree, it is even possible to perform analyses that at first seem to require functional forms. For instance, once a steady state has been identified, one may use the plots of each flux against each contributing variable and determine the slope(s) of the flux at the steady-state metabolite concentration. Entering all slopes into an appropriately laid-out array yields an

approximate Jacobian matrix, which may then be used for numerical analyses characterizing the model behavior close to the steady state and, in particular, the stability of the steady state. We will discuss this option, as well as sensitivity analysis, within the context of the case study in the *Results* section.

Of course, the nonparametric model cannot entirely replace its parametric analog. For instance, it seems difficult to perform formal investigations of bifurcations and other features of complex dynamics, at least in a direct manner.

Results

Case Study: Nonparametric Modeling of the Fermentation Pathway in Yeast

To illustrate the nonparametric modeling capabilities without being encumbered by the idiosyncrasies of experimental datasets, we created artificial “data” from a model in the literature. This model describes in a simplified manner the anaerobic fermentation pathway in the baker’s yeast *Saccharomyces cerevisiae*. The pathway is comparatively well understood, and a considerable body of *in vivo* measurements of metabolites and fluxes at various steady-states is available [53]. This model was originally proposed by Galazzo and Bailey [54] and subsequently converted into a power-law model by Curto *et al.* [55; 56; 57]; it has been used on numerous occasions to demonstrate new modeling and optimization techniques [58; 59; 60; 61; 62; 63].

For the illustration here, we use Curto’s version of the model to generate datasets, which under typical conditions would have been obtained experimentally in the laboratory. We analyze the data without noise. Thus, we pretend that metabolic time courses had been measured and used in DFE to reveal plots of all fluxes *versus* time or *versus* the system variables that affect them. The system contains only a rather small number of metabolites and fluxes, which renders it a good candidate for illustration purposes.

The model captures the dynamics of the fermentation pathway from glucose uptake to ethanol yield (Figure 8). The pathway has essentially a linear structure, although two minor pathways branch off. It is regulated through negative feedback from glucose-6-phosphate (G6P) on glucose uptake and feedforward activation of the enzyme pyruvate kinase by fructose-1,6-bisphosphate (FBP).

Figure 8. Anaerobic fermentation pathway in *Saccharomyces cerevisiae*. The model of the pathway contains five dependent variables and eight fluxes.

The independent variables and fluxes in the model (see Figure 8) are as follows:

X_1 : Glucose (Glc)	$V_1 : V_{in}$	$V_5 : V_{GAPD}$
X_2 : Glucose – 6 – phosphate (G6P)	$V_2 : V_{HK}$	$V_6 : V_{GOL}$
X_3 : Fructose – 1,6 – bisphosphate (FBP)	$V_3 : V_{PFK}$	$V_7 : V_{PK}$
X_4 : Phosphoenolpyruvate (PEP)	$V_4 : V_{POL}$	$V_8 : V_{ATPase}$
X_5 : ATP		

Figure 9 shows the model scheme in a simplified fashion; although regulation is not explicitly shown, it is taken into account by the model. One notes that flux v_5 splits a 6-carbon molecule into two 3-carbon molecules, which causes a doubled rate of influx into the pools of x_4 and x_5 , and mandates corresponding elements of 2 in the stoichiometric matrix (Eq. 5).

Figure 9. Model scheme of the fermentation model.

$$S = \begin{bmatrix} 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & -1 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & -1 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 2 & 0 & -1 & 0 \\ 0 & -1 & -1 & -1 & 2 & 0 & 1 & -1 \end{bmatrix} \quad (5)$$

The stoichiometric equation of the system, $\dot{X} = SV$, is equivalent to the set of differential equations in Eq. (6).

$$\begin{aligned} \dot{X}_1 &= V_1(X_2) - V_2(X_1, X_5) \\ \dot{X}_2 &= V_2(X_1, X_5) - V_3(X_2, X_5) - V_4(X_2) \\ \dot{X}_3 &= V_3(X_2, X_5) - V_5(X_3, X_5) - V_6(X_3, X_4, X_5) \\ \dot{X}_4 &= 2V_5(X_3, X_5) - V_7(X_3, X_4, X_5) \\ \dot{X}_5 &= V_7(X_3, X_4, X_5) + 2V_5(X_3, X_5) - V_2(X_1, X_5) - V_3(X_2, X_5) - V_4(X_2) - V_6(X_3, X_4, X_5) \end{aligned} \quad (6)$$

The model has a stable steady state with concentration values, in [mM], Glc = 0.03456, G6P = 1.011, FBP = 9.188, PEP = 0.009532, and ATP = 1.128 [56].

Library for the Fermentation Model

In an actual systems analysis, experimental time series data would be used to populate the library. Instead, we use the Curto model and solve it multiple times, every time starting from different initial states of all metabolites. These states are located in a hypercube in $\mathbb{R}^5$ that corresponds to the five substrates represented in the pathway model. Of course, experimental data would be noisy. For clarity, we consider noise-free or well-smoothed data.

The collective result consists of time series data of the substrate concentrations and corresponding fluxes, which in reality would have been obtained from DFE. In addition to time plots, the data allow us to establish flux-versus-substrate trajectories, whose values are recorded in arrays, as it was illustrated in Figure 5. Figure 10 shows examples of trajectories for the fermentation model. In all figures, the metabolite concentrations are in millimolar (mM), and all fluxes are in millimolar per unit of time (mM/min).

Dynamic Simulations

Once the library is set up from the artificial data, the nonparametric model is ready to use. Figure 11 confirms with simulation results that the nonparametric model returns essentially the same results for three distinct initial conditions as a parametric simulation with Curto's model. For comparison, the nonparametric results (lines) are superimposed on the "data" (parametric results).

Jacobian at the steady-state. The library permits the estimation of the Jacobian of the system. Namely, once a steady-state metabolite profile has been determined, one estimates the slopes of the fluxes at this profile with respect to the various variables. The slopes of each flux, $\frac{dV}{dX_i}$, are computed numerically, using the flux values at the steady-state, $V(X_i)$, and the flux values at points in the vicinity of the steady-state, $V(X_i + \Delta X_i)$. Using Eq. (1), the flux slopes together yield $\frac{d\dot{X}}{dX_i}$. The quantities $\frac{d\dot{X}}{dX_i}$ of all system variables constitute the elements of the Jacobian matrix. The most prominent use of the Jacobian is the determination of eigenvalues for local stability analysis. This analysis is directly comparable with the corresponding analysis of a linearized parametric model.

Figure 10. Flux-substrate profiles corresponding to different initial conditions. The dynamics of the system forms trajectories that are recorded and used in the form of look-up tables that substitute for explicit mathematical representations of the fluxes. The panels represent the different fluxes. For each panel, metabolite trajectories from different experiments are shown in different colors. Panel (a) shows V_{in} , V_{POL} and V_{ATPase} , which are single-substrate fluxes. Panels (b), (c) and (d) exhibit V_{HK} , V_{PFK} and V_{GAPD} respectively. These three fluxes have two substrates each. Lastly, panel (e) shows substrate trajectories for V_{PK} . This flux has three substrates so that the flux-substrate trajectory is four-dimensional. For this reason, only the substrate trajectory is shown. According to Curto *et al.* [55], the flux V_{GOL} is proportional to V_{PK} with a proportionality constant of about 0.03 and therefore not shown.

For the fermentation pathway, the eigenvalues of the nonparametric model at the steady-state are $[-1734.0, -333.5, -14.04 \pm 7.581i, -1.635]$. For comparison, the corresponding analysis for the parametric model yields quite similar eigenvalues: $[-1689.0, -341.1, -13.66 \pm 7.853i, -1.843]$.

Sensitivity Analysis

An important component of modeling is sensitivity and gain analysis. In the former case, a parameter is slightly altered and corresponding changes in critical system features are recorded. In the latter case, an independent variable is altered. Most prominent is the effect of changes in parameters or independent variables on the steady-state values of the system. While the nonparametric model obviously has no parameters, it is still possible to study sensitivities if one of the fluxes is either raised or decreased by a small percentage, which mimics a situation where the activity of an enzyme is altered. The results of such an analysis are shown in Figure 12. As one would expect, the system reaches a new steady-state, which depends on the specific flux alteration. Comparisons with Curto's model demonstrate that the nonparametric model matches the corresponding analytical results from the parametric model very well.

Figure 11. Nonparametric and parametric simulations of the fermentation pathway. Three different initial conditions were used; the results are shown in different colors. All simulations converge to the same steady state: Glc: 0.03456, G6P: 1.011, FBP: 9.188, PEP: 0.009532, ATP: 1.128. The initial conditions are: green: [0.0657 1.213 12.86 0.0086 1.410]; blue: [0.0518 0.6066 5.513 0.0148 0.6203]; red: [0.03456 1.011 9.188 0.009532 1.128].

red: [0.0207 1.820 17.46 0.0052 2.199]. The nonparametric simulation results (solid line) match the artificial data (circles) very closely.

Figure 12. Sensitivity analysis. Each of the fluxes was separately perturbed by $\pm 10\%$ for the entire duration of each simulation. With each change in a flux, the simulation starts from the original steady state of the system and moves toward the new steady state. For all the flux perturbations other than V_{POL} and V_{GOL} (not shown here because they are insensitive), the system converges to new steady-states. The

nonparametric model results (solid lines) closely match the parametric results (circles). One should note different scales for the five variables.

Figure 13. Trajectories and steady-states of the system for enzymes with altered K_m . The system dynamics was replaced such that fluxes depict enzymes with increased K_m values. Each row represents simulation results for a flux catalyzed by an enzyme with an altered K_m , where the corresponding substrates respectively from top are: [Glc, G6P, FBP, PEP, ATP]. Each flux was modified to mimic a two-fold increase in K_m . Only V_{ATPase} was simulated for a 30% increase, because larger perturbations lead to instabilities in the parametric and nonparametric models. V_{POL} and V_{GOL} did not show any noticeable changes in steady-states and hence are not shown in here. Solid lines show the nonparametric results, while circles represent parametric results from Curto's model.

Figure 13 exhibits simulation results representing the case where the K_m of an enzyme is altered. To mimic an increase in a K_m , more substrate is needed to compensate for the same flux magnitude. As was explained earlier, for example, doubling the K_m of a reaction corresponds to replacing the substrate concentration S with $S/2$ in the library. As can be seen, the steady-states take new values. Specifically, the substrate of the enzyme with altered K_m shows an increased steady-state value in each row. Again, the nonparametric and parametric model results agree quite closely.

Data Collection from a Bolus Experiment

Model set-up. To test the nonparametric approach under more realistic circumstances, we used the Curto model to simulate an *in vivo* nuclear magnetic resonance experiment, which typically uses a bolus input rather than a constant substrate influx (for a pertinent example, see [64]).

In the original Curto model, the input flux of the system, V_1 , is constant and set to simulate a saturated flux that corresponds to an exterior glucose concentration in excess. To model the transient dynamics of a bolus experiment, we converted external glucose into a dependent variable X_0 , thereby expanding the system by one ODE:

$$\dot{X}_0 = -0.01V_1 \quad (7)$$

The coefficient 0.01 reflects the compensation between the concentration within the large exterior volume of the medium and the much smaller interior volume of the cells. This coefficient quantifies how much a change in the external glucose concentration (e.g., from 4 mM to 3.99 mM) affects the internal glucose concentration (from 0 to 1 mM). For the uptake function, we use a Michaelis-Menten function with allosteric, noncompetitive inhibition, where the inhibitor, G6P, reduces the overall activity of the enzyme. This choice was based on the original article which considered the inhibition as not competitive, because G6P binds to the glucose transporter inside the cell while glucose binds on the outside [53]. Thus, we have

$$V_1 = \frac{V_{\max}X_0}{K_m + X_0} \left(1 + \frac{X_2}{K_I} \right)^{-1}, \quad (8)$$

with $K_m = 4$ mM [65]. $V_{\max}$ and K_I were retrieved from Curto's model as about 16 and 394, respectively. With the exception of these two adaptations, the modified model is the same as in Eq. (6) and the pathway diagram is the same as in Figure 9, where the input arrow now originates at X_0 .

Commensurate with the K_m value for v_1 , we used the modified Curto model to perform experiments with three different bolus amounts, where the initial values corresponded to a concentration of half the K_m , twice the K_m , and close to saturation, respectively. The results are shown in Figure 14. Panel (a) corresponds to a bolus of exterior glucose resulting in an external initial concentration of 2 mM. In panel (b) the initial concentration is 8 mM, and in panel (c) it is 60 mM. In each panel, the left plot shows the dynamic metabolite profiles, while the right plot shows the dynamic flux profiles, which in a real experiment would have been inferred per DFE. In all cases, the input v_{in} initially causes the interior glucose concentration to rise, and this rise migrates throughout the pathway. The plots on the right side show how v_{in} and other fluxes start to decline once the exterior glucose becomes depleted. Correspondingly, each metabolite accumulates until the efflux from its pool becomes larger than the influx to the pool, due to the decline in exterior glucose. Eventually all metabolites become consumed and approach zero concentration.

The three experiments allow us to plot flux-versus-substrate trajectories (Figure 15). Panels (a), (b), (c) and (e) correspond to V_{in} , V_{HK} , V_{PFK} and V_{GAPD} respectively. Panel (d) shows the trajectories for V_{POL} and V_{ATPase} , both of which are single-substrate fluxes with two-dimensional trajectories. Naturally, the trajectories converge since each flux value is a function of a substrate and the substrates approach steady state. In a bolus experiment however, the system only converges to a trivial steady-state. Panel (f) only shows the substrate trajectories since the flux-versus-substrate trajectories for V_{PK} are entities in a four-dimensional space.

Library construction. We took the simulation results of the three bolus experiments as artificial metabolite and flux data and used them to construct the library. The MATLAB function *scatteredInterpolant* (Figure 6) provided a fast and effective interpolation for the data. Two examples of resulting surfaces are shown for v_{in} and v_{HK} in Figure 16 (a) and (b), respectively. Panels (c) and (d) show the flux-versus-substrate surfaces with moderate extrapolations beyond the boundaries of the dataset.

Steady-state simulation. We simulated the system for a constant concentration of exterior glucose within the ranges recorded in the library. This experiment is fundamentally different from the bolus experiments that were used to stock the library. The results are shown in Figure 17 for one of the fixed glucose inputs. Three different sets of initial conditions for the dependent variables were simulated, and the system converged to the same steady-state for all three. This

steady state is essentially identical to that of Curto's model if the same glucose input is simulated.

Figure 14. Simulation results for bolus experiments with different external glucose concentrations. The time courses capture the transient behavior of the system very clearly. The left column exhibits concentration profiles, while the right column shows flux profiles. Amount of glucose bolus: Panel a: 2 mM; panel b: 8 mM; panel c: 60 mM.

Figure 15. Flux-versus-substrate trajectories retrieved from the bolus experiment results. The three experiments lead to different trajectories, which are shown in different colors.

Figure 16. Interpolation and extrapolation of flux-versus-substrate surfaces. Panels (a) and (b) exhibit how the interpolations provide approximations for desired values not included in the library. The algorithm can also provide approximate values slightly outside the dataset boundaries. The quality of this extrapolation often declines with the distance from the boundary (panels (c) and (d)).

Figure 17. Steady-state simulation results based on data from bolus experiments. Although the data in the library were obtained from bolus experiments, the nonparametric simulation results for constant glucose inputs are very similar to their parametric analogues. Three different initial conditions were chosen, and the system was simulated for a constant exterior glucose concentration of 40 mM.

Computation of a steady-state. The computation of steady states is of great import, as most biological systems operate within the vicinity of a normal homeostatic state. For the case of parametric models, some structures, including linear systems, S-systems [66] and Lotka-Volterra systems (e.g., [67]), permit the algebraic calculation of steady-state solutions, while most other nonlinear systems require search algorithms.

Interestingly, the nonparametric model permits not only simulations toward a stable steady state, but even the option of direct computational assessments of steady states. In fact, the flux-versus-substrate library appears to be as effective as an explicit functional form. Here, we applied the MATLAB function *fsolve* to the task at hand. Similar to the traditional case involving explicit ODEs, *fsolve* starts from a given initial condition and at each iteration estimates the system derivatives using system equations and the library. These are used to determine the direction and size of the next step toward optimizing the objective function. It turned out that *fsolve* is quite sensitive to initial guesses. We therefore randomly sampled 10,000 initial conditions from the space of metabolite concentrations in the library, ran *fsolve*,

and recorded the solutions that achieved zero derivatives within a reasonable tolerance. Figure 18 compares the computed steady-states (blue circles) with the true steady-state from the parametric model (yellow circles). The left panel exhibits the metabolites, while the right panel shows the corresponding steady-state flux values.

Figure 18. Steady-state metabolite and flux values. The yellow circles show the true solutions for the constant input $V_{in} = 8$, according to Curto's model, while the blue dots exhibit the ensemble of computed nonparametric solutions, which resulted from different initial values in *fsolve*.

Discussion

The term “nonparametric” is currently used almost exclusively within the context of statistics, where nonparametric tests are sometimes preferred to their parametric analogues because they require fewer assumptions and are often simpler to execute, although they are not always as discerning as parametric tests [68]. It is impossible to date the first use of nonparametric statistical methods, as standard features like histograms and sample means do not require a *priori* choices of parameters and have been used for a very long time. The terminology itself was apparently introduced by Wolfowitz in 1942 to characterize those methods that did not require specialized assumptions regarding the functional forms of the distributions characterizing the populations from which samples were analyzed [69; 70]. As an alternative to “nonparametric,” some authors proposed the term “distribution-free,” while Ury suggested “assumption-free” or ISD (incompletely specified distribution) statistics, openly admitting, however, that neither term was “especially felicitous” [71].

Notwithstanding the terminology, the statistics community at the time was of course not unanimously welcoming and initially considered nonparametric methods as “short cuts for well-established parametric methods” and later as “rough and ready (quick and dirty), inefficient

methods, ... that were wasteful of information" [70; 72]. In today's view, the advantages and drawbacks appear to have found a healthy equilibrium, and nonparametric methods are considered true alternatives to more traditional parametric methods, because they make less stringent demands on the data and sometimes reveal quick, although possibly coarser answers with a lesser amount of calculation. It is even possible that nonparametric methods utilize more information in large datasets, especially if the data do not stem from processes that have well-parameterized representations. Then again, by their very nature, nonparametric methods do not involve parameters that permit succinct descriptions, and they sometimes throw away information to a point where, for instance, differences between two populations can no longer be quantified [73]. It is also openly acknowledged that the interpretation of nonparametric statistical methods is sometimes difficult [74]. Overall, nonparametric methods are recognized as sometimes useful and maybe necessary, and on occasion superior, but certainly not perfect.

Similar to nonparametric models in statistics, the nonparametric dynamic models proposed here do not sweepingly solve all problems in nonlinear compartment modeling. In fact, one could say that they have similar advantages and drawbacks. They use fewer assumptions regarding mathematical formats, which immediately implies that these models cannot be described or compared succinctly with a set of parameter values. Similarly, they may be more difficult to interpret. They do not throw away information, however. For instance, if an appropriate format of a flux with correct kinetic parameter values is known from some outside source, this information can easily be used to reduce the degrees of freedom or to replace a unique analytical solution with a regression solution, which one might expect to be more robust. At any rate, the method and its solutions depend on the quantity and quality of available data, thus yielding a situation that is not genuinely different from parametric models.

Similar to parametric models that are obtained with DFE, the models here require solvability of the linear system of fluxes in the ODE model. This solvability issue can be assessed using the topology of the pathway through characterizability analysis, even in the absence of specific data [49]. Non-characterizable pathways incur estimation issues no matter what methods are applied, because such pathways admit entire spaces of solutions, some of which differ drastically from others [45]. In this situation, the DFE approach offers the advantage of identifying whether any, and if so which, fluxes can be determined uniquely, or what additional information would remedy the situation. For instance, Iwata and colleagues [51] demonstrated that it is possible to estimate a few fluxes with traditional methods and thereby to make a DFE solution unique. Similarly, it was shown that reasonable biological constraints may be able to ameliorate or even solve the under-determinedness of a pathway system [45]. The

same arguments apply to our nonparametric approach, although the infusion of additional information may result in a hybrid model where some fluxes are parameterized and others are not.

Not long ago, the experimental characterization of metabolic time courses was considered a very difficult task, and metabolic time series data—and corresponding models—were extremely rare [75]. This situation has changed dramatically during the past decade, and various methods of molecular biology have rendered it feasible to generate time series data with reasonable effort (*e.g.*, [76; 77; 78; 79]). This type of data generation is crucial here, because the construction of nonparametric models depends on a data scaffold, from which a library of nonparametric flux representations can be established.

In contrast to metabolite concentrations, fluxes of a pathway system are presently difficult to measure, except possibly for influxes and effluxes that communicate with the exterior milieu. However, in analogy to the progress in time series measurements of concentrations, one might hope that flux measurements will follow in due time; *e.g.*, see [80]. Indeed, if it becomes more widely known that a limited set of flux measurements may lead to a complete pathway characterization that does not make mathematical assumptions beyond the stoichiometry of the system, the community of molecular biologists may devote concerted effort toward such measurements in the future.

In addition to the difficulties of obtaining the right data, one could argue that experimental data often cover only a relatively small subspace of what an explicit mathematical function would be able to represent. This argument is certainly true, but one must ask to what degree extrapolations with these functions far outside the data domain are justifiable and biologically relevant.

Once experimental concentration and flux measurements become more readily available, the nonparametric method proposed here may actually become preferable over the tried and true approaches of parametric pathway modeling, at least for initial, unbiased assessments and various types of simulation studies. For deeper analyses, such as formal bifurcation studies, it appears that parametric representations will be difficult to replace. In this case, DFE may be executed to the end, where parametric functions are selected based on the inferred flux profiles.

Acknowledgments

The authors are very grateful to Luis L. Fonseca and James Wade for providing stimulating discussions and valuable feedback. This work was supported in part by the following grants: NSF (MCB-0958172, MCB-0946595, and MCB-1517588; PI: EO; MCB 1411672; PI: Diana Downs; DEB-1241046; PI: Kostas Konstantinides); NIH (1P30ES019776-01A1, Gary W. Miller, PI); and DOE-BESC (DE-AC05-00OR22725; PI: Paul Gilna). BESC, the BioEnergy Science Center, is a U.S. Department of Energy Bioenergy Research Center supported by the Office of Biological and Environmental Research in the DOE Office of Science. This project was furthermore supported in part by Federal funds from the US National Institute of Allergy and Infectious Diseases, National Institutes of Health, Department of Health and Human Services under contract HHSN272201200031C, which supports the Malaria Host–Pathogen Interaction Center (MaHPIC). The funding agencies are not responsible for the content of this article.

References

- [1] E.O. Voit, Modelling metabolic networks using power-laws and S-systems. *Essays Biochem* 45 (2008) 29-40.
- [2] J.M. Beaulieu, and R.R. Gainetdinov, The physiology, signaling, and pharmacology of dopamine receptors. *Pharmacol Rev* 63 (2011) 182-217.
- [3] Z. Qi, G.W. Miller, and E.O. Voit, Computational systems analysis of dopamine metabolism. *PLoS One* 3 (2008) e2444.
- [4] Z. Qi, G.W. Miller, and E.O. Voit, Internal state of medium spiny neurons varies in response to different input signals. *BMC Systems Biology* 4 (2010) 26.
- [5] D.J. Surmeier, S.M. Graves, and W. Shen, Dopaminergic modulation of striatal networks in health and Parkinson's disease. *Curr Opin Neurobiol* 29 (2014) 109-17.
- [6] L. Michaelis, and M.L. Menten, Die Kinetik der Invertinwirkung. *Biochemische Zeitschrift* 49 (1913) 333-369.
- [7] F. Shiraishi, and M.A. Savageau, The tricarboxylic acid cycle in *Dictyostelium discoideum*. I. Formulation of alternative kinetic representations. *J Biol Chem* 267 (1992) 22912-8.
- [8] E.O. Voit, H.A. Martens, and S.W. Omholt, 150 years of the mass action law. *PLoS Comput Biol* 11 (2015) e1004012.
- [9] M. Cascante, A. Sorribas, R. Franco, and E.I. Canela, Biochemical systems theory: increasing predictive power by using second-order derivatives measurements. *J Theor Biol* 149 (1991) 521-35.
- [10] M.A. Savageau, Biochemical systems analysis. I. Some mathematical properties of the rate law for the component enzymatic reactions. *J Theor Biol* 25 (1969) 365-9.
- [11] M.A. Savageau, *Biochemical Systems Analysis: A Study of Function and Design in Molecular Biology*, Addison-Wesley Pub. Co. Advanced Book Program (reprinted 2009), Reading, Mass, 1976.
- [12] E.O. Voit, (Ed.), *Canonical Nonlinear Modeling. S-System Approach to Understanding Complexity*, Van Nostrand Reinhold, NY, 1991.
- [13] E.O. Voit, *Computational Analysis of Biochemical Systems: A Practical Guide for Biochemists and Molecular Biologists*, Cambridge University Press, Cambridge, UK, 2000.
- [14] D.A. Fell, *Understanding the Control of Metabolism*, Portland Press, London, UK, 1997.
- [15] R. Heinrich, and T.A. Rapoport, A linear steady-state treatment of enzymatic chains. Critique of the crossover theorem and a general procedure to identify interaction sites with an effector. *Eur. J. Biochem.* 42 (1974) 97-105.
- [16] H. Kacser, and J.A. Burns, The control of flux. *Symp. Soc. Exp. Biol.* 27 (1973) 65-104.

- [17] D. Visser, and J.J. Heijnen, The mathematics of metabolic control analysis revisited. *Metab Eng* 4 (2002) 114-123.
- [18] M.A. Savageau, E.O. Voit, and D.H. Irvine, Biochemical systems theory and metabolic control theory. I. Fundamental similarities and differences. *Mathem Biosci* 86 (1987) 127-145.
- [19] M.A. Savageau, E.O. Voit, and D.H. Irvine, Biochemical systems theory and metabolic control theory. II. The role of summation and connectivity relationships. *Mathem Biosci* 86 (1987) 147-169.
- [20] E.O. Voit, Biochemical Systems Theory: A review. *Int. Scholarly Res. Network (ISRN – Biomathematics)* Article 897658 (2013) 1-53.
- [21] J.J. Heijnen, Approximative kinetic formats used in metabolic network modeling. *Biotechnol. Bioeng.* 91 (2005) 534-545.
- [22] R.C. del Rosario, E. Mendoza, and E.O. Voit, Challenges in lin-log modelling of glycolysis in *Lactococcus lactis*. *IET Syst Biol* 2 (2008) 136-49.
- [23] F.-S. Wang, C.-L. Ko, and E.O. Voit, Kinetic modeling using S-systems and lin-log approaches. *Biochem. Eng. J.* 33 (2007) 238-247.
- [24] A. Sorribas, B. Hernandez-Bermejo, E. Vilaprinyo, and R. Alves, Cooperativity and saturation in biochemical networks: a saturable formalism using Taylor series approximations. *Biotechnol. Bioeng.* 97 (2007) 1259-1277.
- [25] A. Sorribas, E. Vilaprinyo, and R. Alves, Approximate kinetic formalisms for modeling metabolic networks: does anything work? *Philipp. Inform. Techn. J.* 1 (2008).
- [26] I.-C. Chou, and E.O. Voit, Recent developments in parameter estimation and structure identification of biochemical and genomic systems. *Math Biosci* 219 (2009) 57-83.
- [27] P. Gennemark, and D. Wedelin, Efficient algorithms for ordinary differential equation model identification of biological systems. *IET Syst Biol* 1 (2007) 120-9.
- [28] P. Gennemark, and D. Wedelin, Benchmarks for identification of ordinary differential equations from time series data. *Bioinformatics* 25 (2009) 780–786.
- [29] E.O. Voit, What if the fit is unfit? Criteria for biological systems estimation beyond residual errors. in: M. Dehmer, F. Emmert-Streib, and A. Salvador, (Eds.), *Applied Statistics for Biological Networks*, J. Wiley and Sons, New York, 2011, pp. 183-200.
- [30] E.O. Voit, *A First Course in Systems Biology*, Garland Science, New York, NY, 2012.
- [31] G. Goel, I.C. Chou, and E.O. Voit, System estimation from metabolic time-series data. *Bioinformatics* 24 (2008) 2505-11.
- [32] G.R. Gavalas, *Nonlinear Differential Equations of Chemically Reacting Systems*, Springer-Verlag Berlin, 1968.
- [33] R. Heinrich, and S. Schuster, *The Regulation of Cellular Systems*, Chapman and Hall, New York, 1996.
- [34] B.Ø. Palsson, *Systems Biology: Properties of Reconstructed Networks*, Cambridge University Press, New York, 2006.
- [35] S. Dolatshahi, B. Vidakovic, and E.O. Voit, A constrained wavelet smoother for pathway identification tasks in systems biology. *Computers & Chemical Engineering* 71 (2014) 728-733.
- [36] P.H.C. Eilers, A perfect smoother. *Analytical Chemistry* 75 (2003) 3631-3636.
- [37] M. Vilela, C.C. Borges, S. Vinga, A.T. Vasconcelos, H. Santos, E.O. Voit, and J.S. Almeida, Automated smoother for the numerical decoupling of dynamics models. *BMC Bioinformatics* 8 (2007) 305.
- [38] E. Whittaker, *On a new method of graduation*, Edinburgh Mathematical Society, 1923, pp. 63-75.
- [39] J.M. Varah, A spline least squares method for numerical parameter estimation in differential equations. *SIAM J. Sci. Stat. Comput.* 3 (1982) 28-46.

- [40] E.O. Voit, and J. Almeida, Decoupling dynamical systems for pathway identification from metabolic profiles. *Bioinformatics* 20 (2004) 1670-1681.
- [41] E.O. Voit, and M.A. Savageau, Power-law approach to modeling biological systems; III. Methods of analysis. *J Ferment Technol* 60 (1982) 223-241.
- [42] E.O. Voit, and M.A. Savageau, Power-law approach to modeling biological systems; II. Application to ethanol production. *J Ferment Technol* 60 (1982) 229-232.
- [43] S. Dolatshahi, L.L. Fonseca, and E.O. Voit, New insights into the complex regulation of the glycolytic pathway in *Lactococcus lactis*. II. Inference of the precisely timed control system regulating glycolysis. *Mol Biosyst* 12 (2016) 37-47.
- [44] S. Dolatshahi, L.L. Fonseca, and E.O. Voit, New insights into the complex regulation of the glycolytic pathway in *Lactococcus lactis*. I. Construction and diagnosis of a comprehensive dynamic model. *Mol Biosyst* 12 (2016) 23-36.
- [45] S. Dolatshahi, and E. Voit, Identification of dynamic fluxes from metabolic time series data. *Frontiers in Genetics* 7 (2016) 6.
- [46] A. Albert, *Regression and the Moore-Penrose Pseudoinverse*, Academic Press, New York, London, 1972.
- [47] E.H. Moore, On the reciprocal of the general algebraic matrix. *Bull. Amer. Mathem. Soc.* 26 (1920) 394–395.
- [48] R. Penrose, A generalized inverse for matrices. *Proc. Cambridge Phil. Soc.* 51 (1955) 406-413.
- [49] E.O. Voit, Characterizability of metabolic pathway systems from time series data. *Math. Biosci.* 246 (2013) 315-325.
- [50] E.O. Voit, G. Goel, I.C. Chou, and L.L. Fonseca, Estimation of metabolic pathway systems from different data sources. *IET Syst Biol* 3 (2009) 513-22.
- [51] M. Iwata, F. Shiraishi, and E.O. Voit, Coarse but efficient identification of metabolic pathway systems. *Int. J. Syst. Biol.* 4 (2013) 57-72.
- [52] I.-C. Chou, and E.O. Voit, Estimation of dynamic flux profiles from metabolic time series data. *BMC Syst. Biol.* 6 (2012).
- [53] J.L. Galazzo, and J.E. Bailey, Fermentation Pathway Kinetics and Metabolic Flux Control in Suspended and Immobilized *Saccharomyces-Cerevisiae*. *Enzyme and Microbial Technology* 12 (1990) 162-172.
- [54] J.E. Bailey, S. Birnbaum, J.L. Galazzo, C. Khosla, and J.V. Shanks, Strategies and challenges in metabolic engineering. *Ann N Y Acad Sci* 589 (1990) 1-15.
- [55] R. Curto, A. Sorribas, and M. Cascante, Comparative characterization of the fermentation pathway of *Saccharomyces cerevisiae* using biochemical systems theory and metabolic control analysis: model definition and nomenclature. *Math Biosci* 130 (1995) 25-50.
- [56] M. Cascante, R. Curto, and A. Sorribas, Comparative characterization of the fermentation pathway of *Saccharomyces cerevisiae* using biochemical systems theory and metabolic control analysis: steady-state analysis. *Math Biosci* 130 (1995) 51-69.
- [57] A. Sorribas, R. Curto, and M. Cascante, Comparative characterization of the fermentation pathway of *Saccharomyces cerevisiae* using biochemical systems theory and metabolic control analysis: model validation and dynamic behavior. *Math Biosci* 130 (1995) 71-84.
- [58] A. Sorribas, C. Pozo, E. Vilaprinyo, G. Guillen-Gosalbez, L. Jimenez, and R. Alves, Optimization and evolution in metabolic pathways: global optimization techniques in Generalized Mass Action models. *J Biotechnol* 149 (2010) 141-53.
- [59] P.K. Polisetty, E.P. Gatzke, and E.O. Voit, Yield optimization of regulated metabolic systems using deterministic branch-and-reduce methods. *Biotechnol Bioeng* 99 (2008) 1154-69.
- [60] P.K. Polisetty, E.O. Voit, and E.P. Gatzke, Identification of metabolic system parameters using global optimization methods. *BMC Theoretical Biology and Medical Modelling* 3 (2006) 4.

- [61] N.V. Torres, E.O. Voit, C. Glez-Alcón, and F. Rodríguez, An indirect optimization method for biochemical systems: description of method and application to the maximization of the rate of ethanol, glycerol, and carbohydrate production in *Saccharomyces cerevisiae*. *Biotechnol Bioeng* 55 (1997) 758-72.
- [62] J. Vera, P. de Atauri, M. Cascante, and N.V. Torres, Multicriteria optimization of biochemical systems by linear programming: application to production of ethanol by *Saccharomyces cerevisiae*. *Biotechnol Bioeng* 83 (2003) 335-43.
- [63] E.O. Voit, and T. Radivoyevitch, Biochemical systems analysis of genome-wide expression data. *Bioinformatics* 16 (2000) 1023-1037.
- [64] L.L. Fonseca, C. Sánchez, H. Santos, and E.O. Voit, Complex coordination of multi-scale cellular responses to environmental stress. *Mol. BioSyst.* 7 (2011) 731 – 741.
- [65] T. Gonçalves, and M.C. Loureiro-Dias, Aspects of glucose uptake in *Saccharomyces cerevisiae*. *J. Bacteriology* 176 (1994) 1511-1513.
- [66] M.A. Savageau, Biochemical systems analysis. II. The steady-state solutions for an n-pool system using a power-law approximation. *J Theor Biol* 25 (1969) 370-9.
- [67] L. Edelstein-Keshet, *Mathematical Models in Biology*, Random House, Inc., New York, NY, 1988.
- [68] P.H. Kvam, and B. Vidakovic, *Nonparametric Statistics with Applications to Science and Engineering*, John Wiley and Sons, Hoboken, NJ, 2007.
- [69] J. Wolfowitz, Additive partition functions and a class of statistical hypotheses. *Ann. Math. Statist.* 13 (1942) 247-279.
- [70] E.J. Dudewicz, Nonparametric methods: The history, the reality, and the future (with special reference to statistical selection problems). in: W. Sendler, (Ed.), *Contributions to Stochastics*, Physica Verlag, Heidelberg, 1987, pp. 63-83.
- [71] H. Ury, Letter to the editor. *The American Statistician* 21 (1967) 53.
- [72] G.E. Noether, Needed-a new name. *The American Statistician* 21 (1967) 41.
- [73] G.E. Dallal, Nonparametric Statistics. <http://www.jerrydallal.com/lhsp/npa.htm> (2000).
- [74] T. Hoskin, Parametric and Nonparametric: Demystifying the Terms www.mayo.edu/mayo-edu-docs/center-for-translational-science-activities-documents/berd-5-6.pdf (2016).
- [75] E.O. Voit, Models-of-data and models-of-processes in the post-genomic era. *Math Biosci* 180 (2002) 263-274.
- [76] R.V. Bruggner, B. Bodenmiller, D.L. Dill, R.J. Tibshirani, and G.P. Nolan, Automated identification of stratifying signatures in cellular subpopulations. *Proc Natl Acad Sci U S A* 111 (2014) E2770-7.
- [77] A. El-Aneed, A. Cohen, and J. Banoub, Mass Spectrometry, Review of the Basics: Electrospray, MALDI, and Commonly Used Mass Analyzers. *Applied Spectroscopy Reviews* 44 (2009) 210-230.
- [78] S. Li, Y. Park, S. Duraisingham, F.H. Strobel, N. Khan, Q.A. Soltow, D.P. Jones, and B. Pulendran, Predicting network activity from high throughput metabolomics. *PLoS Comput Biol* 9 (2013) e1003123.
- [79] A.R. Neves, W.A. Pool, J. Kok, O.P. Kuipers, and H. Santos, Overview on sugar metabolism and its control in *Lactococcus lactis* - the input from in vivo NMR. *FEMS Microbiol Rev* 29 (2005) 531-54.
- [80] A.D. Sherry, and C.R. Malloy, *Integration of ¹³C Isotopomer Methods and Hyperpolarization Provides a Comprehensive Picture of Metabolism*, eMagRes, John Wiley & Sons, Ltd, 2007.