Ensemble Kinetic Modeling of Metabolic Networks from Dynamic Metabolic Profiles

Metabolites **2012**, *2*, 891-912; doi:10.3390/metabo2040891

OPEN ACCESS

metabolites

ISSN 2218-1989

www.mdpi.com/journal/metabolites/

Article

Ensemble Kinetic Modeling of Metabolic Networks from Dynamic Metabolic Profiles

Gengjie Jia [1], Gregory Stephanopoulos [2] and Rudiyanto Gunawan [3,*]

[1] Chemical and Pharmaceutical Engineering, Singapore-MIT Alliance, Singapore 117576, Singapore; E-Mail: jiagengjie@nus.edu.sg (G.J.)

[2] Department of Chemical Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139, USA; E-Mail: gregstep@mit.edu (G.S.)

[3] Institute for Chemical and Bioengineering, ETH Zurich, 8093 Zurich, Switzerland

* Author to whom correspondence should be addressed; E-Mail: rudi.gunawan@chem.ethz.ch; Tel.: +41 44 633 21 34; Fax: +41 44 633 12 52.

Received: 14 September 2012; in revised form: 2 November 2012 / Accepted: 5 November 2012 / Published: 12 November 2012

Abstract: Kinetic modeling of metabolic pathways has important applications in metabolic engineering, but significant challenges still remain. The difficulties faced vary from finding best-fit parameters in a highly multidimensional search space to incomplete parameter identifiability. To meet some of these challenges, an ensemble modeling method is developed for characterizing a subset of kinetic parameters that give statistically equivalent goodness-of-fit to time series concentration data. The method is based on the incremental identification approach, where the parameter estimation is done in a step-wise manner. Numerical efficacy is achieved by reducing the dimensionality of parameter space and using efficient random parameter exploration algorithms. The shift toward using model ensembles, instead of the traditional "best-fit" models, is necessary to directly account for model uncertainty during the application of such models. The performance of the ensemble modeling approach has been demonstrated in the modeling of a generic branched pathway and the trehalose pathway in *Saccharomyces cerevisiae* using generalized mass action (GMA) kinetics.

Keywords: ensemble modeling; incremental identification; dynamic flux estimation; independent parameter set; generalized mass action model

---

1. Introduction

Mathematical modeling is one of the cornerstones of metabolic engineering [1]. These models vary in their formulation and complexity depending on the specific applications. For example, flux balance analysis relies on algebraic models of metabolic networks to predict the impact of pathway perturbations (e.g. gene knock-out/knock-in) on the steady-state metabolic flux distribution [2,3]. Meanwhile, kinetic ordinary differential equation (ODE) models have been traditionally used for dynamic optimization of culture conditions in a bioreactor [4]. Regardless of the type of the models, the process of model building is typically iterative, combining wet-lab experiments and *in silico* analysis and optimization [5]. Despite much progress in both experimental and computational fronts, e.g. increasing availability of high quality and system-level data and development of efficient parameter estimation methods, the process of creating mathematical models from biological data is still very challenging [6]. Much of the difficulty of this process, especially for kinetic ODE models, is rooted in the fundamental issue of model identifiability [7], wherein it is not possible to uniquely determine model equations and parameter values from experimental data. As we and many others have shown [8–11], the estimation of unknown parameters by fitting model simulations to biological measurements is typically ill-posed. Consequently, even when the best-fit parameters are obtained, the corresponding model may have little predictive capability; or worse, it could be misleading.

The majority of existing parameter estimation methods for the kinetic modeling of metabolic networks involve a single-step estimation, in which unknown parameters are estimated simultaneously by minimizing model prediction error [6,12,13]. There are a few reasons why such a strategy is often inefficient. Kinetic models of metabolic pathways (or cellular networks in general) typically possess a large number of unknown kinetic parameters, where in some cases, the number of parameters increases combinatorially with the number of metabolites. The large number of unknown parameters means not only that the parameter estimation will involve a vast parameter search space, but also that the parameters may not even be completely identifiable from data. The first effect leads to a large-scale, often numerically intractable, global optimization problem. The latter and arguably the more important consequence implies that the estimation problem has no unique solution (*i.e.* it is ill-posed) and many parameter combinations can fit the data equally well. Multiplicity of solutions to the parameter estimation of kinetic ODE models has been documented in different biological systems [11,14].

The aforementioned issues give the motivation for developing and applying a different framework to construct metabolic and biological models from data, one that can explicitly account for model uncertainty. In this work, an ensemble modeling strategy is employed. Ensemble modeling has previously been applied to address structural uncertainty in the modeling of metabolic and other biological networks. For example, ensemble models of metabolic pathways could be created by enforcing thermodynamic feasibility constraints on the metabolic reactions and used for metabolic control analysis [15–18]. In a modeling study of TOR (target of rapamycin) signaling pathway in yeast, an ensemble of 19 kinetic ODE models was generated, where each model in the ensemble represented a different hypothetical topology of the pathway [19]. The process of creating an ensemble of models from the set of possible components and reactions in a biological network has also recently been automated [20]. In these studies, a comparative analysis of models in the ensemble was conducted to determine the most likely mechanistic explanation for some experimental observations. For nonlinear

discrete time dynamic system, an ensemble modeling approach has also been proposed using the set membership framework, without requiring any prior assumption on the functional form of the model equations [21].

Here, we describe a step-wise model identification approach for the creation of an ensemble of kinetic ODE models from metabolic time profiles. Unlike the ensemble modeling work mentioned above, this approach is applied to tackle the uncertainty in the estimation of kinetic parameters. That is, models in the ensemble will share the same network topology, but differ in their parameter values. In essence, these models represent regions in the parameter space from which model prediction errors are (statistically) equivalent. Such an ensemble can be generated by exploring the parameter space using existing methods such as Metropolis-type random walk Markov chain [22] and the Pareto Optimal Ensemble Techniques (POETs), the last of which is based on multi-objective optimization [14]. However, the search was done over the full parameter set in these techniques, and thus the computational requirement may increase quickly with the number of kinetic parameters. In this work, a new and numerically efficient ensemble modeling procedure is developed based on the incremental identification or dynamic flux estimation (DFE) [23,24] and employing an adaptive efficient Metropolis Monte Carlo sampling [25]. The performance of the ensemble modeling procedure has been demonstrated using models of a generic branched metabolic pathway [26] and the trehalose pathway in *Saccharomyces cerevisiae* [27,28].

2. Ensemble Kinetic Modeling

Ordinary differential equations have been commonly used to model metabolic pathways. The model equations describe the mole balance around metabolites as they are enzymatically transformed from one to another. In this case, the system is assumed to be well-mixed (*i.e.* ignoring spatial distribution of metabolites) [29], leading to the following general form:

$$\dot{\mathbf{X}}\left(t,\mathbf{p}\right)=\mathbf{S}\mathbf{v}\left(\mathbf{X},\mathbf{p}\right), \tag{1}$$

where t denotes the time, $\mathbf{p}$ is the parameter vector, $\mathbf{X}(t,\mathbf{p})$ is the vector of m metabolite concentrations, $\mathbf{v}(\mathbf{X},\mathbf{p})$ denotes the vector of n enzymatic reactions/fluxes, and $\mathbf{S}$ is the $m\times n$ stoichiometric matrix. The metabolic fluxes are further specified as the function of $\mathbf{X}$, for example using a power-law dependence:

$$v_j\left(\mathbf{X},\mathbf{p}\right)=\gamma_j\prod_i X_i^{f_{ji}};\qquad \mathbf{p}=\left\{\gamma_j,f_{ji}\right\};\qquad i=1,\cdots,m;\qquad j=1,\cdots,n; \tag{2}$$

where the parameter γ_j is the rate constant of the j-th flux and f_{ji} is the kinetic order, reflecting the influence of metabolite X_i on the j-th flux (positive: substrate or activation, negative: inhibition). Aside from the power-law function, Michaelis-Menten and Hill equations have also been commonly used to describe the kinetics of enzymatic reactions $\mathbf{v}(\mathbf{X},\mathbf{p})$.

The aforementioned power-law model, also known as generalized mass action (GMA) model, belongs to a widely adopted framework for the modeling and analysis of biochemical processes, the Biochemical Systems Theory (BST) [29–31]. Power-law models have a relatively simple structure that permits algebraic manipulation in logarithmic scale. Furthermore, the nature of network connectivity is directly related with the parameter values of rate constants and kinetic orders, facilitating simultaneous parameter estimation and network structure identification. However, the estimation of parameters of

these models is known to be very challenging, even after the development of over 100 methods [32]. Perhaps this difficulty is not surprising, as the number of parameters in power-law models is often large and this number increases quickly with network complexity (*i.e.* the number of metabolites and interactions). Consequently, the parameters are typically not completely identifiable [10], motivating the application of the ensemble modeling developed in this work.

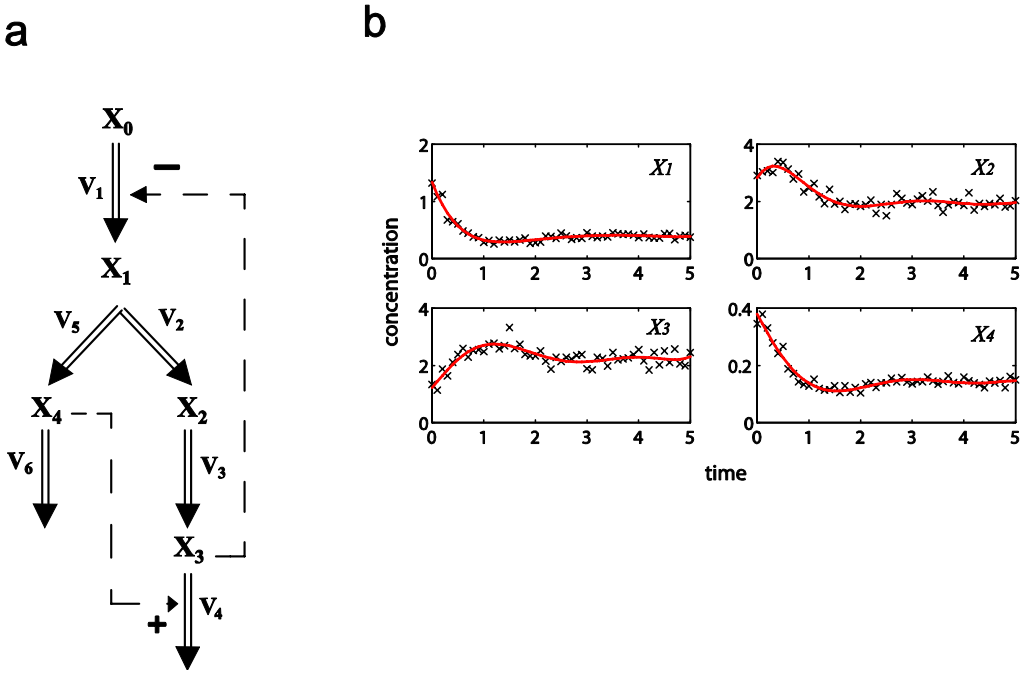
Briefly, the proposed ensemble modeling derives from the incremental identification or dynamic flux estimation method [24,33]. In these methods, the estimation of unknown kinetic parameters from concentration time profiles $\mathbf{X}_M(t)$ is decomposed into a few steps, involving (1) the computation of slopes of time-series data $\dot{\mathbf{X}}_M(t)$, (2) the calculation of dynamic flux profiles $\mathbf{v}(t)$ from $\dot{\mathbf{X}}_M(t)$, and finally (3) the regression of parameters, which can be done one flux at a time. In the original formulation of the incremental identification and DFE, the number of measured species is assumed to be larger than the number of reactions, such that the second step possesses a unique solution. However, since metabolic pathways typically involve more fluxes than metabolites, there now exist (infinitely) many dynamic flux values, each of which is a mathematically valid solution. This is the premise of the new ensemble modeling method. Specifically, models in the ensemble represent a subset of the dynamic flux solutions to $\dot{\mathbf{X}}_M = \mathbf{S}\mathbf{v}$, with additional criteria that the kinetic parameters produce statistically equivalent and biologically relevant model predictions. The construction of the model ensemble is detailed in the Method section.

2.1. A Generic Branched Pathway

The metabolic pathway map in this case study is given in Figure 1a, which describes the transformations among four metabolites with both feedback activation and inhibition. The model of the pathway is written as a GMA model with 13 kinetic parameters, as shown in Equation (3). This model with the reported parameter values (see Table S1 in Supplementary Material) and initial concentrations [26] was used to generate time-course concentration data, contaminated with i.i.d. Gaussian noise with zero mean and 10% coefficient of variation (the ratio of standard deviation to the mean). For validation purpose, two independent datasets were generated in the same manner as above, but with different initial conditions $\left[X_1(t_0)\ \ X_2(t_0)\ \ X_3(t_0)\ \ X_4(t_0)\right] = \left[4\ \ 1\ \ 3\ \ 4\right]$ and $\left[0.2\ \ 0.3\ \ 4.2\ \ 0.01\right]$, respectively. The *in silico* noisy data were smoothened using a *6*-th order polynomial, which gave the best polynomial fit to the data according to adjusted R^2 [34] and Akaike Information Criterion (AIC) [35] (see Figure 1b). Subsequently, a central finite difference approximation was applied to compute the time-slopes of the smoothened data.

$$\begin{aligned} &\dot{X}_1 = v_1 - v_2 - v_5 \\ &\dot{X}_2 = v_2 - v_3 \\ &\dot{X}_3 = v_3 - v_4 \\ &\dot{X}_4 = v_5 - v_6 \end{aligned} \qquad \begin{aligned} &v_1 = \gamma_1 X_0 X_3^{-f_{13}} \\ &v_2 = \gamma_2 X_1^{f_{21}} \\ &v_3 = \gamma_3 X_2^{f_{32}} \\ &v_4 = \gamma_4 X_3^{f_{43}} X_4^{f_{44}} \\ &v_5 = \gamma_5 X_1^{f_{51}} \\ &v_6 = \gamma_6 X_4^{f_{64}} \\ &X_0 = 0.6 \end{aligned} \qquad \begin{aligned} &X_1(t_0) = 1.4 \\ &X_2(t_0) = 2.7 \\ &X_3(t_0) = 1.2 \\ &X_4(t_0) = 0.4 \end{aligned} \tag{3}$$

Figure 1. A generic branched pathway. (**a**) Metabolic pathway map. Metabolic fluxes: double-line arrows, regulatory interactions: dashed arrows with signs; (**b**) The smoothened data (red line) *versus* the noisy data (×).



In this example, the degree of freedoms is 2 (4 metabolites and 6 fluxes). Fluxes v_1 and v_6 were chosen as the independent fluxes, since this selection led to an invertible $\mathbf{S}_D$ and comprised the least number of independent parameters. The involved independent parameters $\mathbf{p}_I$ included the rate constants $\{\gamma_1, \gamma_6\}$ and the kinetic orders $\{f_{13}, f_{64}\}$, which were constrained to within [0, 100] and [0, 5], respectively. The bounds for dependent parameters were set to be the same, *i.e.* $\{\gamma_2,\gamma_3,\gamma_4,\gamma_5\}\in[0,100]$, $\{f_{21},f_{33},f_{43},f_{44},f_{51}\}\in[0,5]$. In addition, the upper bound for allowable metabolic fluxes in this artificial network was set as 5×10^5 mM/min.

Following the ensemble modeling procedure described in the Method section, the initial parameter point for the out-of-equilibrium adaptive Metropolis Monte Carlo (OEAMC) algorithm was taken from the parameter estimation minimizing the flux error function Φ_R (minimum $\sqrt{\Phi_R}$ = 0.130), and the upper 95% confidence bound of the error function value was determined using Monte Carlo approach (viable $\sqrt{\Phi_R}$ < 0.347). Table 1 summarizes the outcome of the ensemble modeling. The multiple ellipsoid-based sampling (MEBS) algorithm produces a model ensemble with 59,928 members within the viable parameter subspace. The corresponding volume of the viable subspace represented only 0.284% of the original parameter space (*i.e.* the space defined by the upper and lower parameter bounds). Figure 2 shows the projections of the viable regions onto the two-dimensional parameter axes of each independent flux. The true parameter values are contained in the viable subspace, and thus belong to the ensemble (red dot in Figure 2). The member models of the ensemble were able to predict the concentration and slope profiles reasonably well (see Table 1), even when the ensemble was constructed using a different error function. The comparison of data and model predictions in Figure 3 demonstrates the equivalence among five randomly selected models in the ensemble. Finally, Figure 4 shows the comparison of model simulations from the same five models and independent (simulated)

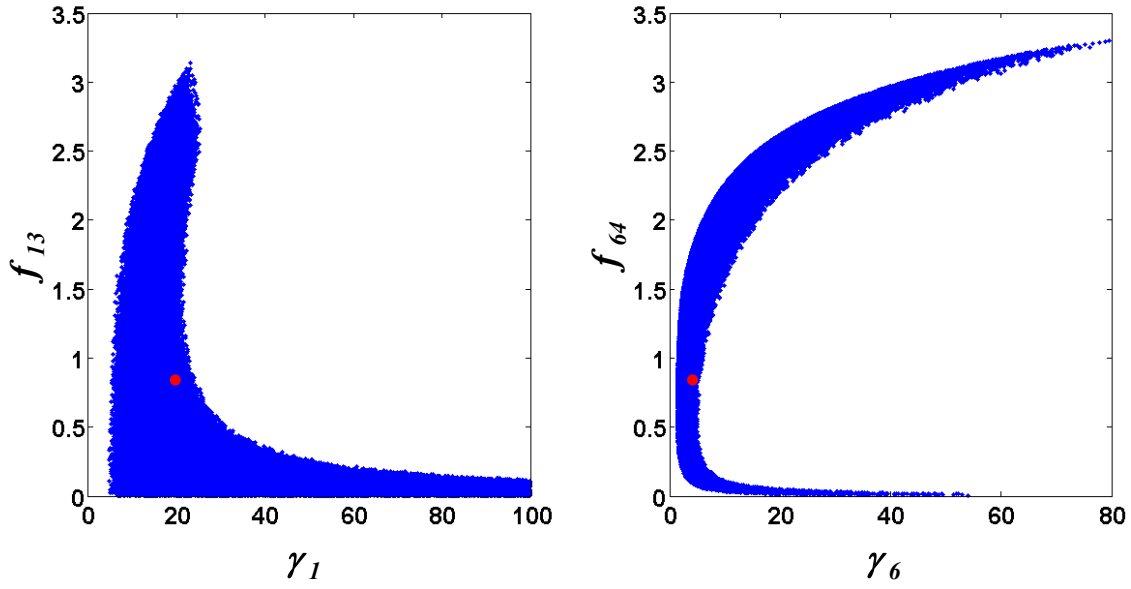
experimental datasets, indicating that these models could predict the systems dynamics under different initial conditions reasonably well.

Table 1. Ensemble kinetic modeling of the branched pathway model using Φ_R.

CPU time (sec) [a]	1664
Calculated volume of initial parameter space (V_{ci}) [b]	2.5×10^5
Estimated volume of viable parameter space (V_{ev}) [c]	710.1 ± 5.1
Ratio of V_{ev} to V_{ci}	$(284.0 \pm 2.0) \times 10^{-3}$%
Range of slope errors $\sqrt{\Phi_S}$ [d]	$[1.370 \times 10^{-1}, 5.081 \times 10^{-1}]$
Range of concentration errors $\sqrt{\Phi_C}$ [e]	$[3.554 \times 10^{-2}, 2.150 \times 10^{-1}]$

a. The CPU time was the total time for the ensemble construction, which was run on a computer workstation with Dual Processors Intel Quad-Core 2.83 GHz.
b. V_{ci} was calculated by simple multiplications of the independent parameter ranges.
c. V_{ev} was calculated by integrating the volumes of an ensemble of ellipsoids that cover the viable parameter space [25].
d. The range of slope error was computed using Equation (14) for all models in the ensemble.
e. The range of concentration error was computed by Equation (15) for all models in the ensemble.

Figure 2. Two-dimensional projections of the viable parameter space onto the parameter axes of each independent flux (v_1: left, v_6: right). The true parameters are marked in red.



Note that besides the Φ_R minimization, the proposed kinetic ensemble modeling approach can also use other error functions. The viable parameter space using the slope error Φ_S, for example, closely resembles that shown in Figure 2 (see Supplementary Material), demonstrating the robustness of the procedure in capturing the model uncertainty.

2.2. The Trehalose Pathway in Saccharomyces cerevisiae

The second case study was taken from the modeling of the glycolysis and trehalose production in the baker's yeast *Saccharomyces cerevisiae*. Figure 5a shows the metabolic pathway and Equation (4) presents the GMA model, describing in a simplified fashion how glucose is converted into trehalose and other products in a cyclic pathway [28]. The notations for the concentrations of metabolites are as follows: extracellular glucose (exGlc) – X_1, intracellular glucose (inGlc) – X_2, glucose 6-phosphate (G6P) – X_3, trehalose (Tre) - X_4, fructose 1, 6-biphosphate (FBP) – X_5, extracellular end-products (ethanol, glycerol and acetate) – X_6, pentose phosphate pathway (PPP) – X_7 and other pathways (Leakage) – X_8. The variables V_{ex} and V_{in} denote the extracellular (5.00×10^{-2} L) and intracellular (7.17×10^{-3} L) volumes of the bioreactor and the cell population, respectively. The time-course concentration data have been obtained using *in vivo* NMR, but only X_1, X_3, X_4, X_5 and X_6 were measured [27]. In the following, we used the dataset from normally grown cells at 30 °C that were fed with a pulse of glucose. The raw experimental data were smoothened using a piecewise cubic spline, the fitting of which was validated by adjusted R^2 [34] and AIC [35] (see Figure 5b). Like before, a central difference approximation was applied to obtain the time-slopes of concentration data.

$$\begin{aligned}
\dot{X}_1 &= -v_1/V_{ex} & v_1 &= f_1(X_1)\\
\dot{X}_2 &= (v_1 + 2v_4 - v_2)/V_{in} & v_2 &= f_2(X_2)\\
\dot{X}_3 &= (v_2 - 2v_3 - v_5 - v_7)/V_{in} & v_3 &= f_3(X_3)\\
\dot{X}_4 &= (v_3 - v_4)/V_{in} & v_4 &= f_4(X_4)\\
\dot{X}_5 &= (v_5 - v_6 - v_8)/V_{in} & v_5 &= f_5(X_3)\\
\dot{X}_6 &= 2v_6/V_{ex} & v_6 &= f_6(X_5)\\
 & & v_7 &= f_7(X_3)\\
 & & v_8 &= f_8(X_5)
\end{aligned} \tag{4}$$

Figure 3. Concentration simulations of five randomly selected models from the ensemble (solid blue, brown, green, red and purple lines) *versus* the noisy data (×).

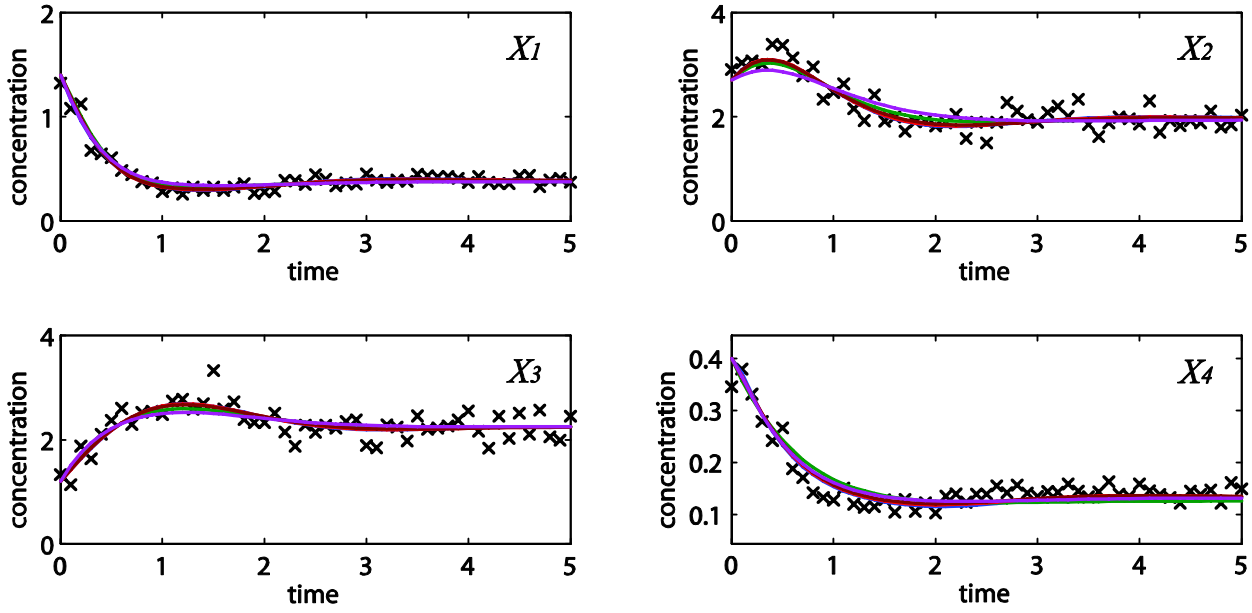
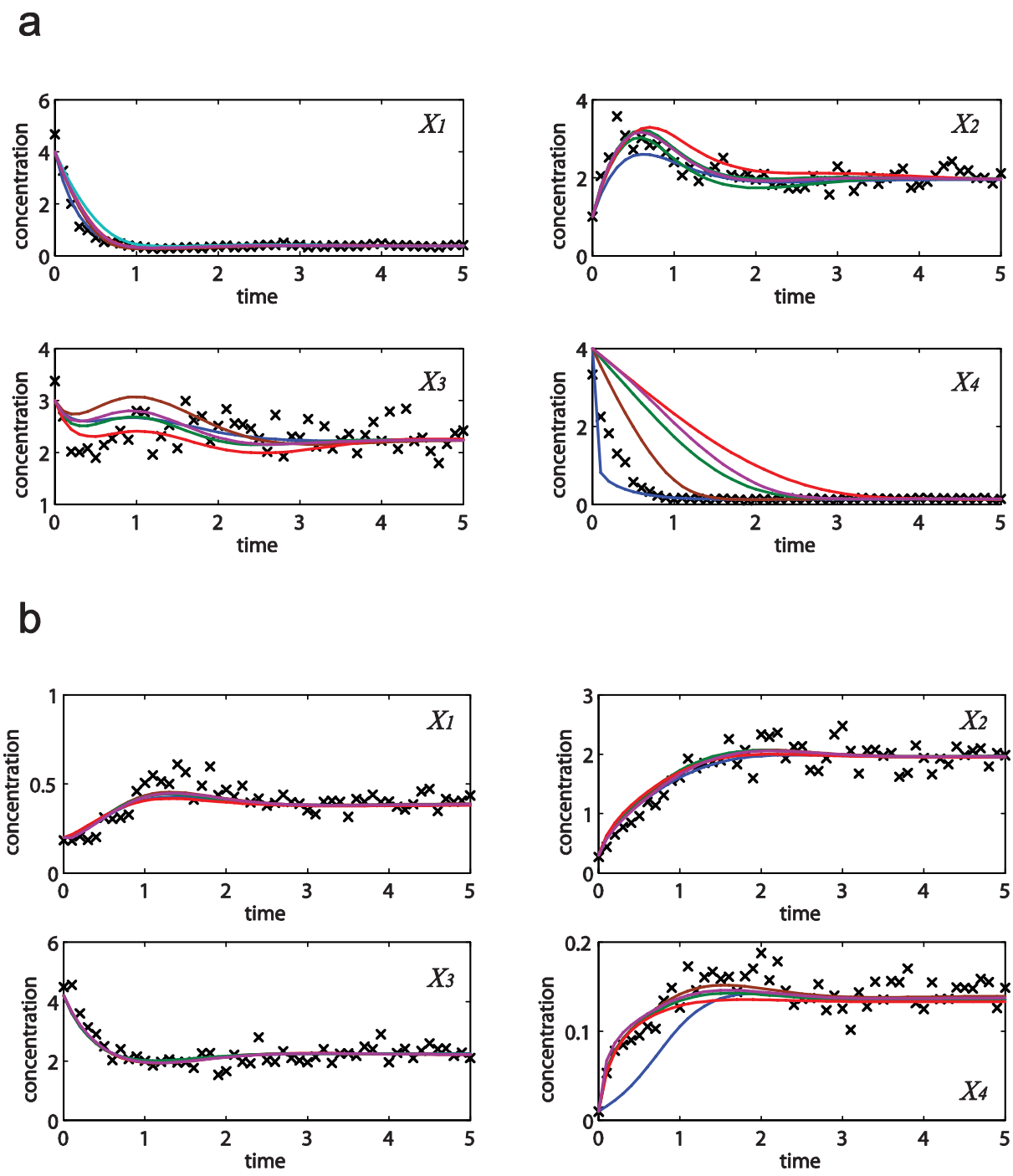


Figure 4. Concentration simulations of the same five models as in Figure 3 (solid blue, brown, green, red and purple lines) *versus* independent datasets (×), with initial concentrations of [4 1 3 4] (**a**) and [0.2 0.3 4.2 0.01] (**b**).



The original ODE model contains 6 metabolites and 8 fluxes, as shown in Equation (4). In this case study, the ODE for X_7 and X_8 are removed, as their concentrations do not affect the other metabolites (*i.e.* they are sinks in the system). While the intracellular glucose X_2 was not measured, its rate of change can be obtained from the measured metabolites by performing an overall mass balance around the pathway, resulting in the following relationship:

$$\dot{X}_2 = (-\dot{X}_1 \cdot V_{ex} - \dot{X}_3 \cdot V_{in} - 2\dot{X}_4 \cdot V_{in} - \dot{X}_5 \cdot V_{in} - \tfrac{1}{2}\dot{X}_6 \cdot V_{ex} - v_7 - v_8)/V_{in} \quad (5)$$

Using this relationship, the model can be reduced to the following equations:

$$\begin{cases} \dot{X}_1 = -v_1/V_{ex} \\ \dot{X}_3 = (v_1 + 2v_4 - 2v_3 - v_5 - v_7)/V_{in} - \dot{X}_2 \\ \dot{X}_4 = (v_3 - v_4)/V_{in} \\ \dot{X}_5 = (v_5 - v_6 - v_8)/V_{in} \\ \dot{X}_6 = 2v_6/V_{ex} \end{cases} \quad \begin{aligned} v_1 &= V_{\max 1} X_1 / (K_{m1} + X_1) \\ v_3 &= \gamma_3 X_3^{f_{33}} \\ v_4 &= \gamma_4 X_4^{f_{44}} \\ v_5 &= \gamma_5 X_3^{f_{53}} \\ v_6 &= \gamma_6 X_5^{f_{65}} \\ v_7 &= \gamma_7 X_3^{f_{73}} \\ v_8 &= \gamma_8 X_5^{f_{85}} \end{aligned} \tag{6}$$

According to Equation (4), we have 3 degrees of freedom (5 measured metabolites and 8 fluxes). Here, fluxes v_4, v_7 and v_8 were chosen as the independent fluxes, by the same rationale as before. Correspondingly, the independent parameters $\mathbf{p}_I$ comprised the rate constants $\{\gamma_4, \gamma_7, \gamma_8\}$ and the kinetic orders $\{f_{44}, f_{73}, f_{85}\}$, which were constrained within [0, 100] and [0, 5], respectively. Note that the glucose transport flux (v_1) was modeled using Michaelis-Menten (MM) kinetics instead of the power law, as this was found to be a better fit to the time profile of X_1 (a constant decrease at high X_1 and an exponential-like time profile at low X_1). The regression of the MM kinetic parameters can also be casted as a linear regression problem as follows:

$$\begin{bmatrix} V_{\max 1} \\ K_{m1} \end{bmatrix} = \left(\begin{bmatrix} X_1 & -v_1 \end{bmatrix}^T \begin{bmatrix} X_1 & -v_1 \end{bmatrix} \right)^{-1} \begin{bmatrix} X_1 & -v_1 \end{bmatrix}^T \begin{bmatrix} X_1 \cdot v_1 \end{bmatrix} \tag{7}$$

where $[X_1 \cdot v_1]$ is the vector of element-wise multiplication of X_1 and v_1. Finally, the upper bound for flux values was set as 5×10^5 mM/min, according to the maximal flux value reported in a similar glycolytic pathway [36].

The initial parameter point for the OEAMC algorithm was again obtained by minimizing Φ_R (minimum $\sqrt{\Phi_R} = 7.64 \times 10^{-2}$) and the upper 95% confidence bound was found using a Monte Carlo approach (viable $\sqrt{\Phi_R} < 0.186$). Table 2 gives the summary of the model ensemble for the trehalose model. The model ensemble was represented by 3423 member models, and the volume of the corresponding viable subspace constitutes 2.59×10^{-3}% of the original constrained parameter space. The slope errors were acceptable, but the concentration errors had a high upper bound. Upon a closer inspection, only a minority of the model (3 out of 3423) had concentration errors larger than 10^2, and removing these, the upper bound for the concentration error reduces to 35.92. This issue is not unexpected as the model ensemble was created based on the flux error function and not the concentration error. In particular, there is no guarantee that parameter values with a small flux error will also provide a low concentration error. However, we note that the divergence between the flux error and concentration error functions occurred only rarely (< 0.1%). Figure 6 shows the projections of the viable parameter subspace onto the two-dimensional parameter axes of each independent flux. Finally, Figure 7 shows a comparison between the concentration predictions of five randomly chosen models from the ensemble and the measured metabolite time profiles, again demonstrating that models in the ensemble can reproduce the data equally well.

Figure 5. The trehalose pathway in *Saccharomyces cerevisiae*. (**a**) Metabolic pathway map. Metabolic fluxes: double-line arrows; (**b**) The smoothened data (red line) *versus* the noisy data (×).

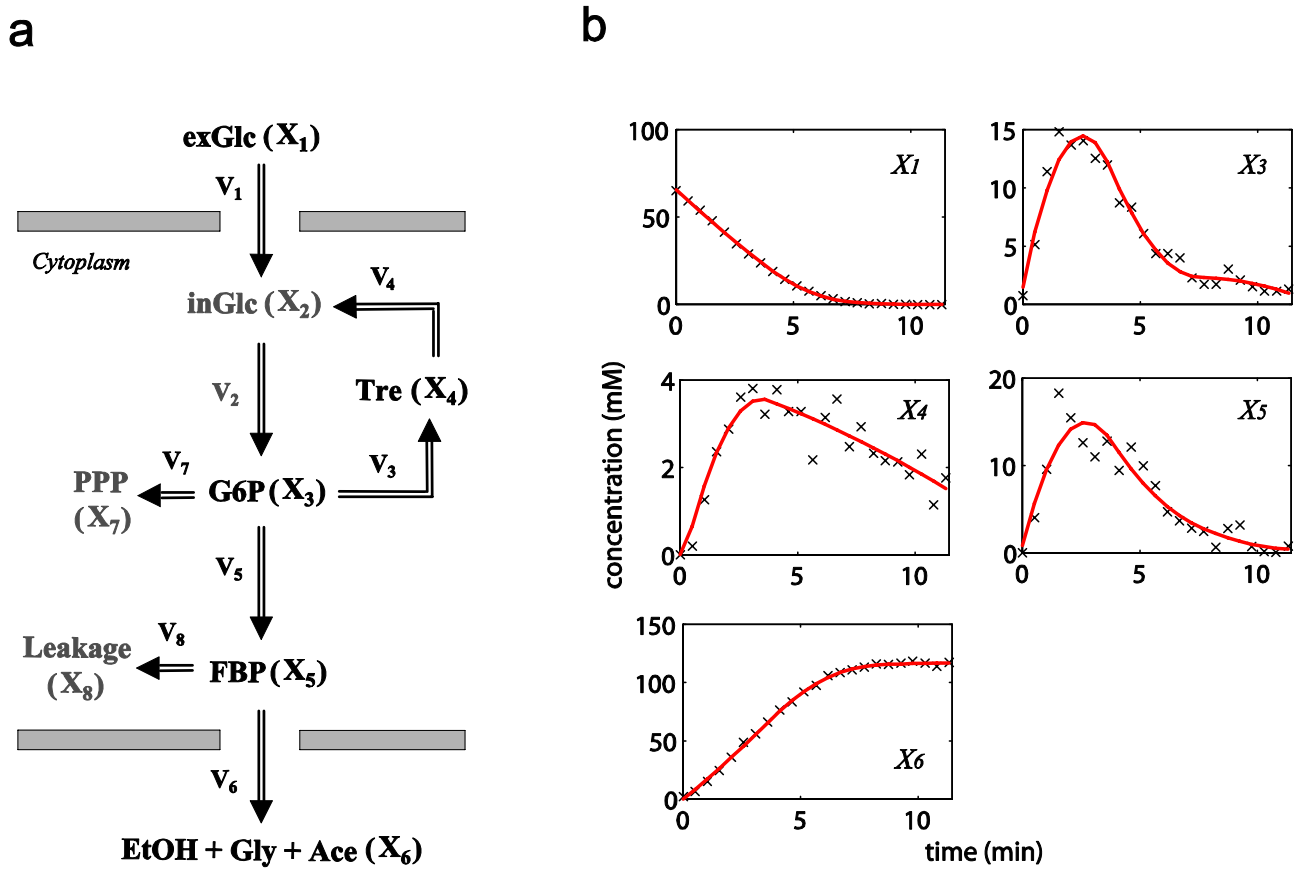


Figure 6. Two-dimensional projections of the viable parameter space onto the parameter axes of each independent flux (v_4: left, v_7: middle, v_8: right).

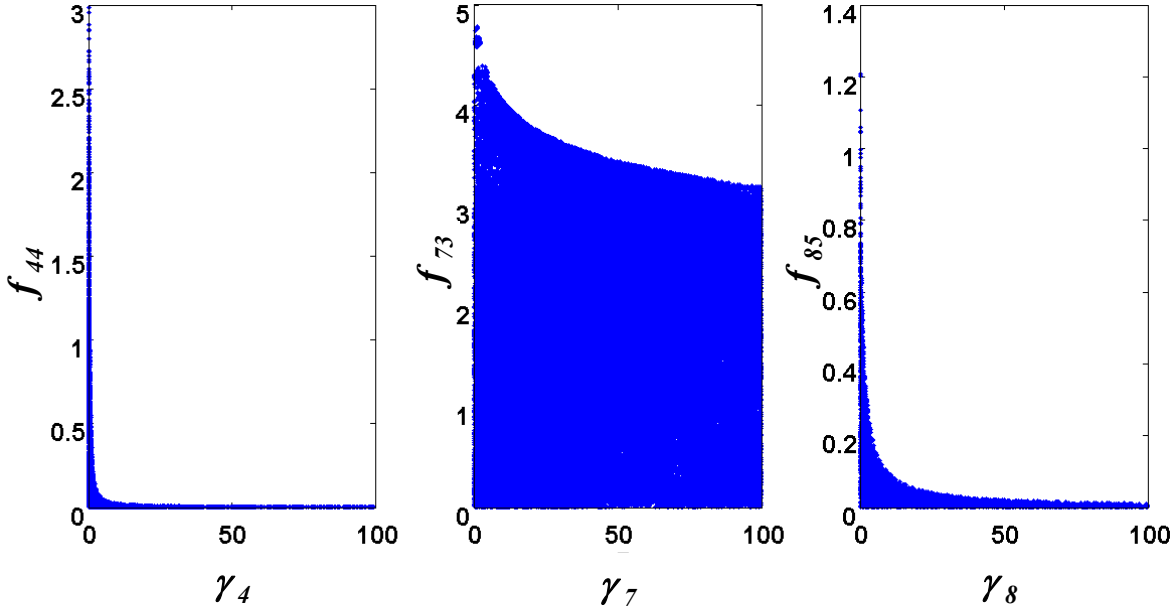


Figure 7. Concentration simulations of five randomly selected models from the ensemble (solid blue, brown, green, red and purple lines) *versus* the experimental data (×).

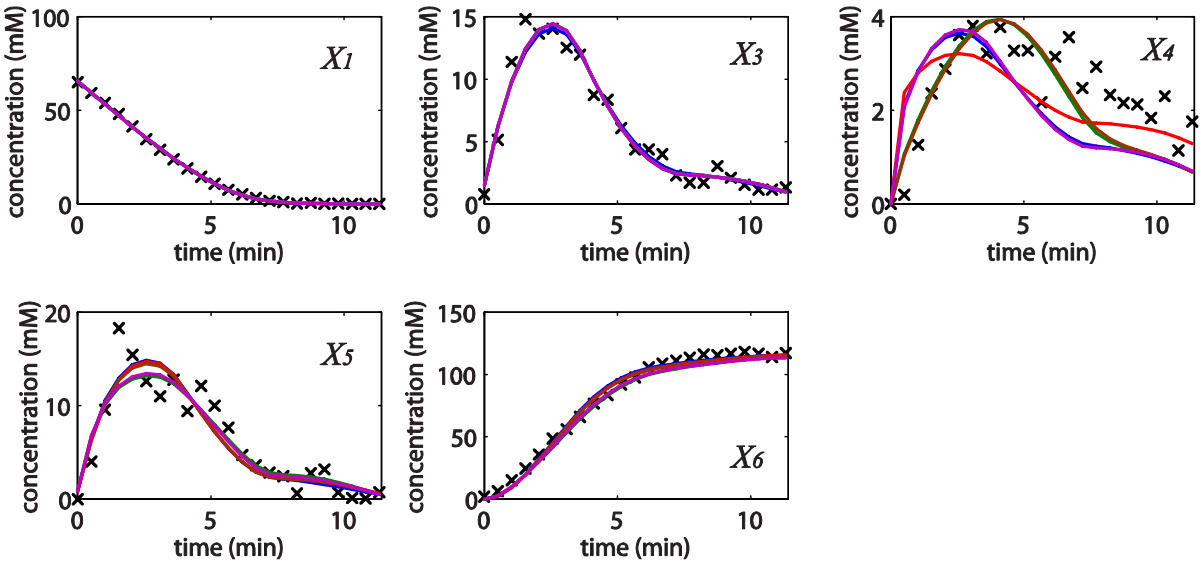


Table 2. Ensemble kinetic modeling of the trehalose pathway model using Φ_R.

CPU time (sec)	6489
Calculated volume of initial parameter space (V_{ci})	1.25×10^8
Estimated volume of viable parameter space (V_{ev})	3237 ± 125
Ratio of V_{ev} to V_{ci}	$(25.90 \pm 1.00) \times 10^{-4}$%
Range of slope errors $\sqrt{\Phi_S}$	[5.825, 46.42]
Range of concentration errors $\sqrt{\Phi_C}$	[1.125, 3.880×10^2]

3. Discussion

The difficulty in simultaneously estimating kinetic parameters of metabolic models is often caused by a lack of complete parameter identifiability [10]. In other words, not all parameters can be uniquely identified and many parameter combinations can give similar goodness-of-fit to the available data [11]. Hence, even when the parameter estimation algorithm could return best-fit values, the resulting model may have little predictive capability; or worse, could be misleading. In the present work, a different approach is taken that directly addresses the issue of model uncertainty through the generation of an ensemble of models. The member models are equivalent in the sense that (1) the models closely approximate the same mass balance equation and (2) the model approximations are statistically equal (to within a 95% confidence level). Although the case studies mainly involved GMA models with power-law flux functions, the ensemble modeling procedure can be used for any form of flux functions, as long as the ODE model follows Equation (1). For power-law and Michaelis-Menten kinetics, the least square regression of the dependent parameters reduces to linear regression, and thus

can be done very efficiently. The main reason to use power-law models here was that they represent some of the most challenging problems in kinetic modeling due to the large parameter space, the lack of complete parameter identifiability, stiff ODEs and high degree of nonlinearity.

In this work, we have used the DOF in estimating dynamic fluxes from time-slopes of concentration data $\dot{\mathbf{X}}(t_k) = \mathbf{S}\mathbf{v}(t_k)$, to restrict the parameter subspace within which the model ensemble is created. Since this DOF is associated with the stoichiometric matrix **S,** the same ambiguity also exists, albeit implicitly, when the original ODE model: $\dot{\mathbf{X}}(t) = \mathbf{S}\mathbf{v}(t)$ is integrated during the parameter estimation. In corollary, there can exist more than one $\mathbf{v}(t)$ that agree with the same $\mathbf{X}(t_k)$. However, in this case, the calculation of $\mathbf{v}_D(t)$ will involve an infinite dimensional vector space (function space). Furthermore, we note that the ambiguity mentioned above is different from the parametric uncertainty that is represented by the ensemble modeling. In particular, the equivalency of models in the ensemble is judged by the error function Φ and different error functions can produce dissimilar model ensembles. As shown in the second case study, a few models of the ensemble created by Φ_R produced large concentration errors Φ_C. This discrepancy is perhaps not surprising as Φ_R is based on the algebraic model $\dot{\mathbf{X}}(t_k) = \mathbf{S}\mathbf{v}(\mathbf{X}(t_k), \mathbf{p})$, while the calculation of Φ_C involves the integration of the ODE model $\dot{\mathbf{X}}(t) = \mathbf{S}\mathbf{v}(\mathbf{X}(t), \mathbf{p})$..

The proposed ensemble modeling method has the advantages that (1) the model ensemble is compactly defined using a small number of independent parameters; (2) the dependent parameters can be efficiently computed from the independent parameters; (3) only biologically-meaningful models are included in the model ensemble; and (4) data uncertainty (noise) is explicitly accounted for. The first two aspects come as courtesy of the step-wise identification approach adopted in the development of the method. The computational cost of constructing the model ensemble is related with the parameter exploration and the computation of the error function. The compactness of the parameter space of the ensemble is therefore particularly important for numerical efficiency and ultimately for practical applications. For OEAMC and MEBS algorithms, the number of required parameter samples during parameter exploration has been shown to increase linearly with the parameter dimension, which in this case is equal to the number of independent parameters [25]. On the other hand, the computational cost of a single evaluation of the error function primarily comes from the least square regression of the dependent parameters and possibly from the integration of the ODE, if the error function requires the simulation of $\mathbf{X}(t)$. For the error function used in the case studies above, this computational cost should increase linearly with the number of dependent fluxes, assuming that the number of unknown parameters in each dependent flux stays about the same.

In the proposed ensemble modeling, the model uncertainty is related to parametric uncertainty that arises from data noise, leaving out the contribution of structural uncertainty (mismatch between the assumed model equations and the true dynamics). Increasing data noise is therefore expected to increase the size of the model ensemble, *i.e.* the volume of the viable parameter subspace, by directly changing the statistics of the error function. However, in this case, higher noise in data will also lead to more uncertainty in the time slopes estimates of the concentration data. Since the direct (error function) and indirect (smoothing and slope calculation) effects of data noise could not be easily separated, we have chosen a Monte Carlo approach in determining the confidence bound of the error function (see Method section).

We have also made the assumption that there exists a unique solution to the computation of $\mathbf{p}_D$ from $\mathbf{p}_I$. For GMA models, this assumption requires that (1) the number of time points exceed the number of parameters $\mathbf{p}_D$ from each flux (not the total number) and (2) the logarithm of the metabolite concentration time profiles appearing in each flux are linearly independent. The first requirement is usually satisfied as the number of parameters involved in every flux ranges only between 2 and 5. The second requirement depends on the experimental conditions, but is again usually fulfilled since each flux depends only on a handful of metabolites and data are contaminated with random noise. If this assumption becomes invalid for one or more dependent fluxes, then these fluxes can be included into the set of independent fluxes, at the cost of increasing the dimensionality and computational time of the parameter exploration step. In such a case, the calculation of dependent fluxes from the independent flux values will require taking a pseudo-inverse of $\mathbf{S}_D$ (see Method).

Constraints on parameters and fluxes are important in restricting the size of the ensemble, in a problem dependent manner. For example, in the first case study, the ensemble hit the lower constraints on both kinetic order parameters (set at 0) and the upper constraint for the rate constant γ_1 (see Figure 2). Meanwhile, parameter constraints affect the second case study more than the first, where the lower and upper constraints of all rate constants and the lower bounds of all kinetic orders limited the viable parameter subspace (see Figure 6). Furthermore, in both case studies, the requirement for positivity of the flux values (*i.e.* lower bounds of the fluxes) was an important constraint, as this was frequently violated during the parameter exploration (data not shown).

The ensemble modeling can be integrated into the iterative model building procedures for biological systems [6]. In this case, the ensemble size will be reduced after every iteration, by removing member models that are not consistent with (additional) time-series concentration data from new experiments. The ensemble of models can also be pruned using steady-state data from knock-out studies and/or thermodynamic constraints [16]. In addition, the benefits of improving the quantification of dynamic fluxes will immediately materialize as such data can be directly used in the proposed method.

Finally, the ability to generate an ensemble of kinetic models also necessitates the development of new methodologies on how to utilize such ensemble. The obvious challenge is how to analyze and/or optimize the system when it is represented by a set of models, not just one model, possibly containing a large number of members. Here, we suggest two strategies: the first involves the generation of a (random) sample of models from the ensemble and in such a case, the results from the analysis and optimization can be represented in the form of a histogram. The second strategy is to take the advantage that the ensemble model generation involves only linear (or log-linear) algebraic equations. In this case, interval or constraint propagation using interval arithmetic can be used to evaluate upper and lower bounds for the system behavior, as done previously for GMA models [37].

4. Method

4.1. Problem Formulation

The ensemble modeling procedure is based on the incremental identification or DFE approach for parameter estimation, where kinetic parameters are estimated in three incremental steps. Initially, given time-course concentration measurements $\mathbf{X}_M(t_k)$, $k = 1, \ldots, K$, the estimation procedure starts

with the computation of time-slopes. Data smoothing is usually applied to improve the numerical estimation of $\dot{\mathbf{X}}_M(t_k)$. The slopes can be estimated using a finite difference approximation of the smoothened data or by differentiating the smoothened curve function, if available. Subsequently, the values of dynamic reaction fluxes are approximated from the mass balance $\dot{\mathbf{X}}_M(t_k)=\mathbf{S}\mathbf{v}(t_k)$. Finally, the kinetic parameters are determined from dynamic flux values using a least square regression $\mathbf{v}(t_k)=\mathbf{v}(\mathbf{X}_M(t_k),\mathbf{p})$, which can now be done for each flux individually. By decomposing the identification problem into smaller easy-to-do subproblems, the step-wise identification can offer a significant reduction in the computational cost of performing the estimation. Furthermore, for power-law flux functions, the third step involve only simple (log-)linear regressions. However, in the original formulation of incremental identification and DFE, one assumes that the subproblems have a unique solution, which is often invalid for a metabolic pathway.

Consider the typical scenario where the number of reactions in the metabolic pathway exceeds that of metabolites (*i.e.*, $m < n$). In this case, there theoretically exist an infinite number of dynamic flux $\mathbf{v}(t_k)$ that can satisfy the mass balance equation $\dot{\mathbf{X}}_M(t_k)=\mathbf{S}\mathbf{v}(t_k)$, each of which represents a valid mathematical solution to the parameter estimation problem. The dimensionality of the dynamic flux solutions is equal to the degree of freedom (DOF) in the mass balance, defined as the difference between the number of fluxes and the number of metabolites: $n_{DOF} = n-m > 0$. Thus, only a subset of n_{DOF} fluxes (called independent fluxes) need to be specified at each time point t_k, while the remaining (dependent) fluxes can be computed from the mass balance equation.

In the following, the flux vector is decomposed into $\mathbf{v}(t_k) = [\ \mathbf{v}_I(t_k)^T\ \mathbf{v}_D(t_k)^T\ ]^T$, where the subscripts I and D denote the independent and dependent subsets, respectively. Similarly, **S** and **p** are restructured as $\mathbf{S} = [\ \mathbf{S}_I\ \mathbf{S}_D\ ]$ and $\mathbf{p} = [\ \mathbf{p}_I\ \mathbf{p}_D\ ]$. As mentioned above, given the values of $\mathbf{v}_I(t_k)$, one can compute the corresponding values of $\mathbf{v}_D(t_k)$, according to:

$$\mathbf{v}_D(t_k)=\mathbf{S}_D^{-1}\left[\dot{\mathbf{X}}_M(t_k)-\mathbf{S}_I\mathbf{v}_I(t_k)\right]. \tag{8}$$

Assuming that **S** has a full row rank, one can choose n_{DOF} independent fluxes such that $\mathbf{S}_D$ is invertible. For numerical efficiency, the independent fluxes are chosen by considering the following: (i) the $\mathbf{S}_D$ is invertible, (ii) the number of the independent parameters $\mathbf{p}_I$ is small, and/or (iii) $\mathbf{p}_I$ values are known *a priori* within a small range. Similar numerical considerations for selecting flux functions have also been discussed elsewhere [38]. Subsequently, by replacing $\mathbf{v}_I(t_k)$ with the flux function $\mathbf{v}_I(\mathbf{X}_M(t_k),\mathbf{p}_I)$ and assuming that the dependent parameters $\mathbf{p}_D$ can be uniquely determined from $\mathbf{v}_D(t_k)$, then the model parameters can be completely defined by assigning the values of the independent parameters $\mathbf{p}_I$. For power-law models, the uniqueness of $\mathbf{p}_D$ is a weak assumption, requiring the least square regression problem $\mathbf{v}_D(t_k)=\mathbf{v}_D(\mathbf{X}_M(t_k),\mathbf{p}_D)$ to be fully or over-determined (see Discussion section).

In the above, we have assumed that time-series data for all metabolites in the model are available. When one or more metabolites are not measured, we can modify the procedure by first rewriting the ODE model, separating the balances associated with those that are measured and those that are not:

$$\dot{\mathbf{X}}(t,\mathbf{p})=\begin{bmatrix}\dot{\mathbf{X}}_M\\ \dot{\mathbf{X}}_U\end{bmatrix}(t,\mathbf{p})=\begin{bmatrix}\mathbf{S}_M\\ \mathbf{S}_U\end{bmatrix}\mathbf{v}(\mathbf{X}_M,\mathbf{X}_U,\mathbf{p}) \tag{9}$$

where the subscripts *M* and *U* refer to the measured and unmeasured metabolites, respectively. The independent fluxes are then selected such that the dependent fluxes can be computed using the following relationship:

$$\mathbf{v}_D(t_k) = \mathbf{S}_{D,M}^{-1}\left[\dot{\mathbf{X}}_M(t_k) - \mathbf{S}_{I,M}\mathbf{v}_I(t_k)\right] \quad (10)$$

where $\mathbf{S}_{I,M}$ and $\mathbf{S}_{D,M}$ are submatrices of $\mathbf{S}_M$, such that $\mathbf{S}_M = [\ \mathbf{S}_{I,M}\ \mathbf{S}_{D,M}\ ]$ following the decomposition of $\mathbf{v}(t_k) = [\ \mathbf{v}_I^T\ \mathbf{v}_D^T\ ]^T$. As expected, the degree of freedoms will increase ($n_{DOF} = n - m^*$, where m^* is the number of measured metabolites), and so will the number of independent fluxes. The independent fluxes should be selected such that $\mathbf{S}_{D,M}$ is invertible and should also include fluxes that appear in $\dot{\mathbf{X}}_U$. The same practical considerations for choosing $\mathbf{v}_I$, e.g. considering the number of and the prior information on $\mathbf{p}_I$, are also applicable. Finally, like before, given the values of $\mathbf{p}_I$, the dependent parameters can be obtained by least square regression of $\mathbf{v}_D(t_k)$. However, since $\mathbf{v}_I(t_k)$ can also depend on $\mathbf{X}_U$, *i.e.* $\mathbf{v}_I(\mathbf{X}_M(t_k), \mathbf{X}_U(t_k), \mathbf{p}_I)$, we will need to simulate $\dot{\mathbf{X}}_U = \mathbf{S}_U\mathbf{v}(\mathbf{X}_M, \mathbf{X}_U, \mathbf{p})$, using the smoothened $\mathbf{X}_M(t)$ as input variables.

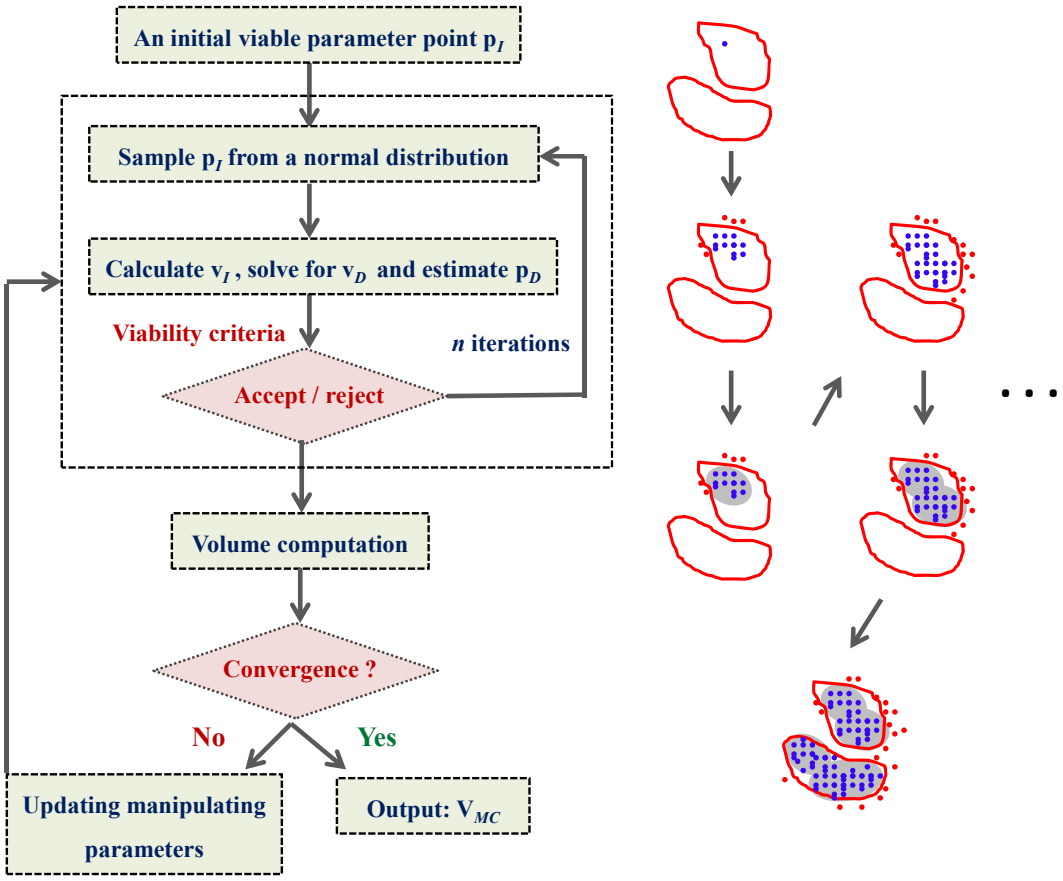
Here, the model ensemble embodies two types of uncertainty: mathematical and statistical. The mathematical uncertainty is related to the aforementioned DOF in the mass balance, while statistical uncertainty is associated with noise in the concentration data. Now, even when different combinations of $\mathbf{p}_I$ and $\mathbf{p}_D$ are obtained from the relationship $\dot{\mathbf{X}}_M(t_k) = \mathbf{S}\mathbf{v}(t_k)$, they may not give the same goodness-of-fit to the concentration measurements $\mathbf{X}_M(t_k)$. Briefly, the difference in the quality of data fitting is due to the fact that the mathematical equivalence above is established based on the slopes of the (smoothened) concentration data, not on the concentrations themselves, and also due to noise in data. Here, the ensemble modeling is performed by exploring the parameter space $\mathbf{p}_I$ and demarcating the viable subset of parameters that satisfy both $\dot{\mathbf{X}}_M(t_k) = \mathbf{S}\mathbf{v}(t_k)$ and two additional criteria: (1) all kinetic parameter values and fluxes are within biologically relevant bounds and (2) the model prediction error is within acceptable statistical bounds. Details of the parameter exploration algorithm and parameter viability criteria are given below.

4.2. HYPERSPACE Toolbox

In this work, the parameter exploration is carried out using the HYPERSPACE toolbox, specifically the out-of-equilibrium adaptive Metropolis Monte Carlo (OEAMC) and multiple ellipsoid-based sampling (MEBS) method [25]. These methods have been shown to be effective in exploring high-dimensional, non-convex and poorly connected viable spaces. Briefly, the OEAMC method provides a coarse-grained global exploration of the viable parameter space. The resulting coarse-grained set in turn becomes the starting point for a fine-grained local exploration offered by the MEBS to further characterize the viable parameter space. The OEAMC algorithm was developed from a combination of Metropolis Monte Carlo sampling [39] and Simulated Annealing [40]. Given an initial viable parameter point, the OEAMC carries out *n* iterations in which new parameter points are sampled from a normal distribution and subjected to the viability criteria. After every *n* iterations, the algorithm determines whether the sampling should be continued depending on a convergence condition. In this case, the viable parameters (blue dots in Figure 8) found so far are grouped into

hyper-ellipsoids of minimum volume (grey areas in Figure 8), which are constructed to enclose the viable points in each cluster. The stopping criterion is then determined from the convergence of the sum of the volumes of these hyper-ellipsoids. Finally, the output from the OEAMC is the set V_{MC} containing coarse-grained viable parameter points. Figure 8 illustrates the procedure of this algorithm.

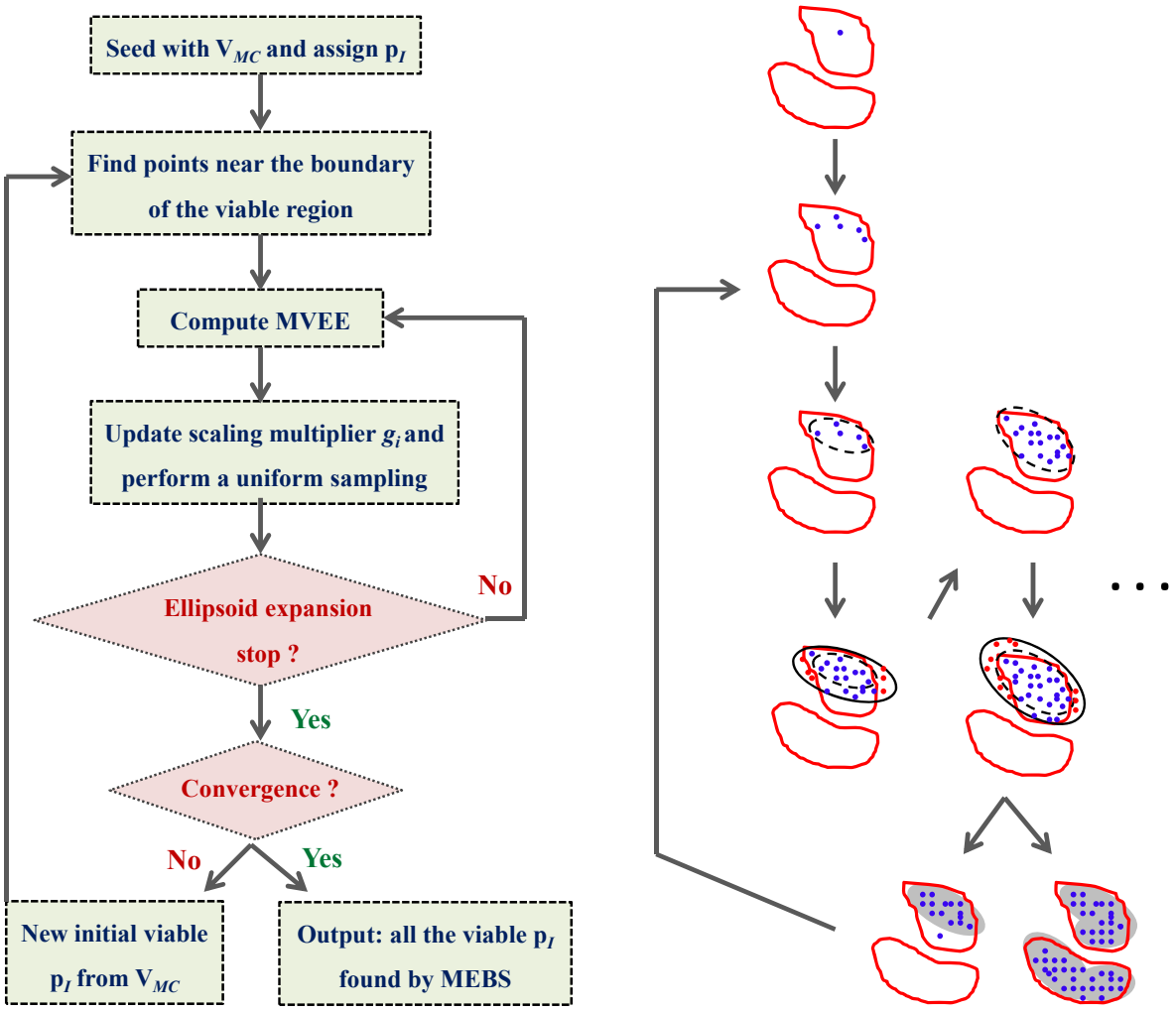
Figure 8. Flowchart of the out-of-equilibrium adaptive Metropolis Monte Carlo (OEAMC) algorithm. On the right, the red closed curves represent hypothetical contour plots of the viable parameter space. The viable points are marked in blue and the nonviable points are marked in red. Finally, the grey areas illustrate the minimum volume enclosing ellipsoids. This figure is adapted from the original publication [25].



The MEBS method is designed to produce fine-tuned hyper-ellipsoids that tightly bound viable regions in the parameter search space, based on another algorithm that has been introduced elsewhere [41]. The ellipsoids' centers, orientations and lengths of axes can be adjusted in order to enclose the viable parameter regions as tightly as possible. Starting with one parameter point of the V_{MC} set, the MEBS searches for viable parameter points near the boundary of the viable region. A Minimum Volume Enclosing Ellipsoid (MVEE, dashed ellipsoids in Figure 9) is then created to cover the local viable region. Subsequently, the MVEE is scaled up by a multiplier g_i (solid curves in Figure 9), and a uniform sample of points is generated inside this scaled ellipsoid. Among these random points, the nonviable points (red points) are discarded, and another iteration of MVEE and another uniform sampling using a new multiplier g_{i+1} are done using the remaining viable ones (blue points). The performance of the algorithm depends strongly on the multiplier g_i, and here we have used the

recommended scaling parameters in the original publication [25]. The iteration is repeated until the scaling multiplier tends to one or a fixed number of iterations is reached. Finally, the whole procedure above is repeated for another viable parameter point from V_{MC} until all parameter points in this set are exhausted. The output of the MEBS is a comprehensive set of viable parameter points. Figure 9 summarizes the procedure of the MEBS algorithm.

Figure 9. Flowchart of the multiple ellipsoid-based sampling (MEBS) algorithm. In the right part of the figure, the red closed curves represent hypothetical contour plots of the viable parameter space defined by some criteria. The viable points are marked in blue and the nonviable points are marked in red. Finally, the grey areas illustrate the minimum volume enclosing ellipsoids. This figure is adapted from the original publication [25].



4.3. Model Viability Criteria

Given any values of the independent parameters $\mathbf{p}_I$, the corresponding dependent fluxes and parameters may not necessarily be biologically relevant, for example the dependent fluxes may become negative or the parameters may assume unrealistic values. Thus, in the ensemble modeling procedure, these cases are excluded by enforcing constraints on the values of fluxes and parameters, as follow:

$$\mathbf{p} \in [\mathbf{L}, \mathbf{U}]; \quad \mathbf{v}(t_k) \in [\mathbf{0}, \mathbf{U}_\mathbf{v}]; \tag{11}$$

where **L** and **U** denote the lower and upper bounds for the parameters, and $\mathbf{U}_\mathbf{v}$ is the maximum value of metabolic fluxes. The second viability criterion is meant to establish equivalence among the member models in terms of their goodness of fit to data. If one makes the assumption that data noise comes from a Gaussian distribution, then the confidence bound of error function Φ can usually be estimated using standard statistical analyses and model sensitivities [42]. When data noise is not Gaussian, the confidence bounds can be estimated using a Monte Carlo approach [43].

In the case studies, the upper confidence bound of the error function Φ was obtained using a Monte Carlo approach. Specifically, 100 sets of time profiles were randomly generated from a Gaussian distribution using the measured concentration data as the mean values. The variance of the data noise was estimated from the residuals of the data smoothing procedure. For each dataset, the same data smoothing and slope calculation were performed and the corresponding parameter estimates were obtained by minimizing the error function (see below). The confidence bound was directly estimated from the set of 100 values of Φ. For example, the 95% upper confidence bound of the upper bound of the error function is approximated by the *5*-th largest Φ in this set.

4.4. Ensemble Modeling Procedure

In the examples, the error function Φ was set to be:

$$\Phi_R(\mathbf{p}, \mathbf{X}) = \frac{1}{mK} \sum_{k=1}^{K} \left[\mathbf{v}_D(t_k) - \mathbf{v}_D\left(\mathbf{X}_M(t_k), \mathbf{p}_D\right) \right]^T \left[\mathbf{v}_D(t_k) - \mathbf{v}_D\left(\mathbf{X}_M(t_k), \mathbf{p}_D\right) \right] \tag{12}$$

where *K* is the total number of measurement time points. This error function is implemented in the last step of the incremental identification, where the dependent parameters $\mathbf{p}_D$ are regressed from the dynamic flux estimates $\mathbf{v}_D(t_k)$. Note that the calculation of this error function was actually done one flux at a time, as the least square regression of $\mathbf{p}_D$ from $\mathbf{v}_D(t_k)$ was performed for each flux function separately. For power-laws, this regression can be performed very efficiently, as the logarithm of the flux function depends linearly on the parameters (leading to a linear least square regression). In this case, $\mathbf{v}_D(t_k)$ was calculated from $\mathbf{v}_I(t_k)$ according to Equation (8), while $\mathbf{v}_I(t_k)$ was computed from the time series data and $\mathbf{p}_I$ using the flux function $\mathbf{v}_I(\mathbf{X}_M(t_k),\mathbf{p}_I)$. In other words, the error function depends only on the independent parameters $\mathbf{p}_I$. The initial parameter point $\mathbf{p}_I$ for the OEAMC algorithm was obtained from the following optimization:

$$\min_{\mathbf{p}_I} \Phi_R(\mathbf{p}, \mathbf{X}) \tag{13}$$

subject to the bounds on the parameters and fluxes as discussed in the previous section. Other error functions can also be used, for example, using the slope prediction error:

$$\Phi_S(\mathbf{p}, \mathbf{X}) = \frac{1}{mK} \sum_{k=1}^{K} \left[\dot{\mathbf{X}}_M(t_k) - \mathbf{S}\mathbf{v}\left(\mathbf{X}_M(t_k), \mathbf{p}\right) \right]^T \left[\dot{\mathbf{X}}_M(t_k) - \mathbf{S}\mathbf{v}\left(\mathbf{X}_M(t_k), \mathbf{p}\right) \right] \tag{14}$$

or the concentration prediction error:

$$\Phi_C(\mathbf{p},\mathbf{X}) = \frac{1}{mK}\sum_{k=1}^{K}\left[\mathbf{X}_M(t_k) - \mathbf{X}(t_k,\mathbf{p})\right]^T\left[\mathbf{X}_M(t_k) - \mathbf{X}(t_k,\mathbf{p})\right] \quad (15)$$

where $\mathbf{X}(t_k,\mathbf{p})$ is the concentration simulation.

The model ensemble procedure starts with finding an initial viable point for the OEAMC algorithm, as discussed above. Next, the upper bound for the error function will be set either by applying standard statistical analysis assuming Gaussian noise or using the Monte Carlo algorithm described in the previous subsection. The OEAMC is then applied to generate the coarse-grained set of viable parameters over the space of the independent parameters. Finally, this set becomes the input to the MEBS algorithm, producing a population of viable parameters $\mathbf{p}_I$ that represents the ensemble of models. Note that while this work concerns with power-law fluxes, the ensemble generation procedure has general applicability to any kinetic models that can be written as $\dot{\mathbf{X}}(t,\mathbf{p}) = \mathbf{S}\mathbf{v}(\mathbf{X},\mathbf{p})$.

5. Conclusions

The kinetic modeling of metabolic networks is challenging, but critical in many applications of metabolic engineering. Particularly, parameter identifiability issue, wherein not all parameters can be uniquely determined from the data, has been identified as a common root cause of the difficulty in this process. This uncertainty in parameters implies that there exist (infinitely) many models that will give statistically equivalent goodness of fit to data. Built on the concept of incremental identification, we have proposed an efficient ensemble modeling procedure that relies on three components: (1) data smoothing and approximation of time-series metabolic concentration data, (2) a compact parameter space defining the model ensemble, and (3) efficient parameter exploration. The applications for the ensemble modeling of a generic branched pathway and the trehalose pathway in *Saccharomyces cerevisiae* demonstrate the efficacy of the proposed method.

Acknowledgments

The authors would like to acknowledge the funding support from Singapore-MIT Alliance and ETH Zurich. We also would like to thank Dr. Adrián López García de Lomana and Prof. Andreas Wagner for their assistance in using the HYPERSPACE toolbox.

Conflict of Interest

The authors declare no conflict of interest.

References

1. Stephanopoulos, G.; Aristidou, A.A.; Nielsen, J.H. *Metabolic Engineering: Principles and Methodologies*; Academic Press: San Diego, USA, 1998.
2. Palsson, B. *Systems Biology : Properties of Reconstructed Networks*; Cambridge University Press: Cambridge, UK, 2006.
3. Varma, A.; Palsson, B.O. Metabolic flux balancing - basic concepts, scientific and practical use. *Nat. Biotech.* **1994**, *12*, 994–998.

4. Gombert, A.K.; Nielsen, J. Mathematical modelling of metabolism. *Curr. Opin. Biotechnol.* **2000**, *11*, 180–186.
5. Gadkar, K.G.; Gunawan, R.; Doyle, F.J. Iterative approach to model identification of biological networks. *BMC Bioinf.* **2005**, *6*, 155–173.
6. Chou, I.C.; Voit, E.O. Recent developments in parameter estimation and structure identification of biochemical and genomic systems. *Math Biosci.* **2009**, *219*, 57–83.
7. Chis, O.T.; Banga, J.R.; Balsa-Canto, E. Structural identifiability of systems biology models: A critical comparison of methods. *Plos One* **2011**, *6*, doi:10.1371/journal.pone.0027755.
8. Nikerel, I.E.; van Winden, W.A.; Verheijen, P.J.; Heijnen, J.J. Model reduction and a priori kinetic parameter identifiability analysis using metabolome time series for metabolic reaction networks with linlog kinetics. *Metab. Eng.* **2009**, *11*, 20–30.
9. Raue, A.; Kreutz, C.; Maiwald, T.; Bachmann, J.; Schilling, M.; Klingmuller, U.; Timmer, J. Structural and practical identifiability analysis of partially observed dynamical models by exploiting the profile likelihood. *Bioinformatics* **2009**, *25*, 1923–1929.
10. Srinath, S.; Gunawan, R. Parameter identifiability of power-law biochemical system models. *J. Biotechnol.* **2010**, *149*, 132–140.
11. Vilela, M.; Vinga, S.; Maia, M.A.; Voit, E.O.; Almeida, J.S. Identification of neutral biochemical network models from time series data. *BMC Syst. Biol.* **2009**, *3*, 47.
12. Mendes, P.; Kell, D. Non-linear optimization of biochemical pathways: Applications to metabolic engineering and parameter estimation. *Bioinformatics* **1998**, *14*, 869–883.
13. Moles, C.G.; Mendes, P.; Banga, J.R. Parameter estimation in biochemical pathways: A comparison of global optimization methods. *Genome Res.* **2003**, *13*, 2467–2474.
14. Song, S.O.; Chakrabarti, A.; Varner, J.D. Ensembles of signal transduction models using pareto optimal ensemble techniques (poets). *Biotechnol. J.* **2010**, *5*, 768–780.
15. Henry, C.S.; Broadbelt, L.J.; Hatzimanikatis, V. Thermodynamics-based metabolic flux analysis. *Biophys. J.* **2007**, *92*, 1792–1805.
16. Miskovic, L.; Hatzimanikatis, V. Modeling of uncertainties in biochemical reactions. *Biotechnol. Bioeng.* **2011**, *108*, 413–423.
17. Tran, L.M.; Rizk, M.L.; Liao, J.C., Ensemble modeling of metabolic networks. *Biophys. J.* **2008**, *95*, 5606–5617.
18. Wang, L.; Birol, I.; Hatzimanikatis, V., Metabolic control analysis under uncertainty: Framework development and case studies. *Biophys. J.* **2004**, *87*, 3750–3763.
19. Kuepfer, L.; Peter, M.; Sauer, U.; Stelling, J. Ensemble modeling for analysis of cell signaling dynamics. *Nat. Biotechnol.* **2007**, *25*, 1001–1006.
20. Schaber, J.; Flottmann, M.; Li, J.; Tiger, C.F.; Hohmann, S.; Klipp, E. Automated ensemble modeling with modelmage: Analyzing feedback mechanisms in the sho1 branch of the hog pathway. *PLOS One* **2011**, *6*, doi:10.1371/journal.pone.0014791.
21. Milanese, M.; Vicino, A. Optimal estimation theory for dynamic systems with set membership uncertainty : An overview. *Automatica* **1991**, *27*, 997–1009.
22. Battogtokh, D.; Asch, D.K.; Case, M.E.; Arnold, J.; Schuttler, H.B. An ensemble method for identifying regulatory circuits with special reference to the qa gene cluster of neurospora crassa. *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 16904–16909.

23. Bardow, A.; Marquardt, W. Incremental and simultaneous identification of reaction kinetics: Methods and comparison. *Chem. Eng. Sci.* **2004**, *59*, 2673–2684.
24. Goel, G.; Chou, I.C.; Voit, E.O. System estimation from metabolic time-series data. *Bioinformatics* **2008**, *24*, 2505–2511.
25. Zamora-Sillero, E.; Hafner, M.; Ibig, A.; Stelling, J.; Wagner, A. Efficient characterization of high-dimensional parameter spaces for systems biology. *BMC Syst. Biol.* **2011**, *5*, doi:10.1186/1752-0509-5-142.
26. Voit, E.O.; Almeida, J. Decoupling dynamical systems for pathway identification from metabolic profiles. *Bioinformatics* **2004**, *20*, 1670–1681.
27. Fonseca, L.L.; Sanchez, C.; Santos, H.; Voit, E.O. Complex coordination of multi-scale cellular responses to environmental stress. *Mol. Biosyst.* **2011**, *7*, 731–741.
28. Chou, I.C.; Voit, E.O. Estimation of dynamic flux profiles from metabolic time series data. *BMC Syst. Biol.* **2012**, *6*, 84-106.
29. Voit, E.O. *Computational Analysis of Biochemical Systems : A Practical Guide for Biochemists and Molecular Biologists*; Cambridge University Press: New York, USA, 2000.
30. Savageau, M.A. Biochemical systems analysis. I. Some mathematical properties of the rate law for the component enzymatic reactions. *J. Theor. Biol.* **1969a**, *25*, 365–369.
31. Savageau, M.A. Biochemical systems analysis: II. The steady-state solutions for an n-pool system using a power-law approximation. *J. Theor. Biol.* **1969b**, *25*, 370–379.
32. Sorribas, A.; Cascante, M., Structure identifiability in metabolic pathways: Parameter estimation in models based on the power-law formalism. *Biochem. J.* **1994**, *298*, 303–311.
33. Marquardt, W.; Brendel, M.; Bonvin, D. Incremental identification of kinetic models for homogeneous reaction systems. *Chem. Eng. Sci.* **2006**, *61*, 5404–5420.
34. Montgomery, D.C.; Runger, G.C. *Applied Statistics and Probability for Engineers*, 4th ed.; Wiley: Hoboken, NJ, USA, 2007.
35. Akaike, H. New look at statistical-model identification. *Ieee T. Automat. Contr.* **1974**, *Ac19*, 716–723.
36. Chassagnole, C.; Noisommit-Rizzi, N.; Schmid, J.W.; Mauch, K.; Reuss, M. Dynamic modeling of the central carbon metabolism of escherichia coli. *Biotechnol. Bioeng.* **2002**, *79*, 53–73.
37. Tucker, W.; Kutalik, Z.; Moulton, V. Estimating parameters for generalized mass action models using constraint propagation. *Math. Biosci.* **2007**, *208*, 607–620.
38. Voit, E.O.; Goel, G.; Chou, I.C.; Fonseca, L.L. Estimation of metabolic pathway systems from different data sources. *IET Syst. Biol.* **2009**, *3*, 513–522.
39. Newman, M.E.J.; Barkema, G.T. *Monte carlo methods in statistical physics*. Clarendon Press: Oxford, UK, 1999.
40. Kirkpatrick, S.; Gelatt, C.D.; Vecchi, M.P. Optimization by simulated annealing. *Science* **1983**, *220*, 671–680.
41. Khachiyan, L.G. Rounding of polytopes in the real number model of computation. *Math. Oper. Res.* **1996**, *21*, 307–320.
42. Beck, J.V.; Arnold, K.J. *Parameter Estimation in Engineering and Science*; Wiley: New York, NY, USA, 1977.

43. Bard, Y. *Nonlinear Parameter Estimation*; Academic Press: New York, NY, USA; London, UK, 1974.