Proceedings of the 2007 IEEE Symposium on Computational
Intelligence in Bioinformatics and Computational Biology (CIBCB 2007)

In vivo Intracellular Metabolite Dynamics Estimation by Sequential Monte Carlo Filter

Jing Yang, Visakan Kadirkamanathan and Stephen A. Billings
Signal Processing and Complex Systems Research Group,
Department of Automatic Control & Systems Engineering,
The University of Sheffield, United Kingdom, S1 3JD
Emails: {j.yang, visakan, s.billings}@sheffield.ac.uk

Abstract— **The quantitative comprehension of a metabolic system in its dynamic state is a prerequisite for purposed strain improvement and enzymatic regulation. It is therefore crucial to accurately obtain the extracellular and intracellular metabolite concentrations *in vivo* in the time scale faster than typical metabolite turn-over rate. Though intracellular metabolite dynamics are addressable by latest rapid sampling technology, the measurements are not satisfactory due to the low percentage of intracellular volume to the total sample volume and often the low concentration levels of most intracellular metabolites. When the examined system is observable, a possible solution to this problem is by means of available statistical estimation approach. Hence, in this paper, the Sequential Monte Carlo filter is applied to estimate the intracellular metabolite concentrations with the knowledge of extracellular metabolite concentrations. The application of this algorithm in a synthetic system with simulated data illustrates the applicability of this approach. All the intracellular metabolite concentrations are accurately estimated and the extracellular states are reconstructed from their noisy measurements. The dynamic flux distributions are also obtained and their underlying biological meanings are described.**

I. Introduction

Metabolic engineering, which manipulates the original metabolic pathway structures and reaction rates through various methods, has been undergoing rapid development since the concept was introduced in 1990s [1] [2]. Especially after the gene sequences of most organisms gradually became available, metabolic engineering enters the 'post-genomic' era [3] [4]. It is then possible to 'knock-in' or 'knock-out' a particular gene or genes in order to activate or inhibit the linked enzyme(s), whose change, ultimately incur variations in metabolic reaction rates and metabolic pathway structures. Metabolic flux, which is often viewed as the quantitative representation of metabolic engineering, has received particular attention. Metabolic balance analysis based solely on metabolic stoichiometric information [5] and metabolic flux analysis based on stoichiometric information with additional constraints from ^{13}C labelling experiment data [6] [7] [8] [9] [10] are the main approaches used for metabolic flux quantification. Meanwhile, metabolic control analysis which intends to understand the underlying enzyme activities by sensitivity analysis of control coefficients, elasticity coefficients etc. arranged in the steady-state representation of a specific metabolic system, has long been viewed as a traditional method for studying enzyme regulations [11] [12] [13].

Among all the above activities in metabolic engineering, it is obvious that flux is a fundamental determinant of cell physiology and it is the crucial part in a metabolic pathway. Nevertheless, the fluxes dealt with in both metabolic flux analysis and metabolic control analysis are only viewed as a standard signal for system efficiency or sensitivity evaluation in (pseudo) steady state. In order to undergo a re-engineering of pathways, it is necessary to obtain a quantified and mechanistic knowledge of regulation phenomena not only on flux phenotype level but also on genome, transcriptome, proteome and metabolome level [14].

Over decades, much time and effort has been devoted to gene sequences, functional genomics analysis, as well as transcription (transcriptomics) and translation (proteomics). However, these methods are unable to provide the change in biological phenotypes coupled with changes in a mRNA or protein. In order to elucidate the regulatory effects of genetic alteration on biological reactions, an analysis constrained exclusively to the metabolites is then necessary [15]. Under some circumstances, only the activity of a specific metabolite or an isolated group of metabolites need identifying and quantifying, which is at the core of 'metabolic profiling' [15] [16]. However, the necessity to analyze a large number of metabolites within the period before an examined metabolic system reaches its steady state is commonly encountered in the quantification of cellular metabolism. Such comprehensive profiling provides not only the regulation and control of biochemical structure, but also the dynamic flux information, both of which are vital for various strain improvement and enzymatic activities investigation.

Due to the high turn-over rate of metabolites, which is usually on a subsecond scale, rapid sampling and fast quenching technology [17] [18] [19] are needed for measuring metabolite concentrations *in vivo*. The cells in usually substrate-limited medium are stimulated by a quick substrate pulse and the response sample volumes afterwards are suspended by immediate rapid sampling and fast quenching facilities [20]. The sample volumes are then separated to extracellular samples and intracellular samples by various-purposed centrifugation and extractions. A combination of enzymatic assays, High Performance Liquid Chromatography (HPLC) and Electrospray Ionization Liquid Chromatographic Tandem Mass Spectrometry (ESI-LC-MS) can be used to quantify the

metabolite concentrations [19] [21].

Although the above technology is capable of providing dynamic metabolite concentration measurements, the potential drawbacks that the intracellular metabolites often lie below 1mM [22] [23] and the intracellular volume usually is less than 3% of the total sample volume, make the intracellular measurements more prone to noise or even undetectable if the specific concentration falls below the limits of detection. Typically, 4-5 samples are obtained per second in a rapid sampling experiment [21]. Therefore, there will be 120-150 samples if the observations during 30 seconds after the substrate impulse are acquired. The corresponding measurements of these samples may cost time and effort even though parallel determination of intracellular metabolites are possible [19].

Here, in addition to what have been achieved in previous literature, where the kinetic information about the interested metabolic system was derived in terms of available extracellular and intracellular metabolite concentrations [24] [25] [26] [20] [27], this paper takes a step further and aims at estimating the dynamics of metabolic intracellular concentrations with the knowledge of extracellular concentrations and system kinetic structures. By formulating the metabolic system as a discrete state space model, *Sequential Monte Carlo filter* (SMC) is applied to obtain the posterior distribution of intracellular metabolite concentrations.

The paper is organized as follows: in Section II, the state space model concept is presented and the application of the SMC filter in deriving latent states in a state space model is described. In Section III, an example system and its relevant Michaelis-Menton kinetics parameters are listed. The detailed SMC filter algorithm is also given. In Section IV, the intracellular metabolite concentrations and the dynamic fluxes estimated by SMC filter are illustrated and compared. The underlying biological meaning is also discussed. In Section V, conclusions and future work are outlined.

II. GENERAL STATE SPACE MODEL AND SEQUENTIAL MONTE CARLO FILTER

A. State space model

A state space representation is a mathematical abstraction of a physical system with a set of unobservable states and observations. Typically, a state space model with additive noise is in the following format:

$$\mathbf{x}_k = f_k(\mathbf{x}_{k-1}) + \mathbf{r}_k \quad (1)$$

$$\mathbf{y}_k = h_k(\mathbf{x}_k) + \mathbf{w}_k \quad (2)$$

where $\{\mathbf{x}_k \in \mathcal{R}^{n_x}, k \in \mathbb{N}\}$ is n_x dimensional state sequence and $\{\mathbf{y}_k \in \mathcal{R}^{n_y}, k \in \mathbb{N}\}$ is n_y dimensional observation sequence, $\mathbb{N}$ is the set of natural numbers. $\{\mathbf{r}_k \in \mathcal{R}^{n_x}, k \in \mathbb{N}\}$ and $\{\mathbf{w}_k \in \mathcal{R}^{n_y}, k \in \mathbb{N}\}$ are i.i.d system state process noise and system observation noise, respectively. $f_k : \mathcal{R}^{n_x} \rightarrow \mathcal{R}^{n_x}$ is the system transition function. $h_k : \mathcal{R}^{n_x} \rightarrow \mathcal{R}^{n_y}$ is the system observation function. Here we denote by $\mathbf{x}_{0:n} \triangleq \{\mathbf{x}_0, \cdots, \mathbf{x}_n\}$ and $\mathbf{y}_{1:m} \triangleq \{\mathbf{y}_1, \cdots, \mathbf{y}_m\}$, the states and the observations up to step n and m, respectively. In most cases, the underlying question is about how to estimate the states $\mathbf{x}_{0:n}$ when only partial or inaccurate observations $\mathbf{y}_{1:m}$ are available [28].

From a Bayesian perspective, the main concern is then focused on the posterior distribution of the states: $p(\mathbf{x}_{0:n}|\mathbf{y}_{1:m})$. Though Kalman filter (for linear state space model with Gaussian noise) [29] and extended Kalman filter (for nonlinear state space model with Gaussian noise) [30] have been extensively adopted in various occasions, their online applications are largely restricted by associated approximation nature and the assumptions of Gaussian noise distribution. The SMC method [31] [32] [33], also termed the particle filter [34], which utilizes a random sample (particle) based representation of the interested posterior distribution, has obtained various applications in online Bayesian estimation.

B. Sequential Monte Carlo filter

From (1) and (2) and considering the initial condition of $\mathbf{x}$: $p(\mathbf{x}_0)$ as $p(\mathbf{x}_0|\mathbf{x}_{-1})$, the state space model can be rewritten from statistical point of view as,

$$p(\mathbf{x}_k|\mathbf{x}_{k-1}) \propto p_{\mathbf{r}_k}(\mathbf{x}_k - f_k(\mathbf{x}_{k-1})) \quad (3)$$

$$p(\mathbf{y}_k|\mathbf{x}_k) \propto p_{\mathbf{w}_k}(\mathbf{y}_k - h_k(\mathbf{x}_k)) \quad (4)$$

The Bayesian filtering distribution $p(\mathbf{x}_{0:n}|\mathbf{y}_{1:n})$ can then be recursively derived as,

$$p(\mathbf{x}_{0:n}|\mathbf{y}_{1:n}) = p(\mathbf{x}_{0:n-1}|\mathbf{y}_{1:n-1}) \frac{p(\mathbf{y}_n|\mathbf{x}_n)p(\mathbf{x}_n|\mathbf{x}_{n-1})}{p(\mathbf{y}_n|\mathbf{y}_{1:n-1})} \quad (5)$$

From (5), it can be derived that,

$$p(\mathbf{x}_{0:n}|\mathbf{y}_{1:n}) \propto p(\mathbf{x}_{0:n-1}|\mathbf{y}_{1:n-1})p(\mathbf{y}_n|\mathbf{x}_n)p(\mathbf{x}_n|\mathbf{x}_{n-1}) \quad (6)$$

which implies that the posterior distribution $p(\mathbf{x}_{0:n}|\mathbf{y}_{1:n})$ can be evaluated iteratively, even though its analytical form may be unavailable in a nonlinear state space model. In such circumstance when direct sampling from the posterior distribution is unobtainable, SMC filter provides an alternative approach for approximating the aforementioned probability density.

SMC filter aims to represent a posterior distribution $\pi(\mathbf{z})$ by a number of samples $\mathbf{z}^i$ drawn from an importance function $q(\mathbf{z})$ with their associated weights $w^i = \frac{\pi(\mathbf{z}^i)}{q(\mathbf{z}^i)}$. From Monte Carlo approximation [35], the expectation of a function $f(\mathbf{z})$ is then given by,

$$E[f(\mathbf{z})] \approx \Sigma_{i=1}^N f(\mathbf{z}^i)\bar{w}^i$$

Here N is the number of samples and $\bar{w}^i = \frac{w^i}{\Sigma_{i=1}^N w^i}$ is the normalized weight of the i_{th} sample. Under weak assumptions, with enough samples, $\Sigma_{i=1}^N f(\mathbf{z}^i)\bar{w}^i$ converges to the expectation of $f(\mathbf{z})$. Hence, a SMC filter [36] is applicable in deriving $p(\mathbf{x}_{0:n}|\mathbf{y}_{1:n})$ from $p(\mathbf{x}_{0:n-1}|\mathbf{y}_{1:n-1})$ recursively as long as the importance function can be written in the format given below,

$$q(\mathbf{x}_{0:n}|\mathbf{y}_{1:n}) = q(\mathbf{x}_{0:n-1}|\mathbf{y}_{1:n-1})q(\mathbf{x}_n|\mathbf{x}_{0:n-1}, \mathbf{y}_{1:n}) \quad (7)$$

The weight w^n is then given by,

$$w^n = w^{n-1}\frac{p(\mathbf{y}_n|\mathbf{x}_n)p(\mathbf{x}_n|\mathbf{x}_{n-1})}{q(\mathbf{x}_n|\mathbf{x}_{0:n-1},\mathbf{y}_{1:n})}$$

The basic bootstrap for a state space model is:

for $k = 1, 2, \cdots$
  Sample $\mathbf{x}_k^i, i = 1, \cdots, N$ from the importance function $q(\mathbf{x}_k|\mathbf{x}_{0:k-1}, \mathbf{y}_{1:k})$;
  Calculate the temporary weight $u_k^i = \frac{p(\mathbf{y}_k|\mathbf{x}_k^i)p(\mathbf{x}_k^i|\mathbf{x}_{k-1}^i)}{q(\mathbf{x}_k^i|\mathbf{x}_{0:k-1}^i,\mathbf{y}_{1:k})}$;
  Normalize the weight u_k^i to get $\bar{u}$: $\bar{u}_k^i = \frac{u_k^i}{\Sigma_{i=1}^N u_k^i}$;
  Calculate weight w_k^i: $w_k^i = w_{k-1}^i \bar{u}_k^i$;
  Normalize the weight $\bar{w}_k^i = \frac{w_k^i}{\Sigma_{i=1}^N w_k^i}$;
  The estimation of $\mathbf{x}_k$: $\hat{\mathbf{x}}_k = \Sigma_{i=1}^N \mathbf{x}_k^i \bar{w}_k^i$
end

Here $\mathbf{x}_k^i$ and $\mathbf{w}_k^i$ represent the sample and the weight for the i_{th} sample in the k_{th} sampling time.

III. SMC FILTER AND METABOLIC SYSTEM DYNAMICS ANALYSIS

Stochastic methods have been applied in bioengineering and bioinformatics for a long time. For example, evolutionary algorithms have extensive application in gene sequencing, protein structure and metabolic pathway optimization [37] [38] [39]. Hidden Markov Model is also widely used in protein modelling and protein sequence analysis [40] [41]. At the metabolic level, there have been various applications of stochastic methods in (pseudo) steady state [8] [42] [43]. However, in a dynamic metabolic system, an appropriate statistical method becomes crucial when considering the underlying drawbacks of rapid sampling experiment under the scenario of the intracellular metabolites with concentrations below limit of detection. Moreover, dynamic flux information, which is important for intended strain improvement can only be obtained by dedicated mathematical approach instead of current laboratory experiments. Hence, when the extracellular data are available, the SMC method, which aims to estimate the latent states by the updated posterior information derived from available measurements, can be applied to estimate the intracellular metabolite dynamics.

A. An example metabolic network and its dynamics

In this paper, an example system in Fig. 1 utilized in [44] is used here in order to illustrate the applicability of the SMC in metabolic dynamics analysis. For the sake of simplicity, the fluxes associated with reactions 'HK', '6PF-1-K', 'FBA', 'GPD', 'GPP', 'ENO' and 'PYK' are represented by $\mathbf{v}_1$, $\mathbf{v}_2$, $\mathbf{v}_3$, $\mathbf{v}_4$, $\mathbf{v}_5$, $\mathbf{v}_6$ and $\mathbf{v}_7$, respectively. And the concentrations of Glucose, F6P, F1,6bP, DHAP, G3P, GAP, PEP, Glycerol and Ethanol are denoted by c_1, c_2, c_3, c_4, c_5, c_6, c_7, c_8 and c_9, respectively. It is assumed that F2,6bP is involved with the reaction 6PF-1-K and is fixed during the reactions. Hence, the concentration of F2,6bP is represented by c_{10}. The Michaelis-Menton kinetics listed in [44] for reactions 'HK', '6PF-1-K', 'FBA','GPD', 'GPP', 'ENO' and 'PYK' are adopted and listed here:

$$\mathbf{v}_1 = v_{max,1}\frac{\frac{c_1}{K_1}(1-\frac{c_2}{c_1K_{1,1}})}{1+\frac{c_1}{K_1}+\frac{c_2}{K_2}}$$

$$\mathbf{v}_2 = v_{max,2}c_{10}\frac{\frac{c_2}{K_2}}{(1+\frac{c_2}{K_2})(1+\frac{L_0}{(1+\frac{c_2}{K_2})^{n_1}})}$$

$$\mathbf{v}_3 = v_{max,3}\frac{\frac{c_3}{K_3}(1-\frac{c_4c_6}{c_3K_{3,1}})}{(1+\frac{c_3}{K_3})(1+\frac{c_4}{K_4}+\frac{c_6}{K_6})}$$

$$\mathbf{v}_4 = v_{max,4}\frac{\frac{c_4}{K_4}(1-\frac{c_5}{c_4K_{4,1}})}{1+\frac{c_4}{K_4}+\frac{c_5}{K_5}}$$

$$\mathbf{v}_5 = v_{max,5}\frac{\frac{c_5}{K_5}(1-\frac{c_8}{c_5K_{5,1}})}{1+\frac{c_5}{K_5}+\frac{c_8}{K_8}}$$

$$\mathbf{v}_6 = v_{max,6}\frac{\frac{c_6}{K_6}(1-\frac{c_7}{c_6K_{6,1}})}{1+\frac{c_6}{K_6}+\frac{c_7}{K_7}}$$

$$\mathbf{v}_7 = v_{max,7}\frac{\frac{c_7}{K_7}(1+\frac{c_7}{K_7})^{n_2-1}}{L_1(\frac{2}{1+\frac{c_3}{K_3}})^{n_2}+(1+\frac{c_7}{K_7})^{n_{PYK}}}$$

The parameters associated with the seven Michaelis-Menton equations are listed in Table I.

TABLE I
PARAMETERS USED IN THE MICHAELIS-MENTON EQUATIONS OF FIG. 1

Parameter	Value	Parameter	Value
$v_{max,1}$	1.0	$v_{max,2}$	1.0
$v_{max,3}$	5.0	$v_{max,4}$	5.0
$v_{max,5}$	5.0	$v_{max,6}$	5.0
$v_{max,7}$	10.0	K_1	1.0
K_2	$1.0/c_{10}$	K_3	1.0
K_4	1.0	K_5	1.0
K_6	1.0	K_7	1.0
K_8	1.0	$K_{1,1}$	1.0
$K_{3,1}$	0.1	$K_{4,1}$	1.0
$K_{5,1}$	1.0	$K_{6,1}$	1.0
L_0	10000	L_1	100
n_1	8	n_2	4

With the mass balance of this dynamic metabolic system, the metabolite concentrations and the associated fluxes in Fig. 1 can be linked by a series of differential equations which can be concisely written in a matrix format:

$$\dot{\mathbf{c}} = \mathbf{S}\mathbf{v} \tag{8}$$

where $\mathbf{c} = \begin{bmatrix} c_1 & c_2 & c_3 & c_4 & c_5 & c_6 & c_7 & c_8 & c_9 \end{bmatrix}^T$ (since c_{10} is assumed fixed in the model, it is not included in the vector $\mathbf{c}$) and $\mathbf{v} = \begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \mathbf{v}_3 & \mathbf{v}_4 & \mathbf{v}_5 & \mathbf{v}_6 & \mathbf{v}_7 \end{bmatrix}^T$. $\mathbf{S}$ is the stoichiometric matrix given by,

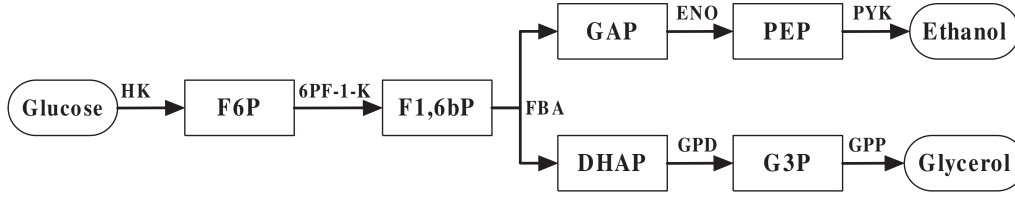


Fig. 1. A synthetic metabolic network system [44]. Abbreviations: F6P, fructose 6-phosphate; F1,6bP, fructose 1,6-bisphosphate; F2,6bP, fructose 2,6-bisphosphate; GAP, glyceraldehyde 3-phosphate; PEP, phosphoenolpyruvate; DHAP, dihydroxyacetone phosphate; G3P, glycerol 3-phosphate. Here 'HK' represents the lumped reactions catalyzed by hexokinase and phosphofructoisomerase; 'ENO' represents the lumped reactions catalyzed by glyceraldehyde dehydrogenase, phosphoglycerate kinase, phosphoglycerate mutase and enolase; 'PYK' represents the lumped reactions catalyzed by pyruvate kinase, pyruvate decarboxylase and alcohol dehydrogenase. 'GPD' represents the two glycerol 3-phosphate dehydrogenase isoenzymes encoded by GPD1 and GPD2; 'GPP' represents the two glycerol 3-phosphatase isoenzymes encoded by GPP1 and GPP2.

$$\begin{bmatrix} -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.5 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & 0.5 & 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

Considering the system observations and the unavoidable noise with system measurements and uncertainties with the system, a nonlinear state space model about the concentrations $\mathbf{c}$ can be formed with the knowledge of Michaelis-Menton equations involved with each reaction,

$$\begin{aligned} \dot{\mathbf{c}} &= f(\mathbf{c}) + \varepsilon \\ \mathbf{y} &= \mathbf{H}\mathbf{c} + \eta \end{aligned} \quad (9)$$

with $\mathbf{c} \in \mathcal{R}^{n_c}$, $\mathbf{y} \in \mathcal{R}^{n_y}$ is the system measurement, n_c and n_y are the dimension of metabolites and measurements, respectively. Since there are three extracellular metabolites, n_y is 3 and n_c is 9 in this case. $\mathbf{H} \in \mathcal{R}^{\mathbf{n_y} \times \mathbf{n_c}}$ is the system observation matrix. Here, H is in the format given below:

$$\mathbf{H} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

ε and η are the noise associated with metabolic reactions and system measurements, respectively. Here

$$f(\mathbf{c}) = \mathbf{S}\mathbf{v}$$

However, due to limited experimental condition, only discrete states are measurable. Hence, the approximated discrete state space model is used here,

$$\begin{aligned} \mathbf{c_k} &= Tf(\mathbf{c_{k-1}}) + \mathbf{c_{k-1}} + T\varepsilon_\mathbf{k} \\ \mathbf{y_k} &= \mathbf{H}\mathbf{c_k} + \eta_\mathbf{k} \end{aligned} \quad (10)$$

where T is the sampling time.

B. SMC filter for intracellular concentration estimation

It is assumed that the noise models $\varepsilon_\mathbf{k}$ and $\eta_\mathbf{k}$ in (10) are truncated Gaussian noise with $\varepsilon_k \propto \bar{\mathcal{N}}(\mathbf{0}, \mathbf{\Sigma_c})$ and $\eta_k \propto \bar{\mathcal{N}}(\mathbf{0}, \mathbf{\Sigma_y})$, where $\mathbf{\Sigma_c}$ and $\mathbf{\Sigma_y}$ are covariance matrices associated with $\varepsilon_\mathbf{k}$ and $\eta_\mathbf{k}$, respectively, $\mathbf{0}$ is the vector filled with 0 with appropriate sizes. For a truncated Gaussian $\bar{\mathcal{N}}(\mu, \beta)$ which only allows nonnegative samples, its probability density format g(s) is,

$$g(s) = \begin{cases} \frac{1}{G_c\sqrt{2\pi\beta}} \exp\left(-\frac{(s-\mu)^2}{2\beta}\right) & s \geq 0 \\ 0 & s < 0 \end{cases}$$

and G_c is in the following format,

$$G_c = \int_0^{+\infty} \frac{1}{\sqrt{2\pi\beta}} \exp\left(-\frac{(s-\mu)^2}{2\beta}\right) \mathrm{d}s$$

Note that the assumption of noise models as truncated Gaussian noise is reasonable since all metabolite concentrations and measurements are nonnegative quantities. Though various proposal distributions are applicable for the SMC filter, here the prior distribution of $\mathbf{c_k}$: $p(\mathbf{c}_k|\mathbf{c}_{k-1})$ is adopted as the proposal distribution, since it can be easily obtained from (10),

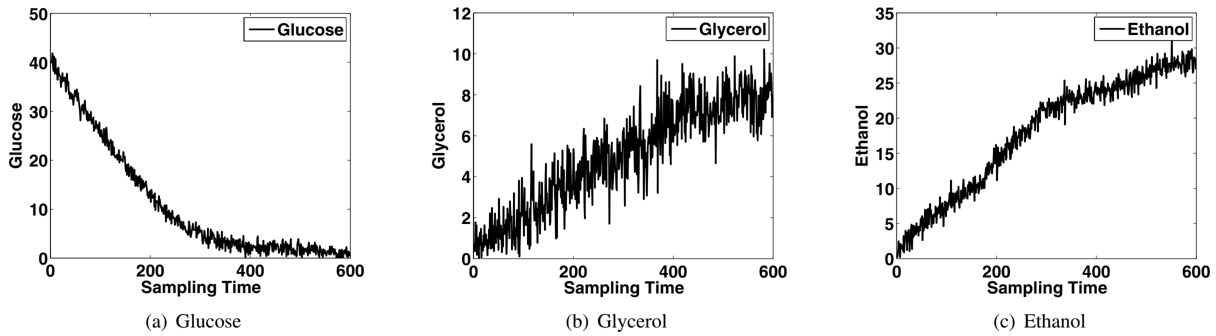
$$p(\mathbf{c}_k|\mathbf{c}_{k-1}) \propto \bar{\mathcal{N}}(Tf(\mathbf{c}_{k-1}) + \mathbf{c}_{k-1}, T^2\mathbf{\Sigma_c})$$

where $\bar{\mathcal{N}}(Tf(\mathbf{c}_{k-1}) + \mathbf{c}_{k-1}, T^2\mathbf{\Sigma_c})$ is a truncated Gaussian distribution with mean $Tf(\mathbf{c}_{k-1}) + \mathbf{c}_{k-1}$ and covariance $T^2\mathbf{\Sigma_c}$. The importance weight w at step k is then given by,

$$w^k = w^{k-1}\frac{p(\mathbf{y}_k|\mathbf{c}_k)p(\mathbf{c}_k|\mathbf{c}_{k-1})}{q(\mathbf{c}_k|\mathbf{c}_{0:k-1}, \mathbf{y}_{1:k})} = w^{k-1}p(\mathbf{y}_k|\mathbf{c}_k) \quad (11)$$

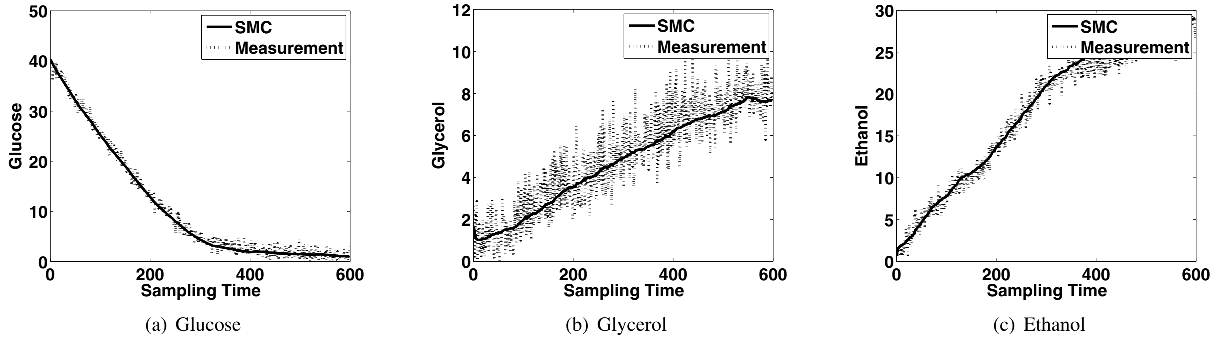
However, in practical situations, it is often the case that a few samples will dominate the weights after several iterations, which greatly affects the diversity required by the SMC algorithm. Hence, when the the sample variance reaches a predefined threshold, resampling is necessary in a typical sequential importance sampling algorithm in order to increase sample diversity. The sample diversity is often measured by sample size S [45], where the sample size at k_{th} step, S^k is,

$$S^k = \frac{1}{\sum_{i=1}^N {w_i^k}^2}$$



(a) Glucose (b) Glycerol (c) Ethanol

Fig. 2. The synthetic measurement data of the extracellular metabolites: Glucose, Glycerol and Ethanol



(a) Glucose (b) Glycerol (c) Ethanol

Fig. 3. Comparison of the measurements and the SMC results of the extracellular metabolites: Glucose, Glycerol and Ethanol. The solid lines (–) represent the estimates from SMC filter and the dotted lines ($\cdots$) represent the simulation data.

TABLE II

THE SMC FILTER ALGORITHM FOR DYNAMIC STATE ESTIMATION

$k = 0$;
Generate N random samples $\mathbf{c}_k^i, i = 1, \cdots, N$;
Set the initial weight $w_k^i = \frac{1}{N}$;
for $k = 1, \cdots, M$
  Generate the new samples $\mathbf{c}_k^i, i = 1, \cdots, N$
  from the proposal distribution $p(\mathbf{c}_k^i | \mathbf{c}_{k-1}^i)$;
  Update the weight $w_k^i = w_{k-1}^i p(\mathbf{y}_k | \mathbf{x}_k^i)$;
  Normalize the weight $\bar{w}_k^i = \frac{w_k^i}{\sum_{i=1}^N w_k^i}$;
  Calculate the sample size $S = \frac{1}{\sum_{i=1}^N {w_k^i}^2}$;
  if sample size S is less than predefined threshold
    Do resampling on the samples $\mathbf{c}_k^i, i = 1, \cdots, N$;
    Set weight $w_k^i = \frac{1}{N}$;
  end
The estimated state is $\hat{\mathbf{c}}_k = \sum_{i=1}^N \mathbf{c}_k^i \bar{w}_k^i$
end

The full algorithm for the SMC filter with resampling procedure is shown in Table II.

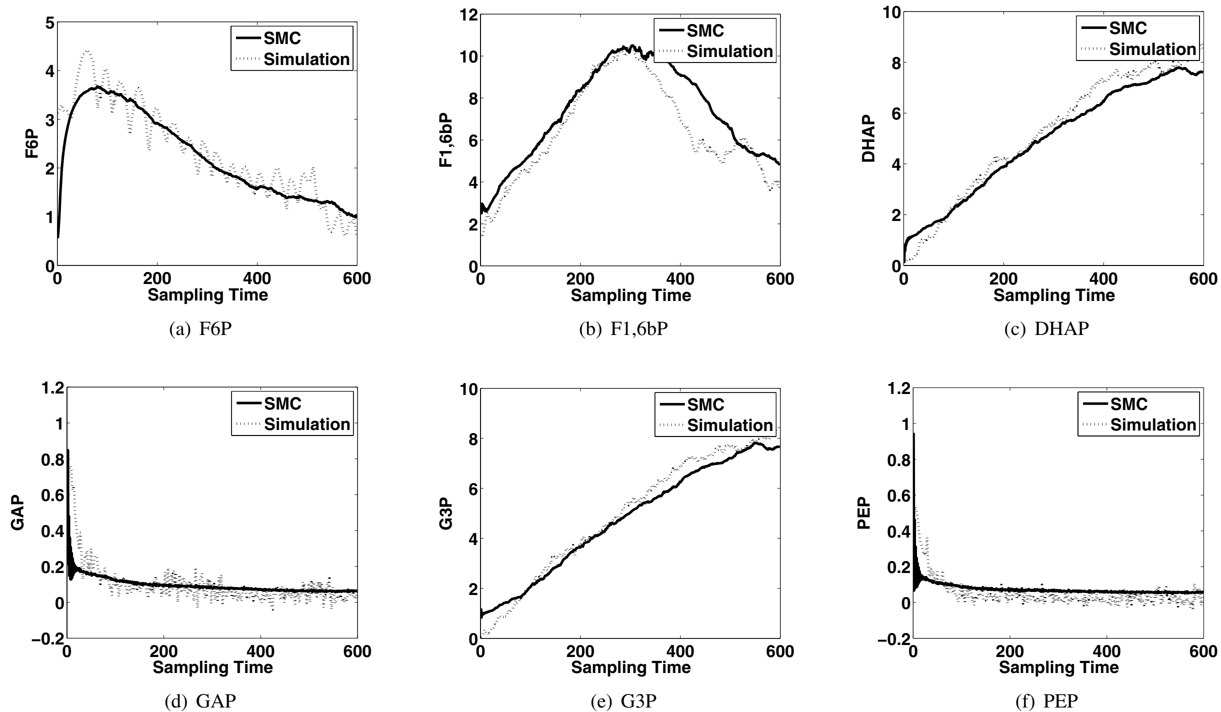
IV. SIMULATION

A. Simulation procedure

In this simulation, the ode45 command in Matlab (The MathWorks) is applied to the system in Fig. 1 described in Section III in order to generate the simulation data. Firstly, the substrate Glucose and the two products Glycerol and Ethanol are fixed at 20.0, 0.0, 0.0 mM, respectively. The initial concentrations for all other intracellular metabolites are set to 1.0 mM. After the system reaches its steady state, the concentrations for F6P, F1,6bP, DHAP, G3P, GAP and PEP are 3.12, 1.48, 0.15, 0.07, 0.62 and 0.49 mM, respectively. Then the Glucose concentration is increased to 40.0 mM in order to simulate the pulse input experiment. Afterwards, the simulation starts and lasts for 120 seconds. Note that the simulation data is already corrupted with noise according to (10). The metabolite F2,6bP is fixed at 1.0mM among the simulation.

B. Data preparation and parameter setting

After the simulation data in 120 seconds are generated, the extracellular measurements obtained are corrupted with additive truncated Gaussian noise with mean 0 and standard deviation 1.0, so as to generate the simulated measurement data. The initial condition for extracellular metabolites are also



(a) F6P (b) F1,6bP (c) DHAP

(d) GAP (e) G3P (f) PEP

Fig. 4. The measurement and simulation results comparison of the intracellular metabolites: F6P, F1,6bP, DHAP, GAP, G3P and PEP. The solid lines (-) represent the estimates from SMC filter and the dotted lines (·) represent the noise-free simulated data.

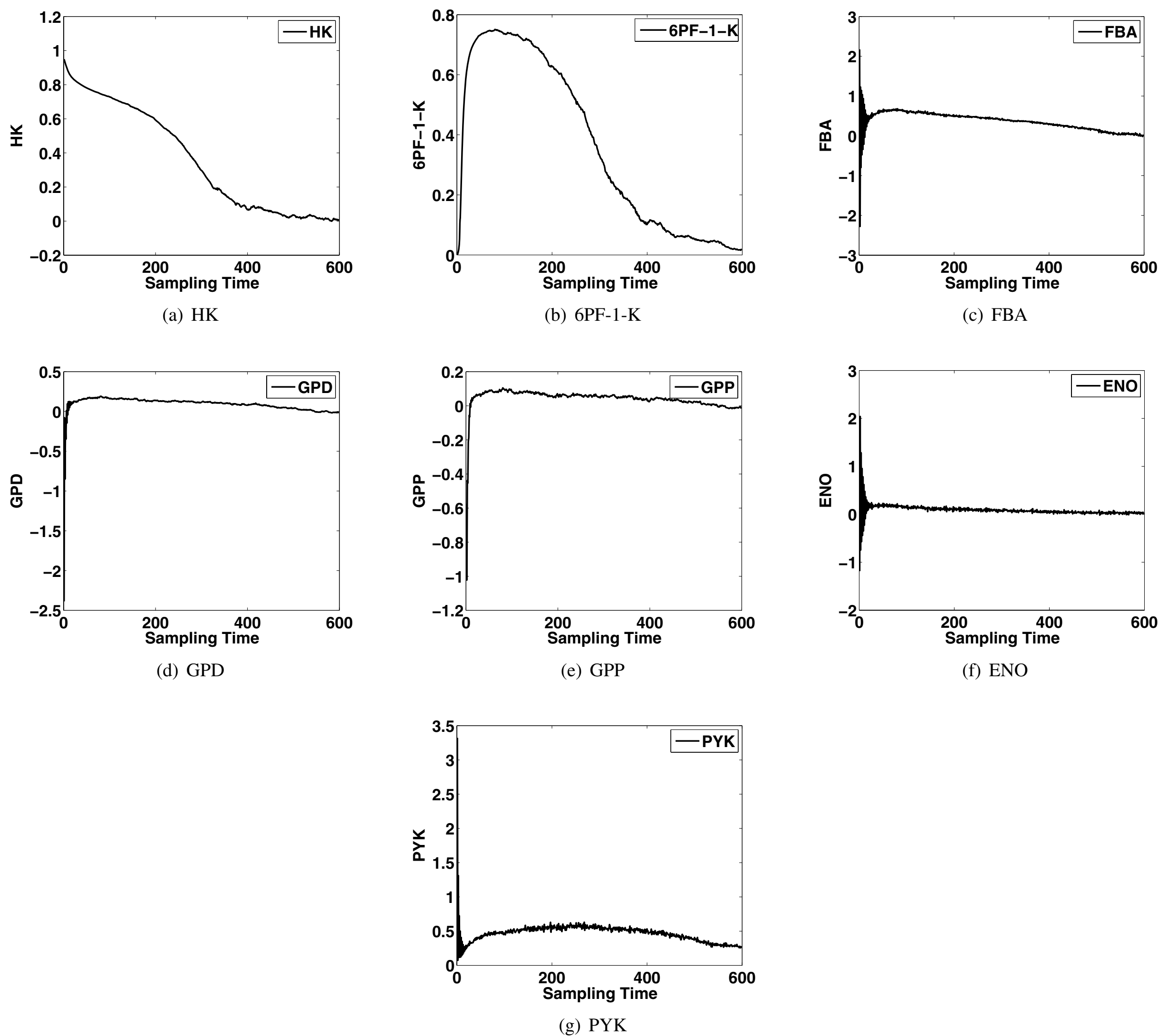
corrupted with the same noise distribution. And the initial condition for intracellular metabolites are assumed unknown and set to reasonable random numbers. The synthetic measurement data are shown in Fig. 2. Since the sampling frequency of a rapid sampling experiment is usually around 4-5 samples per second, here we assume the sampling time is 0.2s, hence, there are five samples to be obtained every second. In reality, the true values of $\mathbf{\Sigma_c}$ and $\mathbf{\Sigma_y}$ in (10) are unknown and are considered as design parameters for the filter. Hence, the covariance matrices $\mathbf{\Sigma_c}$ and $\mathbf{\Sigma_y}$ in the filter are assigned to diagonal matrices with 0.4 and 2.0 on the diagonal, respectively. By trial and error, the sample number is set as 2000. The threshold for resampling is set as one third of the sampling number.

C. Simulation results

The original data and the estimated extracellular and intracellular metabolite concentration data are shown in Fig. 3 and Fig. 4, respectively.

From the result illustration in Fig. 3 and Fig. 4, it is clear that the SMC filter approximates both the extracellular and the intracellular metabolite concentration in most cases. The comparison of the noisy measurements and the SMC results of these three extracellular metabolites, Glucose, Glycerol and Ethanol in Fig. 3 illustrates that the SMC algorithm can reconstruct the original data of the extracellular measurements even though there is large noise with these measurements. Since the metabolite F1,6bP is a critical metabolite, it is apparent that its estimation is more complicated than others. However, the estimation of F1,6bP still follows the original data trend reasonably well. An interesting finding is that the metabolites on GAP, PEP side both stay at lower quantities, however, the product Ethanol generated from PEP increases greatly during the simulation. On the contrary, the DHAP and G3P both increase gradually during the simulation, with around the same speed as the product Glycerol produced by G3P.

Besides the metabolite concentration estimation data, the flux estimation data are also obtained simultaneously. Here, the flux samples are derived by substituting the metabolite concentration data in the Michaelis-Menton equations with metabolite samples received during SMC filter. Then the weights for metabolite samples are utilized as the weights for those flux samples. The estimated fluxes HK, 6PF-1-K, FBA, GPD, GPP, ENO and PYK are illustrated in Fig. 5. It seems that the fluxes undergo various changes in the simulation. The four fluxes, GPD, GPP, ENO and PYK, keep fluctuating around nearly the same central value. The fluxes HK and 6PF-1-K, on the contrary, experience big decrease from around 0.8 to nearly 0.0, The flux FBA, which is affected by F1,6bP, DHAP and GAP, keeps decreasing slowly. Although the flux FBA at steady state is the same as HK and 6PF-1-K by stoichiometric consideration, this study illustrates that its variation at transient state undergoes the route very different from that of the fluxes HK and 6PF-1-K.



(a) HK (b) 6PF-1-K (c) FBA

(d) GPD (e) GPP (f) ENO

(g) PYK

Fig. 5. The fluxes estimated by SMC filter.

V. CONCLUSIONS

The dynamics of a metabolic network before it reaches its steady state has long been of vital importance to comprehensively investigate system properties. Rapid sampling and quenching has become a critical technology in measuring the fast metabolite concentration variations during the transient state. However, the drawbacks of rapid sampling are obvious when considering the instrument limitation and the effort required for data analysis. In this paper, the dynamics of a metabolic network is formulated as a state space model with available knowledge of Michaelis-Menton equations. The SMC approach, which intends to estimate the latent states from available measurements is applied to the metabolic system dynamics analysis. It is illustrated that SMC approach is able to estimate the dynamic intracellular metabolite concentration data with the prior knowledge of noisy extracellular metabolite data. Meanwhile, the algorithm can also reconstruct the original measurements from their noisy counterparts. Moreover, the dynamic fluxes are also obtained simultaneously during the estimation of metabolite states. Their formats among the simulation period show that the stoichiometric balances are no longer applicable in transient states. Though only a simple example is utilized in this paper, the proposed approach can be applied to more complex metabolic network as long as system observability is satisfied.

ACKNOWLEDGMENT

Ms. Yang would like to thank The University of Sheffield and its Department of Automatic Control and Systems Engineering for the financial support to carry out this research

work. Dr. Kadirkamanathan and Prof. Billings would like to thank the EPSRC for the support.

REFERENCES

[1] J. E. Bailey, "Toward a science of metabolic engineering," *Science*, vol. 252, pp. 1668–1675, 1991.

[2] G. N. Stephanopoulos, A. A. Aristidou, and J. Nielsen, *Metabolic Engineering – Principles and Methodologies*. San Diego, USA: Academic Press, 1998.

[3] C. H. Schilling, S. Schuster, B. O. Palsson, and R. Heinrich, "Metabolic pathway analysis : basic concepts and scientific applications in the post-genomic era," *Biotechnol. Prog.*, vol. 15, pp. 296–303, 1999.

[4] J. A. Papin, N. D. Price, S. J. Wiback, D. A. Fell, and B. O. Palsson, "Metabolic pathways in the post-genomic era," *Trends in Biochemical Sciences*, vol. 28, pp. 250–258, 2003.

[5] A. Varma and B. O. Palsson, "Metabolic flux balancing: basic concepts, scientific and practical use," *Bio/Technology.*, vol. 12, pp. 994–998, 1994.

[6] A. Marx, A. A. de Graaf, W. Wiechert, L. Eggeling, and H. Sahm, "Determination of the fluxes in the central metabolism of *corynebacterium glutamicum* by nuclear magnetic resonance spectroscopy combined with metabolite balancing," *Biotechnol. Bioeng.*, vol. 49, pp. 111–129, 1996.

[7] K. Schmidt, J. Nielsen, and J. Villadsen, "Quantitative analysis of metabolic fluxes in *Escherichia coli* using two-dimensional NMR spectroscopy and complete isotopomer models," *J. Biotechnol.*, vol. 71, no. 1-3, pp. 175–189, 1999.

[8] K. Schmidt, L. C. Nørregaard, B. Pedersen, A. Meissner, J. O. Duus, J. O. Nielsen, and J. Villadsen, "Quantification of intracellular metabolic fluxes from fractional enrichment and ^{13}C-^{13}C coupling constraints on the isotopomer distribution in labeled biomass components," *Metab. Eng.*, vol. 1, no. 2, pp. 166–179, 1999.

[9] W. Wiechert and A. A. de Graaf, "Bidirectional reaction steps in metabolic networks: I. modeling and simulation of carbon isotope labeling experiments," *Biotechnol. Bioeng.*, vol. 55, no. 1, pp. 101–117, 1997.

[10] W. Wiechert, C. Siefke, A. A. de Graaf, and A. Marx, "Bidirectional reaction steps in metabolic networks: II. flux estimation and statistical analysis," *Biotechnol. Bioeng.*, vol. 55, no. 1, pp. 118–135, 1997.

[11] H. Kacser and J. A. Burns, "The control of flux," *Symp. Soc. Exp. Biol.*, vol. 27, pp. 65–104, 1973.

[12] H. Kacser, J. A. Burns, and D. A. Fell, "The control of flux," *Bioch. Soc. Trans.*, vol. 23, pp. 341–366, 1995.

[13] D. A. Fell, *Understanding the Control of Metabolism*. Miami: Portland Press, 1997.

[14] H. V. Westerhoff, "The silicon cell, not dead but live!" *Metab. Eng.*, vol. 3, pp. 207–210, 2001.

[15] O. Fiehn, "Metabolomics - the link between phenotypes and genotypes," *Plant Mol. Biol.*, vol. 48, pp. 155–171, 2002.

[16] O. Fiehn, J. Kopka, P. Dormann, T. Altmann, R. N. Trethewey, and L. Willmitzer, "Metabolite profiling for plant functional genomics," *Nat. Biotechnol.*, vol. 18, pp. 1142–1143, 2000.

[17] D. Weuster-Botz and A. A. de Graaf, "Reaction engineering methods to study intracellular metabolite concentrations," *Adv. Biochem. Eng. Biotechnol.*, vol. 54, pp. 75–108, 1996.

[18] U. Theobald, W. Mailinger, M. Baltes, M. Rizzi, and M. Reuss, "In-vivo analysis of metabolic dynamics in saccharomyces cerevisiae: I. experimental observations," *Biotechnol. Bioeng.*, vol. 55, pp. 305–316, 1997.

[19] A. Buchholz, R. Takors, and C. Wandrey, "Quantification of intracellular metabolites in *Escherichia coli* K12 using liquid chromatographic-electrospray ionization tandem mass spectromtric techniques," *Anal. Biochem.*, vol. 295, pp. 129–137, 2001.

[20] J. Hurlebaus, A. Buchholz, W. Alt, W. Wiechert, and R. Takors, "MMT - a pathway modeling tool for data from rapid sampling experiments," in silico *biol.*, vol. 2, pp. 467–484, 2002.

[21] A. Buchholz, J. Hurlebaus, C. Wandrey, and R. Takors, "Metabolomics: Quantification of intracellular metabolite dynamics," *Biomol. Eng.*, vol. 19, pp. 5–15, 2002.

[22] J. B. Russell and G. M. Cook, "Energetics of bacterial growth: balance of anabolic and catabolic reactions," *Microbiol. Rev.*, vol. 59, pp. 48–62, 1995.

[23] B. D. Sanwal, "Allosteric controls of amphilbolic pathways in bacteria," *Bacteriol Rev.*, vol. 34, pp. 20–39, 1970.

[24] S. Vaseghi, A. Baumeister, M. Rizzi, and M. Reuss, "*in vivo* dynamics of the pentose phosphate pathway in *Saccharomyces cerevisiae*," *Metabolic Engineering*, vol. 1, pp. 128–140, 1999.

[25] S. Ostergaard, L. Olsson, and J. Nielsen, "*in vivo* dynamics of galactose metabolism in saccharomyces cerevisiae: metabolic fluxes and metabolite levels," *Biotechnol. Bioeng.*, vol. 73, pp. 412–425, 2001.

[26] M. Rizzi, M. Baltes, U. Theobald, and M. Reuss, "*in vivo* analysis of metabolic dynamics in saccharomyces cerevisiae: Ii. mathematical model," *Biotechnol. Bioeng.*, vol. 55, pp. 592–608, 2001.

[27] D. Visser, G. A. van Zuylen, J. C. van Dam, M. R. Eman, A. Pröll, C. Ras, L. Wu, W. M. van Gulik, and J. J. Heijnen, "Analysis of in vivo kinetics of glycolysis in aerobic *Saccharomyces cerevisiae* by application of glucose and ethanol pulses," *Biotechnology and Bioengineering*, vol. 88, pp. 157–167, 2004.

[28] M. West and P. J. Harrison, *Bayesian forecasting and dynamic models, 2nd edn.* New York: Springer, 1997.

[29] R. E. Kalman, "A new approach to linear filtering and prediction problems," *J. Basic Eng. - T. ASME.*, vol. 82, pp. 35–45, 1960.

[30] R. E. Kalman and B. R. S., "New results in linear filtering and prediction theory," *J. Basic Eng. - T. ASME.*, vol. 83, pp. 95–107, 1961.

[31] S. Arulampalam, S. Maskell, N. Gordon, and T. Clapp, "A tutorial on particle filters for on-line nonlinear/non-Gaussian Bayesian tracking," *IEEE Trans. Signal Process.*, vol. 50, pp. 174–188, 2002.

[32] A. Doucet, N. de Freitas, and N. Gordon, Eds., *Sequential Monte Carlo Methods in Practice*. Springer-Verlag, 2001.

[33] J. Liu and R. Chen, "Sequential Monte Carlo methods for dynamic systems," *J. Amer. Statist. Assoc.*, vol. 93, pp. 1032–1044, 1998.

[34] M. K. Pitt and N. Shephard, "Filtering via simulation: Auxiliary particle filters," *J. Amer. Statist. Assoc.*, vol. 94, pp. 590–599, 1999.

[35] N. Metropolis and S. Ulam, "The monte carlo method," *J. Amer. Statist. Assoc.*, vol. 44, pp. 335–341, 1949.

[36] A. Doucet, S. Godsill, and C. Andrieu, "On sequential Monte Carlo sampling methods for Bayesian filtering," *Statistics and Computing*, vol. 10, pp. 197–208, 2000.

[37] G. B. Fogel and D. W. Corne, Eds., *Evolutionary computation in bioinformatics*. Morgan Kaufmann, 2003.

[38] J. Yang, S. Wongsa, V. Kadirkamanathan, S. A. Billings, and P. C. Wright, "Metabolic flux estimation - a self-adaptive evolutionary algorithm with singular value decomposition,," *IEEE/ACM Trans. Computational biology and Bioinformatics*, in press.

[39] ——, "Self-adaptive evolutionary algorithm based methods for quantification in metabolic systems," in *2004 IEEE Symp. on Computational Intelligence in Bioinformatics and Computational Biology*, 2004, pp. 260–267.

[40] A. Krogh, M. Brown, I. S. Mian, K. Sjolander, and D. Haussler, "Hidden markov models in computational biology: Applications to protein modeling, Tech. Rep. UCSC-CRL-93-32, 1993.

[41] I. Tsigelny, Y. Sharikov, and L. F. T. Eyck, "Hidden markov models-based system (HMMSPECTR) for detecting structural homologies on the basis of sequential information," *Protein Eng. Des. Sel.*, vol. 15, no. 5, pp. 347–352, 2002.

[42] O. Ebenhöh and R. Heinrich, "Evolutionary optimization of metabolic pathways. theoretical reconstruction of the stoichiometry of atp and nadh producing systems," *Bull. Math. Biol.*, no. 63, pp. 21–55, 2001.

[43] K. Schmidt and S. H. Issacs, "An evolutionary algorithm for initial state and parameter estimation in complex biochemical models," in *Preprints of the 6th Int. Conf. on computer applications in biotechnology*, Garmisch-Partenkirchen, Germany, 1995.

[44] L. Wu, W. A. van Winden, W. M. van Gulik, and J. J. Heijnen, "Application of metabolome data in functional genomics: A conceptual strategy," *Metab. Eng.*, vol. 7, pp. 302–310, 2005.

[45] A. Kong, J. S. Liu, and W. H. Wong, "Sequential imputations and bayesian missing data problems," *Journal of American Statistical Association*, vol. 89, pp. 278–288, 1994.