2011 50th IEEE Conference on Decision and Control and
European Control Conference (CDC-ECC)
Orlando, FL, USA, December 12-15, 2011

Parameter Estimation of Biological Phenomena Modeled by S-systems: An Extended Kalman Filter Approach

N. Meskin, *Member, IEEE,* H. Nounou, *Senior Member, IEEE,* M. Nounou, *Senior Member, IEEE,*
A. Datta, *Fellow Member, IEEE,* and E. R. Dougherty, *Senior Member, IEEE*

***Abstract*— Recent advances in high-throughput technologies for biological data acquisition have spurred a broad interest in the development of mathematical models for biological phenomena. S-systems, which offer a good compromise between accuracy and mathematical flexibility, are a promising framework for modeling the dynamical behavior of genetic regulatory networks (GRNs), as well as that of biochemical pathways. In the S-system modeling framework, the number of unknown parameters is much more than the number of metabolites and this makes the parameter estimation task a challenging one. In this paper, a new parameter estimation algorithm is developed based on the Extended Kalman filter (EKF) approach. It is first shown that the conventional EKF approach is not capable of estimating the unknown parameters of S-systems. To remedy this problem, a new iterative extended Kalman Filtering algorithm is developed in which the EKF algorithm is applied iteratively to the available noisy time profiles of the metabolites. The proposed estimation algorithm is applied to a generic branched pathway and the Cad system of *E.coli*. The simulation results demonstrate the effectiveness of the proposed scheme.**

I. Introduction

One of the main research activities in modern molecular biology is to develop mathematical models that can represent the structure and dynamics of biological pathways or genetic regulatory networks [1], [2]. Recent advances in measurement technologies such as mass spectrometry, nuclear magnetic resonance, or protein kinase phosphorylation have provided a wealth of comprehensive time profiles of metabolites or proteins that can be used for biochemical pathway modeling and proteomics. These metabolite profiles are simultaneous measurements of biochemicals which can be obtained as single snapshots or as a sequence of snapshots [3]. For example, *in vivo* nuclear magnetic resonance (NMR) measurements can provide dense time-sequences of a few metabolites for a few seconds or minutes [4], [5]. In order to mathematically represent this information, it is first required to specify a mathematical modeling framework, and then to develop computational methods to fit the measured information to the selected modeling framework. In this paper, we develop an iterative parameter estimation method based on the Extended Kalman filter (EKF) algorithm to estimate the parameters of the S-system model representing the biological phenomena.

S-systems are proposed in [6] as a canonical nonlinear model to capture the dynamical behavior of a large class of biochemical pathways. They are characterized by a good trade-off between accuracy and mathematical flexibility [7]. In this modeling approach, nonlinear systems are approximated by products of power-law functions, which are derived from multivariate linearization in logarithmic coordinates. It has been shown that this type of representation is a valid description of biological processes in a variety of settings. For a more elaborate discussion of S-systems, the interested reader is referred to [6], [8], [9].

The problem of estimating the S-system model parameters has been addressed by several researchers [10]–[18] using optimization approaches such as alternating regression, simulated annealing, Newton-flow, cooperative coevolutionary algorithm and genetic algorithm. These approaches assume the availability of noise-free data. Only in the cases of [12], [16] and [17], even though the developed approaches do not account for the presence of noise, the effect of noise on the performance of these approaches has been investigated. In addition, these three approaches require several data sets with different initial conditions for estimating S-system parameters. Recently, an approach that combines genetic algorithms and *linear* Kalman filtering has been proposed in [19] for parameter estimation in gene regulatory networks modeled by polynomial systems.

In this paper, we adopt a recently developed stochastic estimation algorithm for nonlinear systems, namely the extended Kalman filter. To utilize the EKF algorithm for parameter estimation of nonlinear dynamical systems, unknown parameters of the system are modeled as a stationary process with identity state transition that is driven by a process noise [20], and the nonlinear system is considered as a nonlinear map parameterized by the vector of unknown parameters. Then, the EKF algorithm estimates the parameters using the available measurements. In order to apply the EKF algorithm for parameter estimation of biological phenomena modeled using the S-system framework, it is assumed the *a priori* knowledge about the metabolic pathways and experimentally measured time profiles of the metabolites are available. From the metabolic pathways, one can identify the structure of the

N. Meskin is with the Electrical Engineering Department, Qatar University, Doha, Qatar (email: nader.meskin@qu.edu.qa).

H. Nounou is with the Electrical and Computer Engineering Program, Texas A&M University at Qatar, Doha, Qatar (email: hazem.nounou@qatar.tamu.edu).

M. Nounou is with the Chemical Engineering Program, Texas A&M University at Qatar, Doha, Qatar (email: mohamed.nounou@qatar.tamu.edu).

A. Datta is with the Department of Electrical and Computer Engineering, Texas A&M University, College Station, TX, 77843, USA (email: datta@ece.tamu.edu).

E. R. Dougherty is with the Department of Electrical and Computer Engineering, Texas A&M University, College Station, TX 77843 USA, and also with the Translational Genomics Research Institute, Phoenix, AZ 85004 USA (email: edward@ece.tamu.edu).

978-1-61284-799-3/11/$26.00 ©2011 IEEE

S-system [8]. Then, the EKF algorithm can be utilized to estimate the unknown parameters in the identified S-system.

Due to the fact that the number of unknown parameters is much more than the number of metabolites, the conventional EKF algorithm is not capable of estimating the unknown parameters of S-systems. This motivated us to develop a new iterative EKF (IEKF) algorithm, with a resetting feature, which can be utilized for parameter estimation within the S-system framework. In the proposed IEKF algorithm, the EKF algorithm is applied iteratively to the available time profiles of the metabolites concentration. The effectiveness of the proposed scheme is demonstrated using two case studies in which the unknown parameters of a generic branched pathway and the Cad system of *E.coli* are estimated.

The main advantages of the IEKF algorithm with respect to the aforementioned methods in the literature, which are used to estimate the S-system model parameters, are **1)** many biological systems have stochastic features and the IEKF is based on the stochastic estimation framework, so that stochastic characteristics of the system and the measurement noise can be incorporated in the estimation algorithm; **2)** due to the iterative feature of the IEKF, only one data set is sufficient for estimating the parameters of the given S-system while the proposed methods in [12] and [16] require several data sets, and **3)** the proposed IEKF has the potential to be extended to deal with the case of partial measurements where some of the metabolites concentrations are not measured due to the difficulty or the cost associated with their measurements.

II. Problem Formulation

Suppose that we are interested in identifying a mathematical model of a given biological phenomenon in which we have **1)** *a priori* knowledge about the metabolic pathways, and **2)** experimentally measured time profiles of the metabolites. One possible solution to this model identification problem is to obtain an S-system representation of the given biological system. This problem can be solved in two steps. First, using the metabolic pathways, one can obtain the structure of the S-system, which represents the biological phenomenon of interest following the well-known steps developed in the literature ([8], Chapter 3). Then, the measured time series of the metabolites are used to estimate the unknown parameters in the S-system. In this paper, we adopt the extended Kalman filtering approach for parameter estimation.

Consider the following S-system dynamics,

$$\dot{x}_i = \alpha_i \prod_{j=1}^{N+m} x_j^{g_{ij}} - \beta_i \prod_{j=1}^{N+m} x_j^{h_{ij}}, \qquad i = 1, 2, ..., N \quad (1)$$

where $\alpha_i > 0$ and $\beta_i > 0$ are rate coefficients and g_{ij} and h_{ij} are kinetic orders and there exist $N + m$ variables x_i (genes/metabolites concentration), where the first N variables are the actual state of the system and the remaining m variables are independent. It is assumed that rate coefficients and kinetic orders are unknown. It is assumed that all the variables $x_i, i = 1, ..., N + m$ are measured. Let us denote $w = [\alpha_1, ..., \alpha_N, \beta_1, ..., \beta_N, g_{11}, ..., g_{N,N+m}, h_{11}, ..., h_{N,N+m}]^T$ as the unknown parameter vector and $u = [x_{N+1}, ..., x_{N+m}]^T$ as the vector of independent variables. The above S-system can be written as,

$$\dot{x} = f(x, u, w) \quad (2)$$

where $x = [x_1, ..., x_N]^T \in \mathbb{R}^N$, $u \in \mathbb{R}^m$ and $w \in \mathbb{R}^q$ with $q = N(2 + 2(N + m))$ as the number of unknown parameters. It is assumed that the unknown parameter vector w belongs to a pre-specified set $\mathcal{X}_w \subset \mathbb{R}^q$, which can be obtained using any *a priori* knowledge about the range of each of the parameters. Using the Euler approximation, we can write the discrete time representation of equation (2) as,

$$x[k + 1] = F(x[k], u[k], w) \quad (3)$$

where $F(x[k], u[k], w) = x[k]+T_s f(x[k], u[k], w)$, T_s is the sampling time, $x[k]$ is defined as the sampled continuous-time state $x(kT_s)$ and $u[k] := u(kT_s)$.

A. S-system Structure Identification

Using the metabolic pathways, one can specify the kinetic orders that are zero. For instance, if x_j does not directly affect x_i, the corresponding kinetic orders g_{ij} and h_{ij} are zero. Moreover, based on experience [11], the kinetic orders g_{ii} are set to zero to omit a direct reinforcing effect of a metabolite on its own production. For example, consider the generic branched pathway (shown in Figure 1) which has four dependent variables $x_1, ..., x_4$ and one independent variable x_5 [11]. As shown in this figure, the production of x_1 depends on the independent variable x_5 with an inhibition effect exerted by x_3. Hence, the kinetic orders g_{13}, g_{15} and h_{11} are only non-zero. Similarly, one can determine for each metabolite which kinetic orders are non-zero and which are negative such as g_{13}. For the branched pathway shown in Figure 1, the following S-system can be obtained [11]:

$$\begin{aligned} \dot{x}_1 &= \alpha_1 x_3^{g_{13}} x_5^{g_{15}} - \beta_1 x_1^{h_{11}} \\ \dot{x}_2 &= \alpha_2 x_1^{g_{21}} - \beta_2 x_2^{h_{22}} \\ \dot{x}_3 &= \alpha_3 x_2^{g_{32}} - \beta_3 x_3^{h_{33}} x_4^{h_{34}} \\ \dot{x}_4 &= \alpha_4 x_1^{g_{41}} - \beta_4 x_4^{h_{44}} \end{aligned} \quad (4)$$

where $w = [\alpha_1, ..., \alpha_4, \beta_1, ..., \beta_4, g_{13}, g_{15}, g_{21}, g_{32}, g_{41}, h_{11}, h_{22}, h_{33}, h_{34}, h_{44}]$. Based on the fact that the rate coefficients α_i's and β_i's and all the kinetic orders except g_{13} are positive, the parameter pre-specified set is defined as $\mathcal{X}_w = \mathbb{R}_+^8 \times \mathbb{R}_- \times \mathbb{R}_+^9$.

B. Parameter Estimation

Given the measured time series $x[k], u[k]$, $k = 0, ..., n$, we can define $y[k] = x[k + 1]$, $k = 0, ..., n - 1$, and therefore, equation (3) can be rewritten as the following nonlinear mapping [20],

$$y[k] = F(x[k], u[k], w) \quad (5)$$

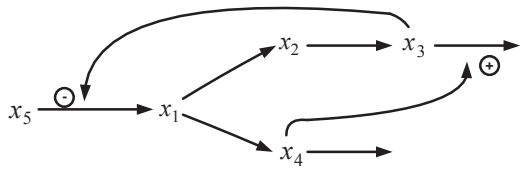

Fig. 1. Generic branched pathway with four dependent variables.

where $x[k]$, $u[k]$ are considered as inputs, $y[k]$ is the output, and the nonlinear map F is parameterized by the vector w. For parameter estimation, a new state-space representation is written as follows:

$$\begin{aligned} w[k+1] &= w[k] + r[k] \\ y[k] &= F(x[k], u[k], w[k]) + e[k] \end{aligned} \tag{6}$$

where the parameters $w[k]$ correspond to a stationary process with identity state transition matrix, driven by process noise $r[k]$. The training data set used in this estimation problem can be written as $\{x[k], u[k], y[k]\}$. In Section III, it will be shown how the EKF algorithm can be utilized to estimate the parameter vector w.

III. Extended Kalman Filter

In this section, the extended Kalman filter is used to estimate the unknown parameters in a given S-system. In EKF algorithm, the state distribution is approximated by a Gaussian random variable and the "first order" linearization of the nonlinear dynamics is used for propagation of state distribution. The EKF parameter estimation equations for system (6) are as follows [20]:

1. Initialization:

$$\begin{aligned} \hat{w}[0] &= \mathbb{E}[w] \\ P_{w[0]} &= \mathbb{E}[(w - \hat{w}[0])(w - \hat{w}[0])^T] \end{aligned}$$

2. Time update

$$\begin{aligned} \hat{w}^-[k] &= \hat{w}[k-1] \\ P^-_{w[k]} &= P_{w[k-1]} + R^r[k-1] \end{aligned}$$

where the innovation covariance $R^r[k-1]$ is set as

$$R^r[k-1] = (\lambda^{-1}_{RLS} - 1) P_{w[k-1]}$$

and $\lambda_{RLS} \in (0, 1]$ is often referred to as the *forgetting factor*.

3. Measurement update equations:

$$\begin{aligned} \mathcal{K}[k] &= P^-_{w[k]} C_k^T (C_k P^-_{w[k]} C_k^T + R^e[k])^{-1} \\ \hat{w}[k] &= Proj\Big[\hat{w}^-[k] + \mathcal{K}[k](y[k] - F(x[k], u[k], \hat{w}^-[k]))\Big] \\ P_{w[k]} &= (I - \mathcal{K}[k] C_k) P^-_{w[k]} \end{aligned}$$

where $Proj[w]$ is the nearest point on the set $\mathcal{X}_w$ to w, $C_k = \frac{\partial F(x,u,w)}{\partial w}|_{\hat{w}[k]}$ and the covariance matrix $R^e[k]$ can be selected as a fixed matrix.

Due to the fact that the number of unknown parameters in the S-system ($q = N(2 + 2(N + m))$) is much more than number of measured states (N), the conventional EKF algorithm is not capable of estimating the parameters of S-systems with too many unknown parameters. This can be due to the lack of global observability of the system (6) when the dimension of w is much more than the dimension of y. To remedy this problem, we propose the following iterative EKF algorithm (IEKF) for estimating the parameters of the S-system.

IEKF Parameter Estimation Algorithm:

1) Set the initial guesses of the parameters and their covariance, i.e., $\hat{w}[0]$ and $P_{w[0]} = P_0$.
2) Set the epoch counter to one, i.e., $i = 1$.
3) Apply the EKF algorithm to the training set $\{x[k], u[k], y[k]\}$ and calculate the last estimate of parameters $\hat{w}[n]$ and its covariance $P_{w[n]}$.
4) Calculate $\hat{y}[k]$ as

$$\begin{aligned} \hat{x}[k+1] &= F(\hat{x}[k], u[k], \hat{w}[n]), \quad k = 0, ..., n-1 \\ \hat{y}[k] &= \hat{x}[k+1] \end{aligned}$$

with initial condition $\hat{x}[0] = x[0]$.
5) Find the root mean square error (RMS) of the generated response of the S-system $\hat{x}[k], k = 0, ..., n-1$ as $e[i] = \sqrt{\sum_{k=0}^{n-1} |y[k] - \hat{y}[k]|^2 / n}$.
6) If $e(i) < \delta_E$ or $i > i_{max}$, then Stop. If not, re-initialize the EKF algorithm with the last estimate of parameters $\hat{w}[n]$, i.e., $\hat{w}[0] = \hat{w}[n]$.
7) If $|e[i] - e[i-1]| < \delta_R$, set the covariance of the parameters to its initial guess (i.e., $P_{w[0]} = P_0$), increase the epoch counter $i = i + 1$, and go to step 3. If not, set $P_{w[0]} = P_{w[n]}$, increase the epoch counter $i = i + 1$, and go to step 3.

Here, δ_E is the desired RMS of the estimation, i_{max} is the maximum number of epochs, and δ_R is the covariance resetting threshold which speeds up the convergence of the algorithm. In order to tighten the resetting criteria after each reset, δ_R can be adjusted as $\delta_R = \frac{\delta_{R0}}{2^j}$, where j is the resetting counter, and is incremented after each resetting.

IV. Case Studies

A. A Generic Branched Pathway

In this section, we demonstrate the efficacy of the iterative EKF parameter estimation algorithm presented in Section III by applying it to the S-system (4) with typical parameter values shown in Table 1. It is assumed that the independent variable x_5 is fixed with a value 1; hence, the kinetic order g_{15} cannot be estimated and, therefore, we set $\alpha_1 = 20 \times 1 = 20$. The total number of unknown parameters q for the S-system (4) is 17. Initial guesses of all the parameters are set to 1 except g_{13} and the initial guess for g_{13} is set to -1 due to the a priori information that it is negative. The initial covariance matrix $P_{w[0]} = 0.1I$. The S-system sampling time is selected as 0.1 second and the S-system (4) is simulated for 5 seconds. Hence, 50 training points are generated for parameter estimation. Two different scenarios, namely noise-free and noisy measurements, are considered here.

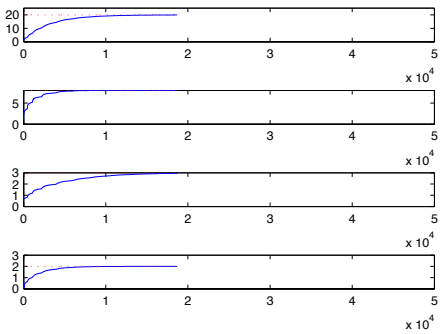

Fig. 2. Estimated parameters for α_i's using IEKF: noise-free measurement (with covariance resetting) .

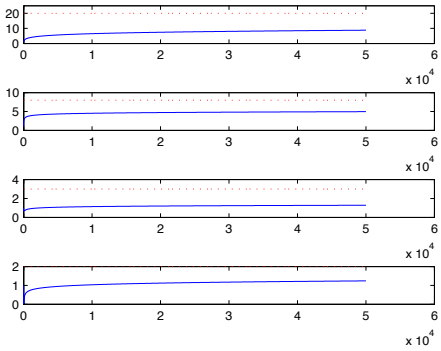

Fig. 3. Estimated parameters for α_i's using EKF: noise-free measurement (without covariance resetting).

1) Noise-Free Measurement:: In this scenario, it is assumed that there exists no noise in the measurement. The RMS and resetting thresholds are selected as $\delta_E = 0.01$ and $\delta_{R0} = 0.001$. The maximum number of iterations is set as $i_{max} = 50000$. Figure 2 depicts the estimated parameters versus epoch for α_i's. The final estimated values of the parameters are summarized in Table I. As shown in this table, the IEKF algorithm is able to perfectly estimate the parameters for noise-free measurement. In order to show the advantage of resetting in the proposed IEKF algorithm, Figure 3 shows the estimated parameters for α_i's without covariance resetting. By comparing Figures 2 and 3, it can be clearly seen that covariance resetting significantly helps speeding up the convergence of estimation.

2) Noisy Measurement: In this scenario, white noise is added to the time profiles of the metabolites. In order to investigate the performance of the proposed estimation algorithm in the presence of measurement noise, two different signal to noise ratio (SNR) values, namely, 20 and 40, are considered. For each measurable metabolite x_i, the signal to noise ratio (SNR_i) is defined as $SNR_i = \frac{\sigma_{x_i}}{\sigma_{n_i}}$ where σ_{x_i} is the variance of metabolite x_i and σ_{n_i} is the variance of the added white noise to the i-th metabolite. The RMS and resetting thresholds are selected as $\delta_E = 0.1$ and

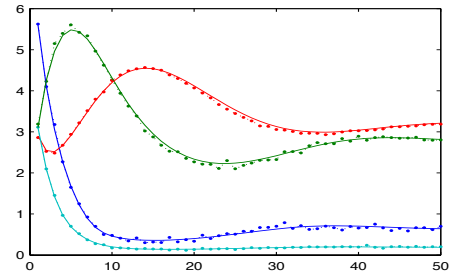

Fig. 4. Response of the estimated S-system (solid line) versus the training set ('.'): SNR 20.

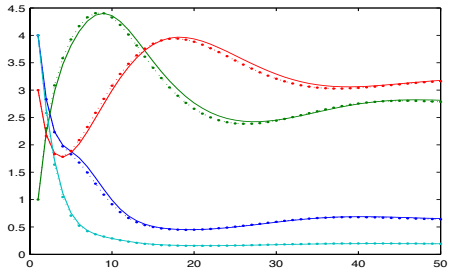

Fig. 5. Response of the estimated S-system (solid line) versus the real one ('.') with initial condition $x_1(0) = 4$, $x_2(0) = 1$, $x_3(0) = 3$, $x_4(0) = 4$: SNR 20.

$\delta_{R0} = 0.001$. The final estimated values of the parameters are summarized in Table I. The final mean square errors between the response of the estimated system and the noise free response of the actual system are also shown in the last column of Table I. As expected, the mean square errors between the response of the estimated system and the noise free response of the actual system increase as the noise level amplifies.

Figure 4 compares the response of the estimated S-system with the training data for the different noise levels. Moreover, Figure 5 depicts the response of the estimated S-system for two different sets of initial conditions and two different SNRs. As shown in these figures, the state trajectories generated by the identified S-system follow the trajectories generated with the actual S-system for both cases with acceptable precision.

B. The Cad system in E.coli

The Cad system is one of the conditional stress response modules in *E.coli*, which is induced only at low pH and a lysine-rich environment [15], [21], [22]. As shown in Figure 6, the main components of the Cad system are the enzyme CadA, the transport protein CadB, and the regulatory protein CadC. The decarboxylase CadA converts lysine into cadaverine in a reaction which consumes H^+. The transport protein CadB imports the substrate, lysine and exports the product, cadaverine. Hence, the intracellular H^+ concentration is reduced and the cell returns back to pH homeostasis. The membrane protein CadC senses the

TABLE I

ESTIMATIONS OF THE VALUES OF THE PARAMETERS OF THE S-SYSTEM (4)

	α_i	g_{i1}	g_{i2}	g_{i3}	g_{i4}	β_i	h_{i1}	h_{i2}	h_{i3}	h_{i4}	Error
True parameter set											
x_1	20.0	0	0	-0.8	0	10.0	0.5	0	0	0	
x_2	8.0	0.5	0	0	0	3.0	0	0.75	0	0	
x_3	3.0	0	0.75	0	0	5.0	0	0	0.5	0.2	
x_4	2.0	0.5	0	0	0	6.0	0	0	0	0.8	
IEKF noise free measurement											0.01
x_1	19.9137	0	0	-0.8025	0	9.9392	0.5021	0	0	0	
x_2	8.0000	0.5000	0	0	0	3.0000	0	0.7500	0	0	
x_3	2.9527	0	0.7559	0	0	4.9399	0	0	0.5042	0.2015	
x_4	1.9998	0.5000	0	0	0	5.9997	0	0	0	0.8000	
IEKF noisy measurement with SNR 40											0.1466
x_1	20.3933	0	0	-0.8337	0	9.8452	0.5071	0	0	0	
x_2	7.5459	0.5182	0	0	0	2.7021	0	0.7860	0	0	
x_3	2.5306	0	0.8076	0	0	4.3094	0	0	0.5505	0.2132	
x_4	2.0357	0.4782	0	0	0	6.0245	0	0	0	0.7880	
IEKF noisy measurement with SNR 20											0.4536
x_1	21.7684	0	0	-0.5631	0	13.6785	0.4096	0	0	0	
x_2	8.9980	0.4659	0	0	0	3.6136	0	0.8303	0	0	
x_3	3.6561	0	0.7949	0	0	6.0291	0	0	0.5113	0.2058	
x_4	1.5911	0.4291	0	0	0	5.2912	0	0	0	0.8241	

external conditions and regulates the stress response by binding directly to the DNA and activating the transcription of cadBA. This ensures that CadA and CadB are produced only under the appropriate external conditions of low pH and lysine abundance. Moreover, as shown in [22], CadC senses the external cadaverine and the accumulation of cadaverine in extracellular medium causes a delayed transcriptional down regulation of cadBA expression. The available time profile data set in [15] is used here for parameter estimation of the S-system model of the Cad system. In this data set, time profiles of a subset of the Cad system components, namely CadA, cadBA transcript, external lysine, external cadaverine and pH are available. Following the approach in [15], the lack of an appropriate data set for CadC, internal lysine and internal cadaverine is handled as follows:

- *CadC*: due to the fact that CadC is the signal protein which senses pH, external lysine and external cadaverine, one can directly consider the effect of these signals on the expression of cadBA. It should be mentioned that the effect of external cadaverine on CadC, and consequently on the expression of cadBA, is not considered in [15].
- *internal lysine and cadaverine*: due to the fact that CadA and CadB are responsible for decarboxylase reaction of lysine and the transport mechanism of lysine and cadaverine, one can couple the decarboxylase reaction and the transport mechanism and consider both CadA and CadB in a single step. It should be noted that the level of CadB is proportional to that of CadA since they are translated from the same mRNA.

Based on the existing qualitative model (Figure 6) and available pathway information, the following S-system can be obtained:

$$\begin{aligned}
\frac{d[CadA]}{dt} &= \alpha_1[cadBA]^{g_{12}} - \beta_1[CadA]^{h_{11}} \\
\frac{d[cadBA]}{dt} &= \alpha_2[Cadav]^{g_{23}}[Lys]^{g_{24}}[pH]^{g_{25}} - \beta_2[cadBA]^{h_{22}} \\
\frac{d[Cadav]}{dt} &= \alpha_3[CadA]^{g_{31}}[Cadav]^{g_{33}}[Lys]^{g_{34}} \\
\frac{d[Lys]}{dt} &= -\beta_4[CadA]^{h_{31}}[Cadav]^{h_{33}}[Lys]^{h_{34}}
\end{aligned} \tag{7}$$

As shown in [22], cadaverine represses the Cad module and hence the cadBA expression, i.e., $g_{23} < 0$. The total number of unknown parameters q for the S-system (7) is 18. Initial guesses of all the parameters except g_{23} are set to 0.1 and initial guess for g_{23} is set to -0.1. The initial covariance matrix $P_{w[0]} = 0.01I$, and the tuning parameters for the EKF are assigned as follows: $\epsilon = 1e-2$, $\kappa = 3-q = -15$, $\beta = 0$ and $\lambda_{RLS} = 1$. The available time profile has a sampling time of 0.5 minutes for 10 minutes, i.e., 21 data samples. Table II summarizes the estimated parameters of the Cad system using the IEKF algorithm and Figure 7 compares the response of the estimated S-system with the real data set. As shown in this figure, our proposed IEKF algorithm can successfully identify an S-system model that captures the dynamics of the Cad system.

V. CONCLUSION

In this paper, we have developed a parameter estimation algorithm based on the extended Kalman filter approach for biological phenomena modeled using the S-system framework. The proposed estimation algorithm is applied to the generic branched pathway as well as the Cad system in *E.Coli* and the estimation performance looks promising. One of the main advantages of EKF for parameter estimation is that it can be used even when only partial measurements of the states are available. This might be achieved based on the

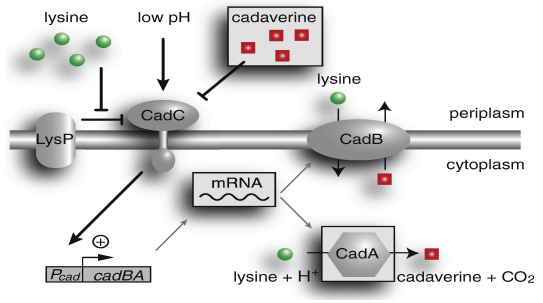

Fig. 6. Qualitative Model of the Cad System in E. coli (simplified) [22].

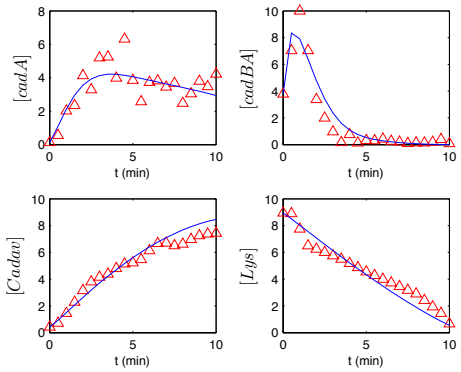

Fig. 7. Response of the estimated S-system (solid line) versus the real data set ($'\triangle'$) for Cad System.

joint extended Kalman filtering approach. However, in this paper, we focus on the cases where all the states of the S-system are measurable and the partial measurement case is left as one of the future research directions of this work.

VI. Acknowledgement

We are grateful to Dr. Christoph Küper, Dr. Eduardo Mendoza and Dr. Orland Gonzalez for providing the data set of the Cad system. This work was supported by the Qatar National Research Fund (a member of the Qatar Foundation) under Grant NPRP08-148-3-051.

TABLE II

Estimated parameters of S-system (7) of Cad system.

	α_i	g_{i1}	g_{i2}	g_{i3}	g_{i4}	g_{i5}
$[CadA]$	1.5399	0	1.3966	0	0	0
$[CadBA]$	0.3411	0	0	-0.7048	0.0638	0.6169
$[Cadav]$	0.3017	0.1755	0	0.0089	0.0062	0
$[Lys]$	0	0	0	0	0	0
	β_i	h_{i1}	h_{i2}	h_{i3}	h_{i4}	h_{i5}
$[CadA]$	0.1573	1.6485	0	0	0	0
$[CadBA]$	1.7299	0	0.0233	0	0	0
$[Cadav]$	0	0	0	0	0	0
$[Lys]$	0.6092	0.7302	0	0.0080	0.2047	0

References

[1] E. Voit, "A systems-theoretical framework for health and disease: Inflammation and preconditioning from an abstract modeling point of view," *Mathematical Bioscience*, vol. 217, pp. 11–18, 2009.

[2] E. O. Voit, F. Alvarez-Vasquez, and Y. A. Hannun, "Computational analysis of sphingolipid pathways systems," *Advances in experimental medicine and biology.*, vol. 688, pp. 264–275, 2010.

[3] R. Goodacre and G. G. Harrigan, *Metabolite Profiling: Its role in biomaker discovery and gene function analysis*. Kluwer Academic Publishers, Dordrecht, The Netherlands, 2003.

[4] A. Neves, R. Ventura, N. Mansour, C. Shearman, M. Gasson, C. Maycock, A. Ramos, and H. Santos, "Is the glycolytic flux in *Lactococcus lactis* primarily controlled by the redox charge?" *Journal of Biochemical Chemistry*, vol. 277, no. 31, pp. 28 088–28 098, 2002.

[5] C. Gerner, S. Vejda, D. Gelbmann, E. Bayer, J. Gotzmann, R. Schulte-Hermann, and W. Mikulits, "Concomitant determination of absolute values of cellular protein amounts, synthesis rates, and turnover rates by quantitative proteome profiling," *Mol. Cell. Proteomics*, vol. 1, pp. 528–537, 2002.

[6] E. O. Voit, *Canonical nonlinear modeling : S-system approach to understanding complexity*, 1991.

[7] R. Gentilini, "Toward integration of systems biology formalism: the gene regulatory networks case," *Genome Inform.*, vol. 16, pp. 215–224, 2005.

[8] E. O. Voit, *Computational Analysis of Biochemical System. A practice guide for biochemists and Molecular biologists*. Cambridge University Press, 2002.

[9] I.-C. Chou and E. O. Voit, "Recent developments in parameter estimation and structure identification of biochemical and genomic systems," *Mathematical Bioscience*, vol. 219, pp. 57–83, 2009.

[10] S. Kikuchi, D. Tominaga, M. Arita, K. Takahashi, and M. Tomita, "Dynamic modeling of genetic networks using genetic algorithm and S-systems," *Bioinformatics*, vol. 19, no. 5, pp. 643–650, 2003.

[11] E. O. Voit and J. Almeida, "Decoupling dynamical systems for pathway identification from metabolic profiles," *Bioinformatics*, vol. 20, no. 11, pp. 1670–1681, 2004.

[12] S. Kimura, K. Ide, A. Kashihara, M. Kano, M. Hatakeyama, R. Masui, N. Nakagawa, S. Yokoyama, S. Kuramitsu, and A. Konagaya, "Inference of S-system models of genetic networks using a cooperative coevolutionary algorithm," *Bioinformatics*, vol. 21, no. 7, pp. 1154–1163, 2005.

[13] P. K. Polisetty, E. O. Voit, and E. P. Gatzke, "Identification of metabolic system parameters using global optimization methods," *Theor. Biol. Med. Model*, vol. 3, no. 4, 2006.

[14] I.-C. Chou, H. Martens, and E. O. Voit, "Parameter estimation in biochemical systems models with alternating regression," *Theoretical Biology and Medical Modelling*, vol. 3, no. 25, 2006.

[15] O. R. Gonzalez, C. Kuper, K. Jung, P. C. Naval, and E. Mendoza, "Parameter estimation using simulated annealing for S-system models of biochemical networks," *Bioinformatics*, vol. 23, no. 4, pp. 480–486, 2007.

[16] Z. Kutalik, W. Tucker, and V. Moulton, "S-system parameter estimation for noisy metabolic profiles using Newton-flow analysis," *IET System Biology*, vol. 1, no. 3, pp. 174–180, 2007.

[17] P.-K. Liu and F.-S. Wang, "Inference of biochemical network models in S-system using multiobjective optimization approach," *Bioinformatics*, vol. 24, no. 8, pp. 1085–1092, 2008.

[18] H. Wang, L. Qian, and E. Dougherty, "Inference of gene regulatory networks using S-systems: a unified approach," *IET System Biology*, vol. 4, no. 2, pp. 145–156, 2010.

[19] L. Qian, H. Wang, and E. R. Dougherty, "Inference of noisy nonlinear differential equation models for gene regulatory networks using genetic programming and Kalman filtering," *IEEE Transactions on Signal Processing*, vol. 56, no. 7, pp. 3327–3339, 2008.

[20] E. A. Wan and R. van der Merwe, in *Kalman Filtering and Neural Networks*, S. Haykin, Ed. John Wiley and Sons, 2001, ch. 5.

[21] C. Kuper and K. Jung, "CadC-mediated activation of the cadba promoter in Escherichia coli," *Journal of molecular microbiology and biotechnology*, vol. 10, no. 1, pp. 26–29, 2005.

[22] G. Fritz, C. Koller, K. Burdack, L. Tetsch, I. Haneburger, K. Jung, and U. Gerland, "Induction kinetics of a conditional pH stress response system in Escherichia coli," *Journal of Molecular Biology*, vol. 393, no. 2, pp. 272–286, 2009.