Chemometrics and Intelligent Laboratory Systems 138 (2014) 57–63

Contents lists available at ScienceDirect

Chemometrics and Intelligent Laboratory Systems

journal homepage: www.elsevier.com/locate/chemolab

An extended fractional Kalman filter for inferring gene regulatory networks using time-series data

Yongqing Zhang [a,*], Yifei Pu [a], Haisen Zhang [b], Yong Cong [c], Jiliu Zhou [a]

[a] College of Computer Science, Sichuan University, Chengdu 610065, PR China
[b] Department of Mathematics, Sichuan University, Chengdu 610064, PR China
[c] College of Chemistry and Chemical Engineering, Lanzhou University, Lanzhou 730000, PR China

ARTICLE INFO

Article history:
Received 9 October 2013
Received in revised form 17 June 2014
Accepted 9 July 2014
Available online 18 July 2014

Keywords:
Gene regulatory network
Extended fractional Kalman filter
Parameter estimation
Time-series data

ABSTRACT

In recent years, inference of gene regulatory networks has received ever increasing attention in the systems biology field. In this paper, for the first time, a fractional gene regulatory algorithm by extended fractional Kalman filter (EFKF) is proposed to estimate the hidden states as well as the unknown static parameters of the model, which can provide insight into the underlying regulatory relations among genes in the biological system. In the proposed method, gene regulatory networks are inferred via evolutionary modeling based on time-series microarray measurements. The gene regulatory network is considered as a fractional order discrete stochastic dynamic model that consists of the gene measurement equation and the gene regulation equation. After specifying the model structure, we apply the EFKF algorithm for identifying both the model parameters and the actual value of gene expression levels. In this paper, the main advantages of using fractional order systems, increasing the flexibility and improving the accuracy of the system state equation in EFKF are highlighted. The performance of the EFKF algorithm is compared with EKF and other nonlinear algorithms in predicting the parameters of gene regulatory networks from synthetic data and real biological data. Extensive computer simulations illustrate that the proposed algorithm outperforms EKF and other methods, and therefore, it can serve as a natural framework for inference gene regulatory networks with a nonlinear structure.

© 2014 Elsevier B.V. All rights reserved.

1. Introduction

Fractional calculus, which is first mentioned by Leibniz and L'Hospital in 1695, has been studied by many mathematicians for a long time. For more details, refer to the books by Oldham and Spanier [1]. During the last 20 years, fractional calculus has started to enter in more and more application fields, including physics, chemistry, materials science, viscoelasticity, electrical circuits, engineering and biology [2–6]. Fractional order models are now being applied to a wide range of problems in bioengineering [2].

Gene expression is the process of generating functional gene products, for example, mRNA and protein. Hence the level of gene functionality can be measured using microarrays or gene chips to produce gene expression data [7]. Measuring the levels of gene expression in different conditions is meaningful in medical diagnosis, treatment, and drug design [8]. Many gene expression experiments produce time-series data with only a few time points owing to the high measurement costs. Therefore, it becomes significant to predict the behavior of gene regulatory networks (GRNs) through modern computing technology. Recently, many algorithms and mathematical models were proposed to predict gene regulatory networks from time-series data [9,10], such as Boolean network [11], Dynamic Bayesian networks [12], neural networks [13], differential equations [14], state-space model [15], and stochastic model [16].

It is well known that gene expression is a complex nonlinear dynamic system. Parameter estimation in nonlinear dynamic systems is extremely important, but also extremely difficult. Some researchers use the S-system model to perform analysis of genetic network [17,18]. It has been successfully used in some biochemical networks, but encounters high-dimensionality problems when used to analyze large-scale genetic networks.

A remarkable feature of time-series gene express data is that the number of the time points is usually much smaller than that of the number of genes. So one of the most significant challenges in GRNs is to build a model that can analyze such a high-dimensional and short-length time-series data. In general, gene expression systems are partially observed. Therefore, a natural way to infer dynamic gene regulatory networks is to employ nonlinear state-space models that consist of two types of equations: system equations and observation equations [19]. The well-known extended Kalman filter (EKF) has been widely used in the state estimation of nonlinear dynamic systems from noisy measurements. Wang [20]has applied EKF to model nonlinear dynamic GRNs via short gene expression time series. The GRNs are considered

* Corresponding author.
E-mail address: zhangyongqingscu@hotmail.com (Y. Zhang).

http://dx.doi.org/10.1016/j.chemolab.2014.07.007
0169-7439/© 2014 Elsevier B.V. All rights reserved.

as a nonlinear dynamic stochastic model that consists of the gene measurement equation and the gene regulation equation. Sun [21]has proposed EKF for estimation of parameters in nonlinear state-space models of biochemical networks. He discussed in detail how to develop a general framework for modeling biochemical networks and how to estimate the parameters in the models. Qian [22]has presented inference of noisy nonlinear differential equation models for GRNs using genetic programming and Kalman filtering. These research works demonstrate that EKF cannot only develop mathematical models but also estimate their parameters in gene regulatory networks. However, the differential order of the system equation in EKF is fixed as an integer which limits its applications and EKF only relies on the current state value. So we proposed a fractional gene regulatory algorithm by fractional calculus. The fractional calculus is a generalization of the traditional differential calculus for a case when integrals and derivatives are in not only integers but also fractional order. It is well known that fractional differential equations are useful because of their nonlocal character [23], i.e., the next state of a system not only depends on its current state but also on its historical states starting from the initial time. This is closer to reality and is therefore the main reason that fractional differential equations have become more popular.

This paper assumes that the GRNs obey a nonlinear fractional differential equation with additive Gaussian white noise. The gene expressions are assumed to evolve following a sigmoid squash function, whereas a linear function is considered for the microarray data. After specifying the model structure, we apply the extended fractional Kalman filter (EFKF) [24] to estimate the model parameters and hidden states of the nonlinear model using time-series data. A synthetic data and two real microarray time-series data from the yeast protein synthesis and SOS DNA Repair network of *Escherichia coli* are used to test our method. Result show that this method is capable of improving the prediction accuracy of microarray time-series dataset. Our major contributions in this study can be summarized as follows. (1) An EFKF is firstly presented to estimate the hidden states and parameters of the nonlinear model in GRNs. (2) The performance of our algorithm is evaluated for time-series data contrasting with the EKF and other nonlinear algorithms. It is demonstrated that our method is more effective than previous methods.

This paper is organized as follows: Section 2 illustrates the fractional gene regulatory algorithm by extended fractional Kalman filter and how to analyze the parameters for EFKF. Simulation results are given in Section 3. The performance of EFKF method is compared with EKF and other nonlinear algorithms. Section 4 contains some concluding remarks.

2. Materials and methods

2.1. System model and problem statement

Let $\mathbb{R}$ be the set of real numbers. The fractional order Grünwald–Letnikov difference of a function $x : \mathbb{R} \to \mathbb{R}$ is given by the following equation [24]:

$$!^{\gamma}x(k) = \frac{1}{\delta^{\gamma}}\sum_{j=0}^{k}(-1)^{j}\binom{\gamma}{j}x(k-j), \tag{1}$$

where $\gamma \in \mathbb{R}$ is a fractional order, and δ is a sampling time later equal to 1, k is the number of samples for which the derivative is calculated. The factor $\binom{\gamma}{j}$ can be obtained from:

$$\binom{\gamma}{j} = \begin{cases} 1, & j = 0 \\ \dfrac{\gamma(\gamma-1)\cdots(\gamma-j+1)}{j!}, & j>0 \end{cases}. \tag{2}$$

According to this definition, it is possible to obtain a discrete equivalent of the derivative (when γ is positive), a discrete equivalent of integration (when γ is negative).

Let n be the number of genes. We assume that the gene expression follows the fractional order discrete stochastic dynamical system:

$$\begin{cases} \Delta^{\gamma}x(k+1) = f(x(k)) + \omega(k), \\ x(k+1) = \Delta^{\gamma}x(k+1) - \sum_{j=1}^{k+1}(-1)^{j}\gamma(j)x(k+1-j), \quad k = 1,2,\cdots,m, \\ y(k) = h(x(k)) + \nu(k), \end{cases} \tag{3}$$

where

$$\gamma_k = diag\left[\binom{\alpha_1}{k}\cdots\binom{\alpha_n}{k}\right],$$
$$\Delta^{\gamma}x(k+1) = \begin{bmatrix} \Delta^{\alpha_1}x_1(k+1) \\ \cdots \\ \Delta^{\alpha_n}x_n(k+1) \end{bmatrix},$$

α_i is the order of system equation with respect to the i-th gene, $i = 1,2,\cdots, n$. m is the total number of data points, $\omega(k)$ and $\nu(k)$ are assumed to be zero-mean Gaussian white noise with covariance $Q(k)$and $R(k)$, respectively, i.e., $\omega(k) \sim N(0, Q(k))$, $\nu(k) \sim N(0, R(k))$. $f(\cdot) : \mathbb{R}^n \to \mathbb{R}^n$ and $h(\cdot) : \mathbb{R}^n \to \mathbb{R}^n$ are some proper nonlinear functions. The nonlinear function $f(\cdot)$ and $h(\cdot)$ can be linearized according to the Taylor series expansion.

Setting $f(x(k)) = [f_1(x_1(k)), f_2(x_2(k)), \cdots, f_n(x_n(k))]^T$. To capture the gene interactions effectively, we assumed [25]

$$f_i(x) = \sum_{j=1}^{n} a_{ij}g_j(x), \quad i = 1,2,\cdots,n, \tag{4}$$

where $\boldsymbol{A} = (a_{ij})_{n \times n}$ represents the nonlinear regulatory relationship among genes; and the nonlinear function $g_j(\cdot)$ is given by

$$g_j(x) = \frac{1}{1+\exp(-x)}, \quad x \in \mathbb{R}^n. \tag{5}$$

A discrete linear Gaussian model for the microarray data is considered which can be expressed at the ith time instant as [20]

$$y_i(k) = x_i(k) + \nu_i(k), \quad i = 1,2,\cdots,n, \quad k = 1,2,\cdots,m, \tag{6}$$

where $y(k) = [y_1(k), y_2(k), \cdots, y_n(k)]^T$. That is, $h \equiv I$, which is the identity matrix.

In our model, the $\theta = [a_{11}, a_{21}, \cdots, a_{n1}, a_{12}, a_{22}, \cdots, a_{n2}, a_{1n}, a_{2n}, \cdots, a_{nn}]$ are parameters to be identified. It is also worth pointing out that we can identify the n state variables as well.

2.2. Extended fractional Kalman filter

The extended fractional Kalman filter (EFKF), which has been studied in [24], is an important and fascinating algorithm in nonlinear theory which is extended to fractional systems. We list this algorithm in the following:

Theorem 2.1. [24]*For the nonlinear fractional order stochastic discrete state–space system given by Eq.(3), the extended fractional Kalman filter (EFKF) is given by the following equations:*

$$\Delta^{\gamma}\tilde{x}(k+1) = f(\hat{x}(k)), \tag{7}$$

$$\tilde{x}(k+1) = \Delta^{\gamma}\tilde{x}(k+1) - \sum_{j=0}^{k+1}(-1)^{j}\gamma_j\hat{x}(k+1-j), \tag{8}$$

where $\tilde{x}(k+1)$ is the state prediction for the system given by Eq.(3), $\hat{x}(k)$ is a state vector estimation at time point k.

$$\tilde{P}(k) = (F(k-1)+\gamma(1))P(k-1)(F(k-1)+\gamma(1))^T + Q(k-1) + \sum_{j=2}^{k}\gamma(j)P(k-j)\gamma^T(j). \quad (9)$$

$\tilde{P}(k)$is a prediction of an estimation error covariance matrix. As it is shown, the prediction of covariance error matrix depends on the value of covariance matrixes in previous time samples.

$$\hat{x}(k) = \tilde{x}(k) + K(k)(y(k)-h(\tilde{x}(k))), \quad (10)$$

$$P(k) = (I-K(k)H(k))\tilde{P}(k). \quad (11)$$

$P(k)$ is an estimation error covariance matrix

$$K(k) = \tilde{P}(k)H^T(k)\left(H(k)\tilde{P}(k)H^T(k)+R(k)\right)^{-1}. \quad (12)$$

$K(k)$ is called the Kalman filter gain vector. where

$$F(k-1) = \left[\frac{\partial f(x(k-1))}{\partial x}\right]_{x=\hat{x}(k-1)},$$
$$H(k) = \left[\frac{\partial h(x)}{\partial x}\right]_{x=\hat{x}(k)}.$$

and the white Gaussian noise sequences $\omega(k)$ and $v(k)$ are assumed to be independent and zero mean:

$$E[\omega(k)] = 0; \quad E[\nu(k)] = 0; \quad E\left[\omega(k)\omega^T(k)\right] = Q(k); \quad E\left[\nu(k)\nu^T(k)\right] = R(k).$$

With the initial conditions

$$x(0)\in\mathbb{R}^N, P(0) = E\left[(\hat{x}(0)-x(0))(\hat{x}(0)-x(0))^T\right].$$

As mentioned above, EFKF estimates the states of fractional order system and in the next section we will describe how to use EFKF to estimate the parameters in Eq. (3).

2.3. The EFKF for parameter identification

The challenging tasks in inference for nonlinear state-space models are to estimate both the states and parameters of the system. One of the methods for this dual estimate is to use EFKF by taking the parameters as additional states and augmenting state equations [26]. So the nonlinear system dynamics can be described by

$$\Delta^{\gamma}x(k+1) = f(x(k),\theta) + \xi(k), \quad (13)$$

with a measurement given by

$$y(k) = h(x(k),\theta) + \nu(k), \quad (14)$$

where $x(k)$ and $y(k)$ are the state vector and the measurement vector, respectively, $\xi(k)$ and $\nu(k)$ are the zero-mean white noise process, and the parameters to be estimated are donated as $\theta = [a_{11}, a_{21}, \cdots, a_{n1}, \cdots, a_{1n}, a_{2n}, \cdots, a_{nn}]$, and n is the number of genes. In order to simplify the algorithm, fractional order γ of each equation is equal in our paper.

Let $X(k) = [x(k)^T, \theta(k)^T]$, $\xi(k) = [\omega^T(k), \eta^T(k)]^T$, $F(X(k)) = [f^T(x(k), \theta(k)), \theta^T(k)]^T$, and $H(X(k)) = h(x(k), \theta(k))$. I_n be the $n \times n$ identity matrix, and the matrix $H = [I_m \quad 0]$ with appropriate dimension.

In order to facilitate the application of the EFKF in the parameter estimation problem, we rewrite Eqs. (14–15) as follows:

$$\begin{cases} \Delta^{\Gamma}X(k+1) = F(X(k)) + \xi(k), \\ X(k+1) = \Delta^{\Gamma}X(k+1) - \sum_{j=1}^{k+1}(-1)^j\Gamma(j)X(k+1-j), \\ y(k) = HX(k) + \nu(k), \end{cases} \quad (15)$$

where

$$\Delta^{\Gamma} = diag\left[\Delta^{\gamma_1}, \Delta^{\gamma_2}, \cdots, \Delta^{\gamma_n}, \Delta^1, \cdots, \Delta^1\right]_{n+n^2},$$

$$\Gamma(j) = diag\left[\binom{\gamma_1}{j}, \binom{\gamma_2}{j}, \cdots\binom{\gamma_n}{j}, \binom{1}{j}, \cdots\binom{1}{j}\right]_{n+n^2},$$

$$F(X(k)) = \begin{bmatrix} A(k)f(x(k)) \\ \theta(k) \end{bmatrix},$$

$$\xi(k) = \begin{bmatrix} \omega(k) \\ \eta(k) \end{bmatrix},$$

with $f(x) = [f_1(x_1), f_2(x_2), \cdots, f_n(x_n)]^T$ and

$$f_j\left(x_j\right) = \frac{1}{1+\exp\left(-x_j\right)}.$$

It is not difficult to see that

$$\frac{\partial F(X(k))}{\partial X(k)} = \begin{bmatrix} Z_1(X(k)) & Z_2(X(k)) & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad (16)$$

where

$$Z_1(X(k)) = A(k)\frac{\partial f(x(k))}{\partial x(k)},$$

$$Z_2(X(k)) = [f(x_1(k))I, f(x_2(k))I, \cdots, f(x_n(k))I],$$

with

$$\frac{\partial f(x(k))}{\partial x(k)} = diag\left\{\frac{\exp(-x_1(k))}{(1+\exp(-x_1(k)))^2}, \frac{\exp(-x_2(k))}{(1+\exp(-x_2(k)))^2}, \cdots, \frac{\exp(-x_n(k))}{(1+\exp(-x_n(k)))^2}\right\}.$$

Based on the EFKF approach, we can identify n state variables and n^2 parameters.

3. Results and discussions

In this section some numerical simulations are given to show the effectiveness of the proposed algorithms. Specifically, at first, it has been applied to synthetic data. Then our method is tested using real microarray time-series data, including the dataset from yeast protein synthesis and SOS DNA Repair network of *E. coli*. Our algorithm has been implemented in the MATLAB programming language.

3.1. Application on synthetic dataset

An 8 gene network is considered for evaluation of the EFKF's ability to predict gene expression profiles. Data with 100 time points are

generated using the model given in Eq. (15). The system noise is taken to be $\xi(k) \sim N(0, 0.2^2 I)$ and observation noise is taken to be $\nu(k) \sim N(0, 0.3^2 I)$. The initial values of the state variables $x(0)$ equal the observed values $y(0)$ and the initial values of parameters $A_{ij}(0)$ are assumed as the arbitrary value between 0 and 1. The initial error covariance matrix is $P(0) = 0.5I$, and I is the identity matrix. The initial value of the fractional order is 0.5.

Table 1 gives the true and estimated values of the system parameters $a_{11}, a_{12}, \cdots, a_{18}, a_{21} \cdots, a_{88}$. We can see that most of the estimate values of EFKF are close to the true values. This shows that EKFK algorithm could work well when modeling nonlinear dynamic gene regulatory networks. At the same time, we also can see some limitations on EFKF for a_{36}, a_{51}, a_{84},which is a large difference between the true and estimated values. One way to overcome the limitations is to choose appropriate covariance matrix of the noise in the parameter in Eq. (15). It is well known that the covariance matrix will affect the convergence rate and tracking performance [27].

3.2. Application on the yeast protein synthesis dataset

In this section, the time-series gene expression data from yeast protein synthesis is considered to test our algorithm. The measurement data is obtained from http://www.wanghaixin.com/biowq2007.html, where related references are also available. The data has 17 equally spaced time points by 10 minute interval. Here, the data for 5 genes (*HAP1*, *CYB2*, *CYC7*, *CYT1* and *COX5A*) are picked because the relations among them have been revealed by biological experiments [28]. The states of 5 genes are represented by $x1, x2, x3, x4, x5$, respectively.

Considering the limited number of original data points, we constructed an interpolated time-series data using a cubic spline interpolation. An interpolation rate is 3 interpolated points between the two closest measurements. The system noise is taken to be $\xi(k) \sim N(0, 0.2^2 I)$ and the observation noise is taken to be $\nu(k) \sim N(0, 0.4^2 I)$. The initial values of the state variable $x(0)$ equal the observed values $y(0)$ and the initial values of parameters $A_{ij}(0)$ are assumed as the arbitrary value between 0 and 1. The initial value of the fractional order is 0.6. The initial error covariance matrix is$P(0) = 0.5I$. Then, based on the EFKF algorithm, we can identify all n^2 the parameters and $n$the state variables.

To keep the paper concise, we only select partial genes displayed in Fig. 1. The true values of gene expression time-series data and its estimated values by EFKF algorithm are illustrated in Fig. 1(A). The blue lines are true values, while the green lines are estimated values. This figure shows that the estimated model makes a fairly good prediction of the gene expression data. The square errors between the true values and estimated values of gene expression are obtained in Fig. 1(B). It showed that at the beginning the square errors display fluctuation, but they quickly tend toward the zeros with the increase in time points. In order to further evaluate the algorithm, our method and EKF are compared with the error covariance of parameters a_{ij} as a fidelity criterion. Fig. 1(C) depicted the estimation results of the parameters a_{11}, a_{12}, a_{13} by EFKF and EKF methods. We can observe that, when time points increase, the error variances of the parameters by EFKF are smaller than EKF. This is mainly because EFKF increase power-law long-term memory and nonlocal character [23]. The value of EFKF depends not only on its current state but also on its historical states starting from the initial time.

3.3. Application on the E. coli dataset

In this section, our method has been applied to analyze the well-known SOS DNA repair network in *E. coli*. The expression datasets of the SOS DNA repair system have been downloaded from the homepage of Uri Alon lab [29]. Those include 8 genes namely *uvrD*, *lexA*, *umuDC*, *recA*, *uvrA*, *uvrY*, *ruvA* and *polB*. Each experiment consists of 50 measurements evenly spaced by 6 min interval.

In this experiment, 8 genes have been chosen from Alon's experiment data. The same dataset of genes were used by Ronen [29], Maraziotis [30] and Yang [31]. Though the data contain four time-series datasets, to make comparison fairly, we use the second dataset. Since each gene including the initial concentrations which are all zeros, these initial concentrations have been removed from the dataset. So each of the genes has 49 time points. The system noise is taken to be $\xi(k) \sim N(0, 0.2^2 I)$ and the observation noise is taken to be $\nu(k) \sim N(0, 0.4^2 I)$. The initial values of the state variable $x(0)$ equal the

Table 1
True values and estimated values of parameters a_{nn} using EFKF.

Parameter	a_{11}	a_{12}	a_{13}	a_{14}	a_{15}	a_{16}	a_{17}	a_{18}
True value	0.327	−0.595	−0.306	0.627	1.380	0.954	−0.816	0.870
EFKF	0.256	−0.599	−0.355	0.555	1.315	0.866	−0.845	0.810
Parameter	a_{21}	a_{22}	a_{23}	a_{24}	a_{25}	a_{26}	a_{27}	a_{28}
True value	0.191	−0.558	0.692	0.992	0.261	1.256	−0.451	−0.20
EFKF	0.215	−0.539	0.778	1.017	0.257	1.261	−0.426	−0.206
Parameter	a_{31}	a_{32}	a_{33}	a_{34}	a_{35}	a_{36}	a_{37}	a_{38}
True value	1.547	0.469	0.864	−0.931	0.280	−0.086	0.418	1.147
EFKF	1.481	0.389	0.812	−0.937	0.189	−0.192	0.370	1.047
Parameter	a_{41}	a_{42}	a_{43}	a_{44}	a_{45}	a_{46}	a_{47}	a_{48}
True value	−0.260	0.149	−0.177	−0.123	−0.004	0.486	0	−0.062
EFKF	−0.255	0.120	−0.155	−0.096	−0.006	0.461	0.005	−0.061
Parameter	a_{51}	a_{52}	a_{53}	a_{54}	a_{55}	a_{56}	a_{57}	a_{58}
True value	0.323	0.048	0.148	−0.132	0.124	−0.094	0.125	0.131
EFKF	0.120	0.036	0.135	−0.146	0.109	−0.105	0.111	0.116
Parameter	a_{61}	a_{62}	a_{63}	a_{64}	a_{65}	a_{66}	a_{67}	a_{68}
True value	0.105	−0.164	0.278	−0.167	0.361	−0.112	0.035	−0.107
EFKF	0.104	−0.141	0.332	−0.151	0.413	−0.093	0.044	−0.105
Parameter	a_{71}	a_{72}	a_{73}	a_{74}	a_{75}	a_{76}	a_{77}	a_{78}
True value	−0.225	0.199	0.495	0.076	0.011	−0.245	0.136	0.066
EFKF	−0.230	0.194	0.491	0.070	0.005	−0.250	0.132	0.062
Parameter	a_{81}	a_{82}	a_{83}	a_{84}	a_{85}	a_{86}	a_{87}	a_{88}
True value	−0.112	0.069	0.191	−0.223	−0.036	0.159	−0.257	0.211
EFKF	−0.124	0.063	0.191	−0.135	−0.039	0.151	−0.264	0.302

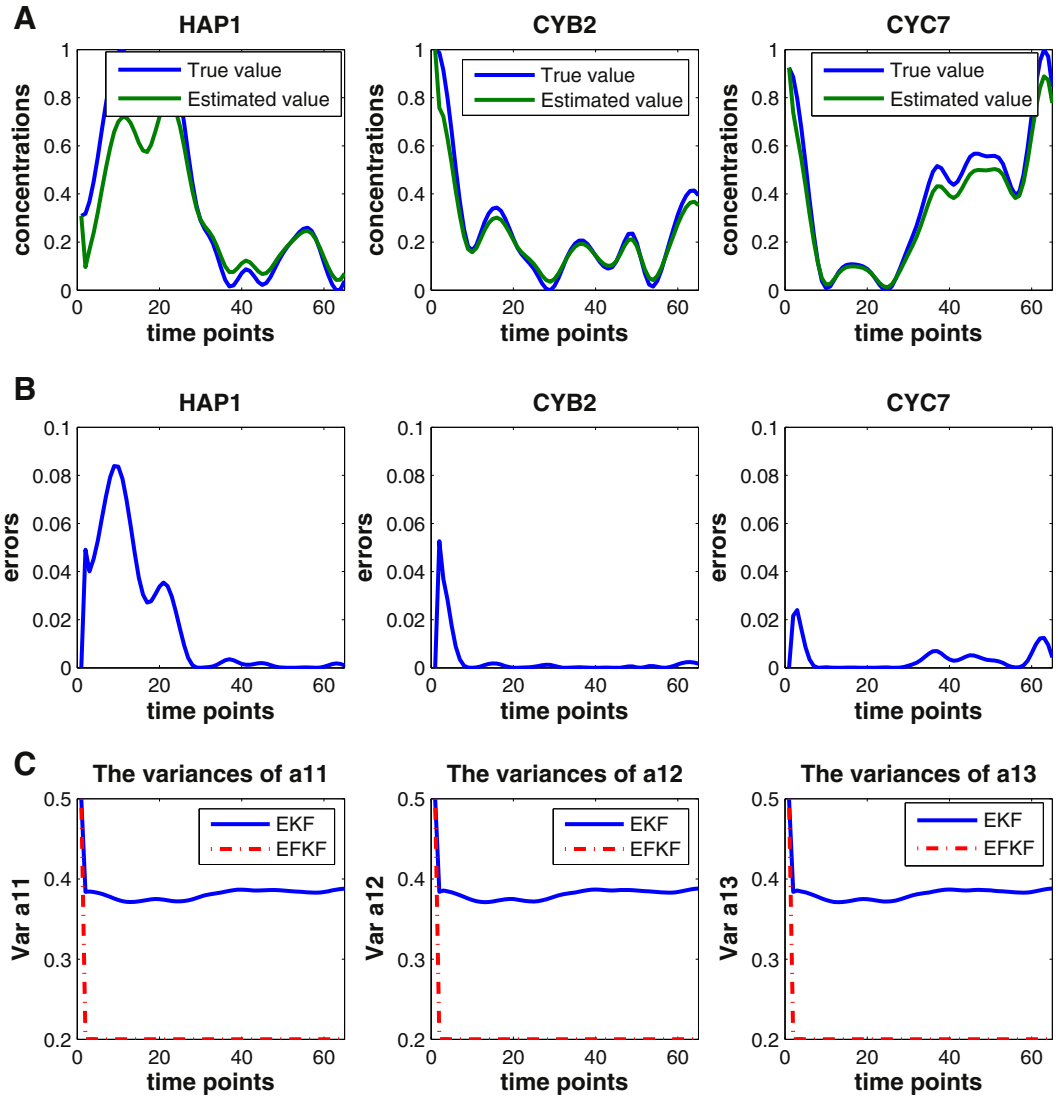

Fig. 1. The estimation result of yeast protein synthesis time-series data by EFKF algorithm. (A) The true values and estimation values of gene expression using the EFKF algorithm. (B) The square error between true values and estimation values of eachgene expression. (C)The variance of estimated parameters a_{11},a_{12},a_{13} by EFKF compared with EKF.

observed values $y(0)$ and the initial values of parameters $A_{ij}(0)$ are assumed as the arbitrary value between 0 and 1. The initial value of the fractional order is 0.9. The initial error covariance matrix is $P(0) = 0.5I$. Then, based on the EFKF algorithm, we can identify all the parameters.

Table 2 illustrates the MSE comparisons of eight genes among EFKF, RNF network [30], FNT model [31] and the model of Ronen [29]. From Table 2, we can see that the prediction of the EFKF model is more accurate than the previously published results. The EFKF could be an important method to infer gene regulatory network. In addition, Fig. 2 depicted the estimation results of the *E. coli* time-series data by our method. From Fig. 2, we can obtain the same conclusions as Section 3.2.

The values of the estimated parameters a_{11}, a_{12}, ···, a_{18} are displayed in Fig. 3, in which each line represents a parameter. It demonstrated that the value of the estimation parameter will be stable as the time increases.

Table 2
MSE comparison among method of EFKF, RNF network, FNT model and Method of Renen.

Target gene	EFKF	RNF network	FNT model	Method of Ronen
uvrD	0.025	0.195	0.1303	0.20
lexA	0.022	0.105	0.0583	0.10
umuDC	0.014	0.200	0.1458	0.21
recA	0.087	0.12	0.1037	0.12
uvrA	0.013	0.115	0.084	0.14
uvrY	0.031	0.42	0.3867	0.45
ruvA	0.050	0.201	0.1895	0.22
polB	0.031	0.302	0.0201	0.31

4. Conclusions

In this paper, a novel approach is presented to infer gene regulatory networks from the time-series data. The gene regulatory network is assumed to follow a nonlinear stochastic differential equation. For the first time, fractional gene regulatory systems are implemented to estimate the parameters which indicate the relationship between the genes by extending fractional Kalman filter. Fractional calculus has been incorporated into chemometrics, mainly because of its long-term memory and nonlocality. An important feature of fractional calculus is that its value depends not only on its current state but also on its historical states starting from the initial time. With the help of the EFKF algorithm, we have identified the hidden states and the parameters of the system. To inspect the performance of the proposed method, both integer Kalman filter and also fractional order Kalman filter have been implemented.

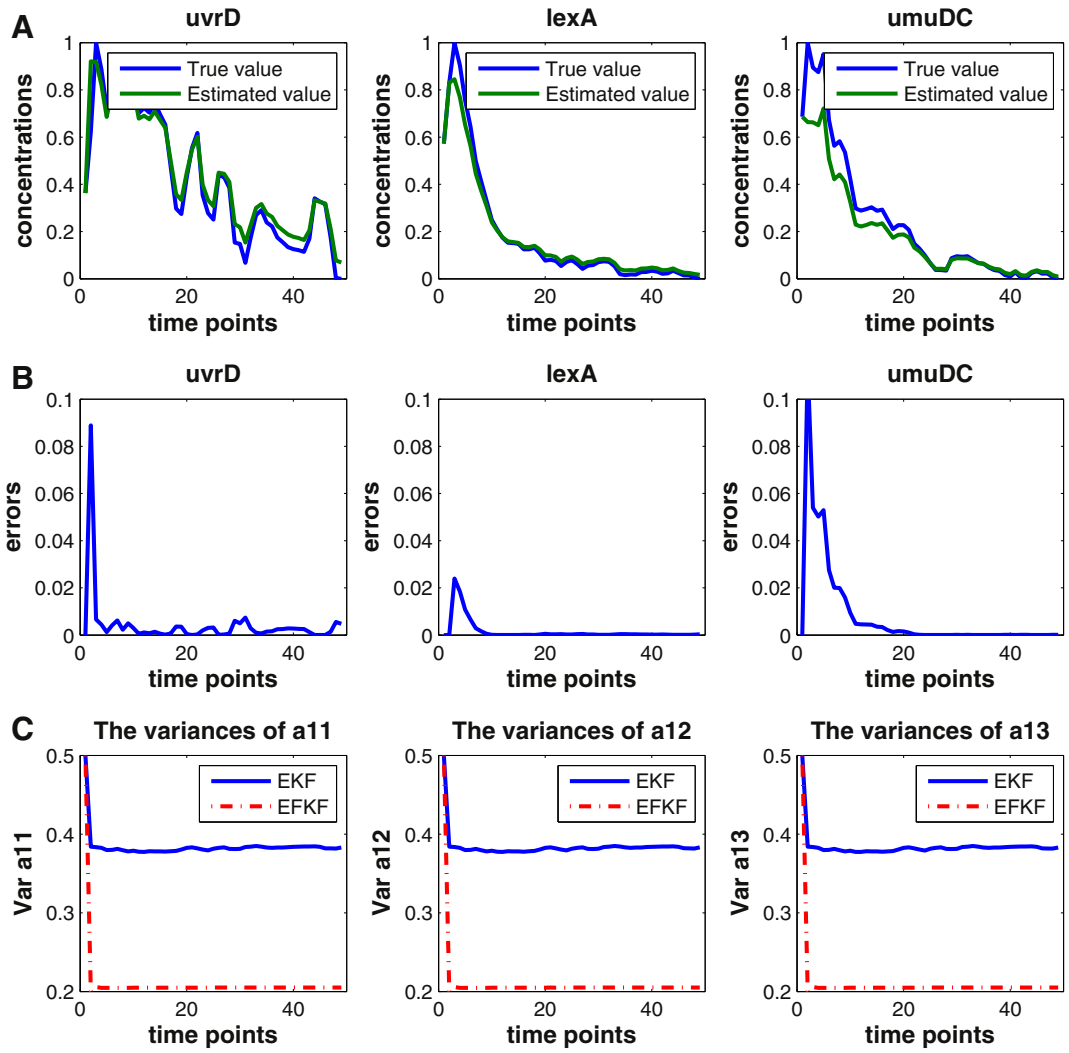

Fig. 2. The estimation result of *Escherichia coli* time-series data by EFKF algorithm. (A) The true values and estimation values of gene expression using the EFKF algorithm. (B) The square error between true values and estimation values of each gene expression. (C) The variance of estimated parameters a_{11},a_{12},a_{13} compared EFKF with EKF.

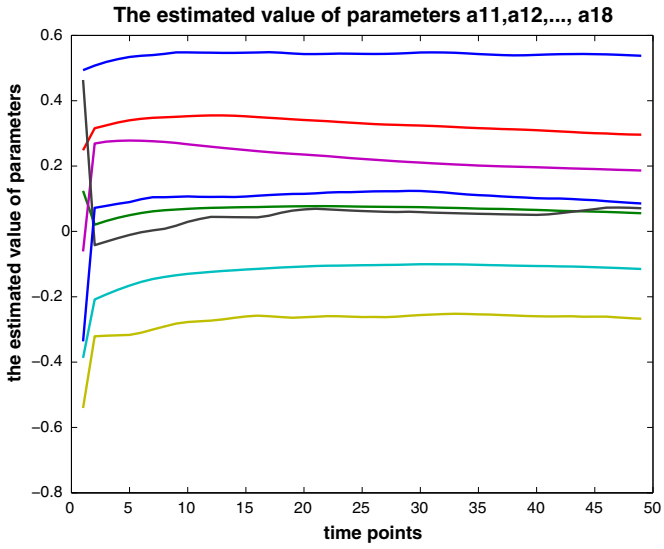

Fig. 3. The estimation result of parameters a_{11},a_{12},$\cdots$,a_{18}, each line represents a parameter.

The EFKF also compared with other nonlinear algorithms. From the simulation results, the performance of the proposed method seems to be satisfactory for inferring the given gene regulatory networks.

Conflict of interest

None declared.

Acknowledgments

The work is supported by the Foundation Franco-Chinoise Pour La Science Et Ses Applications (FFCSA), the China National Nature Science Foundation(Grant Nos.60972131, 61201438, 61302028), the Returned Overseas Chinese Scholars, State Education Ministry of China (Grant No. 20111139), the Sichuan Science and Technology Support Project of China (Grant Nos.2011GZ0201, and 2013SZ0071), the Sichuan Provincial Department of Science and Technology Project (Grant Nos. 2012ZZ023, 2013ZR0010), and the Science and Technology Innovation Youth Foundation of Sichuan Province (2011-014).

References

[1] Keith B. Oldham, Jerome Spanier, The Fractional Calculus, vol. 1047Academic Press, New York, 1974.
[2] Richard L. Magin, Fractional calculus models of complex dynamics in biological tissues, Comput. Math. Appl. 59 (5) (2010) 1586–1593.
[3] Adam Loverro, Fractional calculus: history, definitions and applications for the engineer, Report, Department of Aerospace and Mechanical Engineering, Notre Dame, IN, 46556, 2004.
[4] Igor Podlubny, Fractional differential equations: an introduction to fractional derivatives, fractional differential equations, to methods of their solution and some of their applications, vol. 198Access Online via Elsevier, 1998.
[5] Yi-Fei Pu, Ji-Liu Zhou, Xiao Yuan, Fractional differential mask: a fractional differential-based approach for multiscale texture enhancement, IEEE Trans. Image Process. 19 (2) (2010) 491–511.
[6] Kiani-B. Arman, Kia Fallahi, Naser Pariz, Henry Leung, A chaotic secure communication scheme using fractional chaotic systems based on an extended fractional Kalman filter, Commun. Nonlinear Sci. Numer. Simul. 14 (3) (2009) 863–879.
[7] Xiaobo Zhou, Xiaodong Wang, Edward R. Dougherty, Genomic networks: statistical inference from microarray data, Wiley Encyclopedia of Biomedical Engineering, 2006.
[8] Haiyan Huang, Chun-Chi Liu, Xianghong Jasmine Zhou, Bayesian approach to transforming public gene expression repositories into disease diagnosis databases, Proc. Natl. Acad. Sci. 107 (15) (2010) 6823–6828.
[9] Hidde De Jong, Modeling and simulation of genetic regulatory systems: a literature review, J. Comput. Biol. 9 (1) (2002) 67–103.
[10] Guy Karlebach, Ron Shamir, Modelling and analysis of gene regulatory networks, Nat. Rev. Mol. Cell Biol. 9 (10) (2008) 770–780.
[11] Stefan Bornholdt, Boolean network models of cellular regulation: prospects and limitations, J. R. Soc. Interface 5 (Suppl. 1) (2008) S85–S94.
[12] Kevin Murphy, Saira Mian, et al., Modelling gene expression data using dynamic Bayesian networks, Technical Report, Technical Report, Computer Science Division, University of California, Berkeley, CA, 1999.
[13] Wei-Po Lee, Kung-Cheng Yang, A clustering-based approach for inferring recurrent neural networks as gene regulatory networks, Neurocomputing 71 (4) (2008) 600–610.
[14] Ting Chen, Hongyu L. He, George M. Church, et al., Modeling gene expression with differential equations, Pacific Symposium on Biocomputing, vol. 4, 1999, p. 4.
[15] Fang-Xiang Wu, Wen-Jun Zhang, Anthony J. Kusalik, Modeling gene expression from microarray expression data with state-space equations, Pacific Symposium on Biocomputing, vol. 9, 2004, pp. 581–592.
[16] Yao Wang, Zidong Wang, Jinling Liang, On robust stability of stochastic genetic regulatory networks with time delays: a delay fractioning approach, IEEE Trans. Syst. Man Cybern. Part B Cybern. 40 (3) (2010) 729–740.
[17] Shuhei Kimura, Kaori Ide, Aiko Kashihara, Makoto Kano, Mariko Hatakeyama, Ryoji Masui, Noriko Nakagawa, Shigeyuki Yokoyama, Seiki Kuramitsu, Akihiko Konagaya, Inference of S-system Models of Genetic Networks Using a Cooperative Coevolutionary Algorithm, 2005.
[18] Ahsan R. Chowdhury, Madhu Chetty, Nguyen X. Vinh, Incorporating time-delays in S-system model for reverse engineering genetic networks, BMC Bioinformatics, vol. 14, 2013, p. 196.
[19] Minh Quach, Nicolas Brunel, Florence d'Alché Buc, Estimating parameters and hidden variables in non-linear state-space models based on odes for biological networks inference, Bioinformatics 23 (23) (2007) 3209–3216.
[20] Zidong Wang, Xiaohui Liu, Yurong Liu, Jinling Liang, Veronica Vinciotti, An extended Kalman filtering approach to modeling nonlinear dynamic gene regulatory networks via short gene expression time series, IEEE/ACM Trans. Comput. Biol. Bioinf. (TCBB) 6 (3) (2009) 410–419.
[21] Xiaodian Sun, Li Jin, Momiao Xiong, Extended Kalman filter for estimation of parameters in nonlinear state-space models of biochemical networks, PLoS One 3 (11) (2008) e3758.
[22] Lijun Qian, Haixin Wang, Edward R. Dougherty, Inference of noisy nonlinear differential equation models for gene regulatory networks using genetic programming and Kalman filtering, IEEE Trans. Signal Process. 56 (7) (2008) 3327–3339.
[23] Igor Podlubny, Geometric and Physical Interpretation of Fractional Integration and Fractional Differentiation. arXiv, Preprint Math/0110241, 2001.
[24] Dominik Sierociuk, Andrzej Dzieliński, Fractional Kalman filter algorithm for the states, parameters and order of fractional system estimation, Int. J. Appl. Math. Comput. Sci. 16 (2006) 129–140.
[25] Amina Noor, Erchin Serpedin, Mohamed Nounou, Hazem Nounou, Inferring gene regulatory networks via nonlinear state-space models and exploiting sparsity, IEEE/ACM Trans. Comput. Biol. Bioinf. (TCBB) 9 (4) (2012) 1203–1211.
[26] Ping Li, Roger Goodall, Visakan Kadirkamanathan, Estimation of parameters in a linear state space model using a Rao–Blackwellised particle filter, Control Theory and Applications, IEE Proceedings, vol. 151, IET, 2004, pp. 727–738.
[27] Alex Tremain Nelson, Eric A. Adviser-Wan, Nonlinear Estimation and Modeling of Noisy Time Series by Dual Kalman Filtering Methods, Oregon Graduate Institute of Science and Technology, 2000.
[28] Jane C. Schneider, Leonard Guarente, Regulation of the yeast *CYT1* gene encoding cytochrome c1 by HAP1 and HAP2/3/4, Mol. Cell. Biol. 11 (10) (1991) 4934–4942.
[29] Michal Ronen, Revital Rosenberg, Boris I. Shraiman, Uri Alon, Assigning numbers to the arrows: parameterizing a gene regulation network by using accurate expression kinetics, Proc. Natl. Acad. Sci. 99 (16) (2002) 10555–10560.
[30] I.A. Maraziotis, A. Dragomir, A. Bezerianos, Gene networks reconstruction and time-series prediction from microarray data using recurrent neural fuzzy networks, IET Syst. Biol. 1 (1) (2007) 41–50.
[31] Bin Yang, Yuehui Chen, Mingyan Jiang, Reverse engineering of gene regulatory networks using flexible neural tree models, Neurocomputing 99 (2013) 458–466.