RESEARCH   **Open Access**

A novel cost function to estimate parameters of oscillatory biochemical systems

Seyedbehzad Nabavi[*] and Cranos M Williams

Abstract

Oscillatory pathways are among the most important classes of biochemical systems with examples ranging from circadian rhythms and cell cycle maintenance. Mathematical modeling of these highly interconnected biochemical networks is needed to meet numerous objectives such as investigating, predicting and controlling the dynamics of these systems. Identifying the kinetic rate parameters is essential for fully modeling these and other biological processes. These kinetic parameters, however, are not usually available from measurements and most of them have to be estimated by parameter fitting techniques. One of the issues with estimating kinetic parameters in oscillatory systems is the irregularities in the least square (LS) cost function surface used to estimate these parameters, which is caused by the periodicity of the measurements. These irregularities result in numerous local minima, which limit the performance of even some of the most robust global optimization algorithms. We proposed a parameter estimation framework to address these issues that integrates temporal information with periodic information embedded in the measurements used to estimate these parameters. This periodic information is used to build a proposed cost function with better surface properties leading to fewer local minima and better performance of global optimization algorithms. We verified for three oscillatory biochemical systems that our proposed cost function results in an increased ability to estimate accurate kinetic parameters as compared to the traditional LS cost function. We combine this cost function with an improved noise removal approach that leverages periodic characteristics embedded in the measurements to effectively reduce noise. The results provide strong evidence on the efficacy of this noise removal approach over the previous commonly used wavelet hard-thresholding noise removal methods. This proposed optimization framework results in more accurate kinetic parameters that will eventually lead to biochemical models that are more precise, predictable, and controllable.

1 Introduction

Oscillatory biochemical pathways are an important class of biochemical systems [1,2] that play significant roles in living systems. For instance, "circadian rhythms" are fundamental daily time-keeping mechanisms in a wide range of species from unicellular organisms to complex eukaryotes [3]. One of their most important roles is in regulating physiological processes such as the sleep-wake cycle in mammals [4]. "Cell cycles" are also another vital class of biochemical oscillations. The cell cycle is the sequence of events by which a growing cell replicates all its components and divides into two daughter cells [5]. Inappropriate cell proliferation due to malfunctioning cell cycle control mechanisms can cause development of certain types of cancers [5]. There are also other classes of biochemical rhythms such as cardiac rhythms [6], ovarian cycles [7] and cAMP oscillations [8] that have their own significance in systems biology.

A complete modeling of a biochemical system includes characterization of all nonlinear structures of the network along with the associated kinetic rates. In other words, without fully identifying all the kinetic parameter values, these models are still incomplete even if the full structure of the model has been determined. Few kinetic rates are available directly from experimentation or literature. Most of them, however, have to be estimated by parameter fitting techniques to complete the modeling of the biochemical pathway. Thus, a mathematical framework is needed to fit the kinetic parameters using the observables. Optimization frameworks that focus specifically on estimating parameters

* Correspondence: snabavi@ncsu.edu
Department of Electrical and Computer Engineering, North Carolina State University, Raleigh, NC, USA

associated with biochemical pathways have received much attention in recent years [9-14].

Two main issues in estimating kinetic parameters in biochemical systems are data related issues and computational issues [14]. The measurement dataset used to fit these parameters are usually noisy and incomplete. Measurement datasets are also affected by uncertainties related to experimental conditions such as temperature and light [14]. Much study is done recently to reduce noise for different biochemical signals [15-17]. Mostacci et al. [15] proposed a denoising method for mass spectrometry data by integrating wavelet soft thresholding and principal component analysis. Weng et al. [16] suggested a noise removal approach for oscillatory ECG signals based on a recently developed method known as empirical mode decomposition. Ren et al. [17] also developed a method of denoising biochemical spectra by introducing a new thresholding function integrated with the "translation invariant" approach to lower the root mean square error (RMSE) in the measurements in comparison to the traditional soft and hard thresholding methods.

The computational issues include the challenges optimization algorithms face when identifying an optimal fit to measurement data. There are problems with optimization methods such as slow convergence toward global optima, complicated error surfaces and lack of convergence proofs [14]. Much study has been done to address these issues in parameter estimation in biochemical systems [12,13,18-21]. Zhan et al. proposed a method to reduce the computational time of each trial by integrating the spline functions theory with nonlinear programming to eliminate the need of solving the system of ordinary differential equations (ODEs) [21]. Rodriguez-Fernandez et al. [12] suggested a hybrid optimization method to speed up the convergence toward the global optima. A variety of different algorithms has also been adapted to perform the inverse problem. A comprehensive list of such studies is provided in [14].

Furthermore, heuristic approaches have been developed to address the optimization problem in fitting parameters in oscillatory systems [9-11]. These methods improved the optimization by constructing error functions based on the features extracted from the data. Locke et al. [11] proposed a cost function based on the comparison of entrained period, phase and strength of oscillation for the circadian clock in *Arabidopsis thaliana*. Also, Zeilinger et al. [10] performed another parameter estimation approach for the *A. thaliana* model by investigating amplitudes of some species in dark/light cycles, periods under dark and light conditions and the period of one mutant phenotype under constant light. In [9], Bagheri et al. built up an optimization process to model *Drosophila melanogaster* circadian clock by defining three cost functions based on free running period, light/dark entrained period, differences in amplitude and differences in the phase of the components in the system. These methods are more applicable for problems where characteristics in the system and/or data can be exploited to improve the performance of the parameter estimation. These methods, however, require more information about the system than purely data-driven comparison methods. For instance, the cost function proposed in [9] needs the period information of both the light and dark cycles of their investigated model, which requires a greater level of first principles knowledge. These methods are also model specific, which makes it difficult to apply them to general oscillatory systems. For example, the dark/light cycle characteristics that were introduced in parameter fitting problem of [10] may not be a suitable feature for parameter fitting of non-circadian biorhythms.

This article focuses on the problem of estimating the kinetic parameters in oscillatory biochemical systems. We show that periodicity in the measurements of oscillatory systems results in irregularly surface properties of the LS cost function leading to numerous local minima. These multiple local optima cause premature convergence of even robust optimization algorithms. This eventually results in incorrect estimates, bad predictions of dynamics, and incorrect acceptance of functional hypotheses. This, compounded with uncertainties or noisy measurements leads to a difficult estimation problem to solve.

We develop a parameter estimation framework to address these issues by integrating information of oscillatory systems in the modeling process (parameter estimation and denoising). This periodic information is used to build a cost function with better surface properties. Our proposed cost function takes advantage of the basic properties of these oscillatory systems, which allows us to generalize our cost function to a variety of biochemical systems with sustained oscillations. The proposed cost function also needs less first principles knowledge to generate the cost function in comparison to the previous methods that was developed for oscillatory systems [9-11]. We verified for three oscillatory biochemical systems that our proposed cost function results in increased ability to estimate accurate kinetic parameters as compared to the traditional LS cost function. We combined this cost 6 function with an improved denoising method that also leverages periodic characteristics embedded in the measurements to effectively reduce noise. The results provide strong evidence on the efficacy of this noise removal approach over the previous commonly used wavelet hard-thresholding noise removal method. This proposed optimization framework results in more accurate kinetic parameters that

will eventually lead to biochemical models that are more accurate, predictable, and controllable.

2 Methodology

This study considers deterministic, nonlinear oscillatory biochemical pathways described by ODEs as shown in (1):

$$\dot{\mathbf{x}}(t) = \mathbf{f}(\mathbf{x}(t), \mathbf{p}) \quad t_0 < t < t_e,$$
$$\mathbf{x}_0 = \mathbf{x}(t_0). \tag{1}$$

Here, $\mathbf{x} \in \mathbb{R}^{m \times 1}$ is the state vector of the m components of the pathway, $\mathbf{p} \in \mathbb{R}^{n \times 1}$ is the vector of n kinetic parameters, $\mathbf{f} \colon \mathbb{R}^{m \times 1} \to \mathbb{R}^{m \times 1}$ is a nonlinear vector function, $\mathbf{x}_0 \in \mathbb{R}^{m \times 1}$ is the vector of the initial component concentrations at time t_0 and $t_0 < t < t_e$ represents the time of interest.

Optimization describes the approach of estimating the kinetic parameters (**p**) of the system described in (1) that cannot be measured directly using a set of experimental data. The criteria for verifying the quality of the estimates is often an error function such as Φ as shown in (2). This function quantifies the ability of the estimates to reproduce the same results as the measurements. This objective function is minimized such that $\mathbf{p} = \hat{\mathbf{p}}$ results in the minimum value of Φ. In that way, $\hat{\mathbf{p}}$ is called the estimated point.

$$\hat{\mathbf{p}} = \arg \min_{\mathbf{p}} \Phi(\mathbf{p}) \tag{2}$$

One of the most common cost functions is the *least square (LS) estimator* [22]. This estimator is based on the sum of the squares of the point by point errors between measured experimental data and the simulated measurements from the estimated model as described in (3):

$$\Phi(\mathbf{p}) = \sum_{i=1}^{N_x} \sum_{j=1}^{N_m} (x_{ij} - \hat{x}_{ij}(\mathbf{p}))^2. \tag{3}$$

Here, x_{ij}, is the measurement at time j of the ith state of the system, $\hat{x}_{ij}$ is the reproduced data at time j for the ith state of the system given some parameter $\mathbf{p}$, N_m is the number of time points where measurements are obtained and N_x is the number of measured outputs (in this manuscript, they are considered to be the measured states of the system).

The objective of this article is to propose a method to estimate the kinetic parameters for a given oscillatory biochemical system of the form (1) using the noisy measurements of the system states. We first captured periodic information of the measurements. This information is used to improve noise reduction and generate an error cost function with better optimization properties. The next step implements a modified wavelet hard thresholding denoising approach that uses the previously obtained periodic information of the measurements to further reduce uncertainties in noisy data. We then generate our proposed cost function by integrating the periodic information obtained in the first step with the simulated data and measurements. We searched the surface of the proposed cost function with a series of optimizers in a hybrid manner. Hybrid methods use global optimization followed by local 8 [24]. We used a frequency-based method called optimization [12,23]. Global and local optimization algorithms were used in succession to further improve optimization results. A block diagram of our approach for parameter estimation of oscillatory systems is shown in Figure 1. The following sections outline each of the blocks of this diagram.

2.1 Fundamental frequency estimation

The fundamental frequency is an essential metric for assessing the underlying oscillatory characteristics in a signal and is a critical step in developing our proposed cost function and noise removal method. The fundamental frequency is the oscillation frequency of the continuous data. The measurements are samples of this continuous-time signal. If one assumes a periodic waveform of $x(t)$ such that:

$$x(t) = x(t + kT) \quad \forall k \in \mathbb{Z}, \tag{4}$$

the smallest value of $T \neq 0$ for which (4) is valid is the *"fundamental period"* of oscillation. The inverse of the fundamental period is the fundamental frequency (f_0). Several approaches has been proposed to estimate f_0 [24]. We used a frequency-based method called *component frequency ratio* [24] to extract the fundamental frequency of the measured data due to the fact that the time-series methods may not be adequate for biochemical measurements due to their low rate of sampling and low temporal resolution. This method starts with transforming the data to the Fourier domain by taking their Fourier transform. The locations of the peaks in the spectrum are then identified. The peaks in the frequency spectrum are the harmonics of the fundamental frequency. The final step is to find the greatest common factor of these frequencies in which peaks occur.
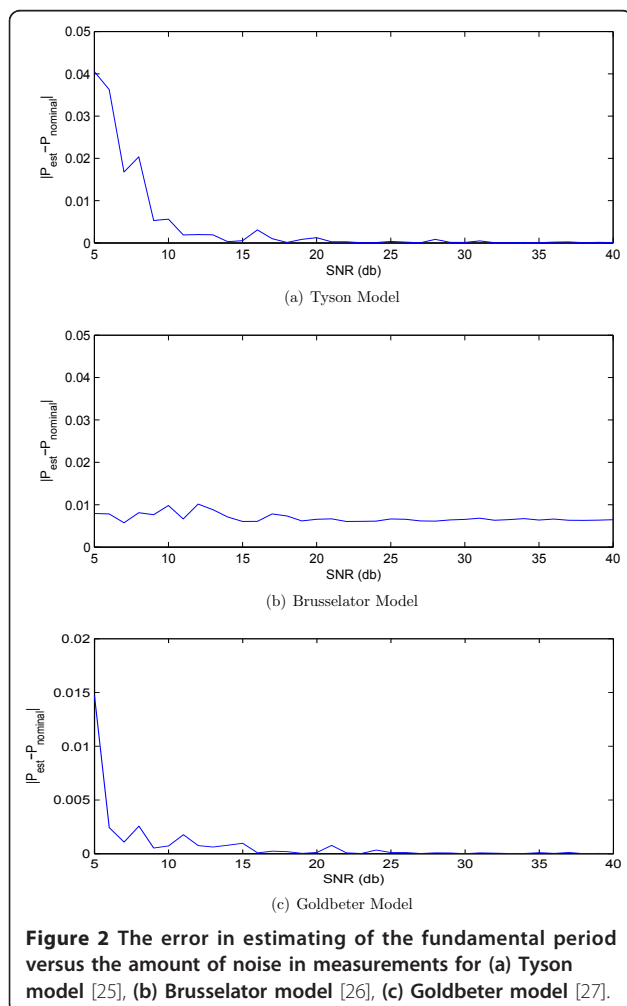

Figure 1 The implemented process of parameter estimation for oscillatory biochemical systems.
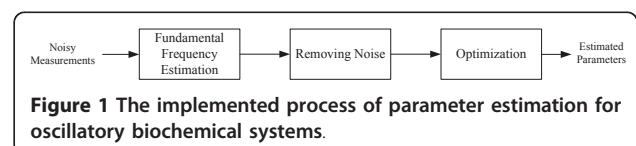
2.1.1 Effect of noise on estimation of f_0

This section investigates the effect of noise on estimation of f_0. We considered three model systems identified from the literature: the two-state Tyson model [25], the two-state Brusselator model [26] and the five-state Golbeter model [27]. We considered the measurements of the states of these models with the sampling rate equal to 1 (sample/hour). Then, we added AWGN noise with various SNRs to these signals and we estimate their fundamental period using the method *component frequency ratio*. Figure 2 shows the absolute error between the estimated and the nominal fundamental period of the three models for various amount of additive noise.

Figure 2 shows that the method used to estimate the fundamental period is robust enough to the additive noise.

2.2 Removing noise

One common approach to reduce noise in measurements is wavelet hardthresholding [28], which employs a thresholding function over the wavelet coefficients of the noisy data samples. The motivation of using wavelets is that it provides an appropriate basis to separate noise from signal in the wavelet domain. The small wavelet coefficients are more likely to be noise and large coefficients are more likely to be components of the original signal. Thus, noise could be eliminated approximately from the signal by thresholding the wavelet coefficients [29]. The steps of the noise removal procedure using this method are shown in Figure 3:

2.2.1 Improving the hard-thresholding method in oscillatory systems

Samples of oscillatory signals contain repetitive patterns if they are taken over multiple periods. Thus, we hypothesized that is possible to take advantage of data oscillation to improve the denoising of the samples provided that their fundamental frequency is given or can be estimated. We modified the denoising procedure of oscillatory signals by adding two additional steps to the traditional hardthresholding method as depicted in Figure 4. Two assumptions have been made about the noisy oscillatory data. First, the fundamental period of the data is not an integer multiple of the sampling rate. Otherwise, it is not possible to increase the resolution of the data by shifting them in this method. Second, we assumed to have the measurements of more than one period of the data. Otherwise, there will be no way to estimate the fundamental period of the measurements.

The first step in Figure 4 is shifting all samples to the first period of the data. This is based on the following steps:
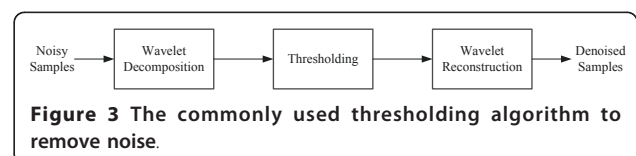
1. Partition the measurements $X(nT_s)$ based on their calculated fundamental period to the sets of X_k according to (5)

$$X_k(nT_s) = X(nT_s) \quad k\mathcal{T} \leq nT_s < (k+1)\mathcal{T}, \ 0 \leq k < m, \quad (5)$$

where T_s is the sampling period and $\mathcal{T}$ is the fundamental period of the measurements X.

2. Shift each x_k by the value of $k\mathcal{T}$. This will result in a single period of the measurements with higher resolution. The shifted versions of x_k's are calculated based on (6)

$$\bigcup_{k=0}^{N-1} x_k(nt - k\mathcal{T}) \quad (6)$$

Figure 2 The error in estimating of the fundamental period versus the amount of noise in measurements for (a) Tyson model [25], **(b) Brusselator model** [26], **(c) Goldbeter model** [27].

Figure 3 The commonly used thresholding algorithm to remove noise.

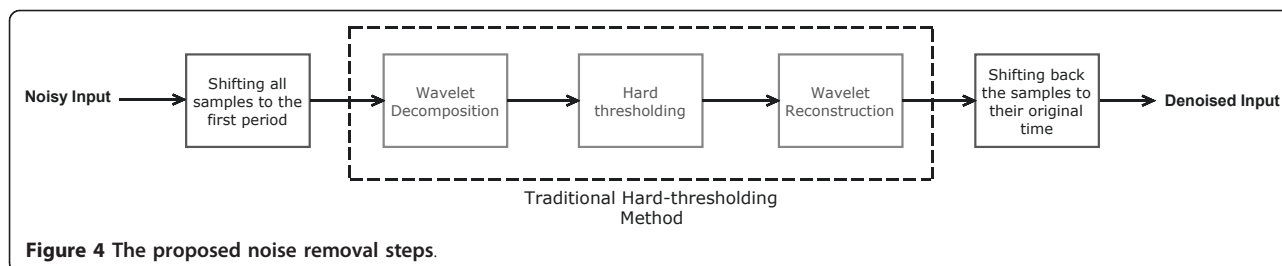**Figure 4 The proposed noise removal steps**.

Figure 5 illustrates the samples of a sine function of $x(t) = \sin 2\pi t$ with the rate of 2 (sample/sec) and its shifted version. We see that Figure 5b shows only the first period of the sine function but with higher resolution.
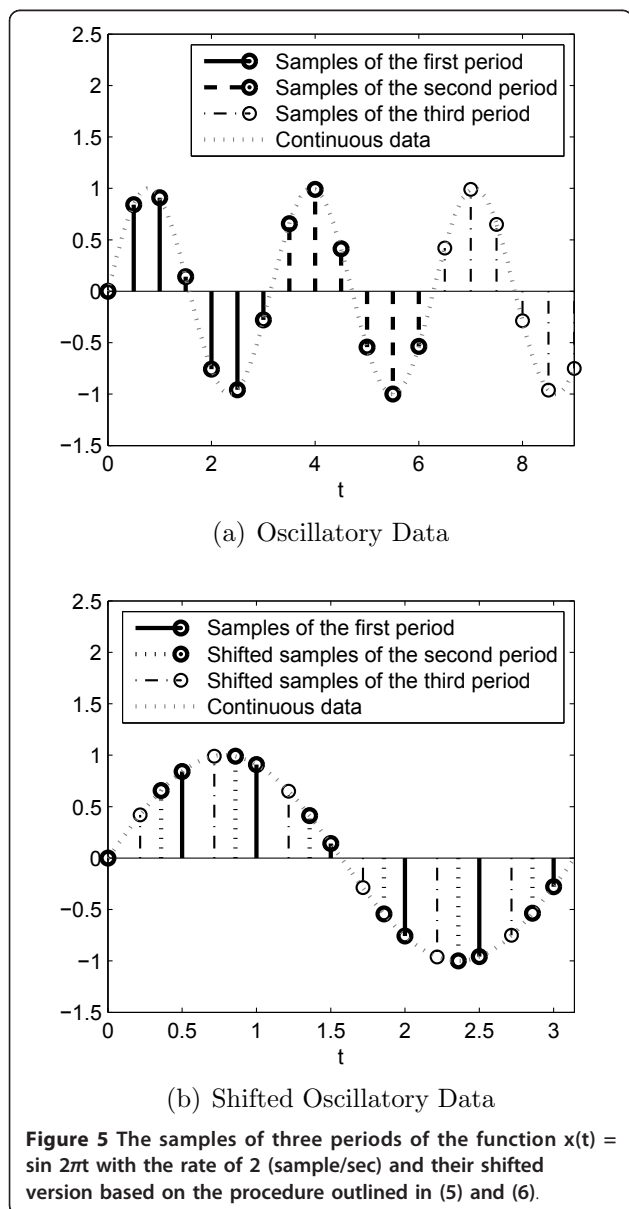

(a) Oscillatory Data

(b) Shifted Oscillatory Data

Figure 5 The samples of three periods of the function x(t) = sin 2πt with the rate of 2 (sample/sec) and their shifted version based on the procedure outlined in (5) and (6).

Figure 6 shows the denoising process for the *in-silico* measurements of the [M] component with sampling rate = 1 (sample/hour) in the model of the circadian clock in *D. melanogaster* proposed by Tyson [25]. The noise in the measurements is additive white Gaussian noise (AWGN) with SNR = 20 dB. Figure 6b shows the shifted version of the noisy measurements of Figure 6a using a calculated fundamental period of 24.21.

Wavelet decomposition, thresholding and reconstruction are then applied to this "shifted version" of the noisy data. MATLAB was used to implement a three level wavelet decomposition using the "*Daubechies 6*" wavelet and the threshold value equaling 0.3. The wavelet type, number of levels, and the threshold value were chosen empirically and may vary from system to system. The results are shown in Figure 6c. The final step is to reconstruct the original signal by shifting the samples back to their respective periods (Figure 6d).

We compared the performance of the proposed denoising method and the traditional wavelet hardthresholding by taking the samples of the [Pt] component with sampling rate = 1 (sample/hour) in the Tyson model of circadian clock in *D. Melanogaster* [25]. Then, we added AWGN noise with SNR = 20 to the dataset in 200 trials. We then removed noise using two approaches: the traditional wavelet hardthresholding method [29] and our proposed 12 method. Figure 7 compares three errors for each of the 200 trials. (1) The RMSE between the noisy data and the original dataset (the original error), (2) the RMSE between the denoised data resulting from the traditional thresholding method and the original dataset (Approach 1), and (3) the RMSE between the denoised data resulting from the proposed denoising method and the original dataset (Approach 2). This figure shows that our proposed method of denoising is more effective at removing noise than the wavelet hardthresholding method, consistently lowering the RMSE between the original signal and the denoised signal.

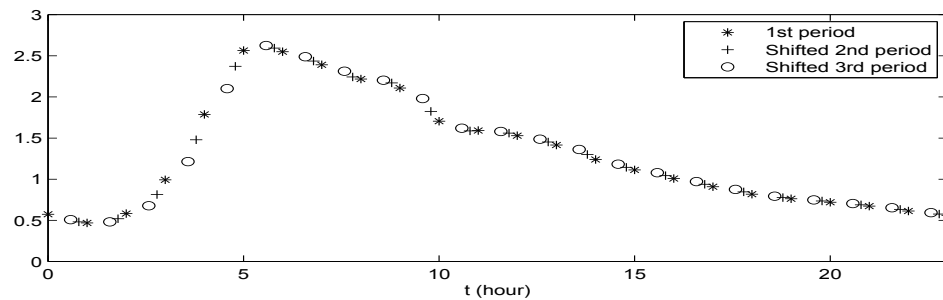### 2.2.2 The effect of error in estimating f_0 on proposed denoising method

This section investigates the impact of the inaccuracies of the fundamental period estimate on the proposed denoising method. We considered the samples of the
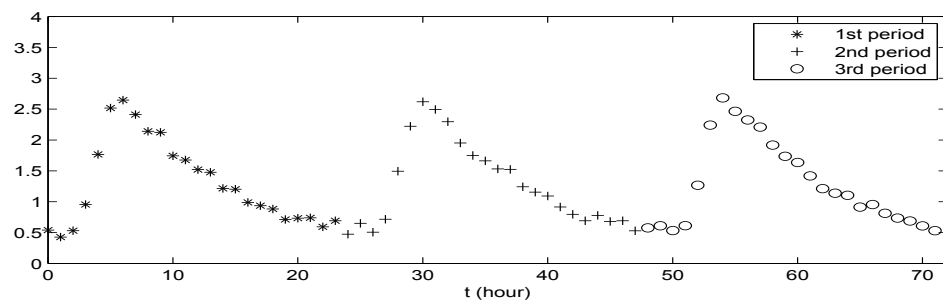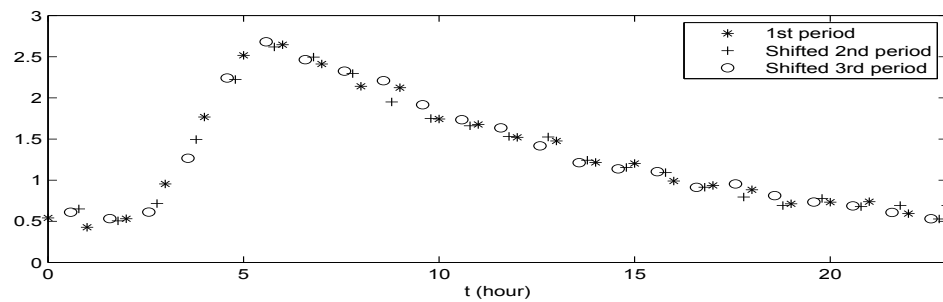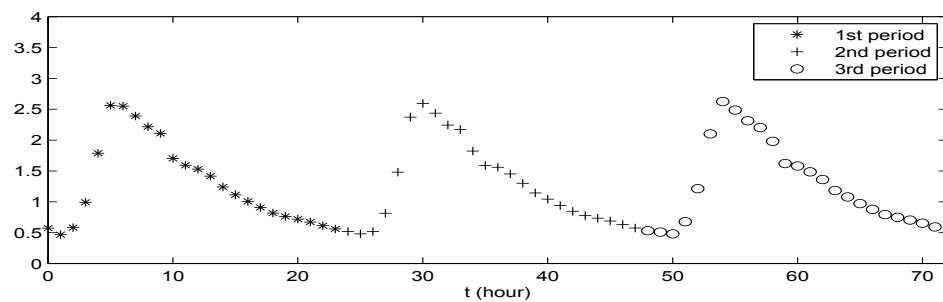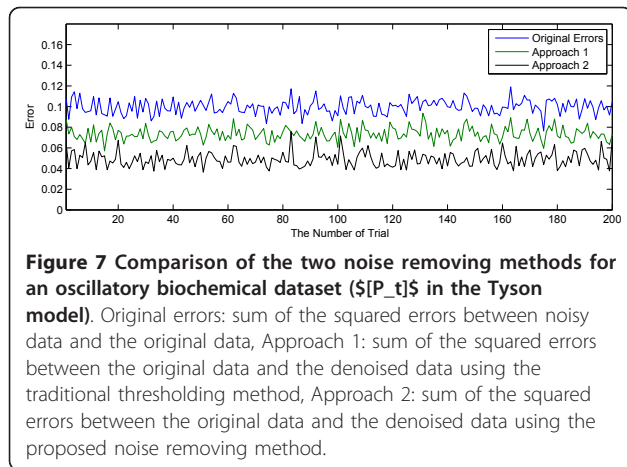
Figure 6 The proposed noise removal steps (a) the original noisy measurements, (b) shifted version of the noisy data based on fundamental period, (c) the thresholding results over shifted version of the data, (d) moving back all samples to their original time.

components of the Tyson [25], Brusselator [26], and Goldbeter model [27] with sampling rate = 1 (sample/hour) and AWGN noise with SNR = 20. Then, we denoised the data with using the traditional wavelet

thresholding (Approach 1) and the proposed denoising method (Approach 2) assuming inaccurate estimated fundamental period. Figure 8 compares the RMSEs of the results of these two methods and the noisy data

Figure 7 Comparison of the two noise removing methods for an oscillatory biochemical dataset ($[P_t]$ in the Tyson model). Original errors: sum of the squared errors between noisy data and the original data, Approach 1: sum of the squared errors between the original data and the denoised data using the traditional thresholding method, Approach 2: sum of the squared errors between the original data and the denoised data using the proposed noise removing method.

for ranges of inaccurate estimated fundamental periods.

Figure 8 shows that the results of the proposed denoising method has lower RMSEs than the traditional wavelet thresholding with small errors in the estimation fundamental period. However, if the fundamental period is estimated with errors approximately more than 0.25 for these models, the proposed method does not yield lower RMSEs. However, Figure 2 shows that the error in fundamental period estimation due to noise is much smaller than the order of error that is considered in Figure 8.
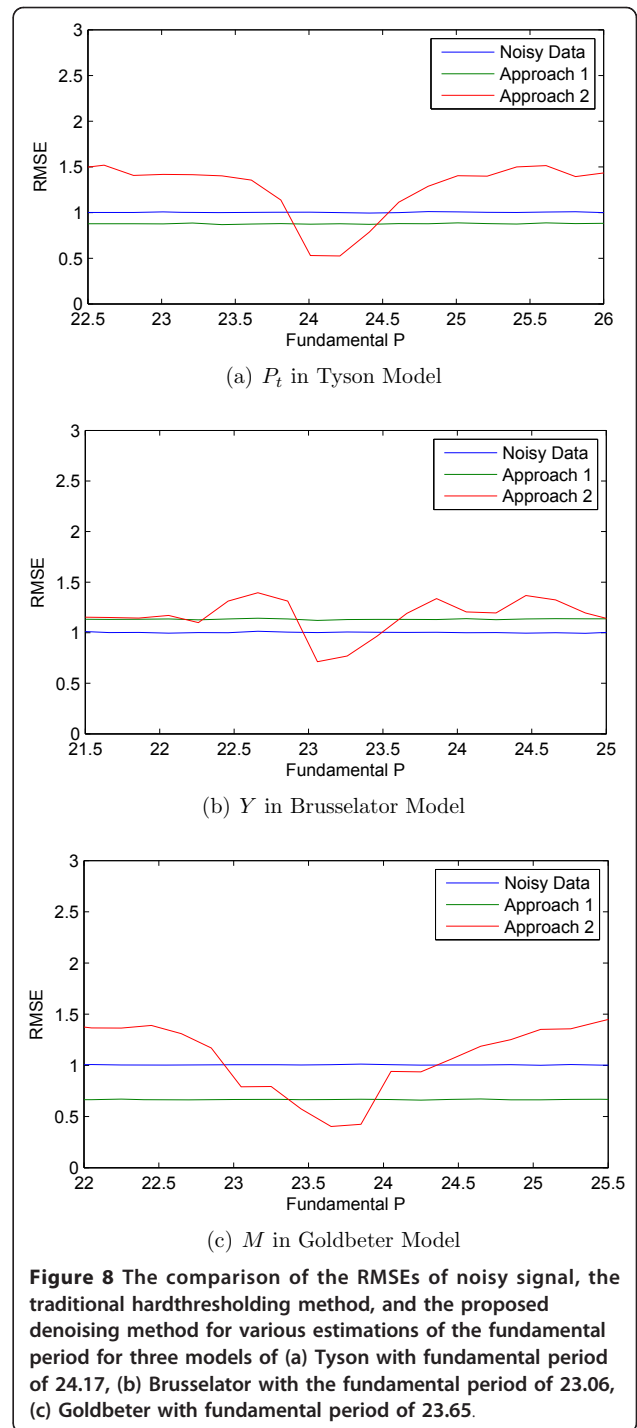
2.3 Optimization
2.3.1 Forming cost function
One big disadvantage of comparing point by point samples to build the LS cost function of (3) for oscillatory systems is the introduction of surface irregularities and numerous local optima. Let us consider a simple example of a sine function described in (7):
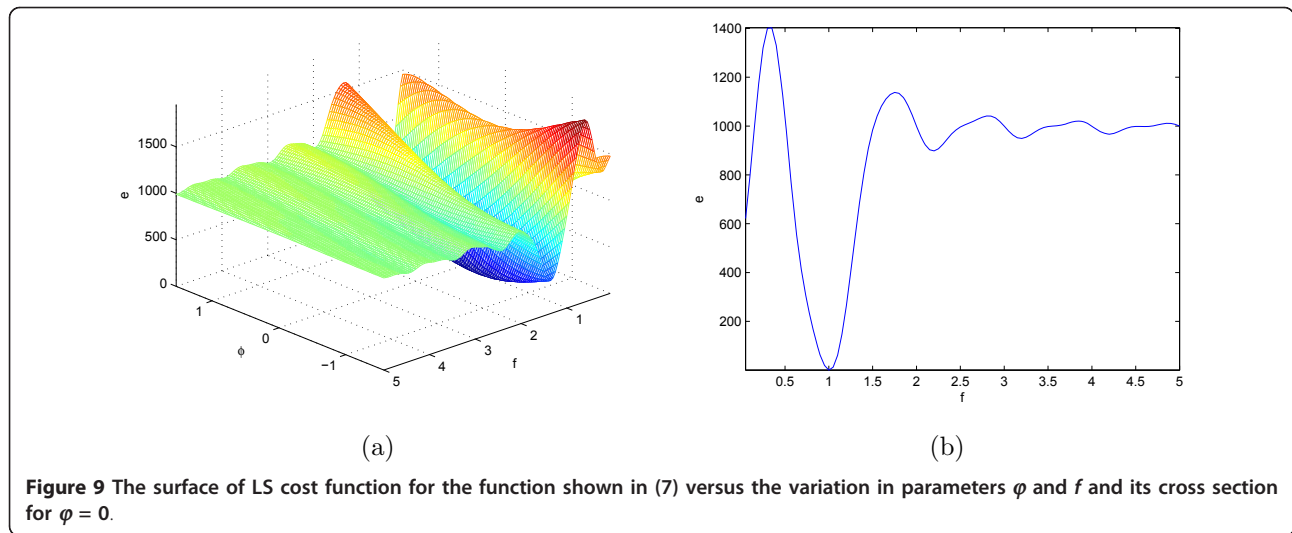
$$\gamma(n) = 1 + \sin(2\pi f n/1000 + \phi), \qquad (7)$$

where $f = 1$ is the frequency and $\varphi = 0$ is the initial phase. Figure 9 illustrates the surface of the LS cost function (3) for ranges of the signal parameters, f and Φ.

Figure 9 shows significant rippling especially along the f direction of the LS cost function. This happens due to the varying degree of overlap between various periods of two oscillatory signals in the LS objective function along the f axis. This potentially results in numerous local basins of attractions that hinder the optimizer's ability to find the global optimum. These ripples are fundamental characteristics of the LS cost function for systems with oscillatory dynamics. This phenomenon can be observed for a large class of oscillatory systems especially along the parameter axes to which the fundamental frequency is more sensitive.

(a) P_t in Tyson Model

(b) Y in Brusselator Model

(c) M in Goldbeter Model

Figure 8 The comparison of the RMSEs of noisy signal, the traditional hardthresholding method, and the proposed denoising method for various estimations of the fundamental period for three models of (a) Tyson with fundamental period of 24.17, (b) Brusselator with the fundamental period of 23.06, (c) Goldbeter with fundamental period of 23.65.
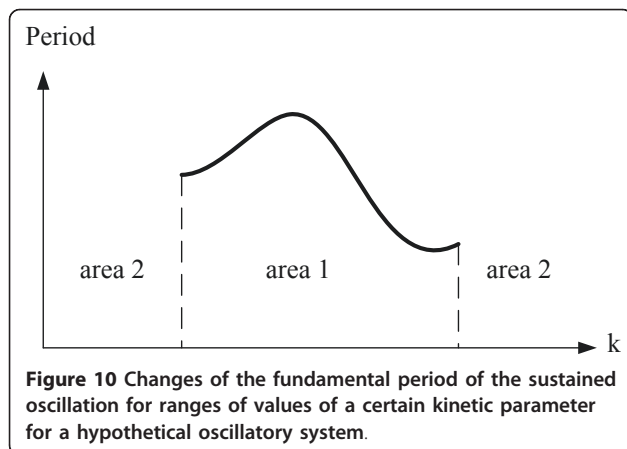
Thus, we hypothesize that we can leverage information embedded in the data to produce a cost function with better surface properties, resulting in fewer local minima. This function is constructed in a piecewise manner based on the oscillatory characteristics of the simulated data at various parameter values. These characteristics are divided into two cases: sustained

Figure 9 The surface of LS cost function for the function shown in (7) versus the variation in parameters φ and f and its cross section for $\varphi = 0$.

oscillations in the simulated data and no sustained oscillations in the simulated data. Sustained oscillation for a specific value of the parameter k is characterized by the fundamental period of the oscillations. A plot describing this is shown in Figure 10. All the parameter values for k in "area 1" produces sustained oscillations. This figure shows that the fundamental period of the sustained oscillation over this range may change. The values in "area 2", on the other hand, lead to dynamics that are not sustained oscillations.

If the simulated data are periodic, we introduced only the samples of one period of the data into the cost function. Likewise, only the samples of one period of the measurements will also be incorporated into this cost function. If the fundamental period of the measured data is not equal to the fundamental period of the simulated data, the signal with the smallest period is padded with zeros until the lengths of the signals are equal. This results in monotonic changes in error with respect to changes in fundamental period of the simulated data.

Figure 10 Changes of the fundamental period of the sustained oscillation for ranges of values of a certain kinetic parameter for a hypothetical oscillatory system.

If the simulated data are nonperiodic as in area two of Figure 10, all time point measurements and the simulated data will be included in the cost function, resulting in the same cost function as the traditional LS objective function. Equation (8) describes the new proposed cost function for the ODE-based model of an oscillatory biochemical pathway (1).

$$e(\mathbf{p}) = \sum_{i=1}^{N_x} \sum_{j=1}^{N_{z_i}} (z_{ij} - \hat{z}_{ij}(\mathbf{p}))^2, \tag{8}$$

where z_{ij} and $\hat{z}_{ij}$ for periodic $\hat{x}_i$ are calculated as:

$$z_{ij} = \begin{cases} x_{ij} & 0 \le t_j < T_i \\ 0 & T_i < t_j \le \max(T_i, \hat{T}_i) \end{cases}', \tag{9a}$$

$$\hat{z}_{ij} = \begin{cases} \hat{x}_{ij} & 0 \le t_j \le \hat{T}_i \\ 0 & \hat{T}_i < t_j \le \max(T_i, \hat{T}_i) \end{cases}. \tag{9b}$$

Otherwise, z_{ij} and $\hat{z}_{ij}$ for non-periodic $\hat{x}_i$ are calculated as:

$$z_{ij} = x_{ij} \tag{10a}$$

$$\hat{z}_{ij} = \hat{x}_{ij}. \tag{10b}$$

Here, x_{ij} is the measurement at time t_j of the ith state of the system, $\hat{x}_{ij}$ is the simulated data at time t_j for the ith state of the system. z_{ij} and $\hat{z}_{ij}$ are the truncated and zero padded x_{ij} and $\hat{x}_{ij}$, respectively, for the oscillatory $\hat{x}_i$. For non oscillatory $\hat{x}_i$, z_{ij}, and $\hat{z}_{ij}$ are equal to x_{ij} and $\hat{x}_{ij}$, respectively. N_{z_i} is the length of the z_i and $\hat{z}_i$. N_x is the number of states of the system, T_i is the fundamental period of the measurements (x_i), which was

computed using the component frequency ratio approach and $\hat{T}_i$ is the fundamental period of the simulated data $(\hat{x}_i)$, which is estimated for each candidate parameter value. $\hat{T}_i$ was estimated using the YIN approach [30], which is a modified version of the time-domain autocorrelation method.

Figure 11 illustrates how the proposed cost function compares two signals with different fundamental periods.

Figure 12 shows the surface of the proposed cost function of (8) for the sine function of (7). The global minimum of the proposed cost function also occurs at $f = 1$ and $\varphi = 0$ similar to the LS cost function of (3) shown in Figures 9. However, visual inspection of these two figures shows that the surface of proposed cost function is smoother than the surface of the LS cost function for the example of (7). We hypothesize that
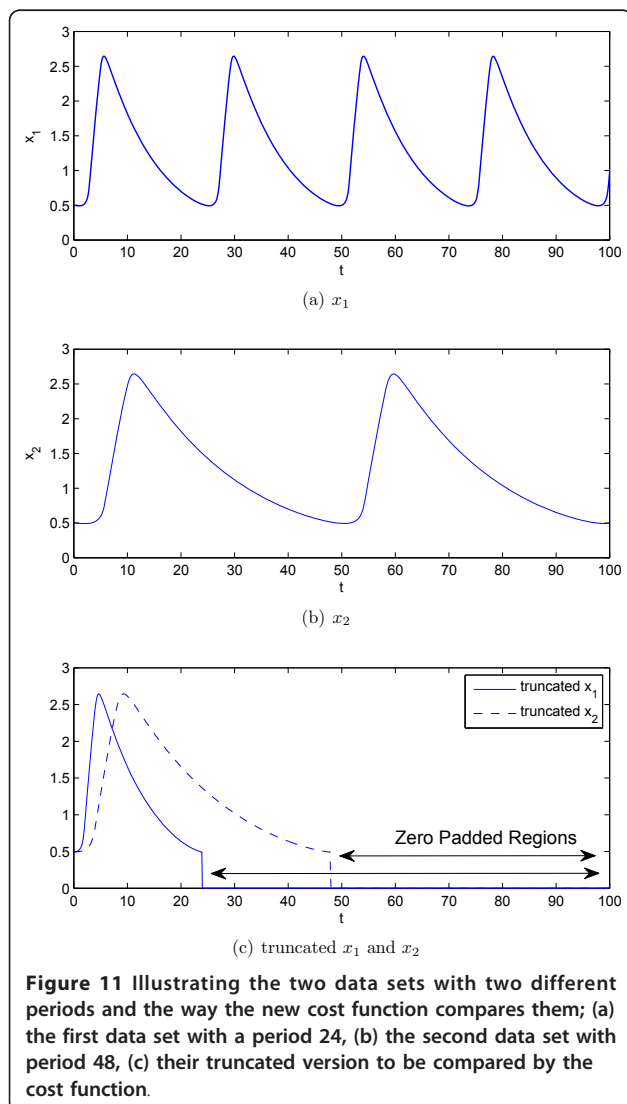
Figure 11 Illustrating the two data sets with two different periods and the way the new cost function compares them; (a) the first data set with a period 24, (b) the second data set with period 48, (c) their truncated version to be compared by the cost function.

this improvement of the cost function surface will improve the performance of the optimization search algorithm.

The effect of error in estimating f_0 on the performance of the cost function The performance of the proposed cost function (8) is not affected significantly by errors in the estimation of the fundamental frequency of the measurements. This is because of the fact that the measurements used in (8) have a certain sampling rate. Basically, if the error of the estimated fundamental period is small with respect to this sampling rate, it will not affect the number of samples that lies in one fundamental period of the data. Also, adding or reducing one sample in the summation of (8) obviously will not change the performance of the proposed cost function dramatically.

2.3.2 The optimization method
The optimization of the proposed cost function was performed using a hybrid approach. Hybrid methods, i.e. the combinations of global and local search methods, have been shown to yield results with smaller errors than global searches individually [12,23]. The global search algorithm that we adopt in this study is the "Genetic Algorithm", which is a widely-used approach of a class of global search methods called *evolutionary strategies* [31]. We used two consecutive local search methods of MATLAB [32] in this research. The first one was the derivative-based, constrained routine of fmincon, and the second one was the derivative-free routine of fminsearch that is based on the simplex algorithm [33].
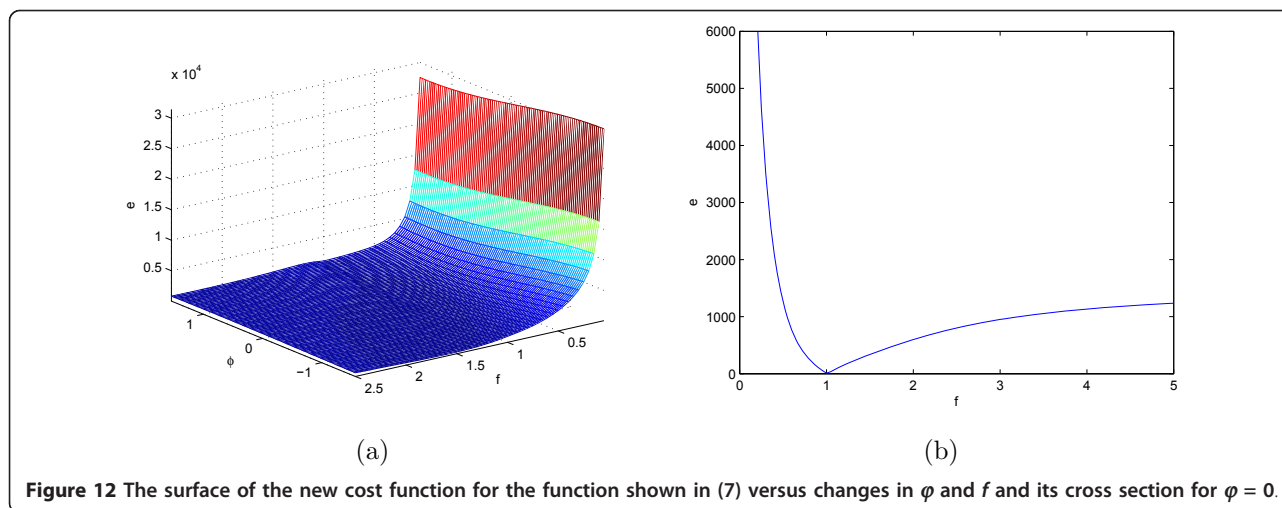
3 Results
This section shows the results of the optimization process that was illustrated in Figures 1 using two cost functions: the LS cost function of (3) and our proposed cost function of (8). We used three model of Tyson [25], Brusselator [26], and Goldbeter [27]. We compare 15 independent runs of the optimization process for parameter estimation for each oscillatory model. We add AWGN noise with SNR = 20 to the data. We use our proposed noise removal method to remove noise. The surface of the two cost functions will be shown and compared for these three systems. Results at all the intermediate steps of the optimization will be presented for each of the 15 runs:

1. The global optimization (MATLAB ga routine).
2. The first local optimization (MATLAB fmincon routine)
3. The second local optimization (MATLAB fminsearch routine)

3.1 Comparison of two different cost function
The two cost functions of (3) and (8) are two different functions of the kinetic parameters which do not

Figure 12 The surface of the new cost function for the function shown in (7) versus changes in φ and f and its cross section for $\varphi = 0$.

necessarily yield the same value for a given parameter set. Thus, a true comparison of the estimated parameters obtained from the two objective functions will require the LS score shown in (11) to equate the quality of the respective estimates. Equation (11) is basically the LS cost function summed only over the samples taken from the first fundamental period of the measurements. Introducing the measurements of only one period in computing the score creates a fair metric that shows the quality of estimated parameter sets.

$$\text{score}(\mathbf{k}) = \sum_{i=1}^{Nx} \sum_{j=1}^{N_{T_i}} \left(\hat{x}_{ij}(\mathbf{k}) - x_{ij} \right)^2. \tag{11}$$

Here, N_x, x_{ij}, and $\hat{x}_{ij}$ are defined as (3) and N_{T_i} is the number of samples that are extracted in $(0 < t < T_i)$ assuming T_i is the fundamental period of the x_i.

3.2 Parameter estimation results for two-state Tyson model

The two-state Tyson model (BIOMD0000000036 in Bio-Models database [34]) is a mathematical model of the circadian clock in wild-type fruit flies, *D. melanogaster*. This organism has circadian clocks similar to mice and bread molds. This model, shown in (12), consists of two states and nine kinetic parameters. The nominal values of the parameters of this system are shown in Table 1.

$$\frac{dM}{dt} = \frac{v_m}{1 + (P_t(1-q)/2P_{crit})^2} - k_m M \tag{12a}$$

$$\frac{dPt}{dt} = v_p M - \frac{k_{p1} P_t q + k_{p2} P_t}{J_p + P_t} - k_{p3} P_t \tag{12b}$$

$$q = \frac{2}{\sqrt{1 + 8K_{eq}P_t}} \tag{12c}$$

Figure 13 shows the surfaces of the LS cost function and the proposed cost function of (8) for pairwise combinations of parameters k_m and J_p and k_{p3} and P_{crit} over specific ranges. Characteristics of these parameters are representative of the characteristics of all kinetic parameters of the Tyson model (results are not shown). The values of the remaining parameters are held constant at their nominal values in all figures.

We see through visual inspection that our proposed cost function produces a smoother surface than that of the LS cost function for different values of the parameters k_m, k_{p3}, P_{crit}, and J_p. Figure 14 shows the cross-sections of the cost functions above (dashed lines) together with the fundamental period of the data (solid line) for ranges of values in the same order of magnitude as the nominal value.

Figure 14a shows that the system produces sustained oscillations only for k_m in the range [0.03 0.44]. The fundamental period of the sustained oscillations falls from 58 to 6.6 along this range. This radical change in the fundamental period produces irregularities in the LS cost function over this interval. However, the proposed cost function maintains good surface properties in spite of this extreme change in the fundamental period of the system. This emphasizes that our proposed cost function addresses the issue of surface irregularities of the LS cost function caused by introducing multiple periods of the data in calculating the error. Figure 14b shows similar results.

Figure 14c, d shows that the fundamental period for different values of P_{crit} is between 15.4 and 25.4 which is less than the changes in fundamental period that

Table 1 The results of optimization with minimum score for Tyson model.

Parameter	Nominal value	Estimation of the proposed cost function	Estimation of the LS cost function
v_m	1	**1.1372**	0.9472
k_m	0.1	0.1049	0.1097
v_p	0.5	0.4668	0.4740
k_{p1}	10	**15.88**	**21.48**
k_{p2}	0.03	**0.0936**	**0.0927**
k_{p3}	0.1	0.0766	0.0615
K_{eq}	200	**692.64**	**922.16**
P_{crit}	0.1	0.1076	**0.1477**
J_p	0.05	0.0511	**0.0738**
Score	0.1378	0.1084	0.2441

The values of kinetic parameters with minimum score derived from optimization using the proposed cost function and the LS cost function for the Tyson model. The bold values are the ones that are estimated incorrectly (error more than 10%). The score values are calculated based on (11)
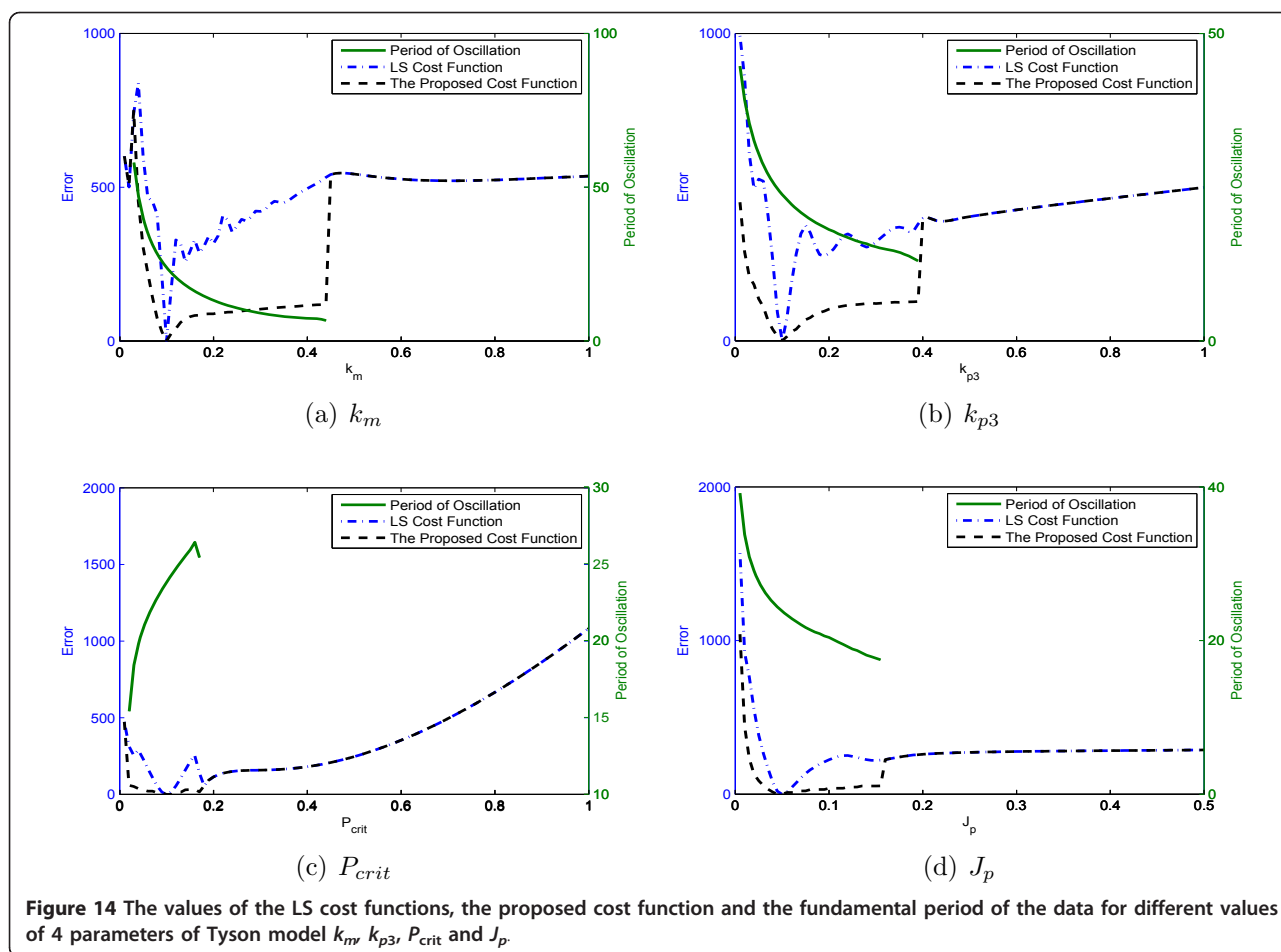
shown in Figure 14a, b. The LS cost function still shows varying levels of surface irregularities particularly along the P_{crit} axis. The proposed cost function again shows smoother surface characteristics under these conditions as well.

### 3.2.1 Results of parameter estimation

We assumed the measurements to be 100 samples of both $[M]$ and $[P_t]$ components with the rate of one

sample per hour and the AWGN noise of SNR = 20. We removed the noise using the proposed approach before the optimization step. The RMSE between the noisy samples and their real values of the samples were 0.0989. This was suppressed to 0.0413 after denoising. The population size was set to 200 and number of generations equals 50 for the ga routine. We calculated N_{T_i} from (11) to be 24 for the Tyson model. The computed

(a) LS cost function versus k_m and J_p

(b) The proposed cost function versus k_m and J_p

(c) LS cost function versus k_{p3} and P_{crit}

(d) The proposed cost function versus k_{p3} and P_{crit}

Figure 13 The comparison of the surface of the two cost functions, left column: LS cost function, right column: the proposed cost function, for different values of k_m, J_p, k_{p3} and P_{crit} in Tyson system, while changing two values of parameters, the rest of parameters are locked in their nominal value.

Figure 14 The values of the LS cost functions, the proposed cost function and the fundamental period of the data for different values of 4 parameters of Tyson model k_m, k_{p3}, P_{crit} and J_p.

scores for the estimated parameters from the 15 runs of optimization are shown in Figure 15 at the three steps of the hybrid optimization process. The mean, median and the minima of the computed scores at each level for the two cost functions are also shown in Table 2. Figure 15 and Table 2 show visually and numerically that the
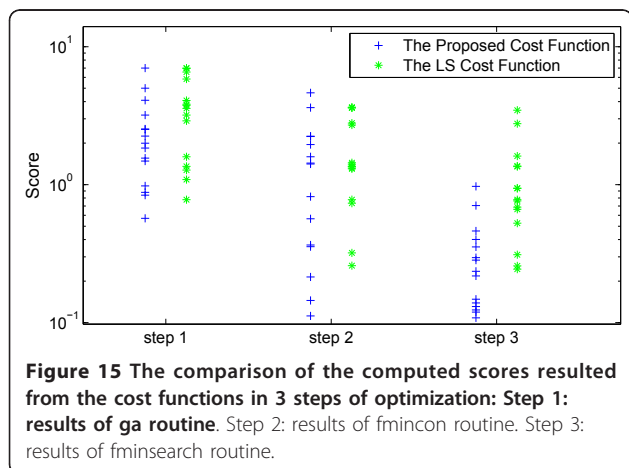
optimization routine performs better using the proposed cost function than the LS cost function at all steps. These results are also consistent with our visual inspections of the cost functions in Figures 13 and 14.

The optimized results with the lowest score out of 15 runs for the LS cost function and the proposed cost function are shown in Table 1.

The estimate results in the lowest score using noise-free measurements produces six of nine kinetic parameters with less than 10% errors (results not shown). Table 1 shows that the noisy case results in four of nine estimated parameters with more than 10% error. In both cases, proposed cost function yields more accurate

Figure 15 The comparison of the computed scores resulted from the cost functions in 3 steps of optimization: Step 1: results of ga routine. Step 2: results of fmincon routine. Step 3: results of fminsearch routine.

Table 2 Statistics of optimization results for Tyson model.

	Step 1		Step 2		Step 3	
	Proposed	**LS**	**Proposed**	**LS**	**Proposed**	**LS**
Mean	2.4497	3.5838	1.4465	1.7760	0.3131	1.1116
Median	1.9998	3.5731	1.4117	1.3595	0.2354	0.7788
Min	0.5706	0.7788	0.1118	0.2589	0.1084	0.2441

Mean, median, and minimum of the score values shown in Figure 15

results in comparison to the LS cost function. The large number of inaccuracies for the noisy case is more a result of system sloppiness versus inaccuracies of the estimation procedure [35,36], which results in 21 a wide range of parameters with similar system dynamics. It is evident that our proposed cost function was able to produce better overall system dynamics than the traditional LS cost function, which is clearly conveyed by the lower overall error. Our proposed method, similar to the LS cost function, only takes into account the accuracy of dynamics. Thus, the sloppiness can results in moderate level of parameter accuracy. Recently, Apgar et al. proposed an experiment design framework to improve estimates of sloppy parameters in biochemical models [37]. This, however, is beyond the scope of this article.

3.3 Parameter estimation for two-state Brusselator model

The Brusselator model was proposed by Prigogine for theoretical analysis of autocatalytic reactions [26]. This model consists of two states and four kinetic parameters as shown in (13). The nominal values of the parameters of this system are shown in Table 3.

$$\frac{dX}{dt} = k_1 A + k_2 X^2 Y - k_3 BX - k_4 X, \tag{13a}$$

$$\frac{dY}{dt} = -k_2 X^2 Y + k_3 BX, \tag{13b}$$

$$A = 0.5, \quad B = 3. \tag{13c}$$

Figure 16 shows the values of the two cost functions together with the fundamental period of the data (the green trajectories) for different values of four parameters of the system.

Figure 16a shows the fundamental period of sustained oscillation falls from 45.9 to 4.3 for k_1 in the range [0.7 2.8]. This change in the fundamental period again produce irregularities in the LS cost function over this interval. The proposed cost function, on the other hand, maintains good surface properties in spite of this change in the fundamental period of the 22 system. This further verifies that the proposed cost function is able to address the irregularities of the LS cost function

resulting from sustained dynamics embedded in the dynamics used to evaluate the cost function.

3.3.1 Results of parameter estimation

This section shows the results of 15 runs of optimization for the Brusselator model using 100 samples of only [Y] component with sampling rate = 1 (sample/hour) and AWGN noise with SNR = 20. We removed the noise using the proposed denoising approach. The RMSE between the noisy samples and their real values was 0.0971, which was suppressed to 0.0570 after denoising. The population size was set to 100 and the number of generations equals 50. The computed scores for the estimated parameters from the 15 runs of optimization are shown in Figure 17 and Table 4. We calculate $N_{T_i} = 23$ for calculating the score of (11) for the Brusselator model. The results again demonstrate visually and numerically that the optimization routine performs better using the proposed cost function than the LS cost function in all steps even in presence of noise. These results are also consistent with our visual inspections of the cost functions in Figure 16.

The derived results with the lowest score out of 15 runs for the LS cost function and the proposed cost function are in Table 3.

Table 3 shows that the resulting overall error for the proposed cost function is lower than that of the LS cost function. All four parameters were estimated incorrectly using the LS cost function, while they were estimated almost accurately using the proposed cost function.
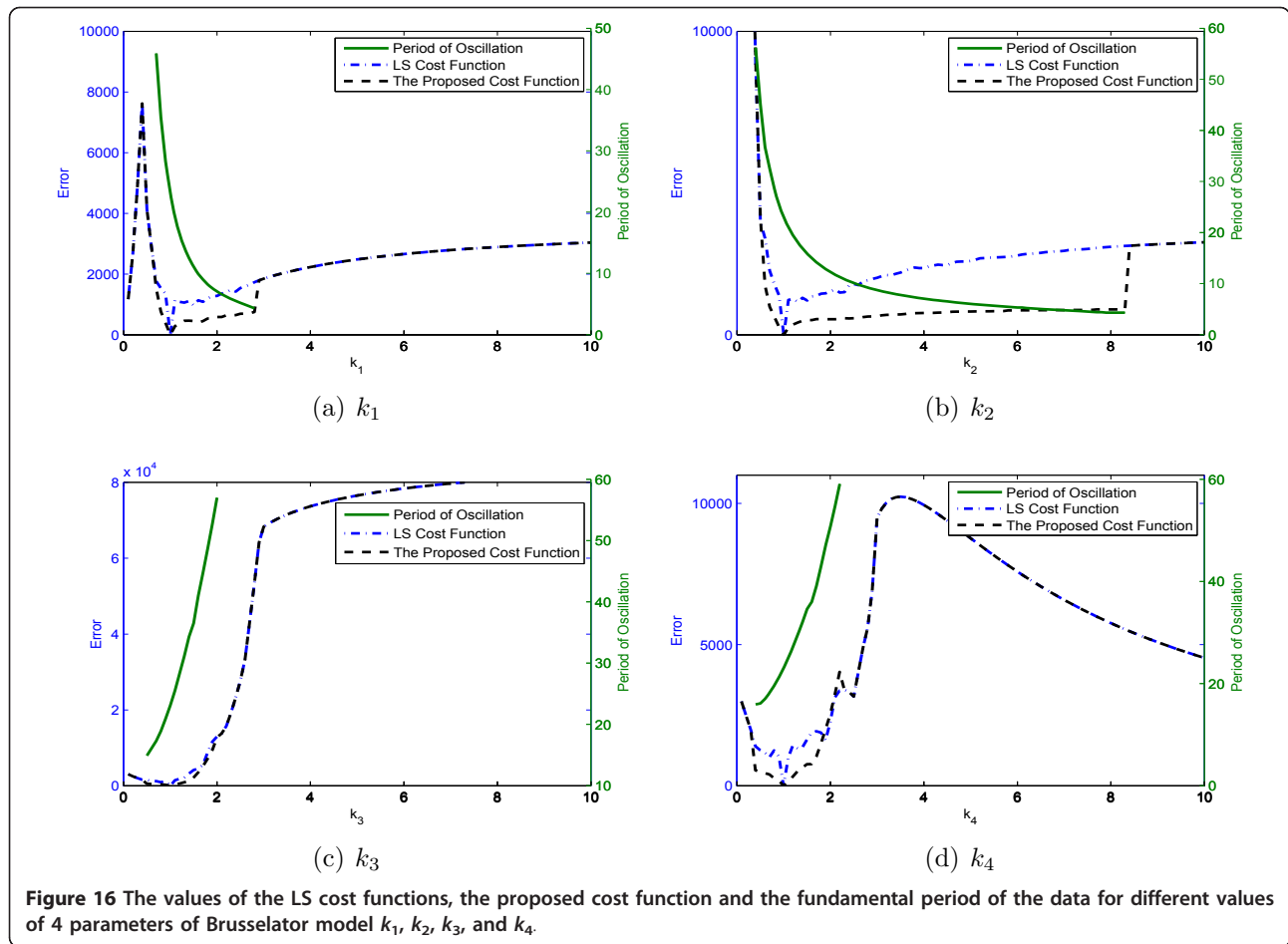
3.4 Parameter estimation results for five-state Goldbeter model

The *D. melanogaster* circadian model of Goldbeter [27] was investigated in the third study. This model is also available in BioModels database [34] (BIOMD0000000016). Here, the circadian oscillations of PER is modeled with five states: PER mRNA [M], PER protein [P0], the mono-phosphorylated form [P1], the bi-phosphorylated form [P2] and nuclear PER [PN]. This five-state model has 18 kinetic parameters. The ODE model of the system is shown in 14. The nominal values of the 18 kinetic parameters of this system are available in Table 5.

Table 3 The results of optimization with minimum score for Brusselator model.

Parameter	Nominal value	Estimation of the proposed cost function	Estimation of the LS cost function
k_1	1	0.9912	**1.4064**
k_2	1	0.9112	**0.8492**
k_3	1	0.9526	**0.8944**
k_4	1	0.9335	**1.6732**
Score	0.3128	0.2763	0.7619

The values of kinetic parameters with minimum score derived from optimization using the proposed cost function and the LS cost function for the Brusselator model. The bold values are the ones that are estimated incorrectly (error more than 10%)
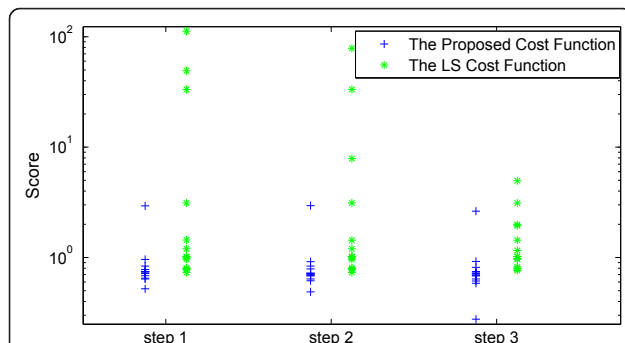
Figure 16 The values of the LS cost functions, the proposed cost function and the fundamental period of the data for different values of 4 parameters of Brusselator model k_1, k_2, k_3, and k_4.

$$\frac{dM}{dt} = v_s \frac{K_I^n}{K_I^n + P_N^n} - v_m \frac{M}{K_m + M} \tag{14a}$$

$$\frac{dP_0}{dt} = k_s M - V_1 \frac{P_0}{K_1 + P_0} + V_2 \frac{P_1}{K_2 + P_1} \tag{14b}$$

$$\frac{dP_1}{dt} = V_1 \frac{P_0}{K_1 + P_0} - V_2 \frac{P_1}{K_2 + P_1} - V_3 \frac{P_1}{K_3 + P_1} - V_4 \frac{P_2}{K_4 + P_2} \tag{14c}$$

$$\frac{dP_2}{dt} = V_3 \frac{p_1}{K_3 + p_1} - V_4 \frac{P_2}{K_4 + P_2} - k_1 P_2 + k_2 P_N - v_d \frac{P_2}{K_d + P_2} \tag{14d}$$

$$\frac{dPN}{dt} = k_1 P_2 - k_2 P_N \tag{14e}$$

Figure 18 shows the values of the two cost function together with the fundamental period of the data (the green trajectories) along different values of four parameters of the system.

Figure 17 The comparison of the computed scores resulted from the cost functions in 3 steps of optimization: Step 1: results of ga routine. Step 2: results of fmincon routine. Step 3: results of fminsearch routine.

Table 4 Statistics of optimization results for Brusselator model.

	Step 1		Step 2		Step 3	
	Proposed	**LS**	**Proposed**	**LS**	**Proposed**	**LS**
Mean	0.8857	14.8167	0.8688	9.4746	0.8177	1.5517
Median	0.7336	1.0179	0.7116	1.0179	0.7107	1.0179
Min	0.5207	0.7323	0.4879	0.7323	0.2763	0.7619

Mean, median, and minimum of the score values shown in Figure 17

Table 5 The Result of Optimization with Minimum Score for Goldbeter Model.

Parameter	Nominal value	Estimation of the proposed cost function	Estimation of the LS cost function
v_s	0.76	0.6980	**0.6275**
K_I	1	0.9996	0.9400
n	4	**4.9920**	**5.7232**
v_m	0.65	0.5972	**0.5559**
K_m	0.5	0.5056	**0.7412**
k_s	0.38	0.3677	0.3732
v_1	3.2	**3.8093**	3.1552
K_1	2	**2.6488**	1.8288
v_2	1.58	**3.1221**	**1.2836**
K_2	2	**4.4760**	**1.1952**
v_3	5	**8.8000**	4.5100
K_3	2	**4.6696**	2.0120
v_4	2.5	**5.6410**	**2.8120**
K_4	2	**7.0128**	**3.0704**
v_d	0.95	0.9713	0.9614
K_d	0.2	**0.2413**	**0.2250**
k_1	1.9	1.7541	**2.1944**

The values of kinetic parameters with minimum score derived from optimization using the proposed cost function and the LS cost function for the five-state Goldbeter model. The bold values are the ones that are estimated incorrectly (error more than 10%)

It could be seen in all figures that the changes in period of the oscillation does not produce significant irregularities in the LS cost function surface, which is different than previous examples. Figure 18b, for instance, shows the changes of period for k_2 in the interval [0.4 2]. However, there are not multiple basins of attractions along the k_2 direction in spite of these changes in fundamental period. This is due to the fact that the LS cost function changes over orders of magnitudes along this parameter direction in a way that the produced ripples has little effect on the monotonicity of the LS cost function. This extreme change in the LS cost function (approximately from 400 to 2200 for k_2 over the interval [0.4 2]) happens because the peak to peak magnitude of the sustained oscillations of the simulated data also increases in order of magnitudes along this parameter direction. For example, the peak of the $[P_2]$ increases from 0.25 to 1.5 for k_2 over the interval [0.4 2].

The proposed cost function still shows good surface characteristics although it was not much different than the already favorable characteristics of the LS cost function. Thus, it is expected that both of these cost functions would perform almost similar in the optimization process.

3.4.1 Parameter estimation results
This section shows the results of 15 optimization runs using 100 samples of $[M]$, $[P_0]$, $[P_1]$, $[P_2]$, and $[P_N]$ components with the sampling rate = 1 (sample/hour) and AWGN noise with SNR = 20. We suppresse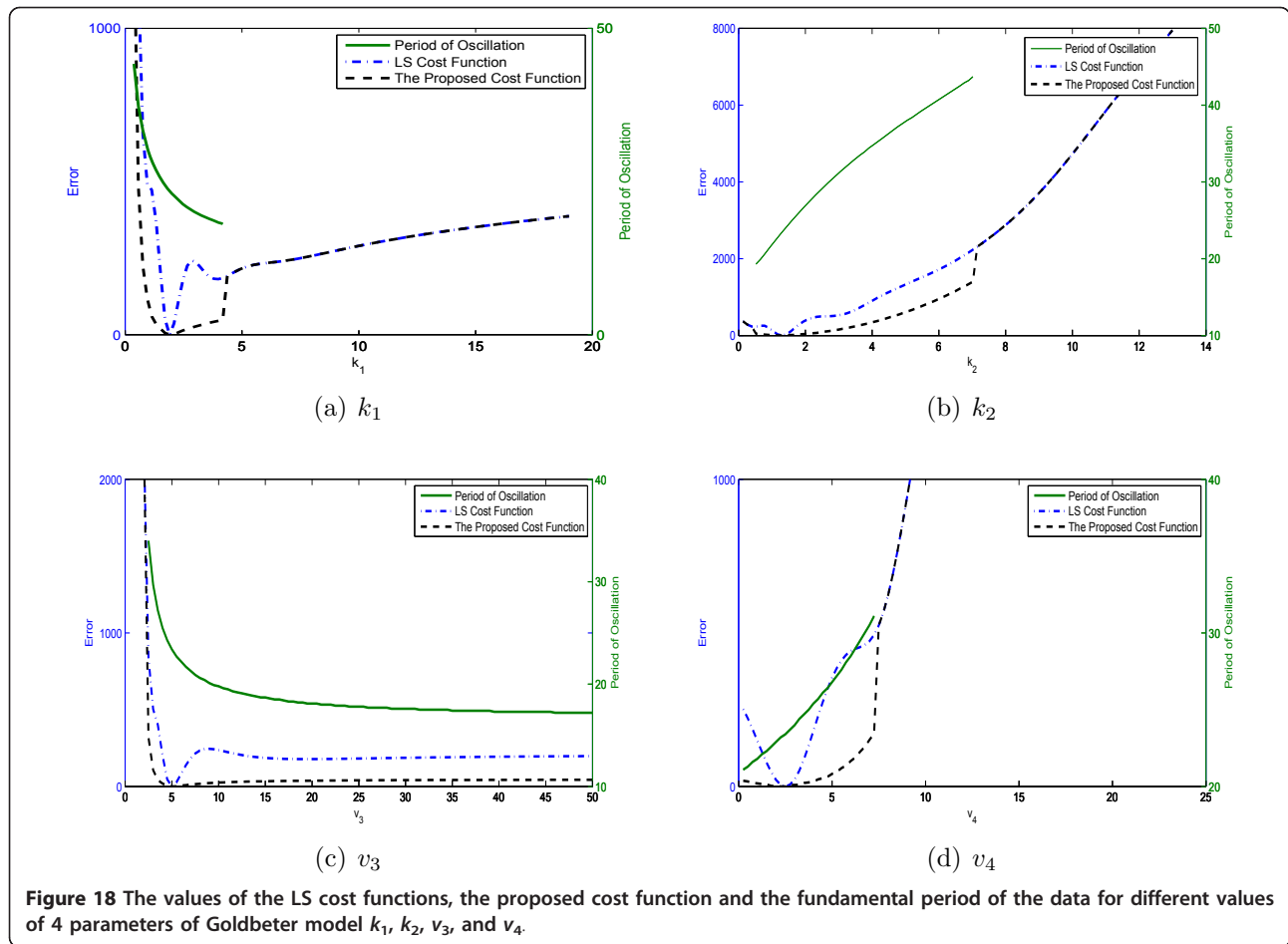d the noise using the proposed denoising approach. The RMSE between the noisy samples and their real values were 0.1012, which was suppressed to 0.04906 after denoising. We calculated $N_{T_i} = 23$ for the score in (11). The results of 15 optimization runs are shown in Figure 19 and Table 6. This shows that the performances of the LS cost function and the proposed cost function are almost the same in all steps. These results are also consistent with our visual inspections of the cost functions in Figure 18.

The derived results with the minimum score out of 15 runs for the LS cost function and the proposed cost function are shown in Table 5.

Table 5 shows that 8 out of 18 parameters were estimated within 10% of their nominal value for the proposed cost function as opposed to 7 out of 18 for the LS cost function. This shows a wide range of parameters have similar dynamics. This is due to system sloppiness that was also mentioned for the Tyson model. Our proposed cost function takes into account the accuracy of dynamics, which is similar to the LS cost function. Therefore, this may results in moderate accuracy in parameter values because of the sloppiness.
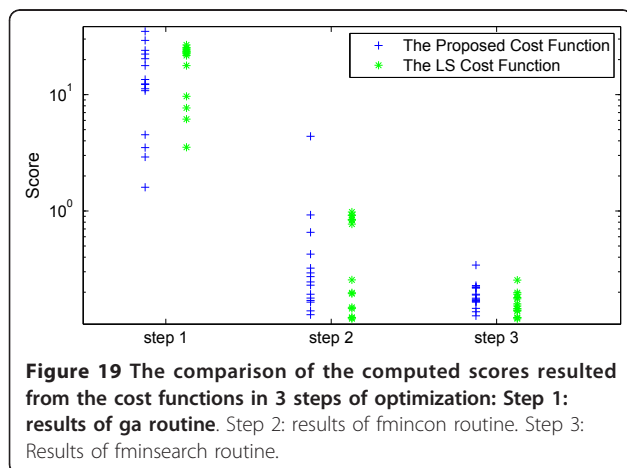
4 Conclusions
This article addresses the issue of kinetic parameter estimation in oscillatory biochemical systems. We showed that the LS cost function for oscillatory systems results in surface characteristics that potentially hinder the performance of optimization routines used to estimate kinetic parameters. Thus, we suggested a new cost function with more favorable surface properties which leads

Figure 18 The values of the LS cost functions, the proposed cost function and the fundamental period of the data for different values of 4 parameters of Goldbeter model k_1, k_2, v_3, and v_4.

to improved results for parameter estimation. This cost function integrates temporal information with periodic information embedded in measurements used to estimate these parameters. This generalized cost function also needs less first principles knowledge to generate the cost function in comparison to the previous developed

methods for oscillatory systems. We tested our cost function using three benchmark oscillatory biochemical pathways and compared our proposed objective function with the traditional LS cost function in several optimization runs using noisy measurements. The comparison of the results verified that the optimization performed more effectively using our 26 proposed cost function as compared to the traditional LS cost function. Furthermore, we introduced a wavelet hardthresholding approach for noise removal. This modified approach is able to suppress noise in oscillatory data better than the traditional wavelet thresholding approach. This, together with the proposed objective function will result in more

Figure 19 The comparison of the computed scores resulted from the cost functions in 3 steps of optimization: Step 1: results of ga routine. Step 2: results of fmincon routine. Step 3: Results of fminsearch routine.

Table 6 Statistics of optimization results for Goldbeter model.

	Step 1		Step 2		Step 3	
	Proposed	**LS**	**Proposed**	**LS**	**Proposed**	**LS**
Mean	14.7428	18.8913	0.5812	0.4939	0.1914	0.1585
Median	12.3454	22.7683	0.2456	0.2553	0.1778	0.1446
Min	1.5970	3.5281	0.1282	0.1183	0.1255	0.1183

Mean, median and minimum of the score values shown in Figure 19

accurate kinetic parameters that will eventually lead to biochemical models that are more precise, predictable and controllable. There are, however, unsolved issues with sloppiness of biochemical pathways [35,36], which require further investigation especially for oscillatory biochemical pathways.

Competing interests

The authors declare that they have no competing interests.

References

1. A Goldbeter, *Biochemical Oscillations and Cellular Rhythms the Molecular Bases of Periodic and Chaotic Behaviour*, (Cambridge University Press, Cambridge, 1996)
2. C Fall, E Marland, J Tyson, *Computational Cell Biology*, (Springer, New York, 2002)
3. J Perez-Martin, Growth and development eukaryotes. Current Opinion Microbiol. **13**(6), 661–662 (2010). doi:10.1016/j.mib.2010.10.007
4. J Yan, H Wang, Y Liu, C Shao, Analysis of gene regulatory networks in the mammalian circadian rhythm. PLos Comput Biol. **4**(10), e1000193 (2008). doi:10.1371/journal.pcbi.1000193
5. K Collins, T Jacks, N Pavletich, The cell cycle and cancer. PNAS: Proc Natl Acad Sci. **94**(7), 2776–2778 (1997). doi:10.1073/pnas.94.7.2776
6. J Boullin, JM Morgan, The development of cardiac rhythm. Heart. **91**(7), 874–875 (2005). doi:10.1136/hrt.2004.047415
7. J Perry, *The Ovarian Cycle of Mammals*, (Oliver and Boyd, Edinburgh, 1971)
8. M Zaccolo, T Pozzan, cAMP and Ca2+ interplay: a matter of oscillation patterns. Trends Neurosci. **26**(2), 53–55 (2003). doi:10.1016/S0166-2236(02)00017-6
9. N Bagheri, M Lawson, J Stelling, F Doyle, Modeling the Drosophila melanogaster circadian oscillator via Phase optimization. J Biol Rhythms. **23**(6), 525–537 (2008). doi:10.1177/0748730408325041
10. M Zeilinger, E Farre, S Taylor, S Kay, F Doyle, A novel computational model of the circadian clock in Arabidopsis that incorporates PRR7 and PRR9. Mol Syst Biol. **2**(58) (2006)
11. J Locke, A Millar, M Turner, Modelling genetic networks with noisy and varied experimental data the circadian clock in Arabidopsis thaliana. J Theor Biol. **234**(3), 383–393 (2005). doi:10.1016/j.jtbi.2004.11.038
12. M Rodriguez-Fernandez, P Mendes, J Banga, A hybrid approach for efficient and robust parameter estimation in biochemical pathways. BioSystems. **83**, 248–265 (2005)
13. V Vyshemirsky, M Girolami, Bayesian ranking of biochemical system models. Bioinformatics. **24**(6), 833–839 (2008). doi:10.1093/bioinformatics/btm607
14. IC Chou, E Voit, Recent developments in parameter estimation and structure identification of biochemical and genomic systems. Math Biosci. **219**(2), 57–83 (2009). doi:10.1016/j.mbs.2009.03.002
15. E Mostacci, C Truntzer, H Cardot, P Ducoroy, Multivariate denoising methods combining wavelets and principal component analysis for mass spectrometry data. Proteomics. **10**(14), 2564–2572 (2010). doi:10.1002/pmic.200900185
16. G Tang, A Qin, ECG de-noising based on empirical mode decomposition. *The 9th International Conference for Young Computer Scientists, 2008. ICYCS* 903–906 (2008)
17. Z Ren, G Liu, L Zeng, Z Huang, S Huang, Research on biochemical spectrum denoising based on a novel wavelet threshold function and an improved translation-invariance method. Proc SPIE. **7280**, 72801Q (2008)
18. M Sugimoto, S Kikuchi, M Tomita, Reverse engineering of biochemical equations from time-course data by means of genetic programming. Biosystems. **80**(2), 155–164 (2005). doi:10.1016/j.biosystems.2004.11.003
19. O Gonzalez, C Kuper, K Jung, JP Naval, E Mendoza, Parameter estimation using simulated annealing for S-system models of biochemical networks. Bioinformatics. **23**(4), 480–486 (2007). doi:10.1093/bioinformatics/btl522
20. P Flaherty, M Radhakrishnan, T Dinh, R Rebres, T Roach, M Jordan, A Arkin, A dual receptor crosstalk model of g-protein-coupled signal transduction. PLoS Comput Biol. **4**(9), e1000185 (2008). doi:10.1371/journal.pcbi.1000185
21. C Zhan, L Yeung, Parameter estimation in systems biology models using spline approximation. BMC Syst Biol. **5**(14) (2011)
22. D Marquardt, An algorithm for least squares estimation of nonlinear parameters. SIAM J Appl Math. **11**(2), 431–441 (1963). doi:10.1137/0111030
23. J Renders, S Flasse, Hybrid methods using genetic algorithms for global optimization. IEEE Trans Syst Man Cybernet Part B, Cybernet. **26**(2), 243–258 (1996). doi:10.1109/3477.485836
24. D Gerhard, Pitch extraction and fundamental frequency history and current techniques. Department of Computer Science, University of Regina, Regina, Canada (2003)
25. J Tyson, C Hong, D Thron, B Novak, A simple model of circadian rhythm based on dimerization and proteolysis of PER and TIM. Biophys J. **77**, 2411–2417 (1999). doi:10.1016/S0006-3495(99)77078-5
26. D Kondepudi, I Prigogine, *Modern Thermodynamics from Heat Engines to Dissipative Structures*, (Wiley, Chichester, 1998)
27. A Goldbeter, A model for circadian oscillations in the drosophila period protein (PER). Proc Royal Soc B, Biol Sci. **261**(1362), 319–324 (1995). doi:10.1098/rspb.1995.0153
28. S Mallat, *A Wavelet Tour of Signal Processing*, (American Press, San Diego, 1998)
29. S Mallat, A theory for multiresolution signal decomposition: the wavelet representation. IEEE Pattern Anal Mach Intell. **11**(7), 674–693 (1989). doi:10.1109/34.192463
30. A Cheveigne, H Kawahara, Yin, a fundamental frequency estimator for speech and music. J Acoust Soc Am. **111**(4), 1917–1930 (2002). doi:10.1121/1.1458024
31. C Moles, P Mendes, J Banga, Parameter estimation in biochemical pathways: a comparison of global optimization methods. Genome Res. **13**(11), 2467–2474 (2003). doi:10.1101/gr.1262503
32. Inc TM, MATLAB: version 7.6.0. Natick Massachusetts (2008)
33. J Lagarias, J Reeds, M Wright, P Wright, Convergence properties of the Nelder-Mead simplex method in low dimensions. SIAM J Optim. **9**, 112–147 (1998). doi:10.1137/S1052623496303470
34. N Le Novère, B Bornstein, A Broicher, M Courtot, M Donizelli, H Dharuri, L Li, H Sauro, M Schilstra, B Shapiro, JL Snoep, M Hucka, BioModels Database: a free, centralized database of curated, published, quantitative kinetic models of biochemical and cellular systems. Nucleic Acids Res. **34**(suppl 1), D689–D691 (2006)
35. R Gutenkunst, J Waterfall, F Casey, K Brown, C Myers, J Sethna, Universally Sloppy Parameter Sensitivities in Systems Biology Models. PLos Comput Biol. **3**(10), 1871–1878 (2005)
36. J Waterfall, F Casey, R Gutenkunst, K Brown, C Myers, P Brouwer, V Elser, J Sethna, Sloppy-model universality class and the Vandermonde matrix. Phys Rev Lett. **97**(15), 150601 (2006)
37. J Apgar, D Witmer, F Whitead, B Tidor, Sloppy models, parameter uncertainty, and the role of experimental design. Mol BioSyst. **6**(10), 1890–1900 (2010). doi:10.1039/b918098b