Parameter Inference for Biochemical Systems that Undergo a Hopf Bifurcation

Paul D. W. Kirk, Tina Toni, and Michael P. H. Stumpf
Division of Molecular Biosciences, Imperial College, London, England

ABSTRACT The increasingly widespread use of parametric mathematical models to describe biological systems means that the ability to infer model parameters is of great importance. In this study, we consider parameter inferability in nonlinear ordinary differential equation models that undergo a bifurcation, focusing on a simple but generic biochemical reaction model. We systematically investigate the shape of the likelihood function for the model's parameters, analyzing the changes that occur as the model undergoes a Hopf bifurcation. We demonstrate that there exists an intrinsic link between inference and the parameters' impact on the modeled system's dynamical stability, which we hope will motivate further research in this area.

INTRODUCTION

Mathematical models in the modern biomedical and life sciences have proved to be particularly useful (1), not only because they can help to increase understanding of the modeled biological system and allow predictions to be made, but also because their use forces model assumptions to be stated explicitly and fully (2). However, there are many challenges associated with mathematical modeling in biology (3). Perhaps the most fundamental of these are model specification (i.e., formulating an appropriate model based upon prior understanding and observations) and—in the case of parametric models—parameter estimation. It is with the latter that this study is concerned.

Once a parametric model has been proposed, it is necessary to infer or estimate its parameters based upon experimental data. Parameter estimation can often prove difficult, with missing data, observational noise, and incomplete model specification being among the numerous problems encountered in practice. However, rather than focusing on these practical issues (which are often highly and subtly investigation-specific), we are here concerned with how the underlying dynamics of the modeled system can affect our ability to perform inference. In particular, we address the apparent need (see, for example, (4)) for further investigation into parameter inferability in nonlinear ordinary differential equation models that undergo a bifurcation.

Informally, bifurcation refers to a phenomenon by which a small change in parameter values may cause a significant and global qualitative change in dynamical behavior. In this study, we shall be principally concerned with how parameter inferability changes in deterministic nonlinear systems that undergo a Hopf bifurcation, although the methods presented here could just as well be applied to investigations into other types of bifurcation. The Hopf bifurcation is of especial interest, however, as it is often cited as a cause of oscillatory behavior in biological systems, and recent research has demonstrated that Hopf bifurcations may occur in (to name but a few) metabolic networks regulated by product-feedback inhibition (5), the plant mitochondrial tricarboxylic acid cycle (6), and models for gene expression (7). The use of global dynamical features is perhaps most prominently illustrated by studies of the yeast cell-cycle (8), where a sequence of Hopf bifurcations (for timing of events) and saddle-point bifurcations (which provide switches taking the system from one stage to the next) appear to capture the underlying dynamics. Since models that exhibit Hopf bifurcations are being proposed and applied to biological systems, it is prudent to investigate the particular challenges that may be faced when inferring their parameters. In practice, we may frequently have to estimate such parameters from finite—often even very short—time series data (e.g., from quantitative Western blots). Fitting a model is preferable to simply measuring (for example) half-times from simple regression analyses, as these do not allow us to account for global features—such as bifurcations—of a biological system's dynamics.

THEORY

We consider a particular chemical reaction system with a Hopf bifurcation (9). This example was formulated as the mathematically simplest reaction system to undergo a Hopf bifurcation, and is well supported by existing mathematical analyses in the literature (9,10). The relative simplicity and tractability of this reaction system makes it particularly suitable for our investigation, while retaining generality and biochemical relevance.

The system involves four reactants (A^0, X_1, X_2, and X_3) and is described by the mechanism shown in Fig. 1. Note that A^0 denotes outer reactants whose concentration, A, is assumed to be constant (see (9) for details).

The system may be described by the ordinary differential equation (ODE) model

$$\begin{aligned}\dot{x}_1 &= (\kappa_A - k_4)x_1 - k_2x_1x_2\\ \dot{x}_2 &= -k_3x_2 + k_5x_3\\ \dot{x}_3 &= k_4x_1 - k_5x_3,\end{aligned} \tag{1}$$

Submitted January 9, 2008, and accepted for publication March 7, 2008.

Address reprint requests to Michael P. H. Stumpf, E-mail: m.stumpf@imperial.ac.uk.

Editor: Steven D. Schwartz.

0006-3495/08/07/540/10 $2.00

doi: 10.1529/biophysj.107.126086

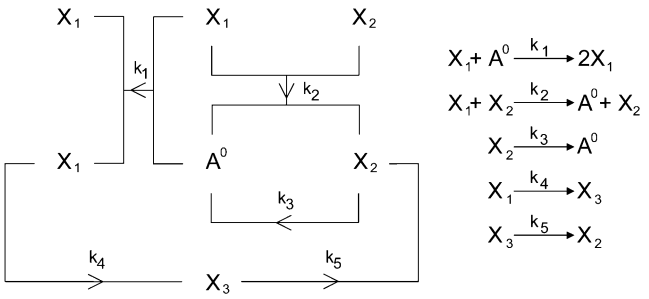

FIGURE 1 (*Left*) Representation of the mechanism of the reaction (adapted from (9)). (*Right*) The reaction described as a system of chemical equations. Each of the k_i ($i = 1, \ldots, 5$) values is a rate constant.

where each x_i is a function of time, t, and represents the concentration of X_i at time t; $\kappa_A = k_1 A$ is the product of the concentration of outer reactants, A, and the rate parameter k_1; and each $k_i > 0$ denotes a rate constant. For our purposes, the values κ_A, k_2, k_3, k_4, and k_5 are regarded as the parameters of the system. For convenience, we shall frequently employ the vector notation $\mathbf{x}(t) = [x_1(t), x_2(t), x_3(t)]$, write the parameters as $\boldsymbol{\theta} = [\kappa_A, k_2, k_3, k_4, k_5]^\top$, and refer to our ODE system using the shorthand $\dot{\mathbf{x}} = g(\mathbf{x}|\boldsymbol{\theta})$. In the usual way, using vector notation also allows us to represent the progression of the system through time as a trajectory in three-dimensional space.

It can easily be shown (see (9)) that this system has two stationary points, located at

$$\mathbf{x}^{(0)} = [0, 0, 0]^\top, \tag{2}$$

$$\mathbf{x}^{(\mathrm{H})} = \left[\left(\frac{\kappa_A - k_4}{k_2 k_4}\right)k_3, \frac{\kappa_A - k_4}{k_2}, \left(\frac{\kappa_A - k_4}{k_2 k_5}\right)k_3\right]^\top. \tag{3}$$

The first of these is stable provided $0 \leq \kappa_A < k_4$, and the second is stable for $k_4 < \kappa_A < k_3 + k_4 + k_5$. We shall be solely concerned with $\mathbf{x}^{(\mathrm{H})}$, which undergoes a supercritical Hopf bifurcation as κ_A passes through the value $k_3 + k_4 + k_5$ (9,10). Fig. 2 (adapted from (11)) provides a simplified phase portrait representation of this three-dimensional supercritical Hopf bifurcation, illustrating the qualitative changes in the dynamics of the system as κ_A passes through the bifurcation point $k_3 + k_4 + k_5$.

Maximum likelihood inference for ODEs

For our mathematical model in Eq. 1, we aim to infer the values of the parameters from a set of observed data. That is, given a set $D = \{\mathbf{y_1}, \ldots, \mathbf{y_M}\}$ of observations at times $t_1, \ldots, t_M$, we wish to infer the true parameter vector $\boldsymbol{\theta}$*****.

Here we consider maximum likelihood estimation of the parameters, which requires us to specify a likelihood function for our model. Given that the ODE is a deterministic and not a probabilistic description of the system, we cannot establish a direct likelihood approach; instead, we define the likelihood through the error between predicted and actual value (analogous to regression/least-squares procedures). Following the likelihood approach for ODE parameter estimation outlined in Williams and Kalogiratou (12) (which is itself a specific case of more general maximum likelihood analysis; see (13), for example), we assume for any observed data point $\mathbf{y_i}$ that $\mathbf{y}_i \sim N(\boldsymbol{\mu}_i(\boldsymbol{\theta}), \Sigma_i)$. Here, $\boldsymbol{\mu}_i(\boldsymbol{\theta})$ is the solution to the system $\dot{\mathbf{x}} = g(\mathbf{x}|\boldsymbol{\theta})$ evaluated at time $t = t_i$, and Σ_i is a covariance matrix. We shall further assume that $\Sigma_i = \Sigma$ is the same for all i.

Assuming independence between observations (an assumption that is straightforwardly relaxed, but helps us to keep notation to a minimum), we therefore obtain the following form for the likelihood function, $L(\boldsymbol{\theta}|D)$, which tells us how the probability of observing the data set $D = \{\mathbf{y_1}, \mathbf{y_2}, \ldots, \mathbf{y_M}\}$ changes with $\boldsymbol{\theta}$:

$$L(\theta|D) = \prod_{i=1}^{M} \frac{1}{(2\pi)^{3/2}|\Sigma|^{1/2}} \times \exp\left(-\frac{1}{2}(\mathbf{y}_i - \mu_i(\theta))^\top \Sigma^{-1}(\mathbf{y}_i - \mu_i(\theta))\right). \tag{4}$$

We are particularly interested in the maximum likelihood estimate for the parameter vector, $\hat{\boldsymbol{\theta}}_{\mathrm{ML}}$, that maximizes the likelihood function, and shall investigate this via the log-likelihood, $\ln(L(\boldsymbol{\theta}|D))$. In our case, we have

$$\ln(L(\theta|D)) = M \ln\left(\frac{1}{(2\pi)^{3/2}|\Sigma|^{1/2}}\right) - \frac{1}{2}\sum_{i=1}^{M}(\mathbf{y}_i - \mu_i(\theta))^\top \Sigma^{-1}(\mathbf{y}_i - \mu_i(\theta)), \tag{5}$$

$$\propto -\frac{1}{2}\sum_{i=1}^{M}(d_\Sigma(\mathbf{y}_i, \mu_i(\theta)))^2, \tag{6}$$

where $d_\Sigma(\mathbf{y}_i, \boldsymbol{\mu}_i(\boldsymbol{\theta})) = \sqrt{(\mathbf{y}_i - \boldsymbol{\mu}_i(\boldsymbol{\theta}))^\top \Sigma^{-1}(\mathbf{y}_i - \boldsymbol{\mu}_i(\boldsymbol{\theta}))}$ is the Mahalanobis distance (14) between $\mathbf{y_i}$ and μ_i (with

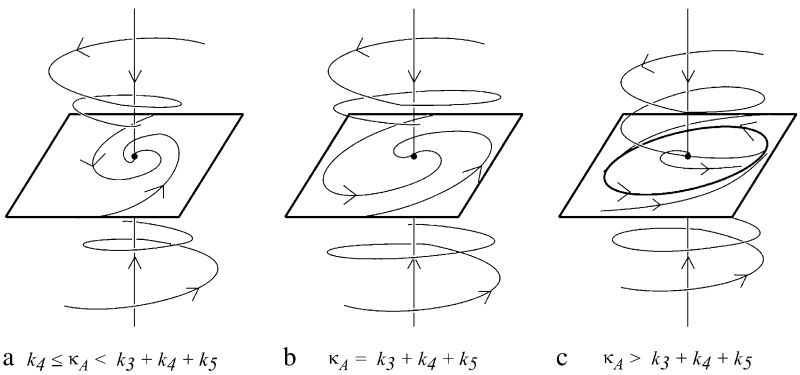

FIGURE 2 Simplified phase portrait for the system showing the change in dynamics as κ_A passes through the bifurcation value of $k_3 + k_4 + k_5$. (*a*) $\mathbf{x}^{(\mathrm{H})}$ is a stable focus; (*b*) as the bifurcation value is approached, the rate at which the trajectory approaches $\mathbf{x}^{(\mathrm{H})}$ decreases; (*c*) $\mathbf{x}^{(\mathrm{H})}$ becomes unstable, but a stable limit cycle bifurcates from the stationary point. As $\kappa_A > k_3 + k_4 + k_5$ increases, the radius of this limit cycle increases.

respect to Σ), and $M\ln(1/(2\pi)^{3/2}|\Sigma|^{1/2})$ is a constant term independent of $\boldsymbol{\theta}$. As we are not interested in model selection or comparison, we can neglect this constant term in this context. Investigating the log-likelihood function, $\ln(L(\boldsymbol{\theta}|D))$, for a given D is therefore equivalent to investigating how the sum of squared Mahalanobis distances between the observations $\mathbf{y_i}$ and the exact solutions $\boldsymbol{\mu}_i(\boldsymbol{\theta})$ changes with $\boldsymbol{\theta}$.

In all the experiments that follow, we shall assume that $\Sigma = I_3$, the 3×3 identity matrix. Under this assumption, the Mahalanobis distance is just the Euclidean distance, and so,

$$\ln(L(\theta|D)) \propto -\frac{1}{2}\sum_{i=1}^{M}(d(\mathbf{y}_i, \mu_i(\theta)))^2, \tag{7}$$

where $d(\mathbf{y}_i,\boldsymbol{\mu}_i(\boldsymbol{\theta}))$ denotes the Euclidean distance between $\mathbf{y_i}$ and $\boldsymbol{\mu}_i(\boldsymbol{\theta})$. It is clear from the above that maximizing the likelihood is equivalent to minimizing the distance function d (often also called the cost function). We can therefore see that investigating the log-likelihood for different parameter values is directly related to the problem of determining how the distance between solutions that start from the same initial point changes with the parameters. Thus, our investigation could be viewed as an inversion of a traditional (and well studied; see (11), for example) problem in nonlinear dynamics: namely, the problem of determining the rate of separation of nearby solutions (given a fixed set of parameters).

METHODS

We initially make the simplifying assumption that the rates k_2, k_3, k_4, and k_5 in Eq. 1 are all known and equal to 1. By Eq. 3, this means that $\mathbf{x}^{(H)}$ is located at $[\kappa_A - 1, \kappa_A - 1, \kappa_A - 1]^\top$ and will undergo a bifurcation as the value of κ_A passes through 3. We shall denote the true κ_A value by κ_A^*.

To obtain the data set D, we numerically simulate observations from the model. In such a simulation study, we are able to control κ_A^*, which enables us to investigate the change in inferability as its value passes through 3, the bifurcation point. It also means that we can either avoid or tightly control practical issues such as observational noise and missing data, so that we may focus specifically on the effect that the Hopf bifurcation has on parameter inferability.

To produce the simulated data, we specify a set of initial conditions and a value for κ_A^*, and then use a numerical ODE solver to evaluate the solution $\mathbf{x}_i = [x_1(t_i), x_2(t_i), x_3(t_i)]$ to the system of differential equations in Eq. 1 at times $t_1, \ldots, t_M$. Unless otherwise stated, our initial conditions are always taken to be $[x_1(0), x_2(0), x_3(0)]^\top = [1, 1, 1]^\top$, and we take $t_1 = 1, \ldots, t_M = M$. This yields the trial data, $D^M_{\kappa_A^*} = \{\mathbf{y}_i = \mathbf{x}(t_i|\kappa_A = \kappa_A^*)\}_{i=1}^M$. To imitate the imperfect nature of real experimentation, we also generate noisy data $\mathbf{y}_i = [x_1(t_i) + \varepsilon_{i1}, x_2(t_i) + \varepsilon_{i2}, x_3(t_i) + \varepsilon_{i3}]$, with each ε_{ij} being drawn from a zero-centered Gaussian distribution.

The numerical ODE solver that we use throughout this section is the Dormand-Prince $\{4, 5\}$ method (a member of the Runge-Kutta family of ODE solvers; see (15) for details of the Runge-Kutta method and (16) for the implementation).

RESULTS

To highlight the generic impact of a Hopf bifurcation on parameter inference from time-course data, we consider a concrete example in detail and under initially idealized conditions of very long time-series. In Appendix A, we show that these results persist for much shorter, realistic time series.

Dynamics and inferability

To understand the effect of the bifurcation on likelihood estimation, we evaluate and then plot the log-likelihood, $\ln\left(L\left(\kappa_A|D^{1000}_{\kappa_A^*}\right)\right)$, against κ_A, for a range of values of κ_A^*. Although the number $M = 1000$ would be an unrealistic number of observations in practice, additional experiments demonstrate that qualitatively similar results are also obtained for the more realistic value $M = 10$ (see Appendix A).

Fig. 3 shows six log-likelihood plots for a range of values of κ_A^*. As expected, the maximum likelihood (ML) parameter value is always the true value. However, the appearance of the plots obtained when $\kappa_A^* \leq 3$ are noticeably different to those obtained when $\kappa_A^* > 3$.

For $1 < \kappa_A^* \leq 3$, each of the log-likelihood curves consists of:

1. A relatively flat part in the region $1 < \kappa_A < 3$.
2. A relatively steep part with negative gradient in the region $\kappa_A > 3$.
3. A distinct elbow in the curve at $\kappa_A = 3$, which marks the boundary between the other two regions of distinct behavior.

The overall shape of the curve means that, when $1 < \kappa_A^* \leq 3$, all parameter estimates >3 become rapidly (and increasingly) unlikely. On the other hand, parameter estimates in the interval (1, 3] are all relatively likely, with a maximum occurring at $\kappa_A = \kappa_A^*$. This is very satisfying to our intuition, as it means that parameter estimates that yield solutions whose dynamical behavior is qualitatively the same as the true solution are generally more likely than those for which the qualitative dynamical behavior is different. Furthermore, the bifurcation value $\kappa_A = 3$, which marks the critical point of change between the two different types of dynamical behavior, also marks the critical point of changing behavior in the log-likelihood curve.

For $\kappa_A^* > 3$, although we still obtain a maximum at $\kappa_A = \kappa_A^*$, this occurs as a spike in the curve. Apart from this, the log-likelihood curves obtained for $\kappa_A^* > 3$ are all qualitatively similar to the curve obtained for $\kappa_A^* = 3$, in that there is a local maximum at $\kappa_A = 3$ and the gradient of the curve to the right of this local maximum is steeper than to the left. We also note that as $\kappa_A^* > 3$ increases, the log-likelihood of values of κ_A other than $\kappa_A = \kappa_A^*$ decreases, so that the amplitude of the spike appears greater (relative to the level of the rest of the curve). While we would certainly expect to see a maximum at $\kappa_A = \kappa_A^*$, intuition would suggest that values of κ_A close to κ_A^* should also be likely—and certainly more likely than values of κ_A between 1 and 3 (where the solution is qualitatively different).

To understand the observed behavior, we have to think a little more deeply about the nature of the stationary point

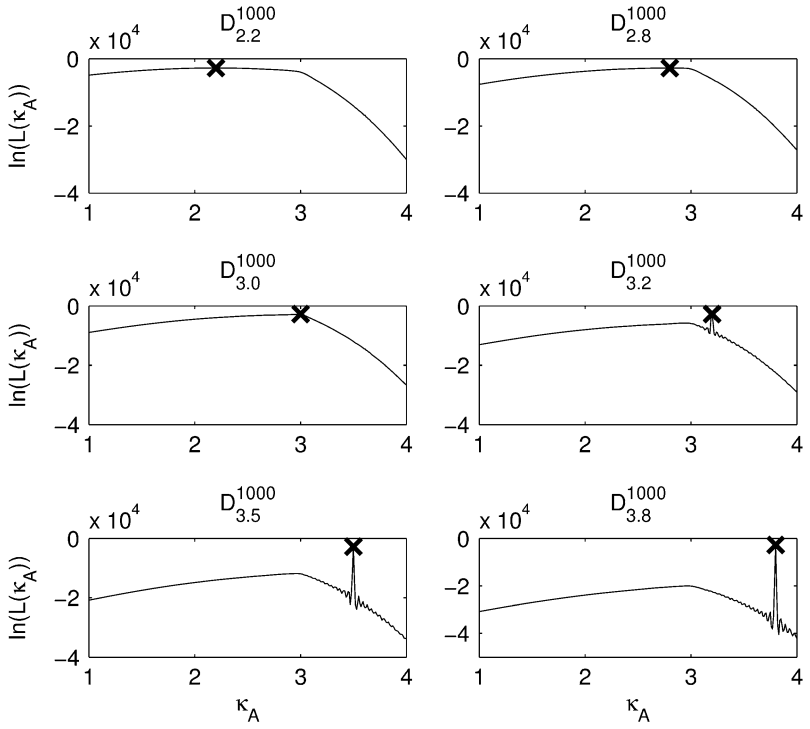

FIGURE 3 Log-likelihood plots for different values of κ_{A}^{*} (as specified by the headings, $D^{1000}_{\kappa_{\mathrm{A}}^{*}}$). In each case, the black cross indicates the maximum value (as expected, in every case this occurs when κ_{A} is the true value). The y axis of each plot shows the value of the log-likelihood, while the x axis shows the value of κ_{A}.

$\mathbf{x}^{(\mathrm{H})*} = [\kappa_{\mathrm{A}}^{*} - 1, \kappa_{\mathrm{A}}^{*} - 1, \kappa_{\mathrm{A}}^{*} - 1]$. The plots in Fig. 4 show a section of the numerical solution to the system $\dot{\mathbf{x}} = g(\mathbf{x}|\kappa_{\mathrm{A}})$ obtained for different values of κ_{A}, and are helpful when considering the effect of the nonlinear dynamics on the likelihood.

Let us first consider the case when $1 < \kappa_{\mathrm{A}}^{*} < 3$. Here, $\mathbf{x}^{(\mathrm{H})*}$ is asymptotically stable, which means that all trajectories that start near to the stationary point will be drawn toward it (see Fig. 2 *a*). In general, for any $\kappa_{\mathrm{A}}^{*} \in (1,3)$, the solution obtained using a parameter estimate $\widehat{\kappa_{\mathrm{A}}} \in (1,3)$ will be qualitatively similar to the true solution. Crucially, this means that the true solution and estimated solution both start at the same point and end at nearby points, and hence the distance between corresponding points in $D^{1000}_{\kappa_{\mathrm{A}}^{*}}$ and $D^{1000}_{\widehat{\kappa_{\mathrm{A}}}}$ will be relatively small. Thus, given the relationship between distance and likelihood (see Eq. 7), we expect all estimates $\widehat{\kappa_{\mathrm{A}}} \in (1,3)$ to be relatively likely.

However, when the estimate $\widehat{\kappa_{\mathrm{A}}} > 3$, we observe a qualitative change in the behavior of the corresponding estimated solution (see Fig. 2 *c*). The (estimated) stationary point at

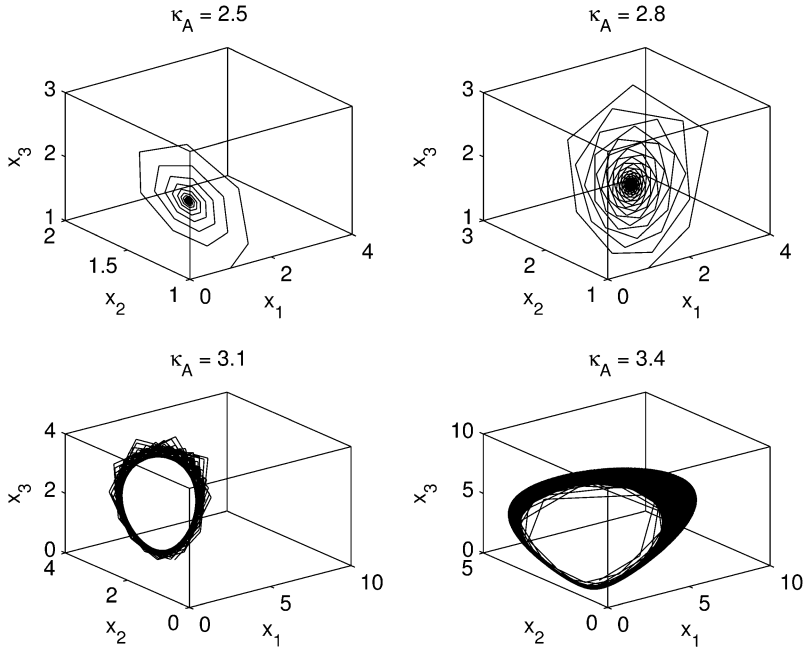

FIGURE 4 The plots here provide a graphical representation of sections of the solutions for different values of κ_{A}. The solutions were evaluated at the times $t = 1, \ldots, 1000$, and are shown with straight lines joining adjacent time points. Note how the scales on the axes grow larger as $\kappa_{\mathrm{A}} > 3$ increases.

$[\widehat{\kappa_A} - 1, \widehat{\kappa_A} - 1, \widehat{\kappa_A} - 1]$ is no longer stable, and the estimated solution is instead attracted toward a stable limit cycle. Two solutions obtained for two different values of $\kappa_A > 3$ will only be close to one another when the two κ_A values are themselves very close together. Trajectories are no longer drawn-inward toward a central point, but are instead forced-outward toward the limit cycle. The effect of this is that, for solutions obtained for any two different values of $\kappa_A > 3$, the distances between corresponding data points sampled from these solutions can become very large very quickly. This is why we observe a spike in the log-likelihood curve: only values of κ_A very, very close to the true value will result in data points that are close to the observed data.

An approximate likelihood

If $\mathbf{y}(t_i) \in D^M_{\kappa_A^* \in (1,3)}$ and i is sufficiently large, we may make the approximation $\mathbf{y}(t_i) \approx [\kappa_A^* - 1, \kappa_A^* - 1, \kappa_A^* - 1]$; i.e., $\mathbf{y}(t_i)$ is close to the stationary point. Suppose we also consider observations $\mathbf{y}_\delta(t_i)$ from the set $D^M_{\kappa_A^*+\delta}$, where $\delta \in \mathbb{R}$ is such that $\delta \neq 0$ and $\delta \in (1 - \kappa_A^*, 3 - \kappa_A^*)$ (this simply says that $\kappa_A^* + \delta$ is a number from (1, 3) that is not equal to κ_A^*). Then, for i sufficiently large, we have $\mathbf{y}_\delta(t_i) \approx [\kappa_A^* + \delta - 1, \kappa_A^* + \delta - 1, \kappa_A^* + \delta - 1]$. Therefore, the squared Euclidean distance between corresponding points $\mathbf{y}(t_i)$ and $\mathbf{y}_\delta(t_i)$ is eventually close to $3\delta^2$.

Numerical simulation suggests that the squared Euclidean distance between corresponding points in $D^{1000}_{\kappa_A^*}$ and $D^{1000}_{\kappa_A^*+\delta}$ is $\sim 3\delta^2$ for the majority of points, so we make the further approximation that this is true for all points (provided $\kappa_A^* + \delta$, $\kappa_A^* \in (1, 3)$). By Eq. 7, this suggests that if $\kappa_A^* \in (1, 3)$, then the log-likelihood function may be approximated by

$$\ln(L(\kappa_A|D)) \approx M \ln\left(\frac{1}{(2\pi)^{3/2}}\right) - \frac{1}{2}\sum_{i=1}^{M} 3(\kappa_A^* - \kappa_A)^2, \quad (8)$$

$$\propto -\frac{3M}{2}(\kappa_A^* - \kappa_A)^2, \quad (9)$$

in the region $1 < \kappa_A < 3$.

The accuracy of this approximation is likely to depend on many factors. For example, we would expect the approximation to be poorer if fewer observations were taken, or if the initial starting point were further away from $\mathbf{x}^{(H)*}$. However, we can see from Fig. 5 that for our idealized case, the approximation provides a good fit to the likelihood function.

Fisher information

To further quantify the sensitivity of the system to changes in κ_A, we now consider the Fisher information, $I(\kappa_A)$ (17). This tells us the amount of information that the observable variable $\mathbf{x} = [x_1, x_2, x_3]^\top$ carries about the value of κ_A. The method used to approximate the Fisher information is given in Appendix B.

Fig. 6 shows that the Fisher information is relatively low for parameter values between 2 and 3, and then rapidly increases once κ_A has passed through the bifurcation point. This reiterates the previous findings and nicely illustrates the difference in parameter inferability either side of the bifurcation point.

Noisy data

Until now, we have been solely concerned with simulating ''perfect'' data that are absent of noise. To demonstrate that

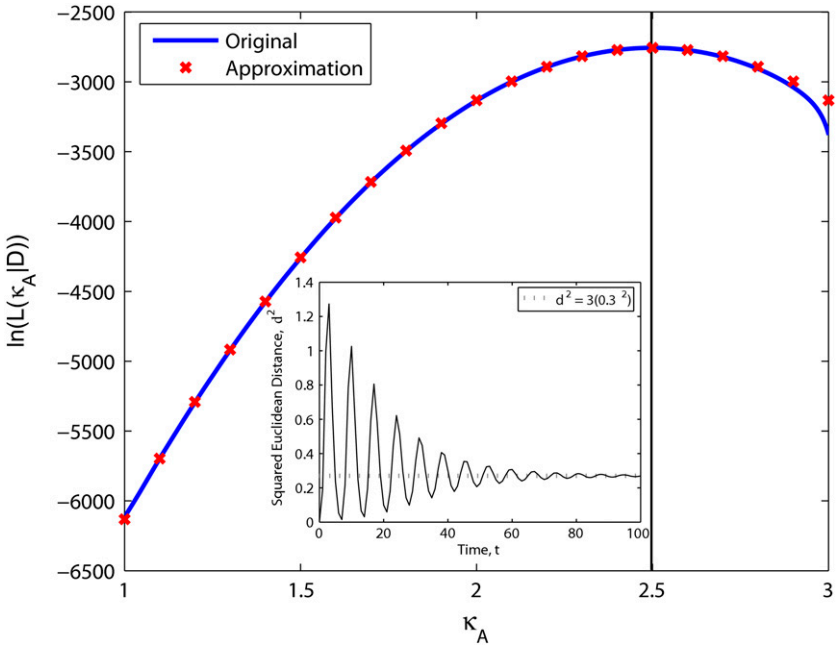

FIGURE 5 Log-likelihood for $1 < \kappa_A < 3$ (the true parameter was set to $\kappa_A^* = 2.5$), together with the approximation given in Eq. 9. The inset shows that the squared Euclidean distance between corresponding points from $D^{1000}_{2.5}$ and $D^{1000}_{2.2}$ approaches $3\delta^2 = 3(0.3)^2$, as predicted.

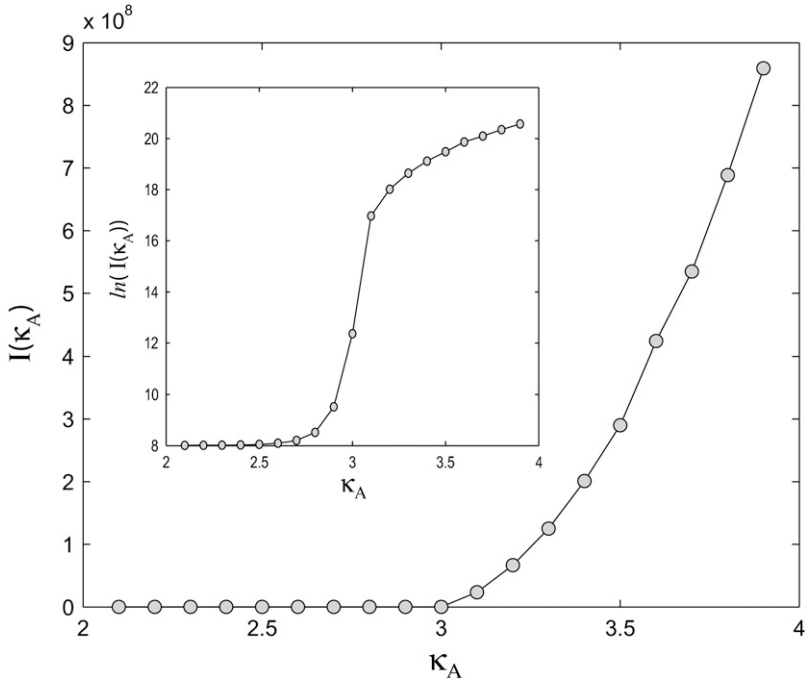

FIGURE 6 Plot showing $I(\kappa_A)$ for a range of values of κ_A. Inset is a plot showing $\ln(I(\kappa_A))$ for the same range of κ_A values.

our findings are also valid in practice, we now try to imitate more realistic experimental conditions by introducing a Gaussian noise term. Our data sets are now of the form

$$D^{M}_{\kappa^*_A} := \{\mathbf{y}_i = \mathbf{x}(t_i|\kappa_A = \kappa^*_A) + [\epsilon_{i1}, \epsilon_{i2}, \epsilon_{i3}]^\top\}^{M}_{i=1}, \qquad (10)$$

where each $\epsilon_{ij} \sim N(0, \sigma^2)$. To obtain the results shown in Fig. 7, we took $\sigma = 10$. This was deliberately chosen to provide a very large noise term, so that we could test the robustness to noise of our previous findings.

Overall, these plots appear qualitatively similar to the ones obtained in the absence of noise (Fig. 3); in particular, the spike for $\kappa^*_A > 3$ persists. Of course, the log-likelihood for any given κ_A value is lower here, reflecting our reduced confidence in the value of any particular parameter value (for example, even when $\kappa_A = \kappa^*_A$, the total distance between the

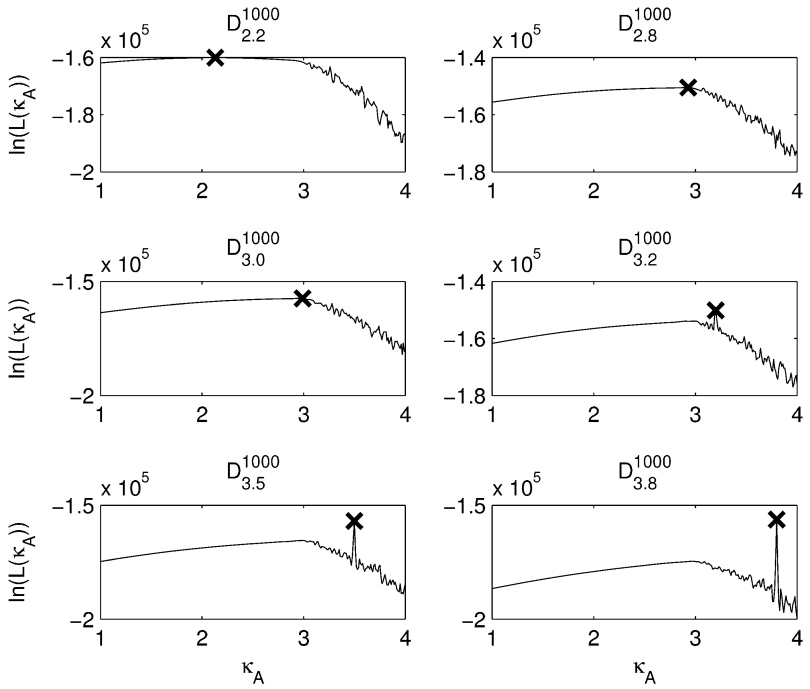

FIGURE 7 Log-likelihood plots obtained using noisy data sets (see Fig. 3). The maximum log-likelihood values (indicated by the *black cross*) in each case are as follows: 2.13, 2.93, 2.99, 3.2, 3.5, and 3.8.

observed and expected data is nonzero, and so the log-likelihood is lower).

Maximum likelihood estimation of other parameters

So far, we have assumed that only the value of κ_A was unknown, and that the values of the other parameters were all known and equal to 1. We shall now return to using noiseless data, but shall consider a case in which both κ_A and k_2 are unknown.

Suppose that the true values of κ_A and k_2 are κ_A^* and k_2^*. Then we denote the Hopf-bifurcating stationary point of the system $\dot{\mathbf{x}} = g(\mathbf{x}|\kappa_A^*, k_2^*)$ by $\mathbf{x}^{(H)*}$, and refer to this as the true stationary point. Since $k_3 = k_4 = k_5 = 1$ are known, we know by Eq. 3 that

$$\mathbf{x}^{(H)*} = \left[\frac{\kappa_A^* - 1}{k_2^*}, \frac{\kappa_A^* - 1}{k_2^*}, \frac{\kappa_A^* - 1}{k_2^*}\right].$$

Similarly, if $\widehat{\kappa_A}$ and $\widehat{k_2}$ are estimated values of κ_A and k_2, then we denote the Hopf-bifurcating stationary point of the system $\dot{\mathbf{x}} = g(\mathbf{x}|\widehat{\kappa_A}, \widehat{k_2})$ by $\widehat{\mathbf{x}^{(H)}}$ and refer to this as the estimated stationary point. Clearly, we have

$$\widehat{\mathbf{x}^{(H)}} = \left[\frac{\widehat{\kappa_A} - 1}{\widehat{k_2}}, \frac{\widehat{\kappa_A} - 1}{\widehat{k_2}}, \frac{\widehat{\kappa_A} - 1}{\widehat{k_2}}\right].$$

Assuming both κ_A and k_2 to be unknown turns out to be particularly interesting for two main reasons:

1. We know that the stationary point $\mathbf{x}^{(H)}$ of the system $g(\mathbf{x}|\kappa_A, k_2)$ is stable provided κ_A is in the open interval $(k_4, k_3 + k_4 + k_5)$. Thus, since we are keeping the other three parameters (k_3, k_4, k_5) fixed, the stability of $\mathbf{x}^{(H)}$ is determined solely by the value of κ_A and not by k_2.
2. Given any estimate $\widehat{\kappa_A}$ of κ_A, there is always a value $\xi_2 = (k_2^*(\widehat{\kappa_A} - 1))/(\kappa_A^* - 1)$ such that if $\widehat{k_2} = \xi_2$, the location of the estimated stationary point will coincide with the location of the true stationary point (regardless of how poor the estimates $\widehat{\kappa_A}$ and $\widehat{k_2}$ may be).

We start by setting $\kappa_A^* = 2.5$ and $k_2^* = 1$, and calculate the log-likelihood on a grid of values for κ_A and k_2 (using exactly the same method as previously, taking the initial point to be [1, 1, 1] and the number of data points, M, to be 1000). We consider $2 \leq \kappa_A \leq 3.9$ and $0.2 \leq k_2 \leq 4$, with a spacing of 0.02 between grid points in both directions. This defines a log-likelihood surface, as shown in Fig. 8. The absolute maximum is marked on the surface with a white dot, and occurs when $(\kappa_A, k_2) = (\kappa_A^*, k_2^*)$ (as we would expect).

To provide more insight, Fig. 9 shows for each value of κ_A the value of k_2, which yields the maximum log-likelihood. We can see from Fig. 9 that there are two approximately linear regions. Shown in red is the line $k_2 = (2)/(3)(\kappa_A - 1)$, and we can see that it provides a good fit to the curve in approximately the region $2 \leq \kappa_A \lesssim 3$. This line corresponds to $\widehat{k_2} = \xi_2$ being the estimate which ensures that $\widehat{\mathbf{x}^{(H)}} = \mathbf{x}^{(H)*}$. This means that even if we estimate κ_A incorrectly, then—provided $\widehat{\kappa_A} \in [2, 3]$—the maximum likelihood estimate of

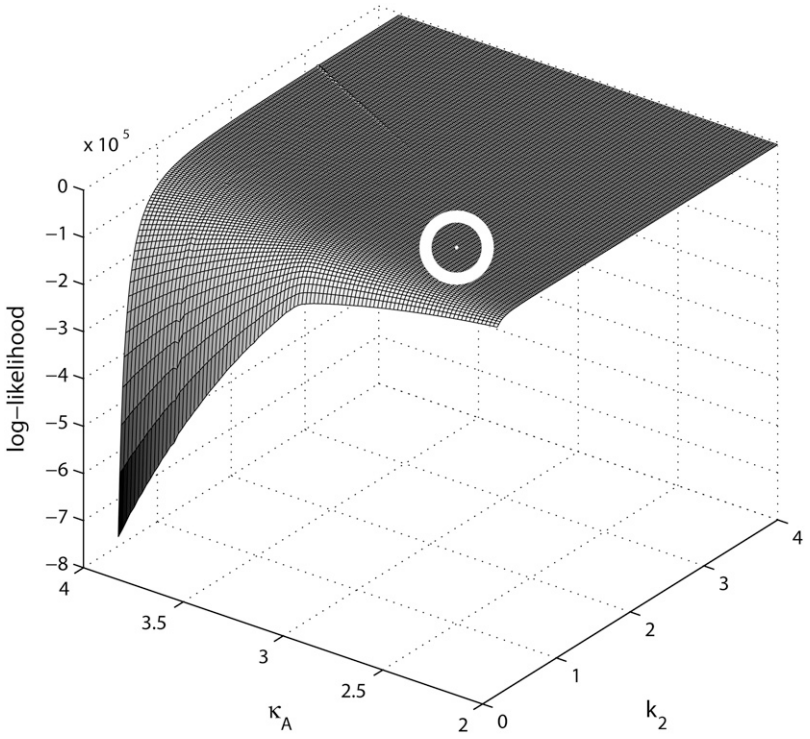

FIGURE 8 Plot of the log-likelihood surface on a grid of κ_A, k_2 values. The absolute maximum value has been highlighted as a circled white dot.

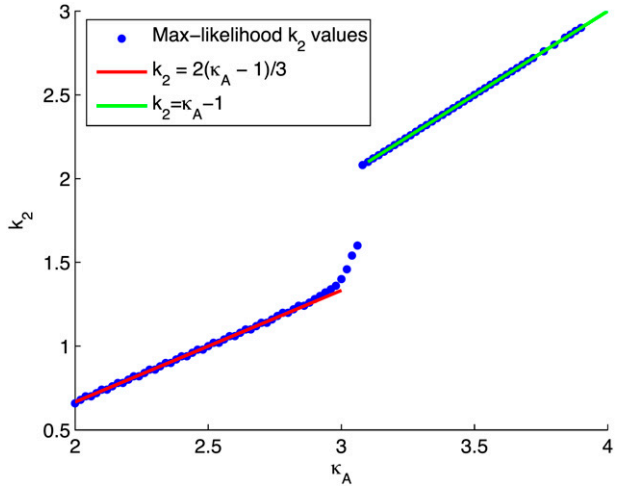

FIGURE 9 Plot showing for each value of κ_A the value of k_2 that yields the maximum log-likelihood (*blue dots*). Shown in green is the line $k_2 = \kappa_A - 1$, and in red is the line $k_2 = (2)/(3)(\kappa_A - 1)$.

k_2 is the one that places the estimated stationary point at the true location (or within a ellipsoid with radii determined by the level of noise in the system). Intuitively, this makes a great deal of sense: provided $\widehat{\kappa_A}$ is in the range $1 \leq \kappa_A \leq 3$, $\mathbf{x}^{(H)}$ is asymptotically stable and so represents the ''destination'' of the solution to the system $\dot{\mathbf{x}} = g(\mathbf{x}|\widehat{\kappa_A}, \widehat{k_2})$ as $t \to \infty$. Thus, if $\widehat{k_2} = \kappa_2$ correctly places the estimated stationary point, then this means that the estimated solution both starts and ends at the same places as the observed solution.

Shown in green is the line $k_2 = \kappa_A - 1$, which provides a good fit to the curve in the region $3.08 \leq \kappa_A \leq 3.9$. This line corresponds to $\widehat{k_2}$ being the estimate that locates $\widehat{\mathbf{x}^{(H)}}$ at [1, 1, 1], i.e., so that the estimated stationary point coincides with the initial starting point of the trajectory. This means that the whole trajectory will just appear at the single point [1, 1, 1]. We can think of this as follows: if κ_A is estimated in such a way that $\widehat{\mathbf{x}^{(H)}}$ would be unstable, then the maximum likelihood value estimate $\widehat{k_2}$ attempts to correct this behavior. Since the value of k_2 cannot alter the stability of $\widehat{\mathbf{x}^{(H)}}$, the best that can be done is to estimate $\widehat{k_2}$ so that the entire solution shrinks down onto a single point. We can also see that there is in fact a small interval of κ_A values (namely (3, 3.08)) in which the unstable solution is favored. Within this interval, the radius of the limit cycle is relatively small, and hence the maximum likelihood solution can be unstable and still be relatively close to the observed solution.

The results from this section appear to demonstrate that—when we perform inference in order to estimate the parameters of an ODE system—not only are we attempting to find the parameter values that best explain the particular observed trajectory, but that also accurately describe the underlying dynamics of the system. In particular, our results suggest that if the true stationary point $\mathbf{x}^{(H)*}$ is stable, then parameter estimates for which $\widehat{\mathbf{x}^{(H)}}$ is stable and correctly located will be preferred over those for which it is not.

We also investigated situations in which κ_A and k_4 were unknown (and the remaining parameters were assumed to be known), and obtained similar results. However, in contrast to the above, both of these two unknown parameters affect the location of the stationary point, which makes the results more difficult to interpret.

DISCUSSION

We have considered the smallest chemical reaction system with a Hopf bifurcation (9), and have investigated how parameter inferability changes in this model. We found that the presence of a Hopf bifurcation in the system has a clear effect on the shape of both the log-likelihood function and the Fisher information considered as a function of the parameters. An analysis of the log-likelihood function for different values of κ_A^* revealed that the function's qualitative behavior changes depending on whether κ_A^* is above or below the critical bifurcation value (see Fig. 3), which we expect to have an effect upon inferability. For $\kappa_A^* > 3$, highly likely estimates for the parameter only occur within a very small interval that contains the true value. Thus, although it might be difficult to find a particularly likely parameter estimate, once we have one, we can be sure that it will be close to κ_A^*. For $1 < \kappa_A^* \leq 3$, any parameter estimate that lies within the interval (1, 3] is relatively likely. So, in this case we have the opposite situation: although it may be easy to find plausible parameter estimates, determining which one is correct is likely to be more difficult.

There is—mirroring the rich literature on dynamical systems, in general, and on bifurcation analysis, in particular—a rich set of phenomena related to estimating parameters of dynamical systems. Parameter estimation generally considers only local dynamics, i.e., its specific aim tends to be to home in on those model parameters most likely to have generated the observed data points as efficiently as possible. Global dynamical effects, such as bifurcations, are hardly ever considered although their importance for understanding biological systems has been amply demonstrated (7,8). Here we have discussed how global dynamics interact with our ability to draw statistical parameter inferences from time-series data. There are a number of obvious ways in which this research can be extended. A relatively simple but important next step would be to consider the effects of having fewer experimental/simulated observations. Preliminary investigation (see Appendix A) suggests that the results presented here remain valid, yet the picture quickly becomes more complicated as data quality and quantity decrease. There is a clear need for a thorough treatment of this issue, as experimental data are often sparse. Another simple extension would be to study the effects of varying or inferring initial conditions. Once these issues have been addressed, the methods used here could be applied to other systems to determine the effects on inferability of other types of bifurcation.

APPENDIX A: INFERENCE FROM SHORT TIME-SERIES RESULTS

Throughout, we have considered data sets that consist of 1000 data points. However, with real experimental data, we would typically expect far fewer observations. In Fig. 10, we show log-likelihood plots obtained using a data set consisting of only 10 observations taken at times $t = 10, 20, \ldots, 100$.

Although now perhaps less clear, the results are still qualitatively similar to those obtained previously. A notable difference is that the maximum value for $\kappa_A^* > 3$ no longer occurs as a true spike in the curve. Instead, values of κ_A near to κ_A^* are also likely, so that the maximum occurs as the peak of a much more gentle hump in the curve. One effect of this is that it is much more difficult to determine the precise bifurcation value of κ_A just from looking at these plots. Another is that for $\kappa_A^* > 3$ we would expect to find it more difficult to pinpoint the precise value of κ_A^* than in our previous cases. However, overall, we can see that the shape of the likelihood curve is roughly similar to previously, with different behavior either side of $\kappa_A \approx 3$.

APPENDIX B: APPROXIMATING THE FISHER INFORMATION

The Fisher information is defined as

$$I(\kappa_A) = E\left(\left(\frac{\partial}{\partial(\kappa_A)}\ln(L(\kappa_A|D))\right)^2 |\kappa_A\right). \qquad (11)$$

Here, the expectation is with respect to the probability density function $f(D|\kappa_A)$, so we may write the expression above as

$$I(\kappa_A) = \int_D \left(\frac{\partial}{\partial(\kappa_A)}\ln(L(\kappa_A|D))\right)^2 f(D|\kappa_A)dD. \qquad (12)$$

Hence, for each κ_A, $I(\kappa_A)$ is the square of the derivative with respect to κ_A of the likelihood function for κ_A, averaged over all observable data sets D. Using standard techniques (Monte Carlo integration (see, for example, (18)), we can clearly approximate the Fisher information by

$$I(\kappa_A) \approx \frac{1}{N}\sum_{j=1}^{N}\left(\frac{\partial}{\partial(\kappa_A)}\ln(L(\kappa_A|D^{(j)}))\right)^2, \qquad (13)$$

where N is a large number and each $D^{(j)}$ is an independent draw from $f(D|\kappa_A)$.Furthermore, from Verdugo and Rand (7), we have

$$\ln(L(\kappa_A|D = \{\mathbf{y}_i\}_{i=1}^M)) \propto -\frac{1}{2}\sum_{i=1}^{M}(d(\mathbf{y}_i, \boldsymbol{\mu}_i(\kappa_A)))^2,$$

where $\boldsymbol{\mu}_i(\kappa_A)$ is the exact solution evaluated at time $t = t_i$.Thus, we have

$$I(\kappa_A) \approx \frac{1}{4N}\sum_{j=1}^{N}\left(\frac{\partial}{\partial(\kappa_A)}\sum_{i=1}^{M}(d(\mathbf{y}_i^{(j)}, \boldsymbol{\mu}_i(\kappa_A)))^2\right)^2, \qquad (14)$$

where $\{\mathbf{y}_i^{(j)}\}_{i=1}^M = D^{(j)}$ is an independent draw from $f(D|\kappa_A)$.

Since we cannot write $\boldsymbol{\mu}_i(\kappa_A)$ analytically, we must make the following further approximation,

$$\frac{\partial}{\partial(\kappa_A)}\sum_{i=1}^{M}(d(\mathbf{y}_i^{(j)}, \boldsymbol{\mu}_i(\kappa_A)))^2 \approx \frac{\sum_{i=1}^{M}(d(\mathbf{y}_i^{(j)}, \boldsymbol{\mu}_i(\kappa_A + h)))^2 - \sum_{i=1}^{M}(d(\mathbf{y}_i^{(j)}, \boldsymbol{\mu}_i(\kappa_A)))^2}{h}, \qquad (15)$$

where $0 < h \ll 1$.

Thus, we have

$$I(\kappa_A) \approx \frac{1}{4Nh^2}\sum_{j=1}^{N}\left(\sum_{i=1}^{M}((d(\mathbf{y}_i^{(j)}, \boldsymbol{\mu}_i(\kappa_A + h)))^2 - (d(\mathbf{y}_i^{(j)}, \boldsymbol{\mu}_i(\kappa_A)))^2)\right)^2. \qquad (16)$$

Using the above, it is possible for us to approximate $I(\kappa_A)$ numerically for any value of κ_A. Results are shown in Fig. 6.

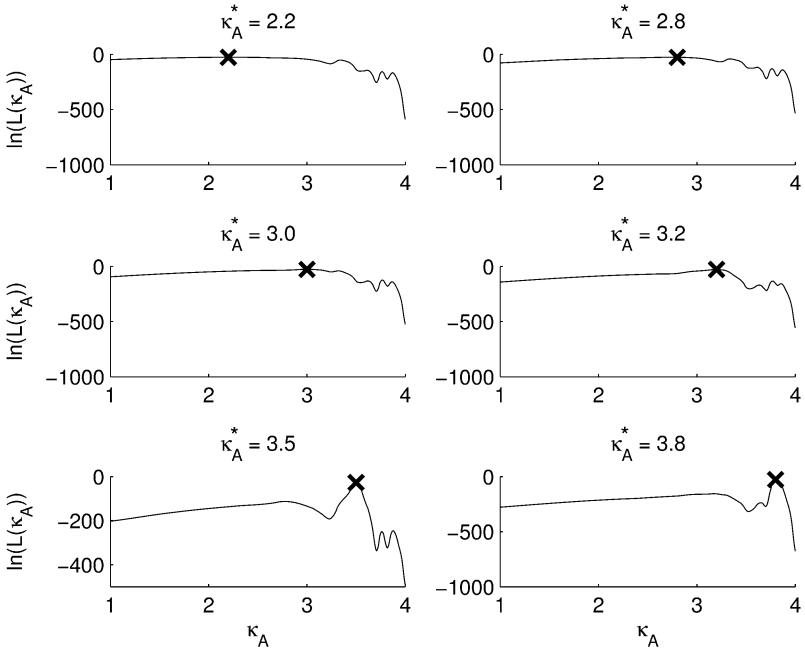

FIGURE 10 Log-likelihood plots obtained using a data set consisting of 10 points, with observations taken at times $t = 10, 20, \ldots, 100$.

REFERENCES

1. May, R. M. 2004. Uses and abuses of mathematics in biology. *Science*. 303:790–793.
2. Haefner, J. 2005. Modeling Biological Systems: Principles and Applications. 2nd Ed. Springer, New York.
3. Reed, M. C. 2004. Why is mathematical biology so hard? *Notices AMS*. 51:338–342.
4. Ramsay, J. O., G. Hooker, D. Campbell, and J. Cao. 2007. Parameter estimation for differential equations: a generalized smoothing approach. *J. Roy. Stat. Soc. B*. 69:1–30.
5. Goyal, S., and N. S. Wingreen. 2007. Growth-induced instability in metabolic networks. *Phys. Rev. Lett*. 98:1–4.
6. Steuer, R., A. N. Nesi, A. R. Fernie, T. Gross, B. Blasius, and J. Selbig. 2007. From structure to dynamics of metabolic pathways: application to the plant mitochondrial TCA cycle. *Bioinformatics*. 23:1378–1385.
7. Verdugo, A., and R. Rand. 2008. Hopf bifurcation in a DDE model of gene expression. *Commun. Nonlinear Sci. Numer. Simul*. 13: 235–242.
8. Battogtokh, D., and J. J. Tyson. 2004. Bifurcation analysis of a model of the budding yeast cell cycle. *Chaos*. 14:653–661.
9. Wilhelm, T., and R. Heinrich. 1995. Smallest chemical reaction system with Hopf bifurcation. *J. Math. Chem*. 17:1–14.
10. Wilhelm, T., and R. Heinrich. 1996. Mathematical analysis of the smallest chemical reaction system with Hopf bifurcation. *J. Math. Chem*. 19:111–130.
11. Kuznetsov, Y. 2004. Elements of Applied Bifurcation Theory. Springer, New York.
12. Williams, J., and Z. Kalogiratou. 1993. Least squares and Chebyshev fitting for parameter estimation in ODEs. *Adv. Comput. Math*. 1:357–366.
13. Bard, Y. 1974. Nonlinear Parameter Estimation. Academic Press, New York.
14. Mahalanobis, P. C. 1936. On the generalized distance in statistics. *Proc. Nat. Inst. Sci. India*. 2:49–55.
15. Press, W., S. Teukolsky, W. Vetterling, and B. Flannery. 1992. Numerical Recipes in C: The Art of Scientific Computing, 2nd Ed. Cambridge University Press, Cambridge, UK.
16. MathWorks. 2004. MATLAB 7. The MathWorks, Natick, Massachusetts.
17. Cover, T., and J. Thomas. 1991. Elements of Information Theory. Wiley-Interscience, New York.
18. MacKay, D. J. C. 2003. Information Theory, Inference, and Learning Algorithms, 1st Ed. Cambridge University Press, Cambridge, UK.