Identifying dynamical systems with bifurcations from noisy partial observation

Yohei Kondo,* Kunihiko Kaneko, and Shuji Ishihara
Graduate School of Arts and Sciences, University of Tokyo, 3-8-1 Komaba, Meguro-ku, Tokyo 153-8902, Japan
(Received 21 August 2012; revised manuscript received 3 February 2013; published 19 April 2013)

We propose a statistical machine-learning approach to derive low-dimensional models by integrating noisy time-series data from partial observation of high-dimensional systems, aiming to utilize quantitative data on biological phenomena in the cell. In particular, the method estimates a model from data at different values of a bifurcation parameter in order to characterize biological functions as bifurcation types that are insensitive to system details and experimental errors. The method is tested using artificial data generated from two cell-cycle control system models that exhibit different bifurcations and the learned systems are shown to robustly inherit the bifurcation types.

DOI: 10.1103/PhysRevE.87.042716 PACS number(s): 87.17.Aa, 05.45.Tp

I. INTRODUCTION

The relevance of dynamical systems to modeling biological phenomena has been increasingly recognized [1,2]. Recent advances in experimental techniques such as live-cell imaging that clarifies molecular activities at high spatiotemporal resolution [3–5] have accompanied this recognition. However, noise, partial observation, and low controllability are still challenges for measuring biological systems in that both the system dynamics and measurement processes are highly stochastic, only a few components in a system are observable, and only a small number of experimental conditions can be examined. These difficulties have hindered model construction from experimental observations on the molecular systems.

To model complex systems such as cellular processes, a full description of all details of the systems is often impractical and not informative. Instead, dynamics can be often captured by simplified systems described by a small number of variables, which are more useful for our comprehension. In the reduced model, the type of bifurcation is a basic feature to be inherited because the qualitative changes of the dynamics against environmental and experimental perturbation are helpful to characterize the underlying mechanism of biological functions that is insensitive to molecular details and experimental errors [6–8]. Thus identification of low-dimensional model systems that inherit the original bifurcation type is an effective step in understanding the dynamics.

Here we propose a statistical machine-learning approach to derive models from time-series data obtained for different conditions, i.e., bifurcation parameter values (Fig. 1). Various techniques for learning nonlinear dynamical systems from time-series data have been developed, especially in the case of noiseless system dynamics [9–12]. Although learning from stochastic dynamics has less theoretical ground, methods considering stochasticity in dynamics have been investigated recently in the framework of statistical theory [13–16]. However, at present it is not certain whether such a statistical framework can be used to estimate a model with a bifurcation. Thus we adopt statistical techniques to obtain the model with the correct type of bifurcation rather than aim at fitting the model parameters to the observations. The expectation-maximization (EM) algorithm and particle smoother [17–19] are combined, as they were suggested to be applicable for nonlinear and stochastic dynamical systems [20]. To obtain models with bifurcations, the learning is conducted by simultaneously using all of the time-series data obtained at different values of a bifurcation parameter. By taking advantage of continuity, we do not need the data from many parameter values; those from a few parameter values are sufficient to predict the dynamics with correct bifurcation types.

Before application to experimental data, verification of the statistical method by artificially generated data is warranted since we can directly compare learned systems with true ones. In particular, it is beneficial to confirm that the learned systems with a small number of variables can approximate the dynamics restricted to low-dimensional subspaces of original high-dimensional systems. Below we test the proposed approach on this point and evaluate the learned systems in comparison with reduced equations of true systems by adiabatic approximation.

II. METHODS

A. Nonlinear state space model

We introduce a nonlinear state space model composed of state and observation equations that describe the system dynamics and observation process, respectively. Let us consider D-dimensional stochastic differential equations that describe the system dynamics and d components in the system that are observed simultaneously. The state equations are discretized in time by the Euler-Maruyama scheme [21]. We write the time evolution of the ith variable at a time point t, $x_i^t (i = 1, \ldots, D)$, as

$$x_i^{t+1} = x_i^t + \Delta t f_i\left(\left\{x_j^t\right\}, s\right) + \sigma_i \xi_i^t \sqrt{\Delta t}, \tag{1}$$

where Δt is an integration time, σ_i is the intensity of the system noise, and s is a bifurcation parameter. The system noise ξ_i^t is sampled from a standard normal distribution. To achieve efficient learning, the function f_i is considered to be expressed by a summation of linearly independent functions as $f_i(\{x_j^t\}, s) = \sum_n^{N_i} k_i^n f_i^n(\{x_j\}, s)$, where N_i is the number of parameters $\{k_i^n\}$ and functions $\{f_i^n\}$. Since our aim is to reproduce bifurcation types of systems subjected to unknown equations, we adopt polynomial basis for the $\{f_i^n\}$ rather than biochemically realistic functions such as Michaelis-Menten equation.

*kondo@complex.c.u-tokyo.ac.jp

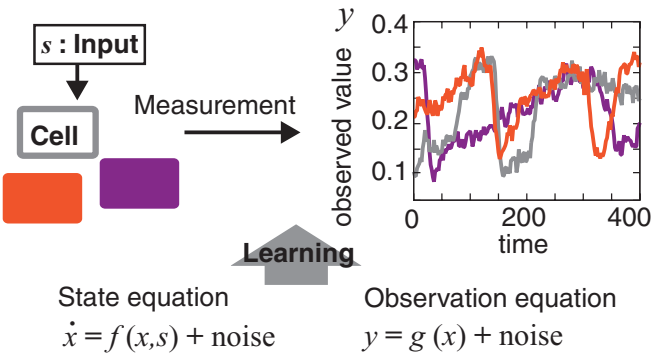

FIG. 1. (Color online) Schematic representation of the proposed method. Time-series data of the molecular activities in individual cells are obtained from measurements under a given input level s, which is regarded as a bifurcation parameter of the system. System and observation equations are then trained to reproduce the time-series data.

The observation value of the ith component at a time point τ, $y_i^\tau (i = 1,\ldots,d)$, is written as

$$y_i^\tau = g_i\left(x_i^\tau\right) + \eta_i \phi_i^\tau, \tag{2}$$

where η_i is an observation noise intensity and ϕ_i^r is sampled from a standard normal distribution. In general, a set of observed time points is a part of the entire set of time points in the numerical integration.

Let us consider a data set of $Y = \{Y_a\}$ $(a = 1,\ldots,A)$, where each Y_a is a time-series sample obtained from an independent measurement at the bifurcation parameter value s_a. The initial condition for the ith component in the ath time-series sample is assumed to obey a Gaussian distribution parametrized by the mean $\mu_{i,a}$ and the variance $V_{i,a}$. Then the parameters to be estimated are $\theta = (\{k_i^n\},\{\sigma_i\},\{\eta_i\},\{\mu_{i,a}\},\{V_{i,a}\})$. In contrast, the bifurcation parameter values $S = \{s_a\}$ are assumed to be known.

B. Maximum likelihood estimation

The learning of dynamical systems is formulated as a maximum likelihood estimation, which is summarized below (further details are given in the Appendix). The likelihood is given by the conditional probability $p(Y|\theta,S) = \prod_a p(Y_a|\theta,s_a)$. However, a straightforward maximization of the likelihood is difficult because it requires the untractable summation of $p(Y|X,\theta,S)p(X|\theta,S)$ with respect to the set of time series of the state variables $X = \{X_a\}$. Thus we employ the EM algorithm to maximize the log-likelihood of a model by a two-step iterative method that alternately estimates the states and parameters [17]. In the first step, the expectation step, the posterior distribution of the time series of a state $p(X|Y,\theta,S)$ is estimated based on the tentative parameter set θ_{old}. In the second step, the maximization step, the expectation value of $\ln p(X,Y|\theta,S)$ is calculated as

$$Q(\theta,\theta_{\text{old}}) = \langle \ln p(X,Y|\theta,S)\rangle_{p(X|Y,\theta_{\text{old}},S)} \tag{3}$$

and the parameter estimation is updated as

$$\theta_{\text{new}} = \underset{\theta}{\text{argmax}}\, Q(\theta,\theta_{\text{old}}). \tag{4}$$

In this step, the optimization problem is reduced to linear simultaneous equations and thus can be solved easily. However, the problem in the expectation step is still analytically unsolvable because the probability distribution of the time series is necessary. This calculation requires a state estimation at all time points including the points at which measurements are not conducted. We therefore obtain a numerical approximation of $p(X|Y,\theta,S)$ using a particle smoother algorithm that performs state estimations of nonlinear models using a Monte Carlo method [18,19]. The particle smoother (a numerical extension of the Kalman smoother) approximates a general non-Gaussian state distribution as a set of particles representing samples from the distribution and evaluates the log-likelihood of the models. Since the use of the particle smoother introduces stochasticity into the learning algorithm, a slight modification of the maximization step is required to ensure convergence of the learning [22]. The optimization function in Eq. (4) is replaced by $Q'_I(\theta) = (1-\alpha_I)Q'_{I-1}(\theta) + \alpha_I Q(\theta,\theta_{\text{old}})$, where I is the iteration index and $\{\alpha_I\}$ is a sequence of nonincreasing positive numbers converging to zero.

C. Artificial data generation

To validate the method, we apply it to artificial data generated from models of a eukaryotic cell-cycle control system since this system provides an illustrative example of cellular dynamics composed of many molecular components [7,23–26]. The cell cycle is a fundamental biological process characterized by repeated events underlying cell division and growth in which key proteins, cyclin and cyclin-dependent kinases, change their concentration periodically and activate various cellular functions such as DNA synthesis.

Two molecular circuit models of the cell-cycle control system in *Xenopus* embryos are adopted as the data generators: that proposed by Tyson and co-workers (the Tyson model) [23,24], and that proposed by Ferrell and co-workers (the Ferrell model) [25,26] (see Sec. 1 in Ref. [27] for the model equations and the parameter values). Both models show an oscillation onset through an increase of synthesis rate of cyclin, which is a bifurcation parameter as adopted in a classic experiment [28]. In contrast, they differ in the type of bifurcation at the onset: The Tyson model exhibits a saddle-node bifurcation on an invariant circle (SNIC), while the Ferrell model exhibits a supercritical Hopf bifurcation. We investigate whether the proposed learning procedure reproduces the correct bifurcation type of each model.

Both data generators are composed of nine variables including cyclin, cell division control protein 2 (Cdc2), and other regulatory proteins. We consider the active Cdc2 and cyclin concentrations to be observable variables since their levels have been observed in previous experiments [25]. The time-series data are generated by numerical calculations of these models as nonlinear Langevin equations at a few values of the cyclin synthesis rate s. Figure 2 exhibits all the time-series data that are used for our estimation below. We simulate noisy observation by adding Gaussian noise to each observation value. Artificial data are prepared for three cyclin synthesis rates across the bifurcation point and for each value of the bifurcation parameter, three independent time-series

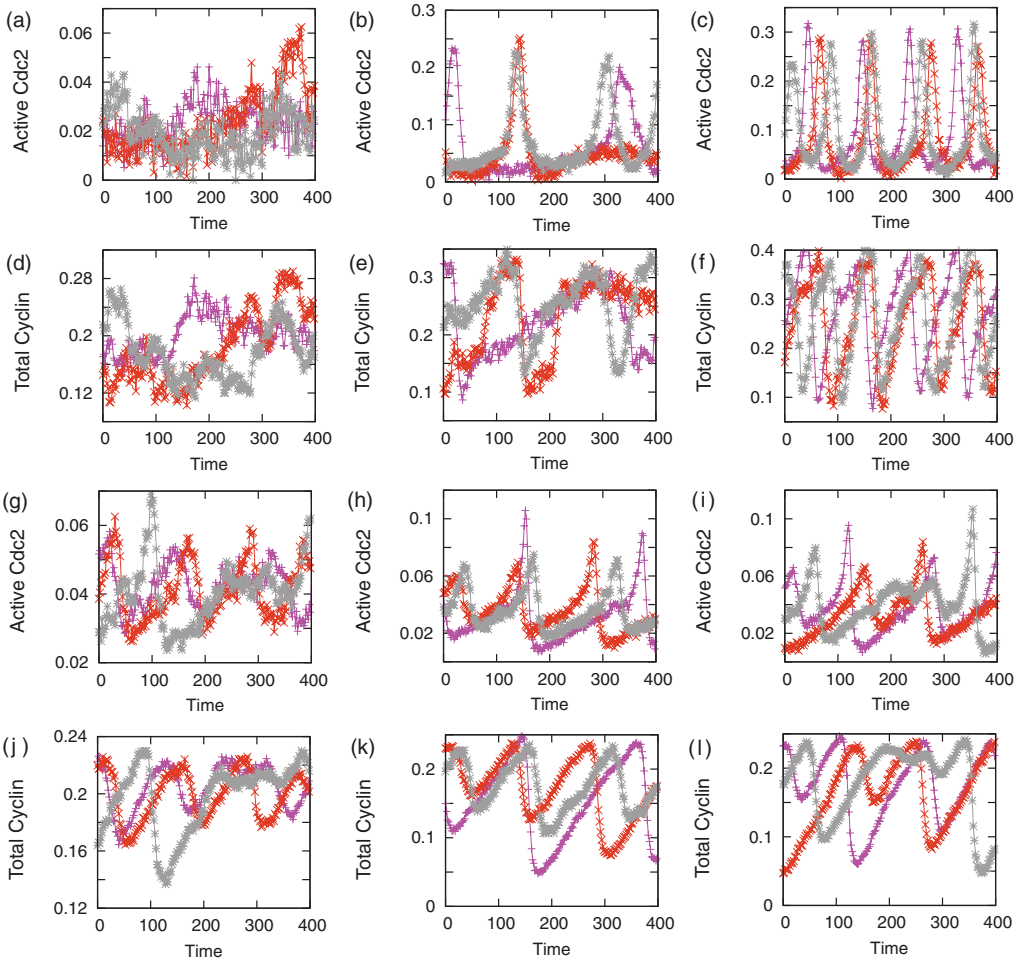

FIG. 2. (Color online) Noisy time series generated from the (a)–(f) Tyson and (g)–(l) Ferrell models through the artificial measurement process. The lines in each figure indicate samples from independent trials. The values of the bifurcation parameter are (a) and (d) $s = 0.002$, (b) and (e) 0.005, (c) and (f) 0.008, (g) and (j) 0.0005, (h) and (k) 0.001, and (i) and (l) 0.0015, respectively. These figures include all the data used for the estimation.

samples are prepared in which the oscillation exhibits large fluctuations in amplitude and period among the samples.

III. RESULTS

Considering a polynomial of degree M, we write the system equations to be learned as

$$f_i\left(\{x_j^t\},s\right) = k_i^1 + k_i^2 x_1^t + k_i^3 x_2^t + \cdots + k_i^{N_i}\left(x_D^t\right)^M. \qquad (5)$$

The observation equations are expressed simply as $y_i^r = x_i^r + \eta_i \phi_i^r$. Accordingly, y_1 (x_1) and y_2 (x_2) represent the observed (true) concentrations of active Cdc2 and cyclin, respectively. The other variables x_i ($i > 2$) represent the true concentrations of unobservable components. We take the constant term in the equation for cyclin to be the bifurcation parameter, i.e., $k_2^1 = s$.

The simplest polynomial form required for reproducing the observed dynamics is determined by starting with linear equations composed of active Cdc2 and cyclin (system dimension $D = 2$ and polynomial order $M = 1$) and increasing the D and M by one. It turns out that $D = 2$ is sufficient for reproducing a given time-series data set as shown below. The polynomial order M is determined by minimizing the information criteria through an optimization of the balance between the goodness of fit and the model complexity [29,30]. The Akaike information criterion (AIC) and Bayesian information criterion (BIC) are evaluated from the log-likelihood, parameter number, and data size for each model (Fig. 3). Both the AIC and BIC show

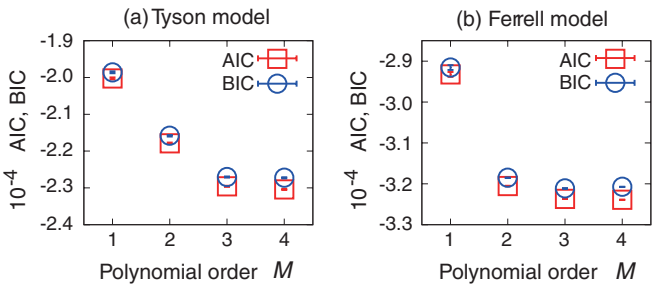

FIG. 3. (Color online) The AIC and BIC are plotted against different polynomial orders $M = 1,2,3,4$ for the (a) Tyson and (b) Ferrell models. For each evaluation, 20 different initial parameters are sampled to avoid the local minima. After the learning, the average and standard deviations are calculated from 100 evaluation trials based on the particle smoother.

TABLE I. Learned parameters of the third-order polynomial systems for the Tyson and Ferrell models. The indices j in $\{k_i^j\}$ $(i = 1,2)$ are defined as $f_i(x_1,x_2,s) = k_i^1 + k_i^2 x_1 + k_i^3 x_2 + k_i^4 (x_1)^2 + k_i^5 x_1 x_2 + k_i^6 (x_2)^2 + k_i^7 (x_1)^3 + k_i^8 (x_1)^2 x_2 + k_i^9 x_1 (x_2)^2 + k_i^{10} (x_2)^3$. The parameters $\mu_{i,a}$ and $V_{i,a}$, i.e., the mean and variance of the initial-state distribution, take different values among the samples (not shown). The estimated initial conditions are nearly identical to the observed ones.

Parameter	Tyson model	Ferrell model
k_1^1	0.00064	−0.00032
k_1^2	−0.334	−0.343
k_1^3	0.0655	0.0657
k_1^4	−0.8	1.05
k_1^5	2.39	−1.25
k_1^6	−0.639	0.107
k_1^7	−25.3	−47.3
k_1^8	31.1	47.8
k_1^9	−13.1	−3.97
k_1^{10}	2.22	−0.157
k_2^2	−0.4	−0.57
k_2^3	0.0612	0.0637
k_2^4	−4.62	−0.9
k_2^5	5.47	2.41
k_2^6	−0.805	0.000818
k_2^7	−9	61.4
k_2^8	22.7	−61.3
k_2^9	−15.3	18.3
k_2^{10}	1.92	−2.73
σ_1	0.00646	0.00199
σ_2	0.00703	0.00205
η_1	0.00334	0.00101
η_2	0.00621	0.00133

a decrease from $M = 1$ to 3, but an increase or insignificant decrease at $M = 4$. Therefore, we analyze models with $D = 2$ and $M = 3$ (see Table I for the learned parameter values).

To check whether the learning procedure can capture the bifurcation of the original data generator system, we compare the bifurcation diagrams of the learned systems with those of the data generators. Figures 4(a) and 4(b) show bifurcation diagrams against cyclin synthesis rate s for the learned systems (red lines) in the Tyson and Ferrell models, respectively. The bifurcation diagrams for the corresponding noiseless data generators are shown by the gray lines. Although the data for the learning are given only at three bifurcation parameter points (indicated by the dashed lines), the learned systems have diagrams quantitatively similar to those of the corresponding data generators. The sudden appearance of a limit cycle with finite amplitude is reproduced for the Tyson model, while the gradual increase in amplitude from the bifurcation point is reproduced for the Ferrell model. These features are characteristics of the SNIC and supercritical Hopf bifurcation. Nullclines of the learned systems in the vicinity of the bifurcation points are shown in Figs. 4(c) and 4(d) for the Tyson model and in Figs. 4(e) and 4(f) for the Ferrell model. The results confirm the SNIC and supercritical Hopf bifurcation

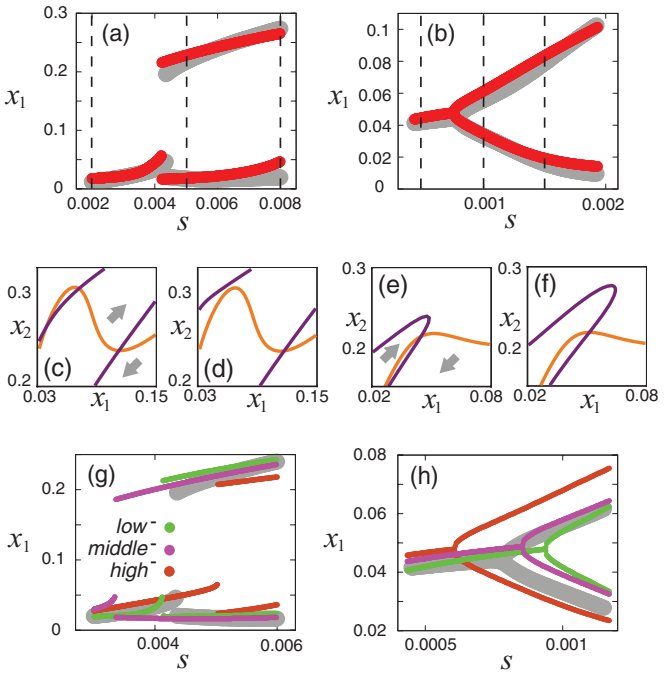

FIG. 4. (Color) Bifurcation diagrams of the (a) Tyson and (b) Ferrell models. The minimum and maximum active Cdc2 concentrations x_1 for the learned systems are plotted against the cyclin synthesis rate s (red); the corresponding concentrations of the data generators are also shown for comparison (gray). The dashed lines indicate points at which the data are given. For the Tyson model, there is another attractor with a tiny basin that is ignored. (c)–(f) The nullclines of the learned systems around the bifurcation points are shown. Purple and orange lines represent nullclines of x_1 (active Cdc2) and x_2 (cyclin), respectively, and the gray arrows indicate the flow directions. (c) and (d) The learned system from the Tyson model exhibits a SNIC and (e) and (f) that from the Ferrell model exhibits a supercritical Hopf bifurcation. The values of the bifurcation parameter are (c) $s = 0.0038$, (d) 0.0044, (e) 0.0005, and (f) 0.0012, respectively. (g) and (h) Bifurcation diagrams using the data at two of the three bifurcation parameter points. The learning that lacks data at the lowest, intermediate, and highest bifurcation parameter values are denoted by low^-, $middle^-$, and $high^-$, respectively.

at the onsets, respectively. Thus each learned system inherits the bifurcation type of the original model through the learning procedure in spite of noisy and partial observations.

When the learning is conducted by using the data on two of the three bifurcation parameter points, the learned systems still exhibit the correct bifurcation types, although the points of oscillation onset and amplitudes are biased [Figs. 4(g) and 4(h)]. Note that identification of bifurcation is possible even by using the data only on one side of a bifurcation point (as indicated by the green lines). These results indicate the interesting possibility that the learning procedure can predict the type of bifurcation that will occur from the data before the bifurcation point only.

We also show here how the high-dimensional phase space structures of the original data generators are mapped onto the lower-dimensional surfaces in the learned systems. Reduced two-variable models are derived by adiabatic elimination following a similar procedure by Novak and Tyson [31] (see

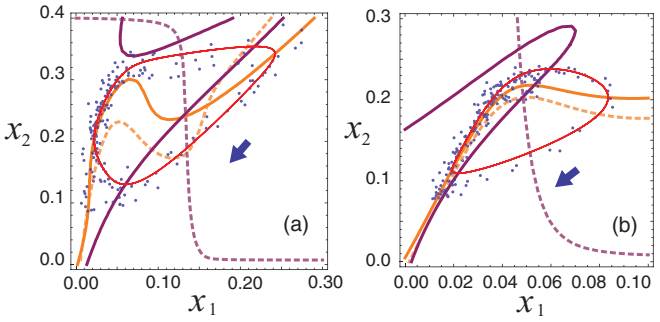

FIG. 5. (Color) Comparison of the learned systems and reduced models for the (a) Tyson and (b) Ferrell models. The purple and orange lines represent the nullclines of x_1 (active Cdc2) and x_2 (cyclin), respectively, for the learned systems (solid lines) and reduced models (dashed lines). A noisy time series from the data generators (blue dots) and the orbits of the learned models (red lines) are also shown. The blue arrows indicate the flow direction. The values of the bifurcation parameter are (a) $s = 0.005$ and (b) 0.0015.

Sec. 2 in Ref. [27] for the detailed procedure and reduced model equations). Like the learned systems, the reduced models are composed of active Cdc2 and total cyclin. Figure 5 shows the nullclines of the learned systems (the solid orange and purple lines) and the reduced models (the dashed lines). In both the Tyson and Ferrell models, the learned system and reduced model nullclines for active Cdc2 have a similar N-shaped form (orange lines), indicating the existence of positive feedback in the molecular circuits. In contrast, those for the total cyclin disagree quite significantly. To check the consistency of the nullclines and dynamics, Fig. 5 also shows a noisy time series from the data generators (blue points) and the orbit of the learned system (red lines). The nullclines of the learned systems are consistent with the dynamics in the data, but the reduced models are not. This failure arises because the dynamics of a component mediating the inhibition from active Cdc2 to cyclin is not fast enough to allow the adiabatic approximation. A higher-order contribution beyond the adiabatic elimination performed here should be included, which requires complicated technical work. Nevertheless, the learning process automatically reproduces the appropriate low-dimensional dynamics and estimates the bifurcation types without knowledge of the detailed high-dimensional model systems.

IV. SUMMARY AND DISCUSSION

Gathering biological data is complicated by intrinsic and observation noises, partial observation, and a small number of possible experimental conditions. We have outlined here a machine-learning procedure based on likelihood maximization that makes use of all the information in the time-series data, including that in the noise. By using synthetic data that share the difficulties found in actual biological data, we demonstrated that the procedure could derive low-dimensional model equations that reproduced the obtained time-series data and captured the bifurcation types of the original systems. These results support the conjecture that the learning procedure will be able to construct reliable low-dimensional models for real time-series data of active Cdc2 and cyclin levels in future studies. Being able to identify the model systems and bifurcation types will provide a useful method for elucidating both the molecular interactions in the circuit and the biological functions of the dynamics. Further, since the dynamics and bifurcation are found widely among various biological processes, the method is expected to be applicable to various cell systems with cell-imaging data.

The present method can be used together with other machine-learning techniques. For example, it was recently shown that compressive sensing exhibits a high performance for learning chaotic systems [12]. Incorporation of prior distributions for model parameters achieves such a sparse optimization in the statistical framework [32] and will enable us to extract an appropriate submodel from an original complex equation by eliminating unnecessary terms. It is advantageous in the case that complex time-series data are obtained typically from a chaotic system.

The systems to be learned are assumed to be reducible to the effectively lower-dimensional ones, which limits the applicability of the method. Careful evaluation of how extensible the method is to higher-dimensional state spaces remains an important future task. Notice that this does not exclude application of the method to spatially extended systems such as reaction-diffusion systems since description of the systems is often simple and the number of parameters to be determined is small. In summary, the proposed method will be an efficient way to capture the essential features of the cellular dynamics by mediating dynamical system modeling with experimental observations.

ACKNOWLEDGMENTS

We would like to thank K. Kamino, N. Saito, and S. Sawai for illuminating comments and stimulating discussions. This work was supported by the Grant-in-Aid MEXT/JSPS (Grant No. 24115503).

APPENDIX: ALGORITHM DESCRIPTION

1. The EM algorithm

Our aim is to find model parameters θ by maximizing the log-likelihood function

$$\ln L(\theta) = \ln p(Y|\theta,S) = \ln \int_X p(Y|X,\theta)p(X|\theta,S), \qquad \text{(A1)}$$

where X, Y, and S are time series of states, data, and bifurcation parameter values, respectively. We employ an EM algorithm that maximizes $\ln P(X,Y|\theta,S)$ (the complete-data log-likelihood function), which is equivalent to maximizing the likelihood in Eq. (A1) [17]. By iterating two steps known as the expectation and maximization steps, the states X and parameters θ are estimated alternately. Since our implementation of the expectation step includes the Monte Carlo method as described below, the stochastic approximation EM (SAEM) algorithm is adopted [22]. The SAEM procedure is described as follows.

(i) Initialize the parameter vector as $\theta = \theta_0$ and set the iteration number I to zero.

(ii) Calculate the posterior distribution of the entire time series of state variable $p(X|Y,\theta,S)$ (the expectation step).

(iii) Rename θ as $\theta_{\rm old}$ and update the parameter vector as

$$\theta_{\rm new} = \underset{\theta}{\rm argmax}\, Q_I(\theta), \tag{A2}$$

$$Q_I = \begin{cases} Q(\theta,\theta_{\rm old}) & (I=0) \\ (1-\alpha_I)Q_{I-1}(\theta) + \alpha_I Q(\theta,\theta_{\rm old}) & (I>0), \end{cases} \tag{A3}$$

where

$$Q(\theta,\theta_{\rm old}) = \langle \ln p(X,Y|\theta,S)\rangle_{p(X|Y,\theta_{\rm old},S)} \tag{A4}$$

and $\{\alpha_I\}$ is a nonincreasing sequence of positive values converging to zero (the maximization step).

(iv) Increment I by one and iterate steps (ii) and (iii) until the estimation of the parameter vector converges.

2. The expectation step

Since different time-series samples are independent stochastic variables, we can write

$$\ln p(X|Y,\theta,S) = \sum_a \ln p(X_a|Y_a,\theta,s_a), \tag{A5}$$

where a is an index of the time-series samples. Each $\ln p(X_a|Y_a,\theta,s_a)$ is evaluated by using a particle smoother algorithm that approximates the non-Gaussian distribution of the state x_i $(i=1,\ldots,D)$ as a collection of many particles, each of which represents a sample from the distribution [18,19]. For the ath time series, let $x_{i,a}^{t,p}$ denote the pth particle for representing x_i^t at time point $t \in T$ and let $y_{i,a}^{\tau}$ denote an observed value at a time point $\tau \in \mathcal{T}$. The procedure of the particle smoother is described as follows, for $a=1,\ldots,A$.

(a) For $i=1,\ldots,D$ and $p=1,\ldots,P$, set the initial states as $x_{i,a}^{0,p} \sim N(\mu_{i,a},V_{i,a})$ and normalize the weights as $w_a^p = 1/P$.

(b) At each $t=0,1,\ldots$, (i) for $i=1,\ldots,D$ and $p=1,\ldots,P$, calculate $x_{i,a}^{t+1,p}$ from $x_{i,a}^{t,p}$ by using the state equations; (ii) if $t=\tau\in\mathcal{T}$, update the weights of the particles as

$$w_a^p = \frac{w_{a,\rm old}^p l_a^{r,p}}{\sum_p w_{a,\rm old}^p l_a^{r,p}}, \tag{A6}$$

where

$$l_a^{r,p} = \prod_i^d p\left(y_{i,a}^r \middle| \{x_{j,a}^{r,p}\}\right); \tag{A7}$$

and (iii) if $P_{\rm eff} = 1/\sum_p (w_a^p)^2 < P_{\rm thresh}$ (i.e., if the effective number of the particles falls below a threshold value), resample the particles according to the new weights. Note that the history of particles $(x_{i,a}^{0,p},x_{i,a}^{1,p},\ldots,x_{i,a}^{r-1,p})$ is resampled in parallel.

(c) Finish when all data points have passed $[t=\max(\mathcal{T})]$ and estimate the log-likelihood as

$$\ln L_a(\theta) = \sum_{\tau\in\mathcal{T}} \ln\left(\frac{1}{P}\sum_p^P l_a^{\tau,p}\right). \tag{A8}$$

Based on the particle smoother, the posterior distribution of the time series of the state is approximated as

$$p(X_a|Y_a,\theta_{\rm old},s_a) = \sum_{p=1}^P w_a^p \delta\left(X_a - X_a^p\right), \tag{A9}$$

where X_a^p indicates a sample path $(\{x_{i,a}^{t,p}\}, i=1,\ldots,D, t\in T)$. On the basis of this approximation, we calculate the average of the complete-data log-likelihood as

$$\begin{aligned} Q(\theta,\theta_{\rm old}) = {} & \sum_{a=1}^A\sum_{p=1}^P\sum_{i=1}^D w_a^p\left(-\frac{1}{2}\ln 2\pi V_{i,a} - \frac{\left(x_{i,a}^{0,p}-\mu_{i,a}\right)^2}{2V_{i,a}}\right) + \sum_{a=1}^A\sum_{p=1}^P\sum_{t\in T}\sum_{i=1}^D w_a^p\left(-\frac{1}{2}\ln 2\pi(\sigma_i)^2\right. \\ & \left. - \frac{\left(\left(x_{i,a}^{t+1,p}-x_{i,a}^{t,p}\right) - \Delta t\sum_{n=1}^{N_i} k_i^n f_i^n\left(\left\{x_{j,a}^{t,p}\right\},s_a\right)\right)^2}{2(\sigma_i)^2\Delta t}\right) + \sum_{a=1}^A\sum_{p=1}^P\sum_{\tau\in\mathcal{T}}\sum_{i=1}^d w_a^p\left(-\frac{1}{2}\ln 2\pi(\eta_i)^2 - \frac{\left[y_{i,a}^{\tau,p} - g_i\left(x_{i,a}^{\tau,p}\right)\right]^2}{2(\eta_i)^2}\right). \end{aligned}$$

3. The maximization step

At the Ith iteration, the parameter-value update is performed by finding the θ for which $\frac{d}{d\theta}Q_I(\theta)=0$. This problem leads to linear algebraic equations that can be solved easily because of the linearity assumption in the state equations as $f_i(\{x_j^t\},s) = \sum_n^{N_i} k_i^n f_i^n(\{x_j\},s)$ and Gaussian statistics of the noises. Only for the variance of the initial condition $V_{i,a}$, we define the minimum value $V_{\rm min}$ to avoid an unnaturally small value resulting from a problem called sample impoverishment [33].

4. Settings

For the decreasing sequence in the SAEM algorithm $\{\alpha_I\}$, we employed $\alpha_I = 1$ $(I \leqslant 30)$ or $1/\sqrt{(I-30)}$ (otherwise).

TABLE II. Predetermined parameters used in the learning algorithm. The interval $[-x,x]$ represents a uniform distribution.

Parameters in the learning algorithm	Value
number of time series A	9
integration time Δt	1.0
entire-time points T	$\{0,1,2,\ldots,400\}$
observed-time points $\mathcal{T}$	$\{0,2,4,\ldots,400\}$
particle number P	1000
threshold for resampling $P_{\rm thresh}$	500
initial dynamics parameters $\{k_i^n\}$	$[-0.001,0.001]$
initial system noise strength σ_i	0.1
initial observation noise strength η_i	0.2
minimum initial variance $V_{\rm min}$	0.001

At the onset of the learning, the means of the initial state distributions $\{\mu_{i,a}\}$ are set as $\mu_{i,a} = y_{i,a}^{\tau=0}$, while the variances $\{V_{i,a}\}$ are set as $V_{i,a} = 10V_{\min}$. The other settings for the learning algorithm are listed in Table II.

[1] K. Kaneko, *Life: An Introduction to Complex Systems Biology* (Springer, Berlin, 2006).

[2] U. Alon, *An Introduction to Systems Biology: Design Principles of Biological Circuits* (Chapman and Hall/CRC, Boca Raton, FL, 2006).

[3] D. J. Stephens and V. J. Allan, Science **300**, 82 (2003).

[4] Y. Shav-Tal, R. H. Singer, and X. Darzacq, Nat. Rev. Mol. Cell Biol. **5**, 855 (2004).

[5] M. Fernández-Suárez and A. Y. Ting, Nat. Rev. Mol. Cell Biol. **9**, 929 (2008).

[6] G. M. Süel, R. P. Kulkarni, J. Dworkin, J. Garcia-Ojalvo, and M. B. Elowitz, Science **315**, 1716 (2007).

[7] B. Pfeuty, T. David-Pfeuty, and K. Kaneko, Cell Cycle **7**, 3246 (2008).

[8] E. M. Izhikevich, *Dynamical Systems in Neuroscience: The Geometry of Excitability and Bursting* (MIT Press, Cambridge, 2006).

[9] K. Judd and A. Mees, Physica D **92**, 221 (1996).

[10] H. U. Voss, P. Kolodner, M. Abel, and J. Kurths, Phys. Rev. Lett. **83**, 3422 (1999).

[11] T. Müller and J. Timmer, Physica D **171**, 1 (2002).

[12] W.-X. Wang, R. Yang, Y.-C. Lai, V. Kovanis, and C. Grebogi, Phys. Rev. Lett. **106**, 154101 (2011).

[13] A. Sitz, J. Kurths, and H. U. Voss, Phys. Rev. E **68**, 016202 (2003).

[14] M. Nagasaki, R. Yamaguchi, R. Yoshida, S. Imoto, A. Doi, Y. Tamada, H. Matsuno, S. Miyano, and T. Higuchi, Genome Inf. **17**, 46 (2006).

[15] R. Yoshida, M. Nagasaki, R. Yamaguchi, S. Imoto, S. Miyano, and T. Higuchi, Bioinformatics **24**, 2592 (2008).

[16] J. Ohkubo, Phys. Rev. E **84**, 066702 (2011).

[17] A. P. Dempster, N. M. Laird, and D. B. Rubin, J. R. Stat. Soc. B **39**, 1 (1977).

[18] N. Gordon, D. Salmond, and A. Smith, IEE Proc.-F **140**, 107 (1993).

[19] G. Kitagawa, J. Comput. Graph. Stat. **5**, 1 (1996).

[20] C. Andrieu, A. Doucet, S. Singh, and V. Tadic, Proc. IEEE **92**, 423 (2004).

[21] P. E. Kloeden and E. Platen, *Numerical Solution of Stochastic Differential Equations*, 3rd ed. (Springer, Berlin, 1999).

[22] B. Delyon, M. Lavielle, and E. Moulines, Ann. Stat. **27**, 94 (1999).

[23] B. Novak and J. J. Tyson, J. Cell Sci. **106**, 1153 (1993).

[24] M. Borisuk and J. J. Tyson, J. Theor. Biol. **195**, 69 (1998).

[25] J. R. Pomerening, S. Y. Kim, and J. E. Ferrell, Cell **122**, 565 (2005).

[26] T. Y.-C. Tsai, Y. S. Choi, W. Ma, J. R. Pomerening, C. Tang, and J. E. Ferrell, Science **321**, 126 (2008).

[27] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevE.87.042716 for preparation of the artificial data (Sec. 1) and the reduced equations by adiabatic approximation (Sec. 2) .

[28] A. W. Murray and M. W. Kirschner, Nature (London) **339**, 275 (1989).

[29] H. Akaike, IEEE Trans. Autom. Control **19**, 716 (1974).

[30] G. Schwarz, Ann. Stat. **6**, 461 (1978).

[31] B. Novak and J. J. Tyson, J. Theor. Biol. **165**, 101 (1993).

[32] M. Kyung, J. Gill, M. Ghosh, and G. Casella, Bayesian Anal. **5**, 369 (2010).

[33] M. Arulampalam, S. Maskell, N. Gordon, and T. Clapp, IEEE Trans. Signal Process. **50**, 174 (2002).