Reconstruction of cell signaling dynamics by statistical time series analysis

（一細胞時系列解析による シグナル伝達機構の抽出と再構成）

Yohei Kondo
(近藤 洋平)

Contents

Chapter 1

General introduction

1.1 Background

1.1.1 Cell signaling

The cell is the fundamental unit of life, and contains molecular systems to realize biological functions that are essential for living things to survive [1]. A unicellular organism performs all the functions required for proliferation, for example, seeking food, avoiding toxin, etc. On the other hand, the cells in a multicellular organism are organized by themselves and differentiate to carry out specific functions required as muscle cell, neuron, and so forth. For both cases, performing the proper functions requires spatial and temporal orchestration of the molecular systems in the cell. To achieve it, the biochemical systems called cell signaling systems are formed through the evolutionary process. The cell signaling regulates the intracellular systems, and perceives and responds to extracellular environment.

Molecular biology studies have identified molecular species involved in the signaling systems such as controlling cell locomotion and cell cycle. In parallel, these studies have also revealed interactions between the molecules; the molecular species, as cell signaling components, activate/inactivate states of other components such as phosphorylation level of protein. As a result, the picture of a regulatory network consisting of the components and the interactions as nodes and edges in which biological information transmits on the pathways was developed [2]. Then, a scientific field called systems biology arose to investigate collective properties of such a network.

Through the studies, investigating the dynamics of signaling is crucial to understand cell behaviors. One reason for this is that cell signaling systems are often demanded to process information on their continuously changing environments. This aspect is clearly demonstrated by studies on a bacterial signal transduction for motor control by which the cell performs chemotaxis through temporal comparison of external chemical concentrations [3, 4, 5].

In addition, there is another important reason. Even when environments stay constant, we frequently observe that signaling systems exhibit temporal variation. The circadian clock is a well-established example of the self-sustained dynamics [6]. This oscillation is synchronized with the day-night cycle through light stimulation but continues with a period similar to 24 hours even under constant darkness.

1.1.2 Dynamical systems approach

In order to deal with the dynamic nature of cell signaling, relevance of dynamical systems to modeling such phenomena has been increasingly recognized [2, 7]. A dynamical system is a mathematical concept where its rule describes time evolution of a state of a system in the phase space. Various phenomena ranging from climate change to chemical reactions have been modeled extensively by dynamical systems [8, 9], because the dynamical-systems approach enables us to reduce a huge set of data into a simple description, and to predict the dynamics quantitatively, and to understand the system of concern based on theoretical tools.

In applications to cell signaling systems, one of classical and best-known examples is the Hodgkin-Huxley model that describes dynamics of electric potentials in neurons [10]. The model was initially considered to explain propagation of the potentials in the squid giant axon based on dynamics of ion channels in cell membrane. Numerical calculations of the model reproduce neuronal excitability, in other words, sharp and transient activation of the potentials underlying the propagation of electrical signal. Furthermore, the study enabled us to understand neurons as electric circuits, and inspired subsequent studies to explain other neuronal phenomena. The success over decades shows power of dynamical systems modeling to understand cellular dynamics, and encourages other attempts, down to this day.

Based on recent accumalation of experimental knowledge, we have come to know how large and complicated signaling networks are. Then, mechanistic models describing the biochemical networks have grown in size and complexity [11, 12]. However, a full description of all details of the systems is often impractical and not informative to elucidate underlying principles of the evolutionary designs. Instead, simplified description by a small number of variables is more useful for our comprehension. Neural cell biology gives once again an illustrative example for this, the FitzHugh-Nagumo equations that are derived from a simplification of the Hodgkin-Huxley model [13, 14]. The aim of the FitzHugh-Nagumo equations is to extract fundamental mathematical properties of the original model to explain the neuronal excitability that plays a key role in information propagation in the neurons. Analysis of the equations leads to clear understanding of a core mechanism related to the neuronal excitability based on a qualitative feature in the phase space. Besides, the excitability appears as a universal concept that helps to con-

struct models of other cell systems exhibiting similar behaviors [15], beyond electrochemical systems.

Therefore, along with efforts of constructing detailed models, many attempts have been made to extract simplified descriptions of the systems that underlie the cell signaling dynamics. In those attempts, a technique to reduce complexity of mechanistic models while keeping the dynamics is required. To this end, adiabatic approximation, decreasing dimension of a state vector by steady state assumption for components with fast dynamics, is of common use [16,17]. The reduced models are analyzed with dynamical-systems techniques. In particular, bifurcation theory has played a significant role, characterizing qualitative changes of dynamics with a smooth variation in the parameter values [18,19,20,21]. Many cell signaling processes, including cell cycle control and neuronal firing, exhibit such qualitative changes like oscillation onset in response to environmental and experimental perturbations. The way of change in system behavior upon parameter values, that is, bifurcation, is considered as a basic property to be reproduced by simplified models. Model simplification while keeping bifurcation types is important to derive a simple model relevant to biological phenomena.

To coordinate the studies on simplified models from mechanistic descriptions, abstract models designed to reproduce specific biological functions are constructed and studied. The model is expected to present a fundamental mathematical property that is required for the specific function; the abstracted property becomes a minimal representation of a core mechanism of the function and helps a model simplification process by advising the destination.

In addition to those approaches focusing mainly on features in phase spaces, an approach focusing on network topologies has also proven as a powerful tool to extract simplified principles. This approach defines network motifs, a set of small simple networks with a few components, which are considered as building blocks of the biological networks [2]. The concept of network motifs is originally motivated by the fact that specific substructures appear in transcriptional networks more often than expected in random networks, and the frequent occurrences of the specific patterns are inferred to be caused by functional advantage of the patterns. Then, relationships between the network motifs and the functions are investigated. The functions include, for example, pulse generation, response acceleration, and so forth. Note that, since the concept of the network motifs is based only on the structures, we can define them for other biological networks in the same way. In fact, network motifs are found later in neural networks, metabolic networks, and so on.

The approaches focusing on a simplified phase space and a network motif to investigate a biological function are clearly not exclusive. Combination of these approaches provide a common language to elucidate design principles of an architecture of cell signaling system across diverse organisms.

1.1.3 Measuring live-cell dynamics

Recently, advances in experimental techniques such as fluorescent probes for live-cell imaging clarify molecular activities in cell systems at unprecedented spatiotemporal resolutions [22, 23, 24]. The high resolutions enable us to measure localization and dynamics of proteins involved in a signaling system inside a single cell. Since correct regulations of the spatiotemporal dynamics are prerequisite for proper functioning, systems biology on the cell signaling receives huge benefits from the high resolutions.

Another important point of such imaging methods is that it allows us to obtain time-series data on dynamics at a single-cell level, not a population level. A key reason why data analysis at a single-cell level is required is that, based on the analysis, we can reveal dynamics hidden by averaging over cell population. For example, such an approach is important to elucidate oscillatory dynamics, as is demonstrated by sequential works on expression of a cancer suppressor gene [25, 26]. These single-cell time series exhibit a sustained oscillation that previously appeared as damped oscillation in assays measuring dynamics of a cell population where the cells have different oscillatory phases because of noise.

Therefore, the observations reveal quantitative dynamics of information transmission in the cell signaling system. Then dynamical-systems approach is expected to elucidate how cellular behaviors are generated from the internal molecular systems more efficiently than ever.

However, it has been also revealed that signaling dynamics in a single cell is highly stochastic, and the stochasticity blurs the observed data and its interpretation. The stochasticity is considered to originally come from finiteness of the signaling molecules, and it has been elucidated how this noise is amplified and propagates through transcriptional and translational processes [27, 28, 29].

Furthermore, measurement noise, partial observation, and a low controllability are still problematic for modeling biological systems. Indeed, measurement processes are also highly stochastic, only a few components in a system are observable, and only a small number of experimental conditions can be examined. The incompleteness in data is an obstacle in model construction from experimental observations on the molecular systems. Thus, instructive criteria and practical methods are required for deriving model equations by systematically integrating the information in the data.

1.1.4 Statistical inference from biological time series

Statistics has developed powerful methods to estimate models under incompleteness in data. In particular, for analyzing dynamics, a method focusing on time-series data is required. This demand has accelerated researches on statistical machine learning to estimate dynamical systems. The commonly-

used modeling framework is called a state space model, which is composed of state equations (a dynamical system) and observation equations that describe system dynamics and observation process, respectively (Fig. 1.1). We can consider both system and observation noises explicitly. The model representation is originally investigated in control theory, but nowadays it has a wide range of applications.

In statistical machine learning, likelihood is the basic function to be maximized by choosing model parameter values in order to estimate them. The likelihood represents probability that a model reproduces observations, which gives a standard measure of goodness of fit. A big advantage of using the statistical concept is that it enables model selection based on information criteria. The criteria give a solution to the problem of trade-off between goodness of fit and model complexity. It is helpful for studying cell signaling dynamics for which there is, in general, a variety of candidate models with different sets of components and interactions. In addition to the likelihood maximization approach, we can consider prior distributions for models and parameters, and posterior distribution of them in Bayesian approach. By setting prior distributions, information from other sources is included; by calculating posterior distributions, we can obtain reasonable estimation even when the amount of data is insufficient to determine a unique solution.

The estimation procedures are, with a sufficient amount of data, easy in linear dynamics, linear observation equations, and Gaussian noises. On the other hand, in practical problems, it is often necessary to estimate nonlinear systems. Such an estimation is rather difficult because, through the nonlinearity, probability distributions to be estimated, for example, state distributions, turn out to be non-Gaussian and cannot be handled analytically. However, as computational resources grew, a variety of techniques for numerical approximation of the general distributions has been developed for practical use. This is critical for applications to the cell signaling systems that show complex dynamics.

In application to cell signaling systems, the statistical framework of state space models has been effective for inference of transcriptional networks by using gene expression data from microarray experiments [30, 31, 32]. The earlier studies mainly focused on the estimation of causal, not mechanistic, relation between levels of mRNA molecules, aiming at drawing the structures of transcriptional regulatory networks. Simple model representations enable estimation of the networks without detailed biochemical knowledge, while the quantitative accuracy of models and the interpretability of parameters are limited. On the other hand, some of recent studies tried to handle rather mechanistic models such as Michaelis-Menten equation and to estimate biochemical parameters [33, 34]. Such methods are indispensable for quantitative analysis of biochemical systems with a large number of components, because the complicated models inevitably contain many uncertain parameters whose values are difficult to be measured directly. In

addtion, non-parametric approach was also examined to model complicated equations flexibly [35].

In contrast to data obtained by biochemical techniques, the application of estimation from live-cell imaging data is just beginning at present [36,37]. However, the high temporal resolution of the live-cell imaging techniques seems to be suitable for estimating dynamical systems, and then fruitful applications are expected in future.

In particular, it is pointed out that such applications are helpful especially for single-cell dynamics [36]. The primary reason is that the statistical framework provides a way to estimate noise in system dynamics, separately from measurement noise, which is difficult without employing statistical approaches. The intrinsic noise in biochemical kinetics at a single-cell level often requires stochastic modeling to quantitatively explain heterogeneity of cells that appears even in a clonal population [38]. At least in some cases, the stochasticity underlying the heterogeneity is considered to have biological roles, for example, facilitation of evolutionary processes [39]. To elucidate biological functions of the heterogeneity, the statistical approach will be hopeful.

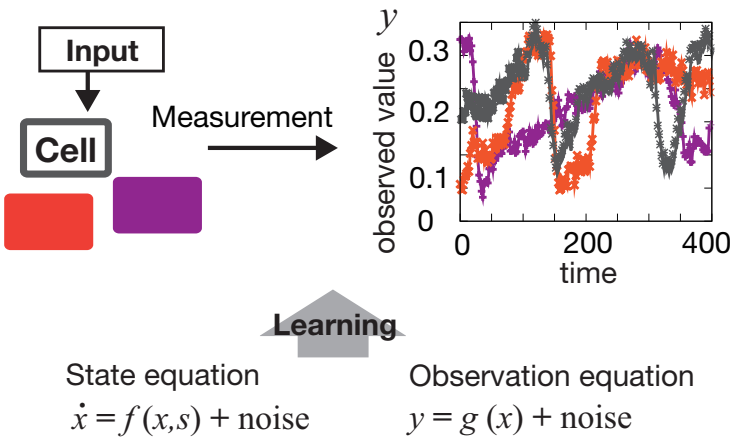


Figure 1.1: Schematic representation of the statistical learning of a state space model. Time-series data of the molecular activities in a cell system are obtained from measurements under a given input level. System and observation equations are then trained to reproduce the time-series data.

1.2 Reconstruction of human-understandable dynamical systems

We hope to promote the understanding of cell signaling dynamics based on the dynamical systems approach especially with the use of simplified phase space. The approach requires a mechanistic model to be reduced. However, such a description of a system is hard to be reliable in detail, not only with regard to parameter values but also to model equations themselves.

To overcome the difficulty, we employ the statistical machine learning techniques to construct a simple, low-dimensional, human-understandable model of a system directly from data, not through a detailed mechanistic description. Such a simplification has often been made implicitly without statistical framework, but without standard measure to evaluate simplified models from incomplete data, it is difficult to justify them and to select a better model. The employment of the statistical framework provides a reasonable solution to the problem.

To this end, we propose to set dynamical-systems models that are not directly based on realistic biochemistry but are suitable to extract simplified principles such as bifurcation type and network motif. In other words, our models, to be estimated, are chosen appropriately for analysis by systems biology terms. In this thesis, we combine dynamical-systems analysis and statistical machine learning, aiming to facilitate understandings of cell behaviors based on the signaling systems.

1.3 Overview of this thesis

This thesis is organized as follows: In Chapter 2, we focus on a basic feature of dynamical systems called bifurcation that characterizes qualitative changes of the dynamics. A statistical machine-learning method is designed to derive low-dimensional models with bifurcation from partial observations of noisy high-dimensional dynamical systems. The method is tested by using artificial data generated from two cell-cycle control models that exhibit different bifurcations, to demonstrate that the learned systems inherit the bifurcation types. We also discuss the relevance of this procedure to model reduction and compare with a standard method, adiabatic approximation. In Chapter 3, we extend the proposed approach to include sparse regularization. The extension enables us to start from a complicated model equation that contains many possible terms, and to extract an appropriate sub-model by eliminating unnecessary terms to reproduce observations. The performance of the algorithm is demonstrated by using artificial data from chaotic dynamics and those from adaptive dynamics. In Chapter 4, we analyze a signaling system that underlies cell-cell communication of a social amoeba. Measurements at a single-cell level have characterized quan-

titatively the adaptive dynamics, and revealed an invariance property, called FCD. Aiming to explain multicellular dynamics from the single-cell level, we examined two strategies. First, based on elementary analysis, we attempt to draw conclusions only from the FCD property. The analysis indicates that a symmetry property required for FCD achieves robustness in multicellular dynamics to cell-density fluctuation. Second, we perform statistical learning of a model by using live-cell imaging data on the signaling system. As a result, we obtain a model that reproduces quantitatively the input-output characteristics of the system. Furthermore, we extract a simplified core system from the model based on a sparse optimization. In Chapter 5, based on the model of the social amoeba at a single cell level, we construct a reaction-diffusion model of the multicellular dynamics in development. The quantitative predictions from the reconstructed model are evaluated in comparison with experimental observations. Other models based on toy models of the signaling system are also constructed and discussed. This thesis concludes in Chapter 6 with summary and discussions towards the understanding of cellular dynamics, and perspectives on future.

Chapter 2

Reduction of dynamical systems with bifurcation

2.1 Introduction

2.1.1 Bifurcation methodology in systems biology

Cell signaling systems often exhibit qualitative changes in the dynamics like oscillation onset, as the environmental conditions are varied. Different modes of signaling dynamics are considered to cause different schemes of cell behaviors that are suitable for the corresponding environments.

In dynamical systems theory, such a change in dynamics is characterized and classified by a mathematical concept, bifurcation. Bifurcation gives a pattern of topological change of flow in the phase space, with a smooth variation of control parameters interpreted as biological conditions. Because of the topological nature, such characterizations are valuable to understand the underlying mechanism of biological functions that are insensitive to biochemical details [18, 19, 20, 21].

The independence from the detailed biochemical mechanisms makes the type of bifurcations to be a basic property that is reproduced by simplified models and provides a criterion to select a better model. In the simplified models, we can see clearly what mathematical properties in the models are responsible to produce behaviors of the original detailed models or experimental observations. In addition, an elemental low-dimensional dynamical system with a correct bifurcation type is useful to obtain a collective behavior of an ensemble of the elements. For example, in neuroscience, elemental single neuron models are employed to simulate large-scale brain dynamics. Those elemental models being simplified while keeping bifurcations are considered to achieve a balance between biological relevance and computational efficiency [20]. Therefore, identification of low-dimensional model systems that inherit the original bifurcation type is a crucial step in understanding

the dynamics in systems biology.

2.1.2 Determining bifurcation types from observations

If a mechanism underlying observed dynamics is not known, we are required to determine a bifurcation type only from experimental data. The determination from data is a common problem not only in cellular dynamics, but in ecological population dynamics, chemical kinetics, and etc. The commonly-used method is to focus on asymptotic behavior of a system near the point of bifurcation [41]. For example, to justify a supercritical Hopf bifurcation, a most frequent type of oscillation onset with a variation of a parameter, we check whether amplitude of the oscillation increases in proportion to square root of absolute deviation of the parameter from the bifurcation point. However, the precise control of parameter and the accurate measurement of system behavior are impractical for cell signaling systems in many cases, unfortunately. Because of difficulties in control and measurement on molecular systems in the cell, the traditional methodology is inapplicable. Then, we need an alternative.

To solve the problem, we employ statistical machine-learning techniques to automatically derive low-dimensional models with bifurcations from data at a few conditions. Various sophisticated techniques for learning nonlinear dynamical systems from observations on dynamics, that is, time-series data, have been developed especially in the case of noiseless system dynamics [42, 43, 44, 45]. Although theoretical ground of learning from stochastic dynamics is weaker, several methods are proposed in the framework of statistical theory recently [34, 46, 47, 48].

However, at present it is not yet examined whether such a statistical framework can be used to estimate a model with a bifurcation. Here we adopt statistical techniques to obtain a model with correct type of bifurcation, through approximation of flow in the phase space by abstract dynamical systems, rather than aiming to fit the mechanistic model parameters to the observations. To this end, the Expectation-Maximization (EM) algorithm and particle smoother [49, 50, 51] are combined as was suggested to be applicable for nonlinear and stochastic dynamical systems [52]. To deal with fluctuation in likelihood evaluation coming from Monte-Carlo procedure in particle smoother, we adopt the stochastic approximation EM (SAEM) algorithm [53]. The employment of the statistical framework enables us to utilize information from noisy time-series data fully, not only from a specific feature like oscillation amplitude as in the traditional approach. For the purpose to obtain models with bifurcations, all of the time-series data at different parameter values across the bifurcation point are simultaneously used for the learning. By taking advantage of continuity, we do not need the data from many parameter values, but those from a few (a couple of) parameter values are sufficient to predict the dynamics with correct bifurcation

types.

Before application to experimental data, verification of the approach by artificially-generated data is helpful since we can directly compare learned systems with true ones. In particular, because reduced descriptions in systems biology are useful, it is beneficial to confirm that the learned systems with a few variables can approximate dynamics restricted to low-dimensional sub spaces of original high-dimensional systems. In other words, we try to automatize a model reduction while keeping the bifurcation type, not through a detailed biochemical description or unrealistically accurate experiments, but directly from noisy partial time-series data. As shown schematically in Fig. 2.1, we now test the proposed approach and evaluate the learned systems in comparison with reduced equations of true systems by adiabatic approximation.

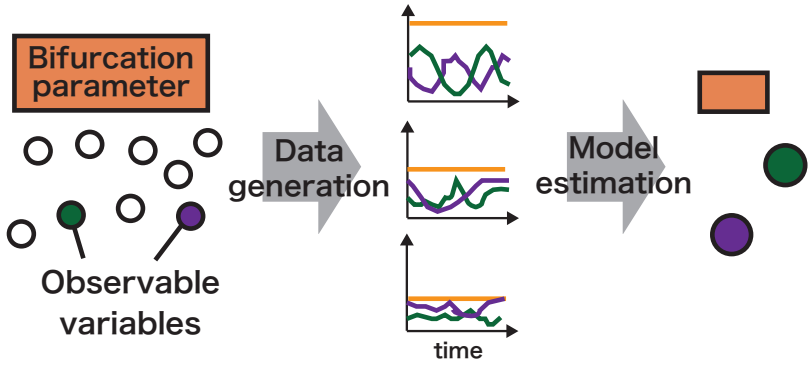


Figure 2.1: A schematic representation of the process of method evaluation. Lower-dimensional model system is trained to reproduce partially-observed dynamics generated from higher-dimensional model.

2.2 Methods

2.2.1 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 a system that is modeled by a D-dimensional stochastic differential equations, and d components in the model can be simultaneously observed. The state equations are discretized in time by the Euler-Maruyama scheme [54]. We write the time evolution of the ith variable at a time point t, $x_i^t (i = 1, \ldots, D)$, as

$$x_i^{t+1} = x_i^t + \Delta t f_i(\{x_j^t\}, s) + \sigma_i \xi_i^t \sqrt{\Delta t}, \tag{2.1}$$

where Δt is an integration time, σ_i is the intensity of the system noise, and s is a bifurcation parameter. 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 $\{f_i^n\}$.

A problem is in determining state equations, since our aim is to reproduce the bifurcation types of systems subjected to unknown equations. Furthermore, in the case of low-dimensional models as are considered below, function form required to represent the dynamics is complicated through projection of the original dynamics into the low-dimensional surface. Therefore, we adopt a polynomial base for the $\{f_i^n\}$, rather than biochemically realistic functions like the Michaelis-Menten equation. In other words, the choice of polynomials is resulting from focusing not on mechanistic modeling but on approximation of flow projected into a low-dimensional phase space.

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

$$y_i^r = g_i(x_i^r) + \eta_i \phi_i^r \tag{2.2}$$

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. Hereafter, θ indicates the parameters to be estimated: $(\{k_i^n\}, \{\sigma_i\}, \{\eta_i\})$.

2.2.2 Machine learning algorithm

The learning of dynamical systems is formulated as a maximum likelihood (ML) estimation, which is summarized below (further details are given in Appendix A). Let us consider a set of data of $Y = \{Y_a\}$ where each Y_a is a time-series data obtained from independent measurement at bifurcation parameter value s_a. 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 as summarized in Fig. 2.2 [49]. In the first step, the E 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 M step, the expectation value of $\log p(X, Y|\theta, S)$ called complete-data log likelihood is calculated as

$$Q(\theta, \theta_{\text{old}}) = \langle \log p(X, Y|\theta, S) \rangle_{p(X|Y,\theta_{\text{old}},S)}, \tag{2.3}$$

and the parameter estimation is updated as

$$\theta_{\text{new}} = \underset{\theta}{\text{argmax}}\, Q(\theta, \theta_{\text{old}}). \tag{2.4}$$

In this step, the optimization problem is reduced to linear simultaneous equations and thus can be solved easily. However, the problem in the E 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 filter algorithm that performs state estimations of nonlinear models using a Monte-Carlo method [50, 51]. The particle filter (a numerical extension of Kalman filter) 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 filter introduces stochasticity into the learning algorithm, a slight modification of the M step is required to ensure convergence of the learning [53]. The optimization function in eq.(2.4) 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 non-increasing positive numbers converging to zero.

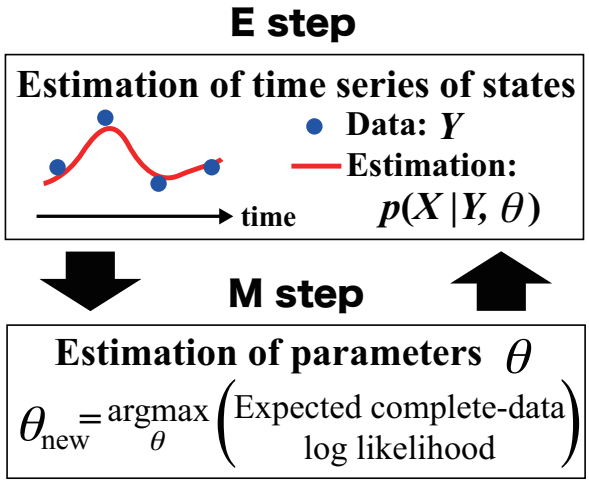


Figure 2.2: Two optimization steps called E step and M step perform estimation of system states and parameters, respectively.

2.2.3 Generation of artificial data

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 [18, 21, 55, 56, 57]. 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) [18, 55], and that proposed by Ferrell and co-workers (the Ferrell model) [56, 57]. 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 [58]. On the other hand, 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. The independent estimations for the two models eliminate the possibility that a specific type of bifurcation is always estimated regardless of the data.

Both data generators are composed of 9 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 [56]. The time-series data is generated by numerical calculations of these models as nonlinear Langevin equations at a few values of the Cyclin synthesis rate, s. 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 samples are prepared in which the oscillation exhibits large fluctuations in amplitude and period among the samples. Figure 2.4 exhibits all the time series data that are used for our estimation below.

Numerical methods of data generation

The model equations and parameter values used in the present study as the data generators are described below. Each data generators, the Tyson and Ferrell models, is numerically integrated with white Gaussian noise by using a stochastic Runge-Kutta II (SRKII) algorithm [59]. To prohibit negative values of the chemical concentrations as a result of noise, each variable is reset to a small positive value ϵ ($= 0.001$) when the value is less than ϵ. We confirmed that the results of the present study are stable so long as ϵ is sufficiently small.

Tyson model

The Tyson model described in Novak & Tyson [55].

$$\frac{d[\text{cyclin}]}{dt} = k_1[\text{AA}] - k_2[\text{cyclin}] - k_3[\text{Cdc2}][\text{cyclin}] + \xi_1(t) \tag{2.5}$$

$$\begin{aligned}\frac{d[\text{Cdc2-cyclin}]}{dt} &= k_{PP}[\text{Cdc2-cyclin-tp}] - (k_{wee} + k_{cak} + k_2)[\text{Cdc2-cyclin}] \\ &+ k_{cdc25}[\text{Cdc2-cyclin-yp}] + k_3[\text{Cdc2}][\text{cyclin}] + \xi_2(t)\end{aligned} \tag{2.6}$$

$$\begin{aligned}\frac{d[\text{Cdc2-cyclin-yp}]}{dt} &= k_{wee}[\text{Cdc2-cyclin}] - (k_{cdc25} + k_{cak} + k_2)[\text{Cdc2-cyclin-yp}] \\ &+ k_{PP}[\text{Cdc2-cyclin-yptp}] + \xi_3(t)\end{aligned} \tag{2.7}$$

$$\begin{aligned}\frac{d[\text{Cdc2-cyclin-yptp}]}{dt} &= k_{wee}[\text{Cdc2-cyclin-tp}] \\ &- (k_{PP} + k_{cdc25} + k_2)[\text{Cdc2-cyclin-yptp}] \\ &+ k_{cak}[\text{Cdc2-cyclin-yp}] + \xi_4(t)\end{aligned} \tag{2.8}$$

$$\begin{aligned}\frac{d[\text{Cdc2-cyclin-tp}]}{dt} &= k_{cak}[\text{Cdc2-cyclin}] - (k_{PP} + k_{wee} + k_2)[\text{Cdc2-cyclin-tp}] \\ &+ k_{cdc25}[\text{Cdc2-cyclin-yptp}] + \xi_5(t)\end{aligned} \tag{2.9}$$

$$\begin{aligned}\frac{d[\text{Cdc25}^*]}{dt} &= k_a[\text{Cdc2-cyclin-tp}]\frac{([\text{Cdc25}_{tot}] - [\text{Cdc25}^*])}{K_a + ([\text{Cdc25}_{tot}] - [\text{Cdc25}^*])} \\ &- k_b[\text{PPase}]\frac{[\text{Cdc25}^*]}{K_b + [\text{Cdc25}^*]} + \xi_6(t)\end{aligned} \tag{2.10}$$

$$\begin{aligned}\frac{d[\text{APC}^*]}{dt} &= k_c[\text{IEP}^*]\frac{([\text{APC}_{tot}] - [\text{APC}^*])}{K_c + ([\text{APC}_{tot}] - [\text{APC}^*])} \\ &- k_d[\text{anti-IE}]\frac{[\text{APC}^*]}{K_d + [\text{APC}^*]} + \xi_7(t)\end{aligned} \tag{2.11}$$

$$\begin{aligned}\frac{d[\text{Wee1}^*]}{dt} &= k_e[\text{Cdc2-cyclin-tp}]\frac{([\text{Wee1}_{tot}] - [\text{Wee1}^*])}{K_e + ([\text{Wee1}_{tot}] - [\text{Wee1}^*])} \\ &- k_f[\text{PPase}]\frac{[\text{Wee1}^*]}{K_f + [\text{Wee1}^*]} + \xi_8(t)\end{aligned} \tag{2.12}$$

$$\begin{aligned}\frac{d[\text{IEP}^*]}{dt} &= k_g[\text{Cdc2-cyclin-tp}]\frac{([\text{IE}_{tot}] - [\text{IEP}^*])}{K_g + ([\text{IE}_{tot}] - [\text{IEP}^*])} \\ &- k_h[\text{PPase}]\frac{[\text{IEP}^*]}{K_h + [\text{IEP}^*]} + \xi_9(t)\end{aligned} \tag{2.13}$$

$$k_{cdc25} = V_{cdc25'}([\text{Cdc25}_{tot}] - [\text{Cdc25}^*]) + V_{cdc25''}[\text{Cdc25}^*] \tag{2.14}$$

$$k_{wee} = V_{wee'}[\text{Wee1}^*] + V_{wee''}([\text{Wee1}_{tot}] - [\text{Wee1}^*]) \tag{2.15}$$

$$k_2 = V_{2'}([\text{APC}_{tot}] - [\text{APC}^*]) + V_{2''}[\text{APC}^*] \tag{2.16}$$

$$\begin{aligned}[\text{Cdc2}] &= [\text{Cdc2}_{\text{tot}}] - [\text{Cdc2-cyclin}] - [\text{Cdc2-cyclin-yp}] \\ &- [\text{Cdc2-cyclin-yptp}] - [\text{Cdc2-cyclin-tp}]\end{aligned} \tag{2.17}$$

Noise terms are represented by $\xi_i(t)$ with white Gaussian statistics $\langle \xi_i(t) \rangle = 0$ and $\langle \xi_i(t)\xi_j(\tau) \rangle = 2\sigma_i^2 \delta_{i,j}\delta(t-\tau)$.

Ferrell model

The Ferrell model is described in Tsai et al. [57].

$$\begin{aligned}
\frac{d[\text{cyclin}]}{dt} &= k_{synth} - k_{dest}[\text{APC}^*][\text{cyclin}] \\
&- k_a[\text{Cdc2}][\text{cyclin}] + k_d[\text{Cdc2-cyclin}] + \xi_1(t) \qquad (2.18)\\
\frac{d[\text{Cdc2-cyclin}]}{dt} &= k_a[\text{Cdc2}][\text{cyclin}] - k_d[\text{Cdc2-cyclin}] \\
&- k'_{wee1}[\text{Cdc2-cyclin}] + k'_{Cdc25}[\text{Cdc2-cyclin-yp}] \\
&- k_{dest}[\text{APC}^*][\text{Cdc2-cyclin}] + \xi_2(t) \qquad (2.19)\\
\frac{d[\text{Cdc2-cyclin-yp}]}{dt} &= k'_{wee1}[\text{Cdc2-cyclin}] - k'_{Cdc25}[\text{Cdc2-cyclin-yp}] \\
&- k_{cak}[\text{Cdc2-cyclin-yp}] + k_{pp2c}[\text{Cdc2-cyclin-yptp}] \\
&- k_{dest}[\text{APC}^*][\text{Cdc2-cyclin-yp}] + \xi_3(t) \qquad (2.20)\\
\frac{d[\text{Cdc2-cyclin-yptp}]}{dt} &= k_{cak}[\text{Cdc2-cyclin-yp}] - k_{pp2c}[\text{Cdc2-cyclin-yptp}] \\
&- k'_{Cdc25}[\text{Cdc2-cyclin-yptp}] + k'_{wee1}[\text{Cdc2-cyclin-tp}] \\
&- k_{dest}[\text{APC}^*][\text{Cdc2-cyclin-yptp}] + \xi_4(t) \qquad (2.21)\\
\frac{d[\text{Cdc2-cyclin-tp}]}{dt} &= k'_{Cdc25}[\text{Cdc2-cyclin-yptp}] - k'_{wee1}[\text{Cdc2-cyclin-tp}] \\
&- k_{dest}[\text{APC}^*][\text{Cdc2-cyclin-tp}] + \xi_5(t) \qquad (2.22)
\end{aligned}$$

$$\begin{aligned}
\frac{d[\text{Cdc25}^*]}{dt} &= k_{Cdc25on}\frac{[\text{Cdc2-cyclin-tp}]^{n_{cdc25}}}{\text{EC50}_{Cdc25}^{n_{cdc25}} + [\text{Cdc2-cyclin-tp}]^{n_{cdc25}}}([\text{Cdc25}_{tot}] - [\text{Cdc25}^*]) \\
&- k_{Cdc25off}[\text{Cdc25}^*] + \xi_6(t) \qquad (2.23)\\
\frac{d[\text{Wee1}^*]}{dt} &= -k_{Wee1off}\frac{[\text{Cdc2-cyclin-tp}]^{n_{wee1}}}{\text{EC50}_{wee1}^{n_{wee1}} + [\text{Cdc2-cyclin-tp}]^{n_{wee1}}}[\text{Wee1}^*] \\
&+ k_{Wee1on}([\text{Wee1}_{tot}] - [\text{Wee1}^*]) + \xi_7(t) \qquad (2.24)\\
\frac{d[\text{Plx}^*]}{dt} &= k_{plxon}\frac{[\text{Cdc2-cyclin-tp}]^{n_{plx}}}{\text{EC50}_{plx}^{n_{plx}} + [\text{Cdc2-cyclin-tp}]^{n_{plx}}}([\text{Plx}_{tot}] - [\text{Plx}^*]) \\
&- k_{plxoff}[\text{Plx}^*] + \xi_8(t) \qquad (2.25)\\
\frac{d[\text{APC}^*]}{dt} &= k_{apcon}\frac{[\text{Plx}^*]^{n_{apc}}}{\text{EC50}_{apc}^{n_{apc}} + [\text{Plx}^*]^{n_{apc}}}([\text{APC}_{tot}] - [\text{APC}^*]) \\
&- k_{apcoff}[\text{APC}^*] + \xi_9(t) \qquad (2.26)
\end{aligned}$$

$$\begin{aligned}
k'_{Cdc25} &= k_{Cdc25}[\text{Cdc25}^*] + k_{Cdc25basal}([\text{Cdc25}_{tot}] - [\text{Cdc25}^*]) \qquad (2.27)\\
k'_{wee1} &= k_{wee1}[\text{Wee1}^*] + k_{wee1basal}([\text{Wee1}_{tot}] - [\text{Wee1}^*]) \qquad (2.28)\\
[\text{Cdc2}] &= [\text{Cdc2}_{tot}] - [\text{Cdc2-cyclin}] - [\text{Cdc2-cyclin-yp}] \\
&- [\text{Cdc2-cyclin-yptp}] - [\text{Cdc2-cyclin-tp}] \qquad (2.29)
\end{aligned}$$

Noise terms are represented by $\xi_i(t)$ with white Gaussian statistics $\langle \xi_i(t) \rangle = 0$ and $\langle \xi_i(t)\xi_j(\tau) \rangle = 2\sigma_i^2 \delta_{i,j}\delta(t-\tau)$.

Parameter values

The parameters used in the Tyson and Ferrell models are listed in Tables 2.1 and 2.2, respectively. In the Tyson model, all the parameter values are the same as those in the original literature [55]. In the Ferrell model, all the parameter values are the same as those in Pomerening et al. [56] except for "$factor$," which is set as in [57].

Parameter	
k_1[AA]/[Cdc2$_{tot}$]	s
k_3[Cdc2$_{tot}$]	1.0
k_{CAK}	0.25
k_{PP}	0.025
$V_{2'}$[APC$_{tot}$]	0.015
$V_{2''}$[APC$_{tot}$]	1.0
$V_{cdc25'}$[Cdc25$_{tot}$]	0.1
$V_{cdc25''}$[Cdc25$_{tot}$]	2.0
$V_{wee'}$[Wee1$_{tot}$]	0.1
$V_{wee''}$[Wee1$_{tot}$]	1.0
k_a[Cdc2$_{tot}$]/[Cdc25$_{tot}$]	1.0
k_b[PPase]/[Cdc25$_{tot}$]	0.125
k_c[IE$_{tot}$]/[APC$_{tot}$]	0.1
k_d[anti IE]/[APC$_{tot}$]	0.095
k_e[Cdc2$_{tot}$]/[Wee1$_{tot}$]	1.33
k_f[PPase]/[Wee1$_{tot}$]	0.1
k_g[Cdc2$_{tot}$]/[IE$_{tot}$]	0.65
k_h[PPase]/[IE$_{tot}$]	0.087
K_a/[Cdc25$_{tot}$]	0.1
K_b/[Cdc25$_{tot}$]	0.1
K_c/[APC$_{tot}$]	0.01
K_d/[APC$_{tot}$]	0.01
K_e/[Wee1$_{tot}$]	0.3
K_f/[Wee1$_{tot}$]	0.3
K_g/[IE$_{tot}$]	0.01
K_h/[IE$_{tot}$]	0.01
$\sigma_{1,\ldots,5}$/[Cdc2$_{tot}$]	0.0024
σ_6/[Cdc25$_{tot}$]	0.0024
σ_7/[APC$_{tot}$]	0.0024
σ_8/[Wee$_{tot}$]	0.0024
σ_9/[IEP$_{tot}$]	0.0024
η_1/[Cdc2$_{tot}$]	0.005
η_2/[Cdc2$_{tot}$]	0.005

Table 2.1: Parameters in the Tyson model

Parameter	
k_{syn}	s
k_{dest}	0.006
k_a	0.1
k_d	0.001
$factor$	10
k_{wee1}	0.05
$k_{wee1basal}$	$k_{wee1}/factor$
k_{Cdc25}	0.1
$k_{Cdc25basal}$	$k_{cdc25}/factor$
k_{cak}	0.8
k_{pp2c}	0.008
$k_{Cdc25on}$	1.75
$k_{Cdc25off}$	0.2
k_{wee1on}	0.2
$k_{wee1off}$	1.75
k_{plxon}	1.0
k_{plxoff}	0.15
k_{apcon}	1.0
k_{apcoff}	0.15
EC50$_{Cdc25}$	40
EC50$_{wee1}$	40
EC50$_{plx}$	40
EC50$_{apc}$	40
n_{cdc25}	4
n_{wee1}	4
n_{plx}	3
n_{apc}	3
Cdc25$_{tot}$	15
Wee1$_{tot}$	15
Plx$_{tot}$	50
APC$_{tot}$	50
Cdc2$_{tot}$	230
$\sigma_{1,\ldots,9}$	0.06
η_1	0.4
η_2	0.4

Table 2.2: Parameters in the Ferrell model

Artificial measurement process

For both the Tyson and Ferrell models, the artificial measurement process is implemented as follows.

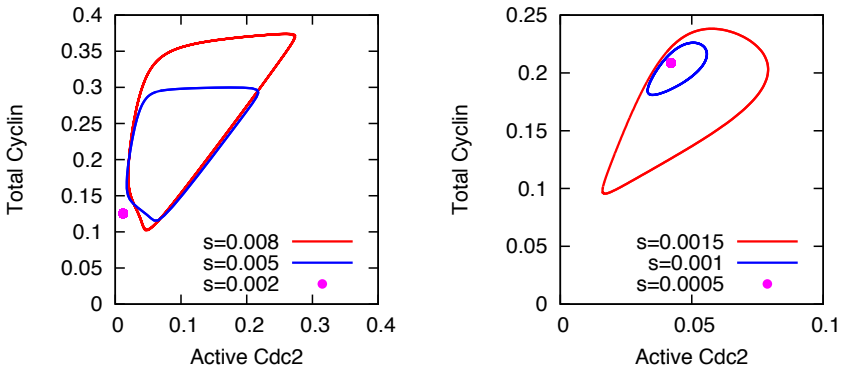
$$y_1 = \text{active Cdc2} = \frac{[\text{Cdc2-cyclin-tp}] + \eta_1\phi_1}{[\text{Cdc2}_{\text{tot}}]} \tag{2.30}$$

$$\begin{aligned} y_2 &= \text{Total Cyclin} \\ &= \big([\text{cyclin}] + [\text{Cdc2-cyclin}] + [\text{Cdc2-cyclin-yp}] + [\text{Cdc2-cyclin-yptp}] \\ &\quad + [\text{Cdc2-cyclin-tp}] + \eta_2\phi_2\big)/[\text{Cdc2}_{\text{tot}}]. \end{aligned} \tag{2.31}$$

Here, $\eta_{1,2}$ is the observation noise intensity, and $\phi_{1,2}$ is sampled from a standard normal distribution.

Orbits of the data generators

Figure 2.3 show noiseless orbits of the data generators. We note that each orbit exhibits no intersection, indicating that the two observable variables are sufficient to abstract the original high-dimensional dynamics.

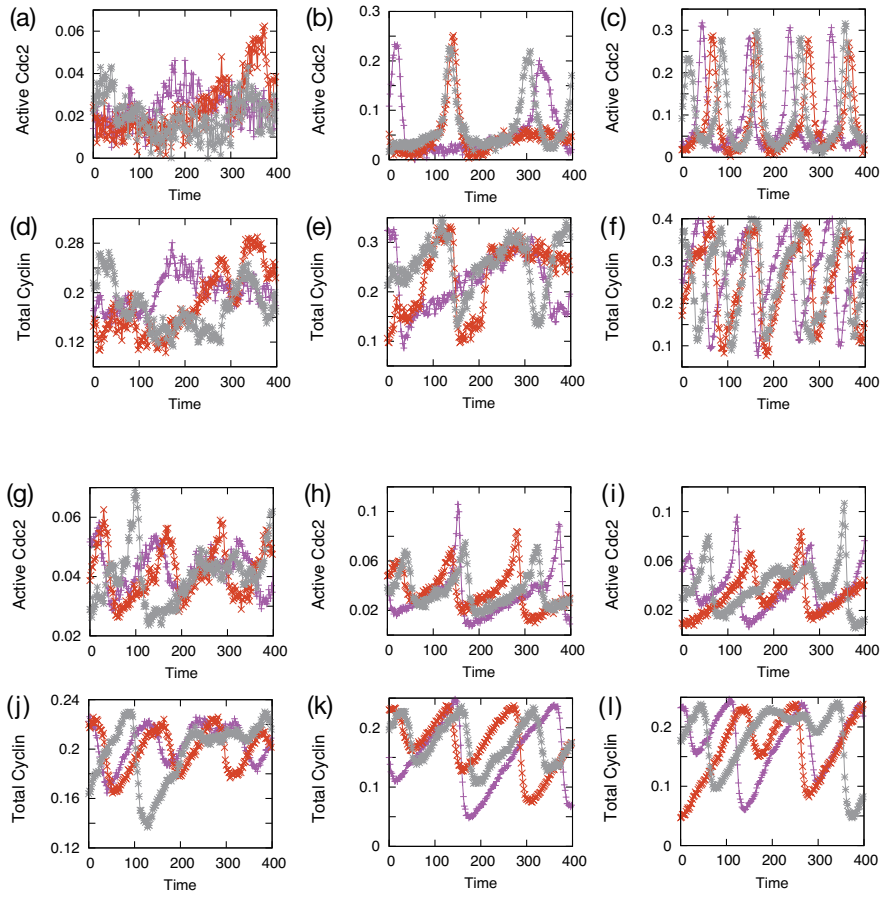


(a) (b)

Figure 2.3: Time series generated from the (a) Tyson and (b) Ferrell models through the artificial measurement process in the case that both system and observation noises are zero.

Data set

The all data used for the learning are shown in Fig. 2.4.



(a)

Figure 2.4: Noisy time series generated from the (a-f) Tyson and (g-l) Ferrell models through the artificial measurement process. The colors indicate results from independent trials. The values of the bifurcation parameter are (a,d) $s = 0.002$, (b,e) 0.005, (c,f) 0.008, (g,j) 0.005, (h,k) 0.001, and (i,l) 0.0015, respectively. Note that the parameter s is defined in Tables 2.1 and 2.2. (Reprinted from Ref. [40], © 2013 American Physical Society)

2.2.4 Setting state and observation equations

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

$$f_i(\{x_j^t\}, s) = k_i^1 + k_i^2 x_1^t + k_i^3 x_2^t + \cdots + k_i^{N_i} (x_D^t)^M \tag{2.32}$$

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., $k_2^1 = s$. Note that the observed orbit in the active Cdc2-Cyclin plane exhibits no intersection, as shown in Fig. 2.3, suggesting that the two variables are sufficient to abstract the original high-dimensional dynamics.

2.3 Results

2.3.1 Model selection based on Information criteria

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 M by one. 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 [60, 61]. 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.2.5). Both the AIC and BIC show a decrease from $M = 1$ to 3, but an increase or insignificant decrease at $M = 4$. Therefore, we analyze models with $D = 2$ and $M = 3$. The learned parameter values are presented in Table 2.3 (see Section A.4 for detailed settings in the learning algorithm).

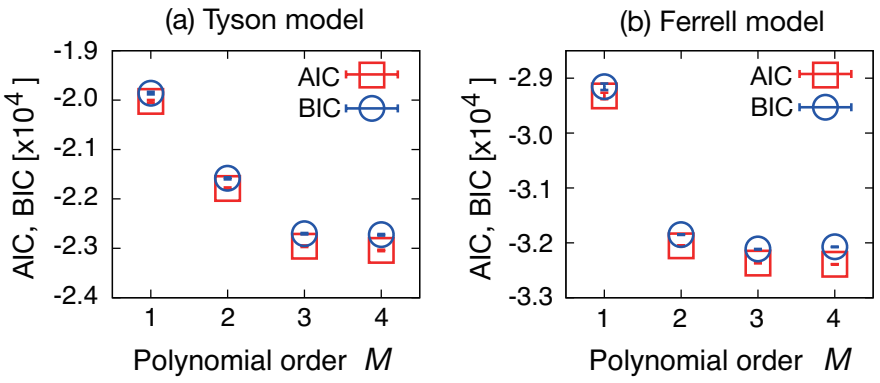


Figure 2.5: The AIC and BIC are plotted against different polynomial orders $M = 1, 2, 3, 4$ for the (a) Tyson and (b) Ferrell models. For each evaluation, 20 different initial parameters are sampled to avoid the local minima. After the learning, the average and standard deviation are calculated from 100 evaluation trials based on the particle filter. (Reprinted from Ref. [40], © 2013 American Physical Society)

Parameter	*Learned (Tyson model)*	*Learned (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_2^9	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

Table 2.3: Learned parameters of the third-order polynomial systems for the Tyson and Ferrell models.

2.3.2 Estimation of bifurcation types

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 2.6(a) and (b) show bifurcation diagrams against Cyclin synthesis rate s (red lines) for the learned systems 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 broken lines), the learned systems have quantitatively similar diagrams 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. 2.6(c) and (d) for the Tyson model and in Figs. 2.6(e) and (f) for the Ferrell model. The results confirm the onset of SNIC and supercritical Hopf bifurcation, 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. 2.6(g) and (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.

2.3.3 Comparison with adiabatic elimination

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 [17]. Like the learned systems, the reduced models are composed of active Cdc2 and total Cyclin.

We reduce the Tyson and Ferrell models to two-dimensional systems by the same procedure as described in [17]. By denoting the non-dimensionalized active Cdc2 and total Cyclin levels as u and v, respectively, the reduced models are written as follows.

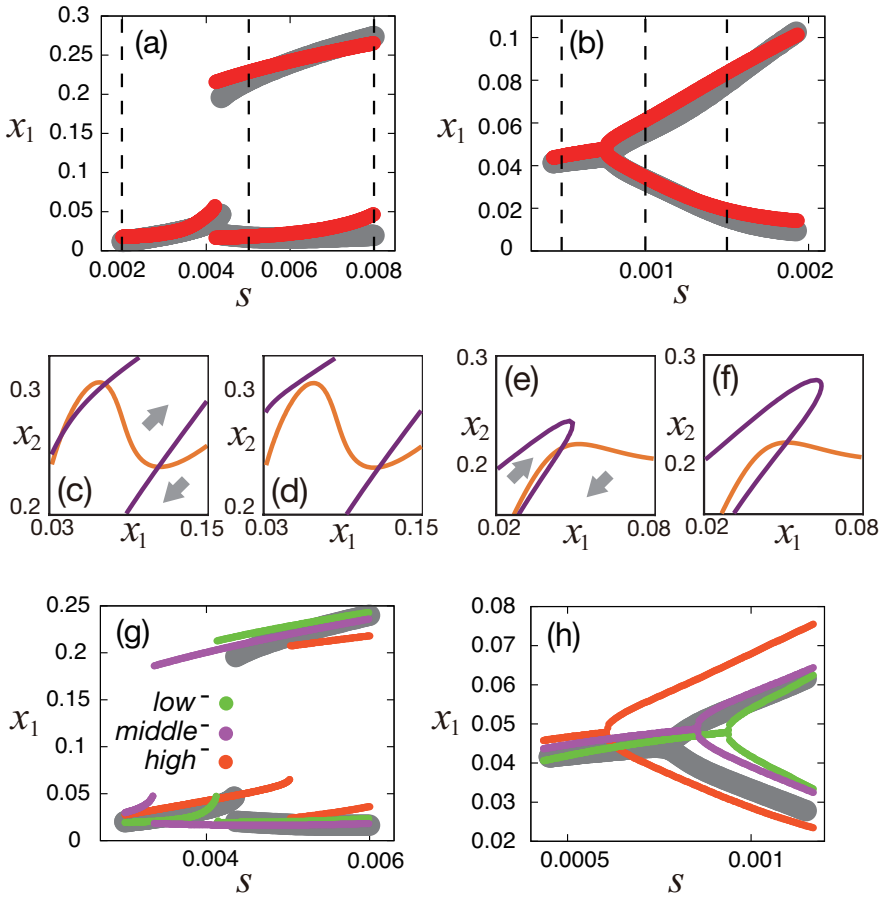


Figure 2.6: Bifurcation diagrams of the (a) Tyson and (b) Ferrell models. The active Cdc2 concentration x_1 for the learned systems is plotted against the Cyclin synthesis rate s (red), and corresponding concentrations of the data generators are also shown for comparison (gray). The broken lines indicate 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 direction. (c,d) The learned system from the Tyson model exhibits SNIC and (e,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,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 as low^-, $middle^-$ and $high^-$, respectively. (Reprinted from Ref. [40],)

The reduced Tyson model takes the form as

$$\begin{aligned} \dot{u} &= \frac{s}{1+k_{PP}/k_{cak}} - (f_{\mathrm{APC}}(u)+f_{\mathrm{Wee}}(u))u \\ &+ f_{\mathrm{Cdc25}}(u)(\frac{v}{1+k_{PP}/k_{cak}} - u), \qquad (2.33a) \\ \dot{v} &= s - f_{\mathrm{APC}}(u)v, \qquad (2.33b) \end{aligned}$$

where $f_{\mathrm{Cdc25}}, f_{\mathrm{Wee}}$, and f_{APC} are the functions corresponding to the adiabatic solutions of eqs. (2.14), (2.15), and (2.16), respectively.

This reduction procedure includes the determination of the level of Cdc2-Cyclin-tp (i.e., the value of u) from the sum [Cdc2-Cyclin]+[Cdc2-Cyclin-tp] (see Appendix A in [17]). This is based on a detailed balance assumption of the phosphorylation reaction between the two molecular species. However, in the Ferrell model, the absence of the reaction makes the original reduction procedure inapplicable. Then, we simply assume [Cdc2-Cyclin-tp] $\sim$ [Cdc2-Cyclin] + [Cdc2-Cyclin-tp], because it is observed that the ratio of [Cdc2-Cyclin] to the summation is small throughout the dynamics within the parameter region we consider. Consequently, the reduced Ferrell model is written as

$$\begin{aligned} \dot{u} &= s - (f_{\mathrm{APC}}(u)) + f_{\mathrm{Wee}}(u))u + f_{\mathrm{Cdc25}}(v-u), \qquad (2.34a) \\ \dot{v} &= s - f_{\mathrm{APC}}(u)v, \qquad (2.34b) \end{aligned}$$

where $f_{\mathrm{Cdc25}}, f_{\mathrm{Wee}}$, and f_{APC} are derived from the adiabatic approximation in the same manner as the Tyson model.

Figure 2.7 shows the nullclines of the learned systems (the solid orange and purple lines) and the reduced models (the broken 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. 2.7 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. 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.

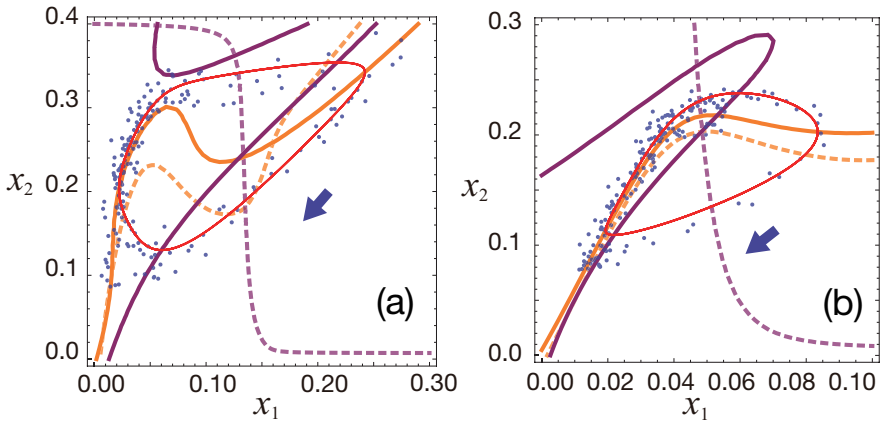


Figure 2.7: Comparison of the learned systems and reduced models for the (a) Tyson and (b) Ferrell models. The purple and orange lines represent the nullclines of x_1 (active Cdc2) and x_2 (Cyclin), respectively, for the learned systems (solid lines) and reduced models (broken lines). A noisy time series 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) $s = 0.006$, (b) 0.0015. (Reprinted from Ref. [40], © 2013 American Physical Society)

2.3.4 Examples of estimation failure

We also investigated how the estimation fails, in order to get insight into what is required to perform a correct estimation. We briefly demonstrate examples of the failure in the case of insufficiency in data or model complexity. First, by using third order polynomials as state equations, we carry out the model estimation from data at only one value of the bifurcation parameter. Figure 2.8 shows the estimated bifurcation diagrams from data after the bifurcation, that is, data at the middle or the highest bifurcation parameter value. Note that the result from the lowest bifurcation parameter value is not shown, because the time-series data at the fixed point could not constrain the coefficients of the third order polynomials. The obtained bifurcation diagrams deviate from that of the data generators. In particular, the bifurcation type is incorrectly predicted from the data at the highest bifurcation parameter value of the Tyson model, while the other cases in Fig. 2.8 preserve the types. In the incorrect case, the limit cycle collides with a saddle point and disappears through the homoclinic bifurcation. Such mispredictions occur more frequently as sampled bifurcation parameter values are farther from a bifurcation point.

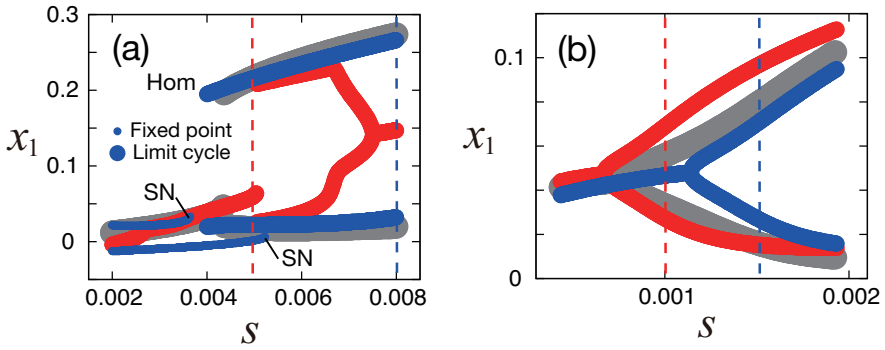


Figure 2.8: Estimation failure caused by insufficient data. The bifurcation diagrams are estimated from data at only one value of the bifurcation parameter. The red and blue broken lines represent points in which the data is given for constructing each bifurcation diagram. The SN and Hom represent saddle-node and homoclinic bifurcations, respectively.

Next, we examine polynomials of degree 1 and 2 as oversimple models in comparison with degree 3 investigated above. Figure 2.9 shows the estimated bifurcation diagrams and the flow from the full data set shown in Fig. 2.4. For the polynomial models of degree 1, any bifurcation is not predicted simply because linear dynamical systems cannot have a limit cycle. However, the estimated flow structure is reasonable, that is, the positions of the stable fixed points and the flow directions around the points reflect the projected dynamics of the data generators.

On the other hand, for the degree 2, while the Hopf bifurcation is successfully predicted for the Ferrell model, the learned system still does not show any bifurcation for the Tyson model. The correct bifurcation type for the Tyson model, SNIC, seems to be basically possible in polynomials of degree 2, although we do not have mathematical proof for this. However, the nullcline of x_1 has a clear N-shaped form in the learned polynomials of degree 3 (see Fig. 2.7(a)), and this fact indicates that the flow of x_1 for a fixed value of x_2 can have three steady points that are impossible for degree less than 3. The N-shaped nullcline is required to reproduce the upper-right-pointing flow in the upper portion of the limit cycle. We infer that, for this reason, the model of degree 2 cannot reproduce limit cycle by learning and, as a consequence, fails to predict the bifurcation type.

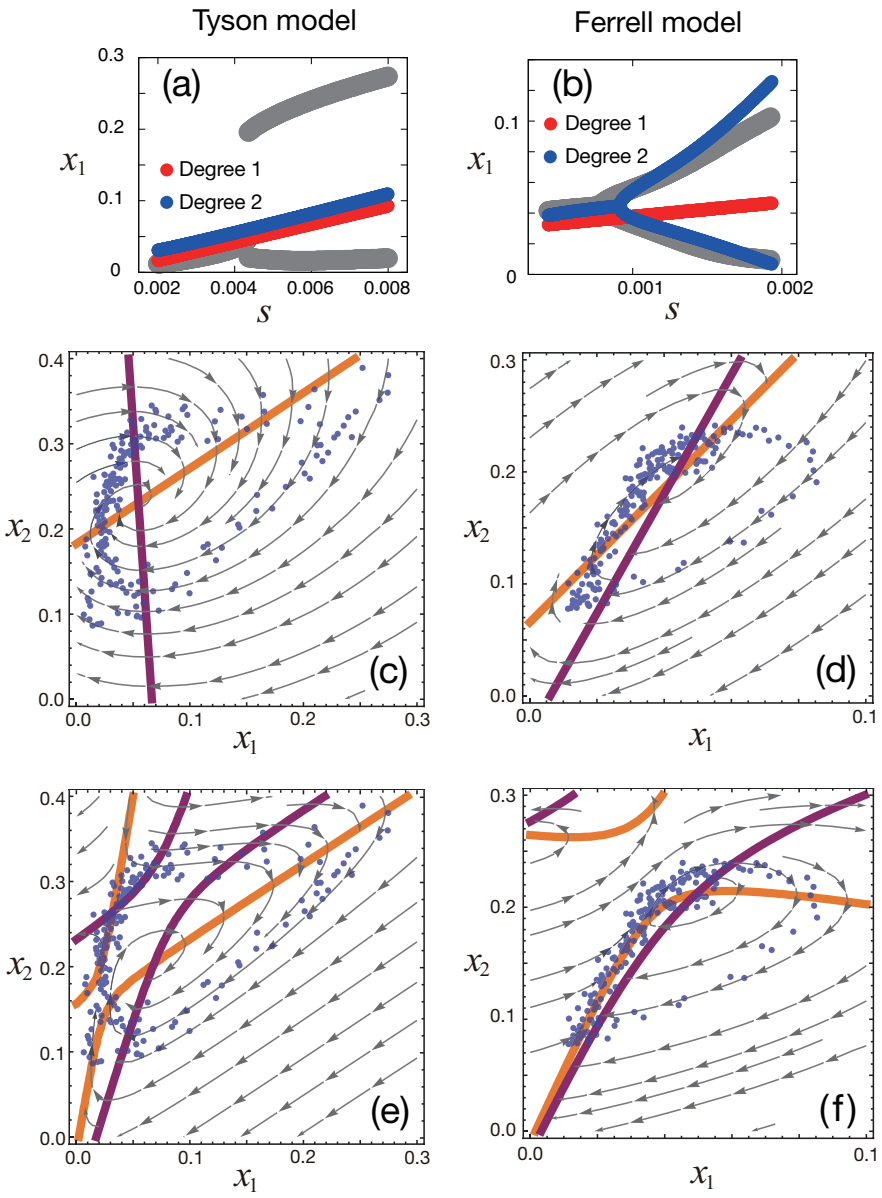


Figure 2.9: Estimation failures caused by insufficient degree of the polynomial dynamical systems. (a,b) The estimated and original bifurcation diagrams are presented for the (a) Tyson and (b) Ferrell models. The red and blue diagrams come from the estimated polynomial models of degree 1 and 2, respectively, while the gray diagrams come from the data generators. (c,d,e,f) The nullclines and the flows of the polynomial of (c,d) degree 1 and (e,f) 2 are presented for the (c,e) Tyson and (d,f) Ferrell models. The purple and orange lines show the nullclines of x_1 (active Cdc2) and x_2 (Total Cyclin), respectively. The gray arrows represent flows in the phase spaces, and the blue dots represent a noisy time-series sample from eahc of the data generators.

2.3.5 Phase sensitivity analysis

We have shown that the polynomial dynamical systems reproduced the bifurcation types of the data generators. This indicates that the learned systems captured topological features of the flow in the phase spaces provided by the data generators. Now, we perform a more quantitative comparison between the learned systems and the data generators from a view point of phase sensitivity. First the concept of phase sensitivity is briefly introduced [20,62,63].

Let us assume that an attractor of a given system is a limit cycle, which gives an oscillator. By choosing a reference point on the limit cycle, any other points can be specified by elapsed time θ from the last passing of the reference. The time θ is called phase of the oscillator. We also assume that a single pulse input is applied to the oscillator, and the input instantaneously changes a value of a state variable. The phase of the perturbed oscillator is delayed or advanced, and the deviation $\Delta\theta$ depends on the phase at which the input is applied. Therefore, we can define a phase response curve (PRC) to a pulse input, as

$$\mathrm{PRC}(\theta) = \Delta\theta \tag{2.35}$$

where θ represents the timing of the input. Furthermore, under sufficiently small inputs, we can decompose the PRC as

$$\mathrm{PRC}(\theta, A) \sim Z(\theta)A \tag{2.36}$$

where A is the height of the pulse. The $Z(\theta)$ is called infinitesimal PRC or phase sensitivity. The concept of phase sensitivity characterizes synchronization behavior of an oscillator to small inputs, and is used in studies on biological rhythms [64,65,66]. Through these studies, techniques to estimate a phase sensitivity from noisy data are sophisticated [67]. Here we note that, in obtaining only a phase sensitivity, there is no reason to use our approach, and the previous techniques can work. We adopt the phase sensitivity here to illustrate quantitative differences between the learned systems and the data generators.

By using the polynomial of degree 3 as the state equations, we calculated the phase sensitivities of the learned systems and the data generators, and found a huge difference between them. To examine the error, we performed 20 trials of estimation by using independently and identically distributed data sets in which each set contains the same amount of data as in the time-series samples in Fig 2.4. Figures 2.10 (a,b,c,d) show the phase sensitivities of the learned systems and the data generators. The phase sensitivities of the polynomial dynamical systems of degree 3 not only show large fluctuations but also are biased from the sensitivities of the Tyson and Ferrell models. On the other hand, the time courses are similar to each other as shown in Figs. 2.10 (e,f,g,h), although the periods of oscillation show significant fluctuations.

These results indicate that the learned systems capture the shape of the limit cycle, but not the flows in the surrounding area of the limit cycle. Since the phase sensitivities from the independent data sets are clearly biased, further increase in the data number might not solve the problem. Instead, we infer that the low dimensionality of the model or the employment of polynomials may cause the mis-prediction. These points have to be elucidated in future.

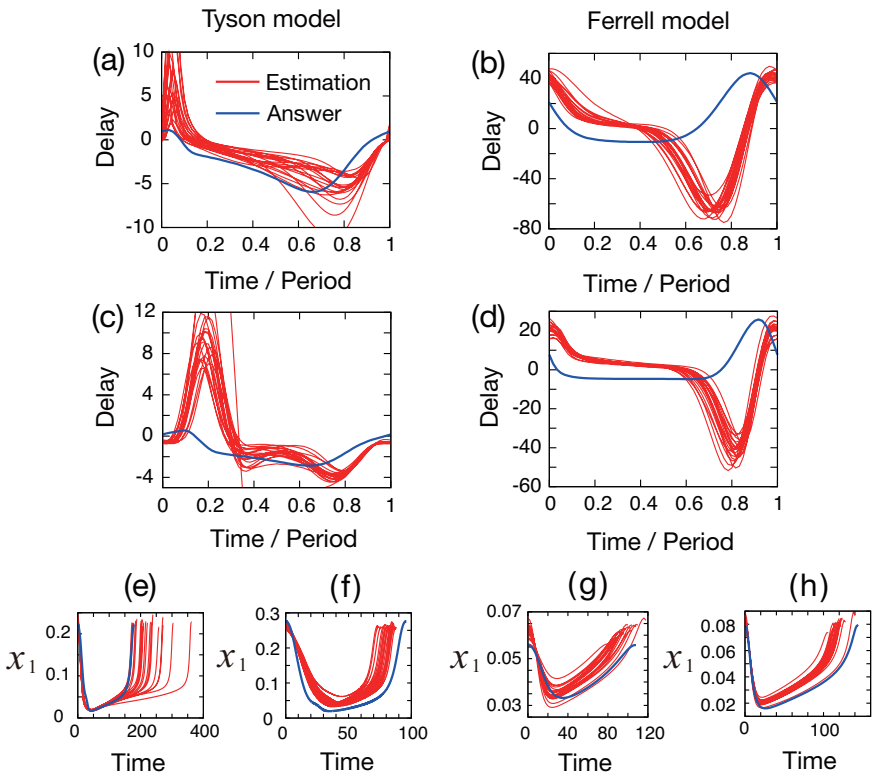


Figure 2.10: Phase sensitivity analysis for learned systems. (a,b,c,d) Sensitivity functions are obtained from 20 learned systems by independent data sets (red) and the data generators (blue), in which pulse inputs are applied to Active Cdc2, x_1. The zero point on the x-axis represents the time point at which x_1 takes the maximum value. The delays, the values on y-axis, are scaled by the oscillation period T and strength of the pulse inputs A, as Delay/$(T \times A)$. The bifurcation parameter values are (a) $s = 0.005$, (b) 0.008, (c) 0.001, and (d) 0.0015, respectively. (e,f,g,h) Noiseless time courses of Active Cdc2 (x_1) over a period are presented for the learned systems (red) and the data generators (blue). The bifurcation parameter values in (e,f,g,h) are the same as in (a,b,c,d), respectively.

2.4 Discussion

Biological data are generally imperfect, due to intrinsic and observation noises, partial observation, and a small number of possible experimental conditions. We can employed 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. In particular, we focused on deriving low-dimensional models with correct bifurcation types, because such models achieve a balance between biological relevance and model simplicity while did not require detailed biochemical knowledge for constructing them. By using synthetic data that share the difficulties found in actual biological data, we demonstrated that the procedure could construct low-dimensional model equations that reproduced the noisy time-series data from partial observations 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. The method is expected to be applicable to analyze data on intracellular processes obtained by imaging techniques for promoting systems biological understandings.

Unfortunately, employment of polynomial as state equations does not always works well. A major risk comes from the fact that polynomials diverge on an unbounded domain. The property impairs extrapolation to a region where data is not given. The problem can be serious, for example, when we try to predict system behavior under a strong perturbation. Bounded functions such as trigonometric functions may help reduce the problem.

We note that the proposed procedure can be interpreted as a reduction method from high- to low-dimensional systems like adiabatic approximation. In particular, in the vicinity of the bifurcation points, the systems are usually reduced to normal forms represented by low-dimensional differential equations with low-order polynomial forms [41]. However, unlike analytical reduction methods that require the original high-dimensional equations, the present learning procedure uses only the time-series data. This is especially advantageous for studying cell dynamics that involve complex molecular interactions. On the other hand, since the learning method has a less theoretical basis for interpreting the obtained equations, it should be complemented by some analytical procedure.

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 [45]. Incorporation of prior distributions for model parameters achieves such a sparse optimization in statistical framework [68], and it will enable us to extract an appropriate sub model from an original complex equation by eliminating unnecessary terms. It is advantageous in the case that complex time-series data is 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.

Chapter 3

Identifying dynamical systems by sparse regularization

3.1 Introduction

The diversity in biochemical interactions between molecules such as protein is one of the major obstacles to derive a reliable model for cell signaling dynamics. Even when all of related cell signaling components in a biological function are known and qualitative interactions between them are clarified, the complexity in the system makes it difficult to determine model equations exactly. For example, negative regulation on x by y can be modeled by, $\dot{x} = -y$, $-xy^2$, etc. Besides, as shown dramatically in recent studies [69,70], molecular crowding in cytoplasm affects not only biochemical parameter values but also function forms of reaction kinetics. This suggests that cell signaling dynamics is hard to be estimated only from *in vitro* experiments which are more controllable than *in vivo* ones. Then, determining model equations is a complicated problem, although the approach is essential to understand cell behaviors.

Many previous studies have defined simplified models as core architectures of systems under investigation and tried to interpret observed cell dynamics in the simplified model [16, 17]. To derive the core architectures, we omit biochemical processes having relatively small effects on the observed behaviors, and replace unknown molecular interactions by phenomenological ones. We can often capture qualitative features of the dynamics by such a simplified picture. However, since the qualitative features still depend on parameter values and function forms in the model equations, appropriate simplifications also require significant information on the underlying biochemistry.

The above approaches focus on explaining system-level dynamics based

on known elementary biochemical processes. Here a complementary approach is provided by a machine-learning technique called sparse regularization that eliminates unnecessary elements of a model while trains the model to reproduce data. That is, the sparse optimization enables us to start with a complex model equation that contains many possible interactions, and to extract an appropriate sub model from the original equations required for explaining observed dynamics.

One common method to perform sparse regularization is to maximize penalized log likelihood with $L1$ norm of a parameter vector θ, as

$$\log L(\theta) - \alpha \sum_i |\theta_i| \tag{3.1}$$

where α means strength of regularization. The penalty term, called Lasso, have a finite differential value by each parameter θ_i in limit of $\theta_i \to 0$, and then the method sets values of unnecessary parameters to zero, unlike the case of $L2$ norm which is also widely used to prevent over-fitting problem.

Although there are several previous studies on learning sparse dynamical systems from time-series data [45], applications for nonlinear dynamical phenomena are limited yet. Then, we examine the sparse estimation of dynamical systems that exhibit chaos and adaptation with input-scale invariance, by using artificial data. To this end, we employ the $L1$ norm as prior distributions, called Bayesian lasso [68], and extend the algorithm in Chapter 2.

3.2 Method: MAP estimation by Bayesian lasso

In this section, we combine the method in Appendix A with Bayesian lasso to perform sparse regularization. Our aim is to extract a simple dynamical system from a more complex equation. For the purpose, the regularization is confined to parameters for deterministic dynamics while intensities of noises are not regularized.

A state space model to be estimated is defined as

$$x_i^{t+1} = x_i^t + \Delta t f_i(\{x_j^t\}, s) + \sigma_i \xi_i^t \sqrt{\Delta t} \tag{3.2}$$

$$y_i^r = g_i(\{x_j^r\}) + \eta_i \phi_i^r \tag{3.3}$$

where x_i $(i = 0, \ldots, D-1)$ and y_i $(i = 0, \ldots, d-1)$ are a state variable and an observed variable of ith component, respectively. The function f_i is assumed to be expressed by the summation of linearly independent functions as

$$f_i(\{x_j^t\}, s) = \sum_{n=1}^{N_i} k_i^n f_i^n(\{x_j^t\}, s) \tag{3.4}$$

where $\{k_i^n\}$ are the coefficients to be estimated. Detailed definitions of the model are in the first section of Appendix A.

The extended algorithm performs maximum a posteriori (MAP) estimation that maximizes

$$L(\theta) \times \prod_i p(\{k_i^n\}|\sigma_i^2)p(\sigma_i^2). \tag{3.5}$$

where θ is the model parameters. Here, the prior distribution for the kinetic parameters $\{k_i^n\}$ is an exponential distribution as

$$p(\{k_i^n\}|\sigma_i^2) \quad \propto \quad (\sigma_i^2)^{-N_i} \exp(-\frac{\alpha}{\sigma_i^2} \sum_n^{N_i} |k_i^n| \langle f_i^n(\{y_i^r\})\rangle) \tag{3.6}$$

where the average $\langle f_i^n(\{y_i^r\})\rangle$ is taken over the all of data points. Note that the factor $\langle f_i^n(\{y_i^r\})\rangle$ is required to equalize the dimensions of the terms in the summation $\sum_n^{N_i}$. The parameter α controls sparseness and can be determined by a cross-validation technique. On the other hand, the prior distribution for the intensity of system noise σ_i^2 is an inverse-Gamma distribution as

$$p(\sigma_i^2) = p(\sigma_i^2|a_i^\gamma, b_i^\gamma) \propto (\sigma_i^2)^{-a_i^\gamma - 1} \exp(-\frac{b_i^\gamma}{\sigma_i^2}) \tag{3.7}$$

with shape parameter a_i^γ and scale parameter b_i^γ.

The prior distributions, eqs. (3.6) and (3.7), change the M step in the EM algorithm as follows. The update of the coefficients in eq. (A.18) for each lth component is altered as

$$\min_{k_l} \frac{1}{2} k_l \cdot A_l k_l - b_l \cdot k_l + \frac{\alpha}{\Delta t} ||k_l||_{L1} \tag{3.8}$$

where k_l is the coefficient vector, and the definitions of the matrix A_l and the vector b_l are the same as in Appendix A. By introducing an augmented coefficient vector Θ as

$$k_l = k_l^+ - k_l^-, \text{ where } \Theta = \begin{bmatrix} k_l^+ \\ k_l^- \end{bmatrix}, \tag{3.9}$$

the minimization problem is converted as

$$\min_{\Theta} \frac{1}{2} \Theta \cdot \begin{bmatrix} A_l & -A_l \\ -A_l & A_l \end{bmatrix} \Theta + \Theta \cdot \left[\frac{\alpha}{\Delta t} \mathbf{1} - \begin{pmatrix} b_l \\ -b_l \end{pmatrix} \right]$$

$$\text{such that } \Theta_i \geq 0 \text{ for all } i \tag{3.10}$$

where $\mathbf{1}$ represent a vector filled with ones. Since the problem takes the form of quadratic programming, the convexity is assured and the global optimal

solution is easily obtained. We calculate the solution of Θ by an active-set algorithm (in particular, we utilized Matlab's quadprog function). Next, the intensity of system noise is updated as

$$(\sigma_l)^2 = \frac{A(|T|-1)(\sigma_l^{\mathrm{ML}})^2 + 2\alpha \sum_n^{N_i} |k_i^n| \langle f_i^n(\{y_i^r\})\rangle + 2b_\gamma}{A(|T|-1) + 2N_l + 2 + 2a_\gamma} \tag{3.11}$$

where σ_l^{ML} is the maximum likelihood estimator obtained through eq. (A.20). The update of the other parameters and the E step remain unchanged.

3.3 Identification of a chaotic system

Chaotic dynamics means non-periodic and bounded behaviors, even though there are some different definitions [71]. Early studies of chaos date back to the work on celestial mechanics at nineteenth century by a French mathematician, Henri Poincaré, and his work suggests a possibility of unexpected complex dynamics by simple systems. Along with the development of computer technology, such complex dynamics becomes visible by numerical calculation of models, and theoretical studies lead to the fundamental concepts like "sensitivity to initial conditions" that makes long-term prediction impossible [72, 73]. Besides, a variety of experiments has demonstrated the existence of chaotic dynamics in the real world. A well-known example in chemical reactions is the Belousov-Zhabotinskii reaction [74, 75, 76].

Although chaotic dynamics can emerges in simple systems, observed chaotic time series often looks complicated to be analyzed. Then, it is worthwhile to develop the machine-learning method to analyze such chaotic dynamics, and there are many studies along this direction. However, estimation of noisy chaotic systems from noisy observation still remains a challenging problem. Here, we consider this problem as a test case to evaluate the learning dynamical systems with Bayesian lasso.

Clearly, estimation of chaotic systems is quite difficult even when noise does not exist, because parametrized model equations like polynomials cannot reproduce detailed structures of observed orbits in general. Then, as the first step, we employ the Rössler model [73], one of the simplest polynomial models exhibiting chaos, as a data generator for examining our method. As a model equation, we use a set of differential equations with polynomial terms of variables that contains the Rössler model as its subset. In this setting, the Rössler model is exactly identified in the best case.

3.3.1 Data generator: Rössler model

The Rössler model, proposed by O. E. Rössler, represents dynamics of an abstract chemical system. The simple model is known to exhibit chaotic

dynamics, and there are previous studies on reconstruction of the model [45, 77]. We generate chaotic time series from the Rössler model with noise,

$$\dot{x_0} = -x_1 - x_2 + \xi_0(t) \tag{3.12a}$$
$$\dot{x_1} = x_0 + ax_1 + \xi_1(t) \tag{3.12b}$$
$$\dot{x_2} = b + x_2x_0 - cx_2 + \xi_2(t) \tag{3.12c}$$

where the noise terms ξ_i $(i = 0, 1, 2)$ satisfy white Gaussian statistics $\langle \xi_i(t) \rangle = 0$ and $\langle \xi_i(t)\xi_j(\tau) \rangle = 2\sigma_i^2\delta_{i,j}\delta(t-\tau)$. The parameter b is chosen as the bifurcation parameter, and time-series data is obtained at $b = 0.4$, 0.6, 1.2, and 1.8. These values of b are taken around the period-doubling cascade of bifurcations from limit cycle to chaos. The other parameters are set as $a = 0.2$, $c = 5.7$ and $\sigma_i = 0.005$. The observation noise is introduced as $y_i = x_i + \eta_i\phi_i$ $(i = 0, 1, 2)$ where $\eta_i = 0.005$ is intensity of observation noise and ϕ_i is sampled from a standard normal distribution. The numbers of independent time-series data are 6, 6, 3, and 3 at $b = 0.4$, 0.6, 1.2, and 1.8, respectively. Figure 3.1 shows typical examples of observed orbit at each value of the bifurcation parameter.

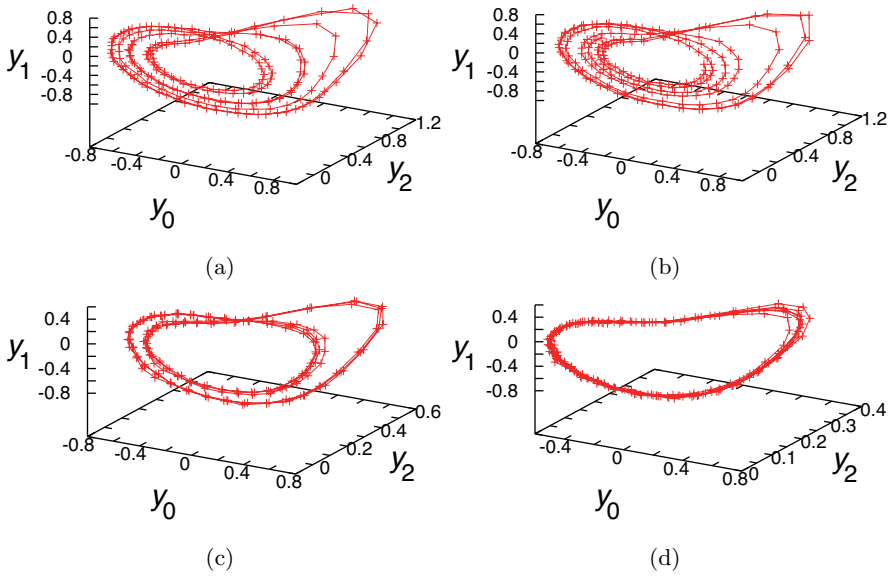


Figure 3.1: Noisy orbits generated from the Rössler model. These calculations are performed by a stochastic Runge-Kutta (SRKII) algorithm [59]. The values of the bifurcation parameter are (a) $b = 0.4$, (b) 0.6, (c) 1.2, and (d) 1.8, respectively. The other parameter values are set as $a = 0.2$, $c = 5.7$, $\tau = 10$, $\sigma_i = 0.005$, and $\eta_i = 0.005$. For reason, the observation values are scaled as $y_i/10$.

3.3.2 Model description

Here we define a state space model composed of state equations and observation equations. By adopting a polynomial of degree 2 with three variables, the state equations in continuous-time form become

$$\begin{aligned} \dot{x_0} &= k_0 + k_1 x_0 + k_2 x_1 + k_3 x_2 + k_4 (x_0)^2 + k_5 (x_1)^2 \\ &\quad + k_6 (x_2)^2 + k_7 x_0 x_1 + k_8 x_1 x_2 + k_9 x_2 x_0 + \text{noise} \qquad (3.13a) \\ \dot{x_1} &= k_{10} + k_{11} x_0 + \cdots + \text{noise} \qquad (3.13b) \\ \dot{x_2} &= b + k_{20} x_0 + \cdots + \text{noise} \qquad (3.13c) \end{aligned}$$

where b is the bifurcation parameter, and the noise terms are assumed to be white Gaussian. Note that we can obtain the Rössler model by setting the parameters as $k_2 = -1$, $k_3 = -1$, $k_{11} = 1$, $k_{12} = a$, $k_{22} = -c$, $k_{28} = 1$ and the other parameters $k_i = 0$. In the estimation process, the state equations are used in discrete-time form by the Euler-Maruyama scheme. Then the form of state equations $\dot{x_i} = f_i(\{x_j\}) + \text{noise}$ $(i = 0, 1, 2)$ are converted to $x_i^{t+1} = x_i^t + \Delta t f(\{x_j^t\}) + \sigma_i \xi_i^t \sqrt{\Delta t}$ where t is each time point, Δt is an integration time, σ_i is an intensity of system noise, and ξ_i^t is sampled from a standard normal distribution.

The observation equations, $y_i = x_i + \eta_i \phi_i$ $(i = 0, 1, 2)$, are assumed to be known except for the intensities of observation noises η_i to be estimated from the data.

3.3.3 ML estimation with full observation

Before the MAP estimation with Bayesian lasso, we perform the maximum likelihood estimation for comparison. In this case, the learning algorithm is identical to that employed in Chapter 2. (see section 4 in Appendix A for the detailed setting of the learning algorithm)

By considering the condition of full observation in which all of the three variables are simultaneously observed, we perform the learning at various settings with respect to initial parameter values sampled from uniform distribution $[-0.01, 0.01]$. The results show that, with sufficiently high initial intensities of system and observation noises, no fine-tuned initial parameter values are required to reach the maximum likelihood estimation. Then we set initial values of the coefficients of polynomial terms to zero, as $k_i = 0$ for all i, unless otherwise stated.

On the other hand, it is found that the integration time Δt strongly affects the estimation results. To characterize the dependence on Δt, an index N_{div} is defined to represent the number of interpolation points between the observation points, as

$$N_{\text{div}} = \frac{N_T - 1}{N_R - 1} \qquad (3.14)$$

where N_T is the number of time points for calculation, and N_R is the number of observed-time points. $N_{\rm div}$ means number of steps of numerical integration between two observation points and can be controlled independently of data. In other words, a value of $N_{\rm div}$ means a fraction of Δt to sampling interval. Figure 3.2 show the bifurcation diagrams of learned systems at different values of $N_{\rm div}$. As the value of $N_{\rm div}$ increases from 2 to 8, the learned bifurcation diagram is improved drastically and a period-doubling cascade clearly appears at $N_{\rm div} = 8$. However, in the particle filter algorithm, the required number of particle for accurate estimation increases exponentially as $N_{\rm div}$ because $N_{\rm div}$ roughly corresponds to the dimension of probability distribution to be approximated at each observation point. Then, with a limited computational resource, excess increase of $N_{\rm div}$ leads to poor estimation. We found that, with further increase of $N_{\rm div}$, the estimation of the system noise particularly gets worse.

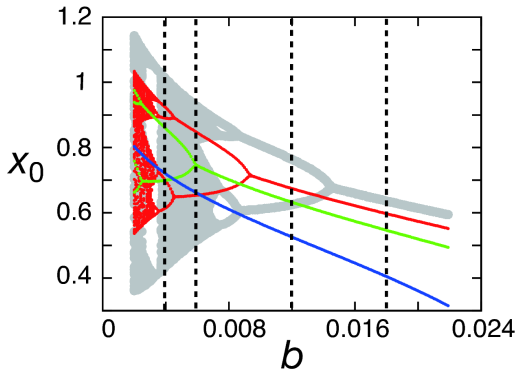


Figure 3.2: The bifurcation diagrams of learned systems at $N_{\rm div} = 2$ (blue), 4 (green), and 8 (red) are shown. 100 peak values of x_0 by noiseless calculation are plotted at each value of the bifurcation parameter b. The diagram of the Rössler model is also shown for comparison (gray). The broken lines indicate the points at which the data are given.

To overcome the dilemma between the particle number and $N_{\rm div}$, a two-step approach is employed. First, we perform the learning with a relatively low value of $N_{\rm div}$ and second, we fixed the estimated intensities of system noises, and re-estimate the other parameters with higher value of $N_{\rm div}$. From the procedure, an improved learned bifurcation diagram is obtained. Figure 3.3 shows the learned parameter values and bifurcation diagram at the best case among 50 trials of the two steps. The used values of $N_{\rm div}^{\rm low}$ and

$N_{\rm div}^{\rm high}$ are 8 and 128, respectively. As shown in Fig. 3.3(a), the procedure accurately estimates the coefficients of terms in the Rössler model within 1% error approximately. However, the coefficients of polynomial terms that do not exist in the Rössler model take small non-zero values, and the extra terms cause significant bias of learned bifurcation diagram in Fig. 3.3(b). In fact, even when the coefficients involved in the Rössler model are set at identical values to the data generator, no further improvement in the bifurcation diagram occurs. Then, we expect that Bayesian lasso improves the estimation by suppressing the unnecessary terms.

Before examining Bayesian lasso, we investigate why the learned bifurcation diagrams deviate in the common direction from the correct one as shown in above examples. A period-doubling bifurcation with decrease of b is delayed in each learned system, and the delay is significant with smaller $N_{\rm div}$ (Fig. 3.2). First an entropic effect is considered. That is, we infer that random fluctuations around the Rössler model result in the asymmetric deviation. To examine this inference, we calculate such fluctuated bifurcation diagrams, and find that the deviations are symmetric, as shown in Fig. 3.4(a). One may think that the result is not conclusive because sufficiently small fluctuations always result in symmetric deviations. However, since the amplitude of deviations in Fig. 3.4(a) is large enough to include the deviation in Fig. 3.3(b), we reject the entropic effect as the main cause of the asymmetric deviation in the bifurcation diagram.

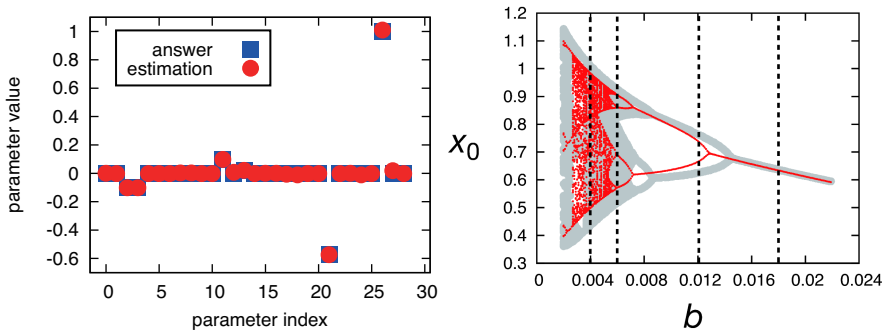


Figure 3.3: The result of the ML estimation with $(N_{\rm div}^{\rm low}, N_{\rm div}^{\rm high}) = (8, 128)$ is summarized. Note that the parameter values are properly scaled in concert with the scalings of time ($t \times 10$) and state variables ($x_i/10$) for practical reason. (Left figure) Blue rectangulars and red circles indicate parameter values of Rössler model and the learned system, respectively. The parameters $\{k_i\}$ are indexed as in eqs. (3.13a), (3.13b) , and (3.13c). (Right figure) Bifurcation diagrams of the learned system (red) and Rössler model (gray) are shown. 100 peak values of x_0 are plotted at each value of the bifurcation parameter b. The broken lines indicate points at which the data are given.

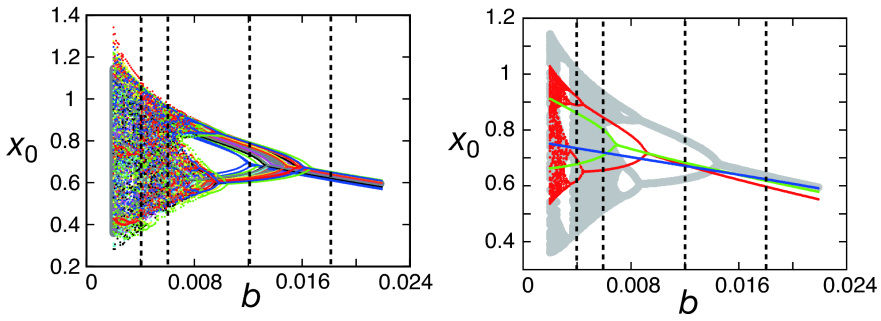


Figure 3.4: (Left figure) 20 bifurcation diagrams of fluctuated Rössler model (colored circles) are shown on the undisturbed bifurcation diagram (thick gray line). Each fluctuated system is generated by adding uniformly-distributed noise ($[-0.001, 0.001]$) to the rate coefficients that reproduce Rössler model. We note that the rate coefficients not involved in Rössler model are also fluctuated. (Right figure) The bifurcation diagrams calculated from estimations with fixed intensities system noises as σ_i ($i = 0, 1, 2$) = 0.02 (Blue), 0.01 (Green), and 0.005 (Red). The value of N_{div} takes 8 in all the cases.

Next it is hypothesized that large values of integration time Δt destabilize calculated dynamics, and then model parameters are estimated to compensate the effect by stabilizing the dynamics. This explains why, in the noiseless case, the learned systems exhibit a more stable dynamics than the data generator at each value of bifurcation parameter b. To examine this hypothesis, bifurcation diagrams are calculated from estimations with different fixed intensities of system noises. Figure 3.4(b) show larger intensity of system noise results in larger delay in period-doubling bifurcations with decrease of b. The results indicate that the estimated coefficients, $\{k_i\}$, stabilize the dynamics to compensate destabilization by a large noise amplitude. The fact indirectly supports the proposed hypothesis.

3.3.4 MAP estimation with full observation

In this section, we present the results of MAP estimation with Bayesian lasso on chaotic time-series data. The same data set as in the previous section is used. In the calculation below, $a_i^\gamma = 0$ and $b_i^\gamma = 0$ ($i = 0, 1, 2$) are employed in the prior distributions for the intensities of system noise, eq. (3.7). These settings correspond to usage of non-informative prior for scale type parameters, $p((\sigma_i)^2) \propto 1/(\sigma_i)^2$, that is derived from the assumption of uniform distribution of $\log(\sigma_i)^2$.

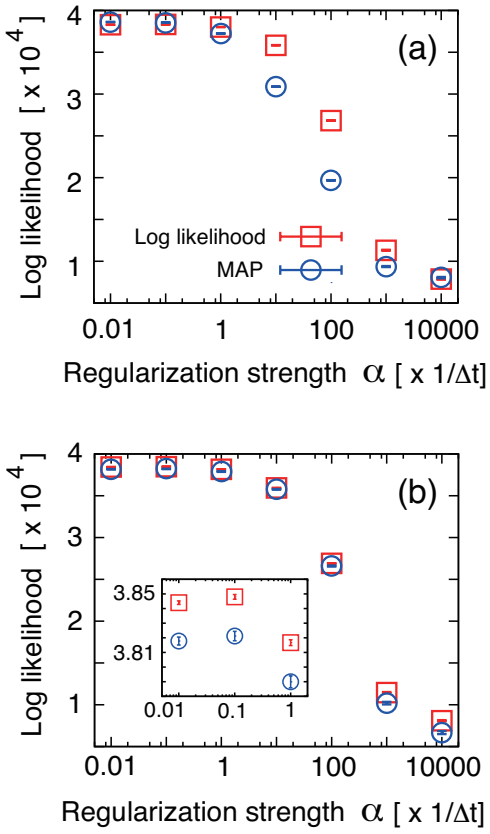
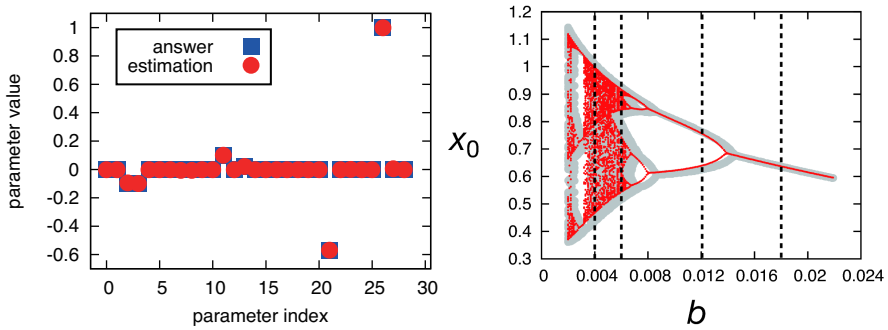


Figure 3.5: Evaluating the learned systems at different values of regularization strength α with $\Delta t = 1/8$. (a) (Log likelihood) $\log L(\theta)$ and (MAP) $\log L(\theta) \prod_i p(\{k_i^n\}|\sigma_i^2)p(\sigma_i^2)$ are presented. The factor $\log \prod_i p(\{k_i^n\}|\sigma_i^2)p(\sigma_i^2)$ is calculated as $\sum_i -2(N_i + a_i^\gamma + 1)\log\sigma_i - (b_i^\gamma + \alpha \sum_i^{N_i} |k_i^n|\langle f_i^n \rangle)\sigma_i^{-2}$ + const, and the constant term is ignored. The error bars indicate standard deviation of fluctuation in likelihood evaluation by the particle filter algorithm. (b) Log likelihood of the learned systems evaluated on two newly generated data sets is presented. Red squares indicate the results for new data set at bifurcation parameter values $b = 0.4$, 0.6, 1.2, and 1.8. The numbers of independent time-series data are 6, 6, 3, 3 at the four values of b. Those values and the number of data are the same as in the data used for the estimation. Blue circles indicate the results for new data set at bifurcation parameter values $b = 0.3$, 0.9, 1.5. The number of independent time-series data are 6 at each value of b. The inset provides an enlarged view between $\alpha/\Delta t = 0.01$ and 1. Before the evaluation, initial conditions are re-estimated to fit the new data while the other parameters are fixed. The error bars indicate standard deviation of fluctuation in likelihood evaluation by the particle filter algorithm.

The first problem to be solved for obtaining reliable estimation result is to choose an appropriate value of regularization strength α in eq. (3.6). Larger value of α leads simpler model in exchange for decrease of likelihood. To address the problem of the optimal value of α, we adopt a cross-validation technique. Learned systems are obtained at different values of α, and each system is evaluated based on likelihood for newly generated data sets from the data generator, as shown in Fig. 3.5.

Figure 3.5(a) shows maximum log likelihood (red squares) and log maximum a posterior probability (blue circles) depending on α. With the increase of α, the likelihood starts to decrease from $\alpha/\Delta t > 0.1$, and reach the minimum value at $\alpha/\Delta t = 10000$. We note that, at $\alpha/\Delta t = 10000$, all coefficients of polynomial terms are eliminated as $k_i = 0$ for all i.

Figure 3.5(b) shows log likelihood for two newly generated data sets. Red squares indicate the results for a new data set at the same values of the bifurcation parameter as in the data used for estimation. On the other hand, blue circles indicate the results for the other data set in which the values of the bifurcation parameters are different from those in the data set for estimation. The evaluated values of likelihood depend on α in the same manner in both cases. As shown in the inset, predictive performance for the new data is maximum at $\alpha/\Delta t = 0.1$, although the difference from the results at $\alpha/\Delta t = 0.01$ is not so large. Then we choose the value 0.1 as an optimal value.



(a) (b)

Figure 3.6: The result of the MAP estimation based on the full observation is summarized. The values of $N_{\rm div}^{\rm low}$ and $N_{\rm div}^{\rm high}$ are 8 and 128, respectively. The definitions of colors and symbols are same as in Fig.3.3.

By using the value of $\alpha/\Delta t = 0.1$, we take the two-step approach, as in the previous section, to estimate the coefficients of polynomial terms with high value of $N_{\rm div}$. The values of $N_{\rm div}^{\rm low}$ and $N_{\rm div}^{\rm high}$ are 8 and 128, respectively. We note that, in the second step of estimation with $N_{\rm div}^{\rm high} = 128$, $\alpha/\Delta t = 0.1 \times |T|^{\rm high}/|T|^{\rm low} \sim 1.6$ is employed. The factor $|T|^{\rm high}/|T|^{\rm low}$ represents

the increase of the number of data points calculated by particle filter at the E step, and then the increase of α by the same order is expected to be required to perform estimation in the same effective regularization. Figure 3.6(a) and (b) show the estimated values of parameters and bifurcation diagram of the learned system, respectively. The unnecessary terms are eliminated successfully, although some negligible values of the coefficients remain, probably because of fluctuation in the learning algorithm or the data set. Correspondingly, the bifurcation diagram of the learned system more accurately reproduces that of the Rössler model than the case of maximum likelihood estimation.

3.3.5 MAP estimation with partial observation

We also examine the MAP estimation based on partial observation by assuming that a state variable x_1 is completely unobservable except at the initial states. The data at the initial conditions is required to determine the scale of x_1. Figure 3.7 shows an example of observed orbit on the y_0-y_2 plane. The mapped orbit on the plane degenerates in a part of the chaotic dynamics, which seems to provide a challenging problem.

As a result of the learning, it is found that transient overestimation of the system noise of the unobservable component becomes huge in the process of iterative optimization, and leads poor estimation. Then, by setting the intensities of system noises to the true values ($\sigma_{0,1,2} = 0.005$), the learning is performed to estimate the coefficients of polynomial terms only. Consequently, the method successfully estimates the parameter values and the period-doubling cascade as shown in Fig.3.8(a) and (b), respectively. By choosing optimal values of a_i^γ and b_i^γ $(i = 0, 1, 2)$, we can expect to obtain as good estimation as this result. Cross-validation techniques may contribute to find the optimal prior for the system noises.

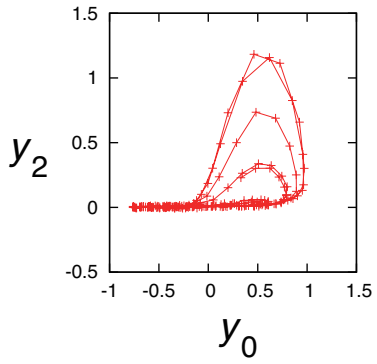
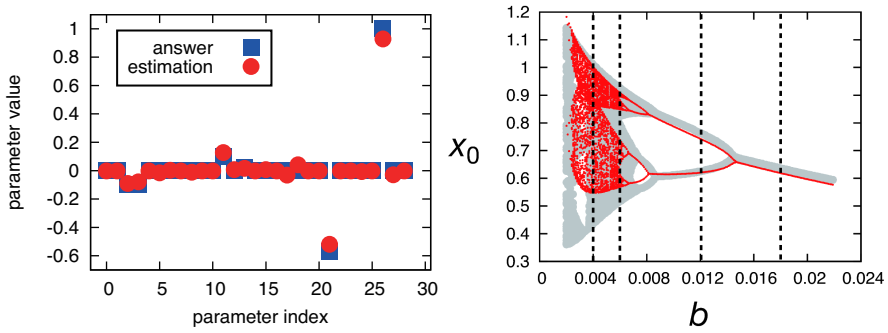


Figure 3.7: A noisy orbit on the y_0-y_2 plain is shown with $b = 0.4$.



(a) (b)

Figure 3.8: The result of the MAP estimation based on the partial observation is summarized. The values of $N_{\mathrm{div}}^{\mathrm{low}}$, $N_{\mathrm{div}}^{\mathrm{high}}$ and α are the same as those used to obtain the result in Fig. 3.6. The definitions of colors and symbols are same as in Fig.3.3.

3.3.6 Discussion

The above-mentioned results indicate that the sparse regularization by Bayesian lasso can improve estimation of chaotic systems in comparison with maximum likelihood estimation. However, the accuracy of the estimation depends on intensity of noise and the sampling interval compared to Lyapunov time, which should be theoretically evaluated in further study.

We examined the case that a part of the model is exactly the same as the data generator, the Rössler model. The assumption does not hold perfectly in many realistic applications. Although we can perform the estimation procedure and obtain a model exhibiting reasonable behaviors in the non-ideal cases, there is a delicate problem. Since characteristics of chaotic dynamics such as attractor dimension and the Lyapunov spectrum depend on micro-structures of the phase space, the characteristics seems to be strongly affected by employed model equations. In such cases, in what sense do we say that a reconstructed chaotic dynamics is similar with the original one? We hope to clarify this in future, but probably the problem does not affect quantitative estimation and prediction performances because noises mask micro-structures of phase spaces anyway.

3.3.7 Note: The Lorenz equations as data generator

Some chaotic attractors appear not through a period-doubling cascade as in the Rössler model. A well-known example of such an phenomena is provided

by the Lorenz model [72] described as

$$\dot{x} = -px + py \tag{3.15a}$$
$$\dot{y} = -xz + rx - y \tag{3.15b}$$
$$\dot{z} = xy - bz. \tag{3.15c}$$

The equations were introduced as a toy model of atmospheric convection, and exhibit chaotic dynamics with $p = 10$, $r = 28$, and $b = 8/3$. Here, with the decrease in r, the chaotic attractor suddenly disappears through a subcritical bifurcation by a mechanism called crisis in dynamical systems theory [71]. To validate the learning in such cases, we performed the maximum likelihood estimation from time-series data around the crisis by adopting the same procedure as was applied to the Rössler model. Polynomials of degree 2 with three variables are employed as state equations, but a slight modification from eqs. (3.13a,b,c) is required due to the different forms to include the bifurcation parameters, as $(k_{11}x_0, b) \to (rx_0, k_{19})$ and corresponding re-allocation of index i in k_is.

As a data set, we use time-series data obtained at $r = 10$, 16, 22, and 28 as shown in Fig. 3.9. Note that the time-series samples reflect the multistability of the Lorenz model at the low values of r. We examined 10 trials of the estimation by independent datasets. In addition, for each dataset, we also performed the estimation only from time-series data at $r = 28$ where chaotic attractor exists, in order to demonstrate advantage of using data at multiple values of the parameter. Figure 3.10(a) and (b) show the estimation results for coefficients of polynomial terms in the Lorenz model and for the other unnecessary coefficients, respectively. Compared with the data at $r = 28$ alone, use of the full datasets improves the results, especially the estimation of the coefficients that should take zero as true values. This indicates that the estimation of dynamical systems is effective even when the crisis occurs. Although the coefficients with zero true values remain finite, the sparse estimation method by L1 norm is expected to eliminate them.

We additionally note that estimated values of some coefficients with zero true values are biased from zero. The biased coefficients are $k_8, k_9, k_{17}, k_{23}, k_{24}$, and k_{25}. Interestingly, they almost correspond to the terms preserving the symmetry in the Lorenz model against the transformation $(x, y) \to (-x, -y)$. The symmetry-preserving terms with zero true values are k_8, k_9, k_{17}, k_{19}, k_{23}, k_{24}, and k_{25}, while there is only one exception, k_{19}. This suggests that the terms violating symmetry in the original system are eliminated easily, compared with the other terms preserving the symmetry.

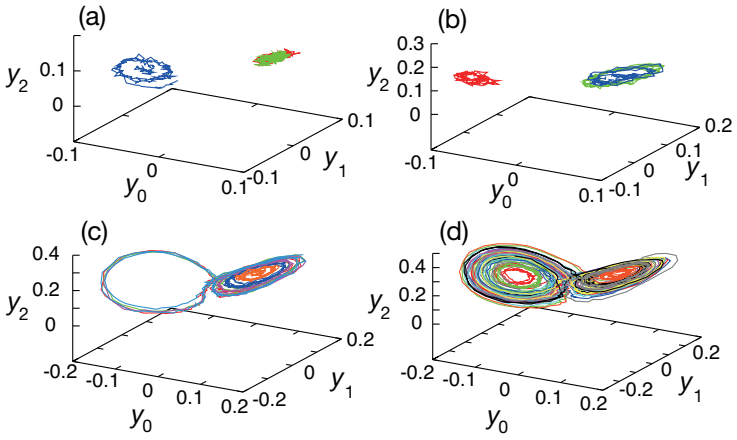


Figure 3.9: Noisy orbits generated from the Lorenz model by setting the initial condition as $(x_0, x_1, x_2) = (10, 10, 0)$. The data are obtained after taking the relaxation interval 600 from the initial condition. These calculations are performed by a stochastic Runge-Kutta (SRKII) algorithm [59]. The colors indicate results from independent trials. The values of the bifurcation parameter are (a) $r = 10$, (b) 16, (c) 22, and (d) 28, and The numbers of independent time-series samples at the values are 3, 3, 6, and 18, respectively. The other parameter values are set as $p = 10$, $b = 8/3$, $\sigma_i = 0.005$, and $\eta_i = 0.005$. For practical reason, the observation values are scaled as $y_i/100$.

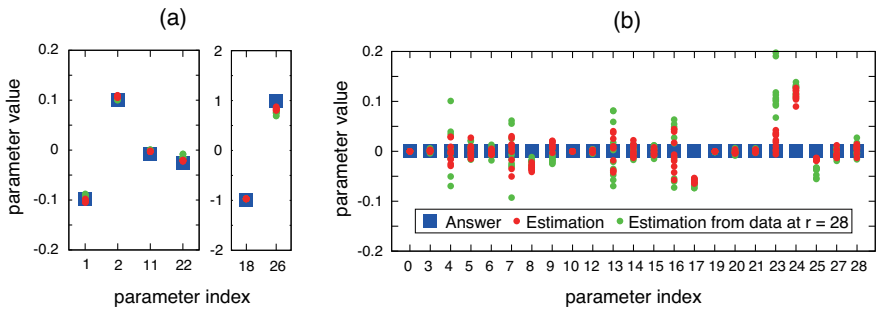


Figure 3.10: Maximum likelihood estimation of the polynomial model from data generated by the Lorenz model. (a) shows estimated coefficients of polynomial terms in the Lorenz model, $\{k_i\}$ $(i = 1, 2, 18, 22, 26)$, while (b) shows the other estimated coefficients. Blue, red, and green symbols are true values, estimated values from the dataset, and estimated values from the data at $r = 28$ only, respectively. Note that the parameter values are properly scaled in concert with the scalings of time $(t \times 100)$ and state variables $(x_i/100)$ for practical reason.

3.4 Identification of an adaptive system

Biological systems from unicellular to multicellular organisms experience various environmental stimuli during their lives, and respond appropriately to those in order to survive. The response often exhibits adaptation, in other words, output arises transiently in response to the input and returns to prestimulus level even when the input still holds [78, 79]. Biological functions of the adaptive behaviors are originally considered as extended dynamic ranges of signaling systems while detection of input temporal changes sensitively. This nature of response is not fully understood yet.

One of the most well-characterized systems to show adaptation is found in *Escherichia coli* cells that regulates flagellar rotation. The system underlies chemotaxis and is essential to search for food and to escape from toxin. Experimental studies revealed a quite simple structure of the system, and several models were proposed to reproduce key features of observed dynamics [80, 81, 82]. The model studies concluded that receptor desensitization by multiple methylation works as negative feedback to suppress signal transduction, and to achieve adaptation. The mechanism of the robust adaptation was extracted as integral feedback [83], and the concept helped to understand other systems like osmo-adaptation in yeast [84]. Besides, it was shown that the adaptive system enables the cells to perform chemotaxis by integrating temporal changes of concentrations of chemicals along with the time course, even when spatial gradient of the chemicals cannot be measured directly because of the short cell length [5].

From the viewpoint of machine learning of dynamical systems, adaptive behaviors provide a challenging problem because, unlike oscillatory or chaotic systems, information extraction from transient dynamics with time-varying input is essential for system identification. However, by considering the ubiquity and the biological significance of adaptation, it is essential to analyze such dynamics to elucidate cell behaviors. Then we apply our algorithm to estimate a model of adaptation by using step stimulus that is most frequently employed to characterize cellular response experimentally.

In order to choose a specific data generator, we focus on a subclass of adaptive behaviors called fold-change detection (FCD) that is recently proposed [85, 86]. The definition is that systems with FCD property response depending on fold change of input, not absolute change. Then FCD is analogous to Weber's law in sensory physiology that states relationship between stimulus and perception as logarithmic, with the exception that FCD is defined for temporal input-output relations. FCD is considered to provide robustness in signaling cascades agaist cell-cell variations of the signaling because such variations sometimes cause fluctuation while keeping fold change of the response as is actually observed [87].

Along with the experimental works, several theoretical studies are performed mainly by control theory in engineering [88,89]. In particular, Shoval

and co-workers show that a symmetry property in dynamical systems is sufficient for the adaptive systems to perform FCD. FCD models can be drawn from the criteria easily.

Below we construct a polynomial model of FCD as shown in Fig. 3.11, and confirm the symmetry can be identified by the machine learning. In particular, the state equations are defined to contain another adaptive system whose the attractor is the same as the polynomial FCD model, in order to verify that the learning integrates information in the transient dynamics, not in the attractor alone.

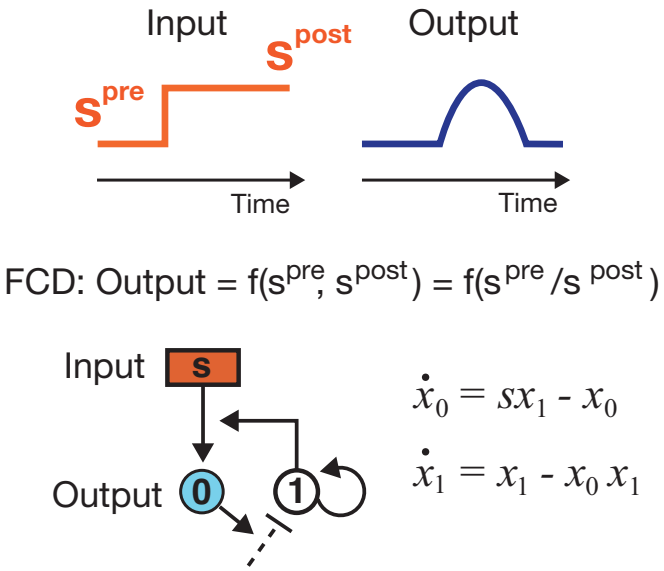


Figure 3.11: A definition of FCD and a polynomial model for FCD.

3.4.1 Data generator: A polynomial model of FCD

The FCD condition in simplified form is as follows [88]. Let us consider a two-dimensional dynamical system under input s as

$$\dot{x_0} = f_0(x_0, x_1, s) \tag{3.16a}$$
$$\dot{x_1} = f_1(x_0, x_1, s), \tag{3.16b}$$

and x_0 is defined as output of the system. It is assumed that the system has a globally stable fixed point. If, for each fold change of input p, there is some differentiable map $\phi_p(x_1)$ with the following properties,

$$f_0(x_0, \phi_p(x_1), ps) = f_0(x_0, x_1, s) \tag{3.17a}$$
$$f_1(x_0, \phi_p(x_1), ps) = \frac{d\phi_p(x_1)}{dx_1} f_1(x_0, x_1, s), \tag{3.17b}$$

then the system has FCD property. Based on the condition, we can easily determine whether a system has FCD property or not.

As the data generator, a polynomial model is constructed to satisfy the condition as

$$\dot{x_0} = c_0 s(t) x_1 - c_1 x_0 + \xi_0(t) \tag{3.18a}$$
$$\dot{x_1} = c_2 x_1 - c_3 x_0 x_1 + \xi_1(t) \tag{3.18b}$$

where the noise terms $\xi_i(t)$ $(i = 0, 1)$ satisfy white Gaussian statistics $\langle \xi_i(t) \rangle = 0$ and $\langle \xi_i(t) \xi_j(\tau) \rangle = 2\sigma_i^2 \delta_{i,j} \delta(t-\tau)$. In the case of noiseless dynamics, eigenvalue analysis indicate that the fixed point with a positive constant input s ($x_0 = c_2/c_3$, $x_1 = (c_1 c_2)/(c_0 c_3 s)$) is stable with positive rate constants $c_{0,1,2,3}$, and then the system show FCD. In addition, time-varying input are defined as

$$s(t) = \begin{cases} s^{\mathrm{pre}} & (t < t_{\mathrm{input}}) \\ s^{\mathrm{post}} + (s^{\mathrm{pre}} - s^{\mathrm{post}}) \exp(-\lambda(t - t_{\mathrm{input}})) & (t \geq t_{\mathrm{input}}). \end{cases} \tag{3.19}$$

The input time series comes from dynamics of a chemical concentration in a perfusion chamber as

$$s(0) = s^{\mathrm{pre}}, \dot{s} = \begin{cases} s^{\mathrm{pre}} - \lambda s & (t < t_{\mathrm{input}}) \\ s^{\mathrm{post}} - \lambda s & (t \geq t_{\mathrm{input}}) \end{cases} \tag{3.20}$$

where λ is flow rate, t_{input} is switching time, s^{pre} and s^{post} are pre- and post-stimulus levels of input, respectively.

By using simple observation equations as $y_i^r = x_i^r + \eta_i \phi_i^r$, we generates 43 time-series data at various pre- and post-stimulus values, s^{pre} and s^{post}, as shown in Table 3.1. The other parameter values are set as, $c_0 = c_1 = c_2 = c_3 = 2$, $\sigma_0 = \sigma_1 = 0.1$, $\eta_0 = \eta_1 = 0.002$, $\lambda = 32$ and $t_{\mathrm{input}} = 2.33$. As shown by a subset of the artificial data set presented in Fig. 3.12, the duration time of each data ($0 \leq t \leq 7$) includes the first peak, while subsequent damped oscillation is not fully observed. The data provide sufficient information to identify the data generator as shown below.

(s^{pre},s^{post})	Number of generated time series
(0.03, 0.15)	9
(0.1, 0.3)	14
(0.1, 1)	10
(0.3, 0.9)	5
(0.3, 3)	5

Table 3.1: Input values and corresponding numbers of generated time series.

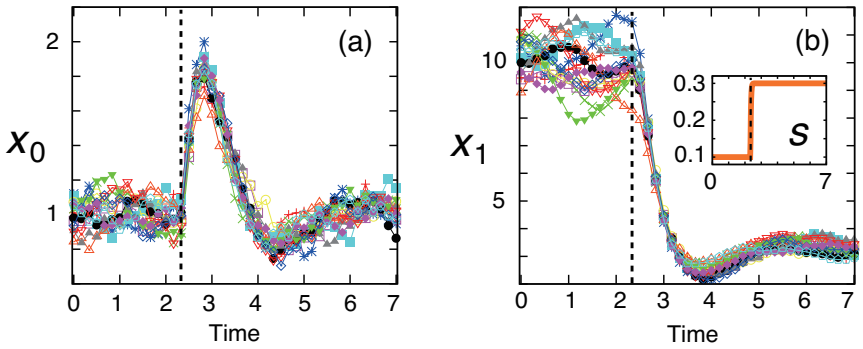


Figure 3.12: Time series of (a) x_0 and (b) x_1 generated from eqs. (3.18a) and (3.18b) at $s^{\mathrm{pre}} = 0.1$ and $s^{\mathrm{post}} = 0.3$. The time-varying input is shown in the inset of (b). The broken lines indicate t_{input}.

3.4.2 Learning from transient dynamics

State equations

First we express state equations in a continuous-time form as

$$\dot{x_i} \quad = \quad s(t)g_i(\{x_j\}) + f_i(\{x_j\}) + \text{noise} \tag{3.21a}$$

where f_i and g_i $(i = 0 \ldots D-1)$ are polynomials of degree M and M', respectively. By considering $M = 2$ and $M' = 1$ in two variables $(D = 2)$, Discrete-time form of the state equations with integration time Δt is wrote down as

$$\begin{aligned} x_0^{t+1} \quad = \quad & x_0^t + \Delta t(k_0 + k_1 x_0^t + k_2 x_1^t)s \\ & + \Delta t(k_3 + k_4 x_0^t + k_5 x_1^t + k_6 (x_0^t)^2 + k_7 x_0^t x_1^t + k_8 (x_1^t)^2) \\ & + \sigma_0^t \xi_0^t \sqrt{\Delta t} \end{aligned} \tag{3.22a}$$

$$x_1^{t+1} \quad = \quad x_1^t + \Delta t(k_9 s + k_{10} s x_0^t + \cdots + k_{17}(x_1^t)^2) + \sigma_1^t \xi_1^t \sqrt{\Delta t} \tag{3.22b}$$

where σ_i is the intensity of the system noise, and ξ_i^t is sampled from a standard normal distribution as previous examples in this thesis.

We emphasize importance to extract information from transient dynamics in this setting. The polynomial state equations contain a linear integral

feedback system,

$$\dot{x_0} = c_0 c_3 s(t) x_1 - c_1 c_2 \tag{3.23a}$$
$$\dot{x_1} = c_2 - c_3 x_0. \tag{3.23b}$$

This system shows adaptation and has the same fixed point with any constant input s as the data generator, i.e., the values of (x_0, x_1) at the fixed point are identical to $x_0 = c_2/c_3$, $x_1 = (c_1 c_2)/(c_0 c_3 s)$. Existence of such a system as a sub model indicates that the data generator cannot be identified by positions of fixed points alone.

In principle, noisy time series around a fixed point at a constant input value contains information required to distinguish the different adaptive systems; however we found that estimation based on such data alone does not work at least in the setting of data number and sampling interval. This is reasonable because estimation of dynamical systems based on the noisy time-series data around the fixed point is known to be difficult even in the case of the Ornstein-Uhlenbeck process [90] that is simpler than the polynomial FCD model considered here. Then, we conclude that integration of information in transient dynamics with varying input is practically essential to identify the data generator by the machine-learning.

Estimation of initial condition for each prestimulus input value

Since the data generator under constant input s has the globally stable fixed point, initial conditions with common prestimulus input value s^{pre} are quite similar to each other, unlike in the case of oscillatory dynamics in which the degree of freedom of the phase exists. Then independent estimation of initial conditions of all time-series data seems to be over-parametrization, and might cause overfitting problem. To prevent the problem, initial conditions with common s^{pre} are estimated by single normal distribution. For each value of s^{pre}, the mean μ and the variance V of the estimated Gaussian are calculated as $\mu = \frac{1}{n}\sum_i \mu_i$ and $V = \frac{1}{n^2}\sum_i V_i$, respectively, where the summation is taken over the initial conditions with the same s^{pre}, and n is number of the condition.

Results

The learning with Bayesian lasso is performed based on adaptive time-series data on x_0 and x_1. In the calculations below, $a_i^\gamma = 10^3$ and $b_i^\gamma = 10$ $(i = 0, 1)$ are employed in the prior distributions for the intensity of system noise, eq.3.7, in order to avoid to be trapped on local maxima although it is not essential for the estimation. (see section 4 in Appendix A for the detailed setting of the learning algorithm)

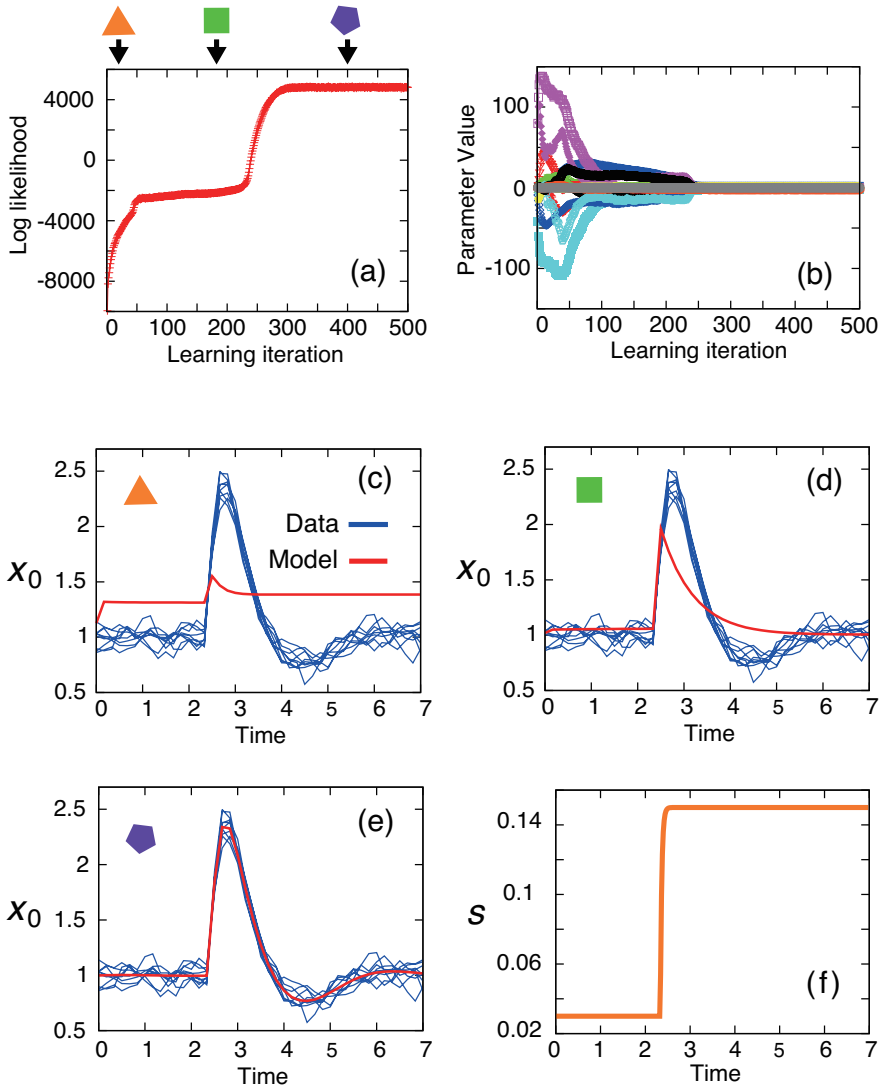


Figure 3.13: The two-phased learning trajectory with $\alpha = 9600$ of (a) log likelihood and (b) parameter values. Colors in (b) indicate different parameters. (c), (d), and (e) show comparison between the model behavior and the data with tentatively estimated parameters at learning iteration $I =$ (c) 20, (d) 180, and (e) 400. Corresponding values of I are indicated in (a) by black arrows and colored polygon for (c) orange triangle, (d) green square, and (e) purple pentagon. The input time series used in (c-e) is shown in (f).

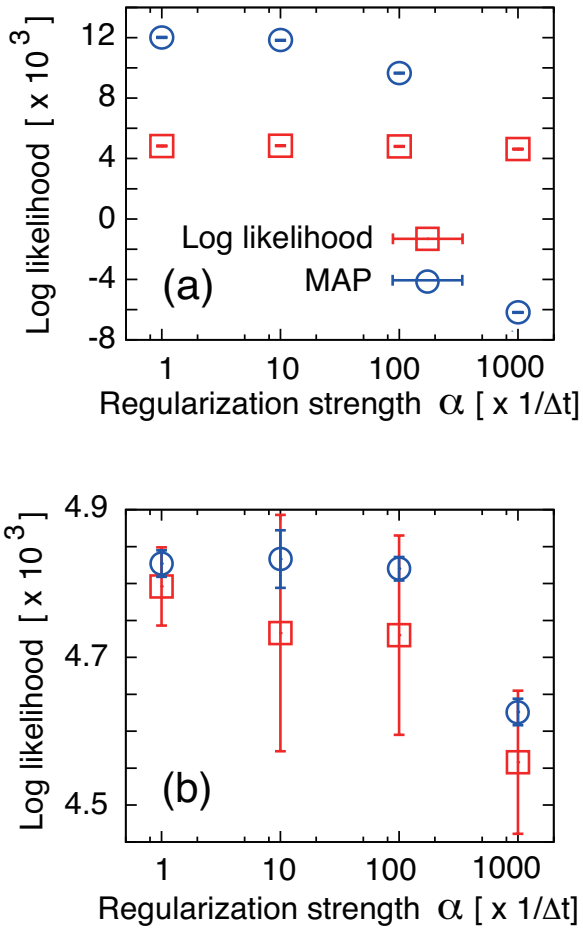


Figure 3.14: Evaluating the learned systems at different values of α with $\Delta t = 1/96$. (a) (Log likelihood) $\log L(\theta)$ and (MAP) $\log L(\theta) \prod_i p(\{k_i^n\}|\sigma_i^2)p(\sigma_i^2)$ are presented. The factor $\log \prod_i p(\{k_i^n\}|\sigma_i^2)p(\sigma_i^2)$ is calculated as $\sum_i -2(N_i + a_i^\gamma + 1)\log \sigma_i - (b_i^\gamma + \alpha \sum_i^{N_i} |k_i^n|\langle f_i^n \rangle)\sigma_i^{-2} + \text{const}$, and the constant term is ignored. The error bars indicate standard deviation of fluctuation in likelihood evaluation by the particle filter algorithm. (b) Log likelihood of the learned systems evaluated on newly generated data sets is presented. Red and blue symbols indicate results from two independently identically distributed (i.i.d.) data set. Before the evaluation, initial conditions are re-estimated to fit the new data while the other parameters are fixed. The error bars indicate standard deviation of fluctuation in likelihood evaluation by the particle filter algorithm.

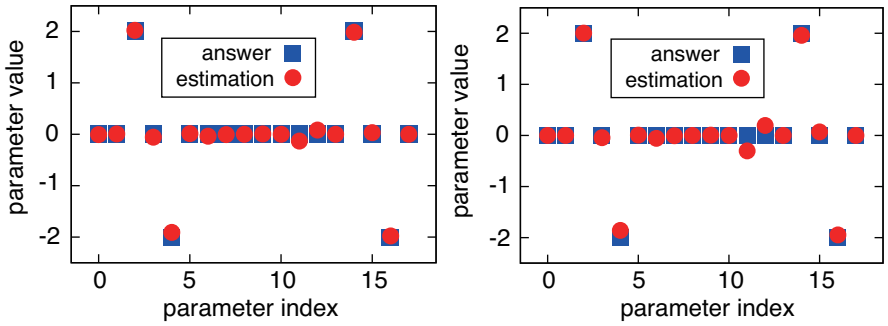
As a result, two-phased learning is observed, where a plateau is visible in the trajectory of log likelihood along with learning iteration, as shown in Fig. 3.13(a). In Fig. 3.13(b), the tentative parameter estimations along with the learning trajectory show long-time stay far from the terminal values. This reflects the two phase in the log likelihood optimization. To understand the learning, we investigate the model behaviors before the plateau, on the plateau, and after the plateau. Although the model parameters are complicated to analyze unless the estimation converges, changes of the model behaviors are interpretable. Figures 3.13(c), (d), and (e) show the noiseless orbit of x_0 with the estimated parameters in comparison with time-series data at the three points on the learning iteration. Although the basal output and transient response in the data are not reproduced by the model before the plateau, the basal output is reproduced on the plateau, and the full output behavior is reproduced after the plateau. The two phases are interpreted as learning the observed attractors in the first phase and the transient dynamics in the second phase. Then it is confirmed that information in transient dynamics is certainly included in the estimation.

Following the success of the learning, we perform the estimations by using various values of the regularization strength α to determine the optimal value as shown in Fig. 3.14. Figure 3.14(a) shows changes in maximum $\log L(\theta)$ and $\log L(\theta) \prod_i p(\{k_i^n\}|\sigma_i^2)p(\sigma_i^2)$ from $\alpha/\Delta t = 1$ to 1000. The values of log likelihood are stable against 1000-fold increase of α. On the other hand, Figure 3.14(b) show the result of cross validation. That is, likelihood of the learned system is evaluated on newly generated artificial data with the same settings as the data used for the learning. Although the evaluated likelihood at $\alpha/\Delta t = 1000$ exhibits clear decrease, the learned systems at $\alpha/\Delta t = 1, 10, 100$ predict the new data equally well, and likelihood-based cross validation alone is insufficient to choose the optimal value between $\alpha/\Delta t =$ 1-100.

Therefore, here we focus on the learned system at $\alpha/\Delta t = 100$ because the system is simpler than the other system with $\alpha/\Delta t =$ 1-10 and can be expected to be similar to the original data generator in which number of polynomial terms with non-zero coefficients is small. As shown in Fig. 3.15(a), the parameter values $\{k_i\}$ are correctly estimated at $\alpha/\Delta t = 100$. However, some unnecessary terms that are not involved in the data generator are also visible, for example, at values of k_{11} and k_{12}. Furthermore, it is found that stronger regularizations by larger values of α do not suppress the unnecessary terms, but amplify them. Figure 3.15(b) demonstrates the case of $\alpha/\Delta t = 1000$ for comparison.

We infer that the unnecessary terms come from time discretization error as in the case of identifying the Rössler model. Then we consider the division number between the observed-time points $N_{\rm div} = (|T| - 1)/(|R| - 1)$ once again, and perform the learning at various values of $N_{\rm div}$ while intensities of system noise $\{\sigma_i\}$ are fixed. The results are summarized in Fig. 3.16.

The unnecessary parameters k_{11} and k_{12} diminish gradually as increase of $N_{\rm div}$, and on the other hand the necessary parameters k_2, k_4, k_{14}, and k_{16} converge the true value, 2. Therefore, we finally obtain a learned system that is nearly identical to the data generator.



(a) (b)

Figure 3.15: The parameter values $\{k_i\}$ $(i = 0, \ldots 17)$ of learned systems (red circle) are shown in comparison with the data generator (blue square) at $\alpha/\Delta t =$ (a) 100 and (b) 1000.

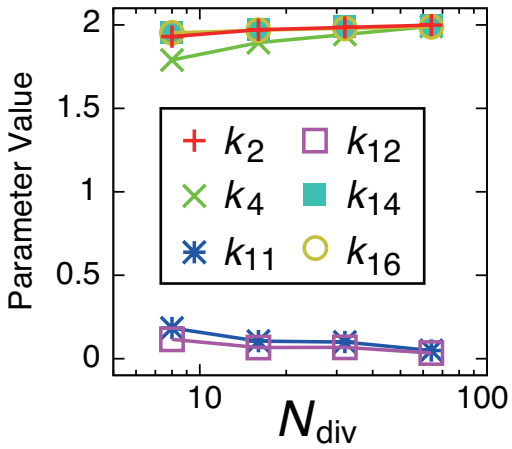


Figure 3.16: The variation of parameter values as increase of $N_{\rm div} = (|T| - 1)/(|R| - 1)$ is shown.

3.4.3 Discussion

The learned system by MAP estimation with Bayesian lasso reproduces sequentially the observed adaptive behavior and the transient response through the two-phased learning trajectory. The results indicate that the method include information not only in the observed attractors, but in the transient dynamics with time-varying input. In fact, the learned systems have never been trapped on the false system described by eqs. (3.23a) and (3.23b) that exhibits the same base output as the data generator. The success is prerequisite for applications to adaptive behaviors that are ubiquitous among the cell signaling dynamics.

In particular, we succeeded to extract the adaptive systems having a type of input-scaling invariance, called FCD. As mentioned in [89], symmetries other than the FCD can be considered and may be advantageous for environmental perception. The human sensory system provide a variety of examples, such as rotational symmetry that is required for robust visual object recognition. However, at the single-cell level, there is little known about what kind of symmetry can be realized or actually exists in the sensory systems. The machine learning may reveal hidden invariance properties in cellular perception as principles in evolutionary design of the signaling systems.

3.5 Summary

The biochemical diversity in cell signaling systems complicates selection of function forms in the model to reproduce the dynamics. On the other hand, a statistical technique, called sparse estimation, has been developed to extract an appropriate sub model from a rather complicated equations. We implemented a sparse estimation method of dynamical systems and examined it with applications to nonlinear dynamical phenomena by using artificial data; one application is for chaotic dynamics and another application is for adaptation with a type of input-scale invariance called FCD. The sparse estimation is a simple extension of the algorithm considered in Chapter 2 by Bayesian lasso. For both cases of chaos and FCD, the method identify the data generator models by eliminating unnecessary terms in state equations. The results for the nonlinear phenomena imply a wide range of potential application of the method to cellular dynamics.

The sparse estimation for dynamical systems focuses mainly on reducing parameter number, not state dimension. However, the reduction of state dimension is, if possible, also useful because it corresponds to a procedure of omitting cell signaling components having relatively small effects on the dynamics, which is frequently performed in systems biology. Model reduction techniques in the direction have been developed in control theory and statistics, and there are applications to models for cell signaling systems.

We hope that a combined approach of reduction of both parameter number and state dimension can automatize model simplification processes.

Chapter 4

A cell-cell communication system in a social amoeba

4.1 Introduction

Populations of cellular organisms often exhibit collective behaviors based on cell-cell communications [91]. The communications rely on intracellular signaling systems that perceive and transmit extracellular signals. Here we aim at understanding and predicting collective dynamics of cells by phenomenological modeling of the input-output characteristics of the signaling systems at single-cell level, even when detailed structures of the intracellular molecular circuits are unknown. Such procedures have a wide range of applications, potentially including all of developmental dynamics of multicellular organisms. In this chapter, we take the first step towards this direction, by investigating a cell-cell communication system in a social amoeba.

4.1.1 Model organism: *Dictyostelium discoideum*

Dictyostelium discoidum is a species of amoeba, which lives in soil and eats bacteria [92]. This model organism is known to have a unique life cycle, called asexual development, described below (see also Fig. 4.1). Upon nutrient depletion, *Dictyostelium* cells undergo transition from vegetative state to starved state through cell cycle arrest and starting developmental transcriptional programs. In the starved state, the cells are capable of sensing an extracellular diffusive chemical, cyclic adenosine 3',5'-monophosphate (cAMP), and respond to the signal by synthesis and secretion of cytosolic cAMP. Then the cAMP signal is relayed from cell to cell, and large-scale wave patterns appear from the multicellular dynamics [93, 94]. After that, the cells move towards the chemical waves, and form aggregates at the signaling centers. The aggregate shows collective movement and differentiation of involved cells, finally turn themselves into a reproductive structure, called

a fruiting body [95].

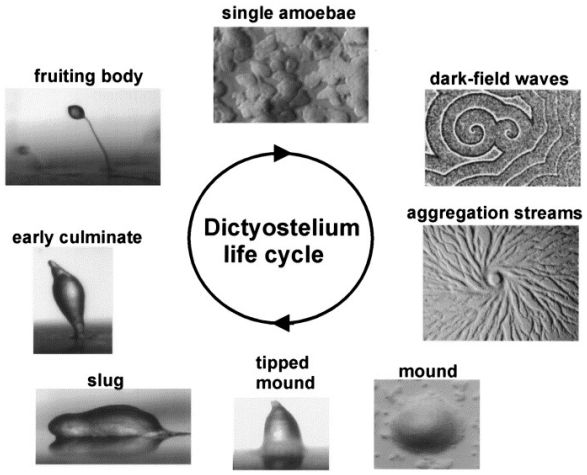


Figure 4.1: Asexual developement of *D.discoideum*. The picture is from Ref. [96].

4.1.2 Signal transduction at a single-cell level

The spatiotemporal dynamics of cAMP in the early phase of the development generates spiral pattern of a wavelength $\sim$ 2 mm in which local cell populations exhibit regular oscillation with a period of approximately 6 min. Both of the spatial and temporal patterns are of fundamental importance for progress of the development, because the spiral waves provide guide signal for aggregation and the periodic variation induces cell differentiation. However, measurement of the underlying cAMP dynamics at a single-cell level was prohibited by technical limitation, which made it hard to understand how the collective dynamics arise.

Recently S. Sawai and his group succeeded in measuring the the single-cell dynamics by using a fluorescence resonance energy transfer (FRET)-based sensor to monitor cAMP in live *Dictyostelium* cells [97]. They also employed a perfusion chamber to reveal cAMP dynamics in isolated cells under various extracellular cAMP concentrations. As a result, they found adaptive behaviors, that is, transient rise and subsequent decay of cytosolic cAMP level, with moderate dose of the extracellular cAMP.

4.1.3 Scale invariance in response dynamics

In the previous study, the adaptive behaviors are examined mainly by step input with zero pre-stimulus level of extracellular cAMP. To characterize the

adaptive system in more detail, we tried to quantify the response by step input with various non-zero prestimulus levels [98]. The result indicates that the adaptive dynamics approximately has a invariance property known as fold change detection (FCD). FCD is defined by the response whose entire time course depends on only on fold changes in input but not on absolute levels. The property is suggested by indirect evidences, and observed in a population-level response in a bacterial signaling. In addition to that, we provide the first direct evidence of FCD at a single-cell level.

We also investigated an effect of cell individuality on FCD of the cAMP relay system. To this end, we sequentially applied two step inputs with the same fold change, but with different absolute levels, to each single cell. While the responses show large variation among cells, the two subsequent responses from each cell are strongly correlated. Therefore, we found that FCD is robust against cell-cell variation. Since the examined cell population is clonal, the cell-cell variation seems to come from fluctuation in biochemical parameter values. This interpretation imply that FCD of the cAMP relay is originally caused by a structural property in the system, not by a precise balance between biochemical parameter values.

4.1.4 Modeling single-cell signaling

We hope to explain the multicellular dynamics based on the experimental results at a single-cell level mentioned above. A straightforward approach is to construct a model to describe the single-cell dynamics, and to calculate collective behaviors of the elemental models. However, the whole structure of the circuit is largely unknown because of the complexity [99], although behaviors of some sub systems are quantitatively elucidated especially around a circuit related to cell polarity formation in chemotaxis [100]. On the other hand, the M&G model, the most well-known mechanistic model for the cAMP dynamics, shows response dynamics far from FCD, at least, with parameter values employed in the original literature [16, 98]. Besides, the receptor desensitization mechanism assumed in the M&G model seems to be inconsistent with a recent experiment on the receprtor property [101]. Therefore, no mechanistic model is available in this case.

Instead, we take another approach based on a phenomenological model for reproducing the input-output characteristics, FCD. Nevertheless, determining a model is still not trivial because we have a variety of models that show FCD. Although all the models show FCD, they have different quantitative properties and possibly make different predictions for multicellular dynamics. To overcome the problem of many potential models, we adopt two strategies focusing on qualitative and quantitative features. First, we investigate what conclusions can be drawed only from FCD property, not from quantitative properties of the models. To this end, we analyze a set of equations for describing multicellular dynamics in which signaling reponse

in each cell is FCD. Instead of solving the equations, we elucidate qualitative features in flows in the phase space that are invariant against certain variable transformations representing cell-density fluctuation. Second, we try to construct a phenomenological model based on quantitative information in data, and, if possible, select a model for FCD. This is performed by using time series of the FRET signals, and The model to be estimated is designed to contain well-known models for FCD as sub models.

4.2 Theory: Robustness in multicellular dynamics based on FCD

A variety of models for FCD has been proposed, but all of them satisfy a symmetry condition defined by eqs. (3.17a) and (3.17b). Therefore, it is beneficial to derive general conclusions that can be drawn only from the symmetry, not from quantitative feature of each model. To this end, we analyze a set of equations for describing multicellular cAMP dynamics of *D.discoideum*. The analysis leads a hypothesis to answer a question what is an advantage of the FCD property in the multicellular dynamics.

4.2.1 Robustness to cell density fluctuation

Let us introduce a model of multicellular dynamics as a reaction diffusion system, by considering secretion, dilution, and diffusion of extracellular cAMP, as

$$\dot{x} = f(x, y, z) \tag{4.1a}$$
$$\dot{y} = g(x, y, z) \tag{4.1b}$$
$$\frac{\partial z}{\partial t} = \rho k_t y - \lambda z + D\Delta z \tag{4.1c}$$

where x, y, and z represent concentrations of an intermediate signaling component, intra- and extra-cellular cAMP, respectively. The parameters ρ, k_t, λ, and D are cell density, secretion rate, dilution rate, and diffusion constant, respectively. We assumed that a number of the intermediate components is just one for simplicity, but the analysis below is easily extended for the case of more than one intermediate component, as long as the single-cell system composed of x and y is FCD.

It is found that, when the intracellular dynamics of x and y satisfy the symmetry condition for FCD, we can remove the cell density ρ from the model by variable transformations $z = \rho Z$ and $x = \phi_\rho(X)$, as

$$\dot{X} = f(X, y, Z) \tag{4.2a}$$
$$\dot{y} = g(X, y, Z) \tag{4.2b}$$
$$\frac{\partial Z}{\partial t} = k_t y - \lambda Z + D\Delta Z. \tag{4.2c}$$

Here the function ϕ_ρ is defined in eqs. (3.17a) and (3.17b). This means that the cell density does not change the topological properties of the multicelluler dynamics in the phase space such as existence of oscillation and number of stable states. Therefore, the symmetry property for FCD make the multicellular dynamics robust against cell density.

Actually it is experimentally observed that the asexual development successfully progresses over a wide range of approximately 100-fold change in cell density [102, 103]. We expect that the symmetry property for FCD at the single-cell level underlies the robust development.

We note that, although the kinetics of cAMP secretion is modeled as $\propto \rho y$ in the above calculation for simplicity, the discussion holds in general as long as the secretion is described by a separated form, $p(\rho) \times q(y)$.

4.2.2 A possible effect of *PdsA* on the robustness

Next, we investigate how the robustness is affected by considering a more realistic equation for kinetics of extracellular cAMP. This is performed by introducing effect of a phosphodiesterase, *PdsA*, to the model. The enzyme is secreted from the cells and degrades extracellular cAMP. Although we have considered the constant degradation rate of extracellular cAMP, λ, in above discussion, the rate can changes depending on cell density and kinetics of secretion of the cAMP-degrading enzyme which remains unclear. Here we assume the effect of *PdsA* as a cell-density-dependent constant rate of cAMP degradation $\lambda(\rho)$, because the regulation of *PdsA* depends on transcriptional control and the time scale is slow compared with the cAMP oscillation (5-6 min). In that case, the robustness based on FCD does not perfectly hold because of the cell-density dependence in λ.

However, when we adiabatically approximate the dynamics of extracellular cAMP z as

$$0 \quad = \quad \rho k_t y(r) - \lambda(\rho) z(r) + D\Delta_r z(r) \tag{4.3}$$

where r represent a position vector, the cell density ρ can be removed from the system by variable transformations $z = \rho Z/\lambda(\rho)$, $x = \phi_{\rho/\lambda}(X)$, and $r = R/\sqrt{\lambda(\rho)}$. This means that the temporal dynamics of cAMP does not change with cell density while the spatial pattern is scaled as $1/\sqrt{\lambda(\rho)}$. The weak robustness to cell density seems to be still helpful for the cells in the asexual development because the topological properties of the multicellular dynamics still remain invariant. Furthermore, the system can control the scale of the spatial pattern independently of the temporal dynamics, which propose a possible role of *PdsA*. It may be interesting to quantify λ and the wavelength of the spiral pattern on *Dictyostelium* cells at various cell densities.

To evaluate the validity of the adiabatic approximation, we separate $z(r, t)$ into the solution to the steady state equation and deviation from the

solution, as $z(r,t) = z^*(r) + \delta z(r,t)$. By Fourier transformation of eq.(4.3), relaxation dynamics of power spectrum of the deviation is written as

$$\frac{d}{dt}|\delta z(\hat{k})|^2 \quad = \quad -2(\lambda(\rho) + D|k^2|)|\delta z(\hat{k})|^2 \tag{4.4}$$

where k represent a wave number vector. The degradation rate λ set a lower limit to the decay rate of the deviation, and the decay time scale defined by typical experimental value of λ (2.5-12.5 min^{-1}) [16, 104, 105, 106] is faster than a time scale of collective cAMP oscillation of the cells ($\sim$0.2 min^{-1}). From the rough evaluation, the adiabatic approximation seems to be reasonable, even though not highly accurate. Therefore, we can expect the robustness based on FCD even under the realistic assumption on the kinetics of extracellular cAMP, which support the hypothesis.

4.3 Estimation

In contrast to the qualitative approach based only on the symmetry property for FCD, model selection from quantitative information in the FRET time series is also performed. To this end, we define a complex model that contains a variety of models for FCD as sub models, and train it to reproduce the observed adaptive time series. After that, we try to extract a sub model by sparse regularization.

4.3.1 Data acquisition

We focus on the cAMP relay system that underlies the multicellular chemical pattern formation in an early phase of the asexual development. By defining the input and the output of the system as extracellular and intracellular cAMP concentration, characteristics of the system are quantified as follows (Fig. 4.2) [98].

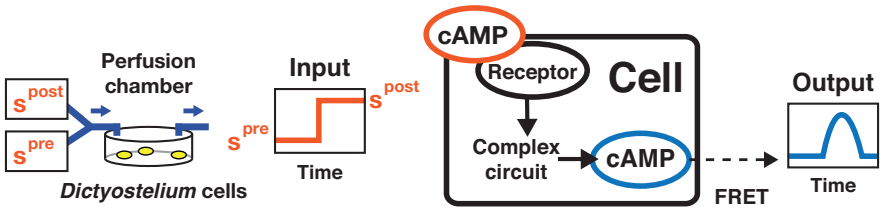


Figure 4.2: Schematic representation of the experimental setup to quantify the input-output characteristics of the cAMP relay system in *D.discoideum* cells.

Experimental setup

The extracellular cAMP concentration (input) is controlled temporally by using a perfusion chamber, while the intracellular cAMP concentration in live *Dictyostelium* cells (output) is measured by a fluorescence resonance energy transfer (FRET)-based sensor [97]. By using the experimental setup, we apply step stimulus of cAMP to each single cell. The step stimulus is defined by switching the input level with sufficiently high rate to make the output independent of the rate. In response to the input, the output exhibits adaptive dynamics, in other words, returns to the pre-stimulus level even when the input persists.

Data preparation

The responses to the step input with zero pre-stimulus level are characterized well in previous studies [97]. In addition to that, we obtain time-series data of the response to step inputs with various pre-stimulus levels. We found that the response dynamics still exhibits adaptation as long as the input level fall below some critical value ($<$ 100-1000 nM). On the other hand, the response dynamics becomes different from the zero pre-stimulus case as long as the pre-stimulus level is sufficiently high ($>$ 10-100 pM). This suggests that the limit of sensitivity of the cAMP signaling system is lower than 10-100 pM, and that noise in input level may be ignored with higher extracellular concentration of cAMP. Then we adopt a range of cAMP concentration (300 pM $\leq s \leq$ 10 nM) for estimation.

Sometimes the FRET signal from the microscope exhibits unnatural fluctuations because of out-of-focus effects or cell damage by the laser. To remove the fluctuations from data, we limited the observation period to 6.5 minutes, and set selection criteria for time-series data. That is, (i) FRET signal before a step input should be within $[F_{\mathrm{low}}, F_{\mathrm{high}}]$, and (ii) time average of FRET signal after four minutes of a step input should be within $[F_{\mathrm{low}}, F_{\mathrm{high}}]$. By adopting $F_{\mathrm{low}} = 1.35$, and $F_{\mathrm{high}} = 1.6$, we collected 18 time-series data for estimation. The examined pre- and post-stimulus input levels (s^{pre} and s^{post}) used in the subsequent learning are listed in Table 4.1, and Fig. 4.3 shows representative time courses of FRET signal.

(s^{pre},s^{post}) [nM]	Number of time series
(0.3, 1.5)	7
(1, 10)	5
(1.5, 7.5)	6

Table 4.1: Input values and corresponding numbers of time-series data.

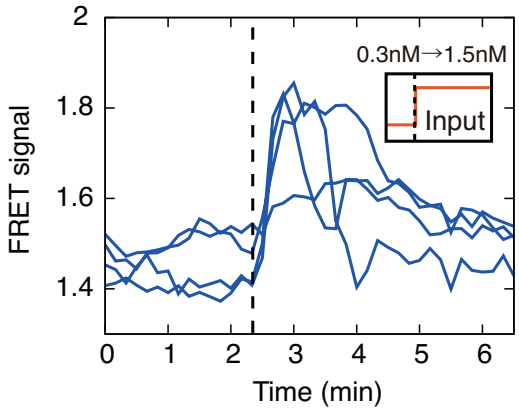


Figure 4.3: Representative time courses of FRET signal with a step input ($s^{\mathrm{pre}} = 0.3$ nM, $s^{\mathrm{post}} = 1.5$ nM, the flow rate in the perfusion chamber $\lambda = 32$ min^{-1}, and the switching time $t_{input} = 2.33$ min). Each line comes from a single-cell measurement.

4.3.2 Model description

State equation: Enzymatic regulatory network including models for FCD

The concept of FCD in cell signaling is motivated by examples of eukaryotic cell signaling systems that show fold-change responses; in the examples, an intermediate signaling component has various basal levels among different cells, but the responses to a stimulus is proportional to the basal levels at an equal rate among the cells [85,87]. Then, in order for the cell signaling to work equally among the cells, FCD is required in the downstream processes.

We give a brief review on model studies for FCD. First, an abstract feedforward network model, called incoherent type-1 feedforward loop (I1-FFL) as one of the network motifs in transcription networks, is shown to achieve FCD [107]. A recent study shows that a feedback type model, a specific kind of nonlinear integral feedback mode, also achieve FCD [88]. Furthermore, the study provides a general condition for dynamical systems that are sufficient for FCD, and also show that a model of bacterial chemotaxis exhibiting FCD-like behavior satisfy the condition approximately. The sufficient condition also seems to be necessary to achieve FCD perfectly, although it was shown that a model violating the condition can exhibit adaptive behaviors similar to FCD [108].

We define an artificial signaling network model that contains the models for FCD, that is, the I1-FFL, a nonlinear integral feedback system, and the approximate FCD model. Note that our model does not cover all of possible models for FCD, because there are models for FCD as many as we want, to

satisfy the condition for FCD.

To represent the artificial signaling network model, we employ an framework of enzymatic regulatory networks [82, 109]. The state equations in continuous-time form are written as

$$\dot{x_i} = A_i(\{x_j\}, s)\frac{(1-x_i)}{(1-x_i)+K_i^A} - I_i(\{x_j\})\frac{x_i}{x_i+K_i^I} + \text{noise} \tag{4.5}$$

where x_i $(i = 0, \ldots, D-1)$ means fraction of a cell signaling enzyme in the active form, K_i^A and K_i^I are Michaelis-Menten constants, A_i and I_i represent activation and inhibition on each component, respectively. The input level s is extracellular cAMP concentration in this context. The functions A_i and I_i are defined as

$$A_i(\{x_j\}, s) = (k_i + k_i^{\mathrm{AA}} x_i)\frac{J_i^{\mathrm{s}} + J_i^{\mathrm{PA}} + \sum_j J_{ij}^{\mathrm{PA}} x_j}{J_i^{\mathrm{SA}} + \sum_j J_{ij}^{\mathrm{SA}} x_j} \tag{4.6a}$$

$$I_i(\{x_j\}) = J_i^{\mathrm{PI}} + \sum_j J_{ij}^{\mathrm{PI}} x_j. \tag{4.6b}$$

The parameters that appear in the equations are defined in Table 4.2. Note that the model defined here is over-parametrized in order to avoid to be trapped at local maxima of an optimization function in the parameter space.

Parameter	Definition
J_{ij}^{PA}	Promotion of activation of x_i by x_j
J_{ij}^{SA}	Suppression of activation of x_i by x_j
J_{ij}^{PI}	Promotion of inhibition of x_i by x_j
$J_i^{\mathrm{PA,SA,PI}}$	Basal regulation of x_i
j_i^{s}	Activation of x_i by input s
k_i^{AA}	Autoactivation of x_i
k_i	Basal activation of x_i

Table 4.2: Definition of parameters in eqs. (4.6a) and (4.6b)

To obtain the state equations in discrete-time form, we first adopted the Euler-Maruyama scheme [54] and converted the equations as $\dot{x_i} = f_i(\{x_j\}) +$ noise $(i = 0, \ldots, D-1) \rightarrow x_i^{t+1} = x_i^t + \Delta t f(\{x_j^t\}) + \sigma_i \xi_i^t \sqrt{\Delta t}$ where t is each time point, and Δt is an integration time, and σ_i is intensity of system noise, and ξ_i^t is sampled from a standard normal distribution.

Observation equation

We define the x_0 as a output variable, and assume that the intracellular cAMP concentration x_{cAMP} is represented as

$$x_{\mathrm{cAMP}} = C_{\mathrm{cAMP}} x_0 + B_{\mathrm{cAMP}} \tag{4.7}$$

where C_{cAMP} and B_{cAMP} are constants to be estimated. The observation equation represents the FRET signal as a function of x_{cAMP}. It is known that the function takes a specific form [110] based on the binding kinetics between cAMP and the FRET sensor protein, *Epac1*, as

$$\text{FRET signal} = (R_{\max} - R_{\min}) \frac{(x_{\mathrm{cAMP}})^n}{(x_{\mathrm{cAMP}})^n + {\mathrm{EC}_{50}}^n} + \text{noise}. \tag{4.8}$$

The noise term is assumed to satisfy white Gaussian statistics, as noise $= \eta\phi^r$ where η is intensity of observation noise and ϕ^t is sampled from a standard normal distribution at each observed-time point r.

The parameters in eq. (4.8) are determined from experiments [98] and a literature [110] as $R_{\max} = 2$, $R_{\min} = 1.375$, $n = 0.74$, and $\mathrm{EC}_{50} = 2.5$ μM. When taken together, all parameters θ to be estimated are C_{cAMP} and B_{cAMP} in eq. (4.7), $\{K_i^A\}$ and $\{K_i^I\}$ in eq. (4.5), the parameters listed in Table 4.2, and intensity of the system and observation noises η. Note that all of the parameters are assumed to have non-negative values.

Input

As expected, time courses of extracelluler cAMP concentration s in the perfusion chamber is experimentally confirmed to follow

$$s(0) = s^{\mathrm{pre}}, \dot{s} = \begin{cases} s^{\mathrm{pre}} - \lambda s & (t < t_{\mathrm{input}}) \\ s^{\mathrm{post}} - \lambda s & (t \geq t_{\mathrm{input}}). \end{cases} \tag{4.9}$$

where flow rate $\lambda = 32$ min^{-1} is employed to obtain the data [98]. Then, we can calculate the input level at any time point based on this equation.

4.3.3 Estimation algorithm

The state space model defined above is optimized by a machine-learning approach. Statistical techniques for estimating stochastic nonlinear dynamical systems have been developed recently, and we can find some examples on time series data from cellular processes [33, 34, 37]. Although the nonlinear inclusions of the parameters in the Michaelis-Menten form make it difficult to apply such methods, a recent study gives a solution based on linearization of the Michaelis-Menten equations by explicit introduction of fast dynamics of the intermediate component [34].

Nevertheless, in this case, the model seems to contain too many parameters without available prior knowledge, to obtain reasonable estimation on the values from the FRET data. On the other hand, our aim is not to know a biochemical mechanism but to build a simple phenomenological model for the FCD behavior. Then, we introduce a penalized log-likelihood function to eliminate unnecessary parameters and maximize it by a simulated-annealing-like approach, as shown schematically in Fig. 4.4. Sparse estimation is also performed to estimate gene regulatory networks to reflect sparseness of such networks [111], while we use it to simplify the model for ease of comprehension. The results by the approach should not be regarded as definitive conclusions, but instead as a step towards rather comprehensive analysis with additional data.

The detailed procedure of the quasi simulated annealing (quasi-SA) is described and validated by using artificial data in Appendix B in the case of optimization of log likelihood. In Appendix B, it is demonstrated that the quasi-SA can estimate an adaptive system from partial observations of an adaptive variable only.

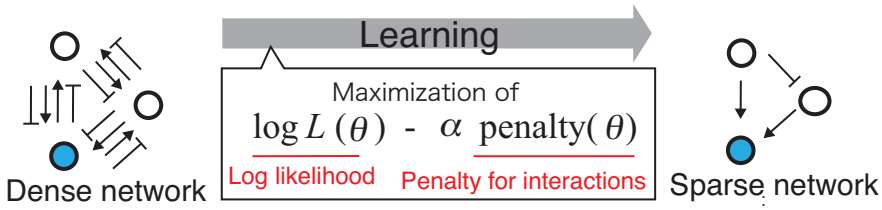


Figure 4.4: Schematic representation of the procedure of extraction of simple systems based on a penalized log likelihood approach.

Penalized log likelihood with L1 norm

Below, we perform maximum likelihood estimation of the enzymatic regulatory networks. After that, we extract a simple system from the optimized densely-connected network model by a penalized log likelihood approach as shown schematically in Fig. 4.4. In particular, we employ a $L1$ norm along with the method in Ch. 3, as

$$O(\theta) \equiv \log L(\theta) - \alpha \sum_{i=0,\ldots,D-1} \sum_{j \in J_i^{\mathrm{all}}} |\theta_j| \tag{4.10}$$

where J_i^{all} means a set of all parameters in the right hand side of the state equation governing dynamics of x_i (eq. 4.5), as

$$J_i^{\text{all}} = (\{J_{ij}^{PA,SA,PD}\}, J_i^{PA,SA,PD}, J_i^s, k_i^{\text{AA}}, k_i).$$

Note that we do not perform non-dimensionalization of each term in the penalty as in Ch.3, because the variables in the state equations are dimensionless quantities. That is, the state variables mean fractions of signaling components in the active forms, and the input is scaled by 1/10 nM so that the level of input s falls within $[0, 1]$ through the estimation. Through the sparse estimation, the values of $K_i^{A,I}$, η, and $\theta_j \in J_i^{\text{all}}$ are updated, while the other parameters are fixed.

First we tried to estimate both system and observation noises, but preliminary results indicate that the system noise is drastically overestimated, so that dynamics of the learned system is far from the observations. It is probably because of cell-to-cell variation in the cAMP signaling system. Such a variation has a long-time correlation compared with measurement time and then is estimated as noisy system dynamics rather than as noise in observation that has no time correlation. Although estimation techniques for models with static cell individuality have been studied recently [112], the number of time-series samples is insufficient to estimate the complex models. Then, we use the state equations without system noise ($\sigma_i = 0$) to avoid the problem.

4.3.4 Results

Reconstruction of input-output characteristics

First, maximum likelihood estimation is performed by the quasi-SA without the penalty ($\alpha = 0$). The enzymatic regulatory network with three components is employed, and 400 sets of initial parameter values are examined (see section B.3 for the detail settings in the quasi-SA algorithm). Figure 4.5(a) shows frequency distribution of log likelihood of the learned systems. Three peaks in the distribution are visible. Figures 4.5(b-e) show typical behaviors of the learned systems at the three peaks in comparison with the data. Note that all of the data used for estimation are presented in these figures. The simulated time courses of FRET signal by the lowest-log-likelihood model do not exhibit adaptation at all (purple lines), and the middle-log-likelihood model adapts only to a subset of the examined inputs (green lines). On the other hand, the highest-log-likelihood model adapts to all of the examined inputs (orange lines), and the peak and resting levels of FRET signal reproduces the data well.

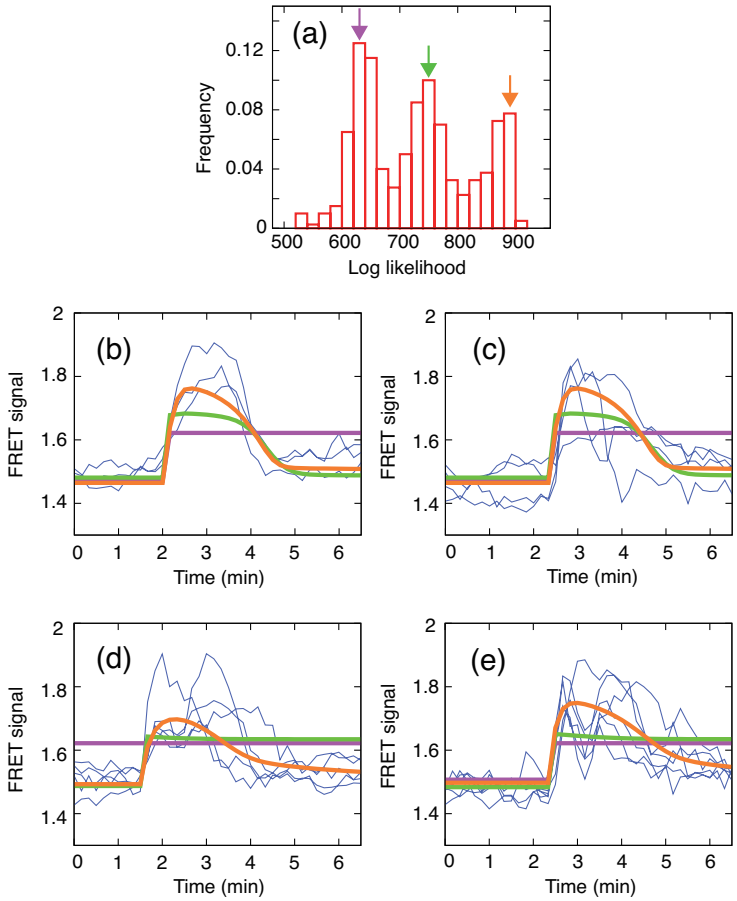


Figure 4.5: Maximum likelihood estimation by the quasi-SA without the penalty ($\alpha = 0$). (a) Frequency distribution of the log likelihood of the learned systems. (b,c,d,e) Time series of the FRET signal simulated by the learned systems at the peaks in the frequency distribution are presented in comparison with the data with $(s^{\mathrm{pre}}, s^{\mathrm{post}}, t_{\mathrm{input}}) = (0.3\mathrm{nM}, 1.5\mathrm{nM}, 2.0\mathrm{min})$ (b), $(0.3, 1.5, 2.34)$ (c), $(1.5, 7.5, 1.5)$ (d), and $(1, 10, 2.34)$ (e). Purple, green, and orange lines represent behaviors of the models at the peaks with the lowest, the middle, and the highest values of log likelihood, respectively, and blue lines represent the data.

Extraction of core architecture

Next, we extract a simple system from the optimized model network. To this end, we resume the optimization procedure from the model with maximum log likelihood (the ML model), at various values of penalty strength α. At each value of α, we perform 10 times of independent optimization trials and obtain different models because of stochasticity in the procedure. Figure 4.6 (a) shows values of log likelihood and the optimization function $O(\theta)$ of the models at each value of α. Decreases in log likelihood occurs twice in $\alpha = 0.1$-10000; the first occurs when α is increased from 10 to 100, and the second does when α is increased from 300 to 3000. At $\alpha \leq 10$, dynamics of the learned systems looks identical to that of the ML model, and at $30 \leq \alpha \leq 300$, it becomes slightly different from the ML model but is still quite similar to the observed dynamics. Figures 4.6 (b-e) present a comparison of dynamics between the models at $\alpha = 10$ and 300. On the other hand, dynamics at $1000 \leq \alpha$ is completely different from the data. These results are interpreted as follows. The learned systems in $\alpha < 10$ contains unnecessary parameters, and the parameters are eliminated at $\alpha = 10$. At $30 \leq \alpha \leq 300$, although some useful parameters are also eliminated, a core system required for reproducing the data remains. At $\alpha \geq 1000$, the core system breaks down, so that dynamics of the models does not reproduce the data at all.

Therefore, we expect to obtain a specific model for FCD by investigating the models at $\alpha = 300$. It is found that, for all of the models, a component is eliminated and the state equations are represented as

$$\dot{x_0} = k_0 \frac{s}{x_1} \frac{(1 - x_0)}{(1 - x_0) + K_0^A} - k_1 \frac{x_0}{x_0 + K_0^I} \tag{4.11a}$$

$$\dot{x_1} = \frac{k_2 x_0 + k_3 s}{k_4 + k_5 x_1} \frac{(1 - x_1)}{(1 - x_1) + K_1^A} - k_6 \frac{x_1}{x_1 + K_1^I}. \tag{4.11b}$$

At least, one of k_2 and k_5 values is zero in each model (typical parameter values of a model with $k_2 = 0$ and a model with $k_5 = 0$ are listed in Table 4.3). Beside, as shown in Fig. 4.7, removal of k_2 or k_5 results in small decrease in likelihood in comparison with that of other parameters k_i ($i = 0, 1, 3, 4, 6$). Correspondingly, removal of the both parameters does not change the dynamics qualitatively. Then we conclude that neither of the two are essential, and define a model with $k_2 = k_5 = 0$ as the core system of the ML model. The core system achieves adaptation by feed forward control through the inhibitor, x_1.

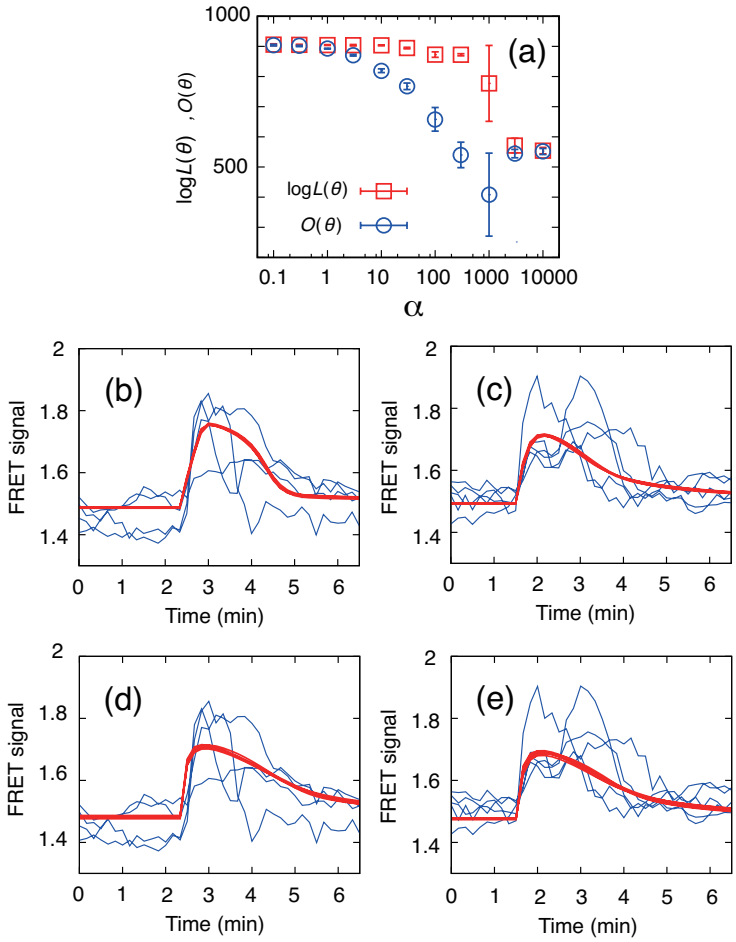


Figure 4.6: (a) The average values of Log likelihood and the $O(\theta)$ of the models optimized at each value of α are calculated by 10 independent trials. The error bars indicate the standard deviations. The values of log likelihood and the $O(\theta)$ are evaluated after 300 learning iterations from the ML model ($I_1 = 0$, $I_2 = 300$, the quantities are defined in Appendix B). The other settings in the quasi-SA algorithm are the same as in the no penalty case. (b,c,d,e) Time series of FRET signal simulated by the learned systems (red lines) and the data (blue lines) are presented at $\alpha = 10$ (b,c) and 300 (d,e), with $(s^{\rm pre}, s^{\rm post}, t_{\rm input}) = (0.3{\rm nM}, 1.5{\rm nM}, 2.34{\rm min})$ (b,d) and $(1.5, 7.5, 1.5)$ (c,e).

Parameters	Model 1	Model 2
k_0	4.75	4.99
k_1	2.34	3.24
k_2	0.0115	0
k_3	0.178	0.229
k_4	1	1
k_5	0	0.139
k_6	0.122	0.169
K_0^A	2.71	1.46
K_0^I	0.0269	0.0323
K_1^A	0.000149	9.87×10^{-7}
K_1^I	0.875	1.23
C_{cAMP}	6.27	6.27
B_{cAMP}	0.185	0.185
η	0.0729	0.0728

Table 4.3: Parameter values of typical models at $\alpha = 300$.

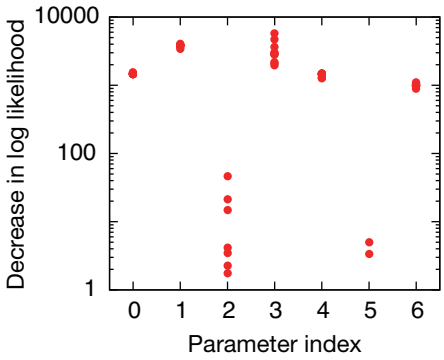


Figure 4.7: Decreases in log likelihood in the models at $\alpha = 300$ when a parameter k_i $(i = 0, \ldots, 6)$ is set to zero.

To understand dynamics of the core system, we attempt further simplification of the model equations. By considering that the Michaelis-Menten constant K_1^A (K_1^I) differ substantially from time average of the substrate quantity $\langle 1 - x_1 \rangle$ $(\langle x_1 \rangle)$, we approximate the state equation of x_1 by

$$\dot{x_1} = \frac{k_3 s}{k_4} - \frac{k_6}{K_1^I} x_1, \tag{4.12}$$

although the later assumption $K_1^I \gg \langle x_1 \rangle$ is less accurate with high level of the input. One can see that the system described by eqs. (4.11a) and (4.12)

has a symmetry property discussed in Section 3.4 as sufficient condition for fold-change detection (FCD), that is, eqs. (3.17a) and (3.17b). Besides, the simplified form corresponds to the I1-FFL model for perfect FCD. Then, we finally extract the I1-FFL as the core architecture that performs FCD in the observed cAMP dynamics.

Additionally, we confirme FCD property of the core system by checking the time-course behaviors. Figs. 4.6 (d) and (e) present a resemblance in dynamics of the models with the step inputs that show the same fold changes, $(s^{\mathrm{pre}}, s^{\mathrm{post}}) = (0.3, 1.5)$ and $(1.5, 7.5)$ (nM). Further, we calculate L2 norm of difference in the outputs in response to the five-fold changes in the inputs in order to compute deviation from the FCD property. Figure 4.8 shows the deviation for models at each α value. The deviation shows sharp decrease when α is increased from 10 to 30, and the model approximate a FCD system.

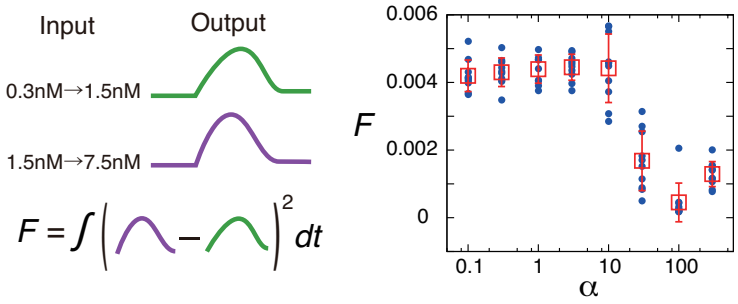


Figure 4.8: L2 norm of difference of the outputs in response to $(s^{\mathrm{pre}}, s^{\mathrm{post}}) = (0.3, 1.5)$ and $(1.5, 7.5)$ (nM), the five-fold changes in the inputs, is presented. The blue circles represent the L2 norms of the models at each α values, and the red symbols represent the mean and the standard deviation. The responses between t_{input} and $t_{input} + 3$ are used to calculate the L2 norms.

4.4 Discussion

Measurements at a single-cell level have characterized quantitatively the adaptive dynamics of the cAMP signaling in *D.discoideum* cells, and revealed FCD property in the system. However, the complexity in the signaling system and the abundance of possible models make it difficult to construct a reliable model for the single-cell dynamics. Towards the goal of explaining multicellular dynamics from the single-cell level, we examined two strategies, that is, qualitative and quantitative ones.

First, we attempted to draw conclusions only from the symmetry property for FCD. To this end, we define a set of equations for describing multicellular dynamics in which signaling response in each cell is FCD. By analyzing the equations, we found that FCD at the single-cell level achieves the

robustness to cell-density fluctuation.

Second, we constructed a phenomenological model to reproduce the input-output characteristics of the cAMP response dynamics, and compared it with the known models for FCD. In particular, the two-step method is adopted to obtain a complex model for quantitative reproduction of the data and a simple core system in the complex model. As a result, we obtained an enzymatic regulatory network model that reproduces the observed adaptive behaviors in the FRET time-series data. Furthermore, the I1-FFL model, one of the network motifs known to achieve FCD, is extracted as the core system.

A potential candidate for a molecular realization of the I1-FFL for FCD is the Ras pathway that is a signaling system in immediately downstream from the receptors. A recent study shows that the response of the pathway to cAMP stimuli is adaptive [113], which is prerequisite for FCD, and proposes that an incoherent feedforward control achieves the adaptation. Besides, quantitative characterization of the response in the study seems to indicate FCD property, although the point has not been discussed yet.

Nevertheless, these results should not be regarded as definitive conclusions because sophistication of the optimization method and the amount of data are insufficient for exhaustive analysis on the complicated model. In particular, the following point is critical and requires further investigation. The ML model is found to achieve adaptation by feed-forward inhibition through the hidden component, but we also extracted systems with feedback adaptation from the sub-optimal models. Since difference in log likelihood between the ML and other sub-optimal models is not large, it is difficult to undoubtedly conclude that the observed adaptation is based on feed-forward control. Our results support the hypothesis that the cAMP relay shows FCD, but determination of the detailed mechanism for FCD remains elusive.

In summary, we successfully related FCD property at the single-cell level with the robustness in the multicellular dynamics, and obtained a phenomenological model to reproduce the input-output characteristics of the cAMP relay system. In the next chapter, the quantitative model will be evaluated in the predictive performance for the multicellular dynamics.

Chapter 5

Multicellular dynamics of a social amoeba

5.1 Introduction

In Chapter 4, we obtained an enzymatic regulatory network model for cAMP relay system in *D.discoideum* cells by optimization based on the FRET time-series data, and the learned results reconfirmed FCD property in the input-output characteristics. In this chapter, we evaluate explanation capability of our knowledge at a single-cell level for multicellular dynamics of *D.discoideum* cells. The evaluation concerns the following two points. The first point is whether a system with FCD exhibits qualitative features of experimentally-observed collective dynamics like synchronized oscillation. In fact, although multicellular systems consisted of FCD cells show robust behavior against cell-density fluctuations, it is not yet confirmed that there is a FCD system actually showing oscillatory and excitatory behaviors. The second point is whether the learned network model reproduces quantitative behaviors of *D.discoideum*. The two points are examined below.

5.2 Toy models with perfect FCD

Here we simply demonstrate that several toy models with FCD reproduce collective dynamics of *D.discoideum* cells qualitatively.

5.2.1 FCD oscillator

As a most simple collective dynamics in the cells, first we consider synchronized oscillation in spatially homogeneous cases. Such an oscillation is experimentally observed in cAMP concentration in a cell suspension [114], and has been investigated for decades.

Let us consider dynamics of concentrations of an intermediate signaling component, intra- and extra-cellular cAMP, denoted as x, y, and z, respectively. A toy model with FCD is introduced as follows,

$$\dot{x} = axy - bx \tag{5.1a}$$

$$\dot{y} = c\frac{z^n}{z^n + x^n} - dy \tag{5.1b}$$

$$\dot{z} = \rho ey - \lambda z \tag{5.1c}$$

where $a, \ldots, e$ are rate constants, and the other parameters ρ and λ are cell density and dilution rate, respectively. The Michaelis-Menten form of the activation function of y gives threshold-like behavior and amplification in the cAMP signaling pathway. When nonlinear amplification ($n > 1$) occurs as is observed [115], the toy model certainly exhibits oscillation due to the positive feedback through the extracellular cAMP, as shown in Fig. 5.1. Then it is shown that perfect FCD system can exhibit an oscillatory behavior.

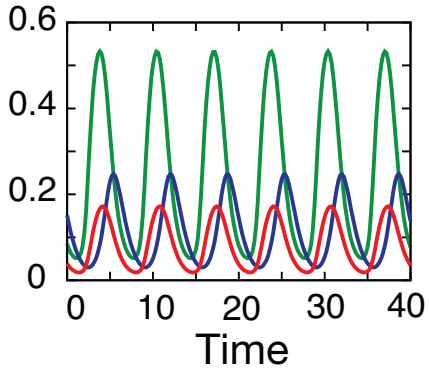


Figure 5.1: Oscillatory time series generated by eqs.(5.1a), (5.1b), and (5.1c) of x (blue), y (green), and z (red). The parameter values are set as $a = 4$, $b = 1$, $c = 1$, $d = 1.5$, $\rho e = 1$, $\lambda = 3$, $n = 2$

Next, the spiral pattern of the cAMP wave is investigated. To this end, we add diffusion of extracellular cAMP z to the model (Fig. 5.2) while assuming homogeneous cell density, as

$$\frac{\partial z}{\partial t} = \rho ey - \lambda z + D\Delta z \tag{5.2}$$

where D the is the diffusion constant. By numerical simulations of the model in one and two dimensional spaces, we found the model to exhibit spatio-temporal chaos as shown in Fig. 5.3. In the chaotic dynamics, the spiral cores are not maintained long time enough to allow for cells to aggregate into the cores.

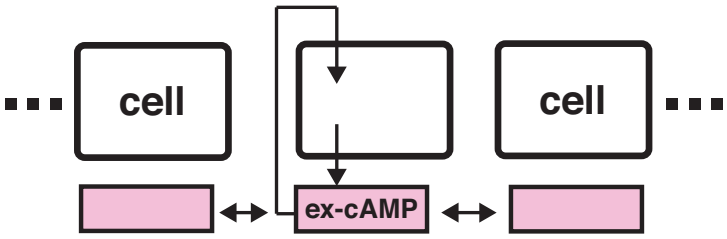


Figure 5.2: Schematic representation of a model of multicelluler dynamics as a reaction-diffusion system.

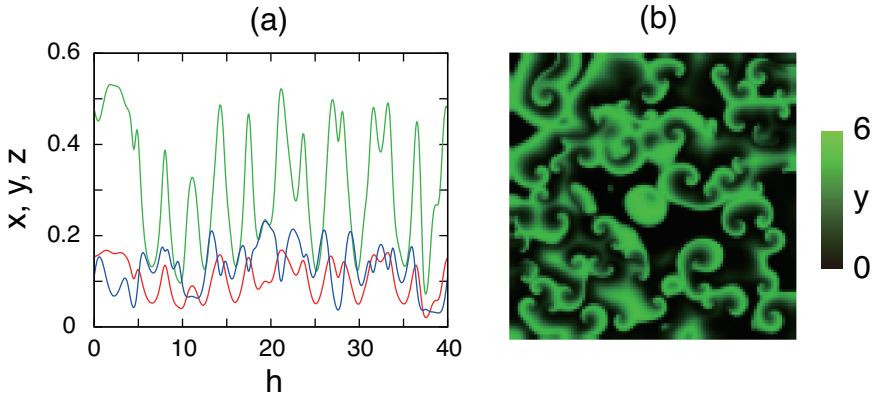


Figure 5.3: Chaotic dynamics generated by eqs.(5.1a), (5.1b), and (5.2) of x (blue), y (green), and z (red) on (a) one and (b) two dimensional space. h is the space variable. The diffusion constant is $D = 0.048$, and the other parameters are set as in Fig. 5.1

5.2.2 FCD excitor

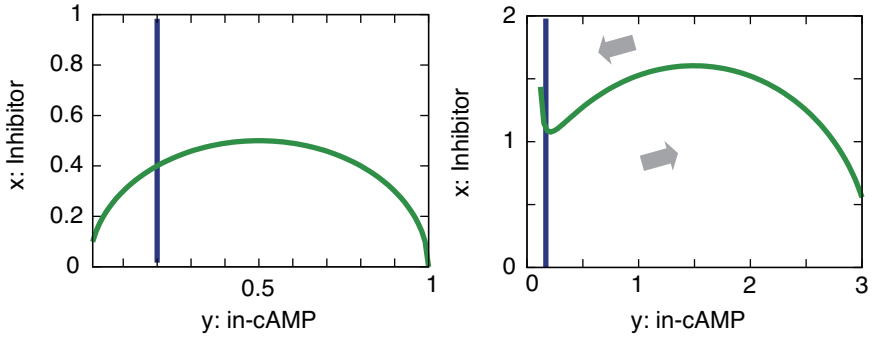
The chaos in cAMP pattern can be avoided by a simple and reasonable modification of the previous model, the FCD oscillator. Here we add a basal synthesis rate δ of intracellular cAMP as

$$\dot{y} = c\frac{z^n}{z^n + x^n} + \delta - dy. \tag{5.3}$$

Figure 5.4 show how the added term changes the nullclines of model. The modified nullclines give a characteristic feature of excitable system as in the FitzHugh-Nagumo equations [13, 14]. As a result, excitation waves propagate stably in the modified model, the FCD excitor (Fig. 5.5(a)). However, the obtained excitation waves seems to have too narrow cAMP pulses compared with the refractory period. To improve this, we also modify the dynamics of x slightly without affecting the nullclines, as

$$\dot{x} = ax - b\frac{x}{y}. \tag{5.4}$$

This modification drastically broadens the width of cAMP pulses (Fig. 5.5(b)).



(a) (b)

Figure 5.4: It is presented how nullclines of x (blue) and y (green) are changed from (a) the FCD oscillator to (b) the FCD excitor. In both cases, the dynamics of z is adiabatically eliminated. The arrows indicate flow direction. The parameter values in (a) are set as $a = 5$, $b = 1$, $c = 1$, $d = 1$, $\rho e = 1$, $\lambda = 1$, $n = 2$. The values are also employed in (b) except for $c = 3$ and $\delta = 0.1$.

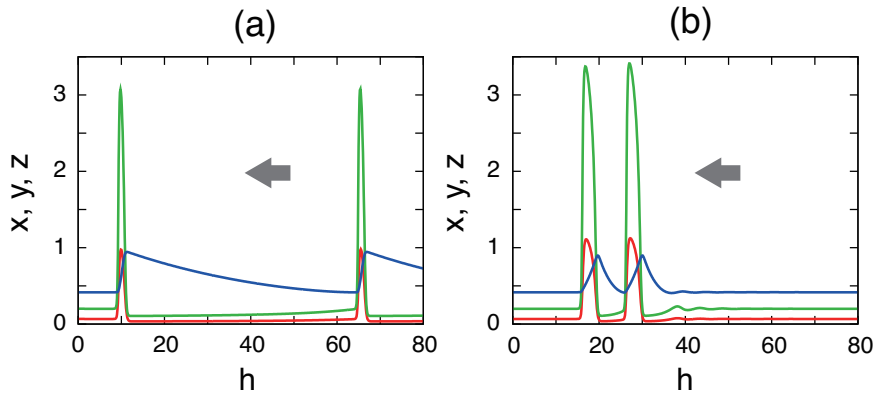


Figure 5.5: Excitation waves generated from (a) the FCD excitor and (b) the modified FCD excitor with eq. (5.4) on one dimensional space. In each case, two waves with minimal spacing defined by the refractory period are presented. x, y, and z are represented by blue, green, and red, respectively. The arrows indicate the propagation directions of the waves. The parameter values are set as $a = 0.05$, $b = 0.01$, $c = 12$, $d = 3$, $n = 2$, $\rho e = 1$, $\lambda = 3$, $D = 0.048$, and $\delta = 0.3$.

5.2.3 FCD chemotaxis

Finally we focus on chemotactic aggregation guided by the cAMP wave in early phase of the asexual development. To construct a toy model for the phenomenon, we take a coarse-grained approach with inclusion of spatial change of the cell density as

$$\frac{\partial \rho}{\partial t} = D_\rho \Delta \rho - \nabla(\rho C(x, y, z)) \tag{5.5}$$

where the diffusion and advection terms represent random and directional motions of the cells, respectively [116,117]. The chemotaxis function $C(x, y, z)$ means velocity of the cells depending on cell states and extracellular cAMP. Then the first problem is to determine the function.

As a simplest approach, the cell velocity is assumed to be proportional to cAMP gradient, as $C(x, y, z) = C(z) \propto \nabla z$. This seems to be reasonable, especially with stationary extracelluler chemical gradients. However, with incoming pulses of a signal, the simple model generates backward motion of the cells in the wavebacks, and results in ineffective chemotaxis. To solve this paradox, Hofer and co-workers consider a model in which extracellular signal inhibits chemotaxis through an intracellular signaling component with a delay [117]. In the Hofer model, chemotactic movement in the waveback stops by the delayed inhibition, and then, net cell movement occurs towards incoming waves. Besides, their model explains cell-stream formation during

the cell aggregation due to instability of the density field under periodic perturbation by the cAMP pulses.

However in the Hofer model, both the chemotaxis and the intracellular signaling (the M&G model [16]) are far from FCD, and robustness to cell-density fluctuation is not assured. Then, we introduce a toy model of chemotaxis with FCD as

$$C(p,a) = C\frac{|p|^{m-1}}{|p|^m + w^m}p \tag{5.6}$$

and

$$\tau_w \frac{\partial w}{\partial t} = a_w z - b_w w \tag{5.7a}$$

$$\tau_p \frac{\partial p}{\partial t} = \nabla z - p. \tag{5.7b}$$

where w and p are a chemotactic inhibitor and a cell polarity, and a_w, b_w, and C are constant parameters. Since fast polarity formation in response to cAMP simulation (in seconds) is observed in the chemotaxis system in *D.discoideum* cells, eq. (5.7b) is adiabatically elimanted as $p = z$. By combining these equations with the FCD excitor model (eqs. (5.4), (5.3), and (5.2)), we can model the cell aggregation in the cAMP wave (note that advection terms in the chemical kinetics are also considered). By assuming the time scale of the chemotactic inhibitor to be comparable with that of cAMP kinetics, we actually observe excitation wave with cell movement towards proper direction in the total model as shown in Fig. 5.6. The wave dynamics is invariant under cell-density fluctuation, which is explained by the fact that variable transformations $(z, x, w) = (\rho Z, \rho X, \rho W)$ remove the cell density parameter ρ from the total model.

5.2.4 Discussion

Above, it is shown that the perfect FCD systems exhibit qualitative features of collective dynamics of *D.discoideum*, that is, synchronized oscillation, propagation of excitation wave, and cell movement with the excitation wave. These results support the hypothesis of robust collective dynamics based on FCD. In the bottom part of the investigation, FCD chemotaxis model is introduced. We found no experimental support for this assumption, partly because experiments to confirm FCD in chemotaxis require extremely precise control of extracellular chemical gradient. We expect to elucidate such property in the chemotaxis of *D.discoideum* in future.

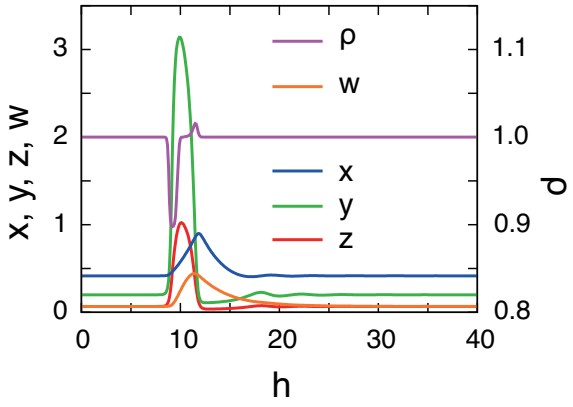


Figure 5.6: Excitation wave generated from the total model with homogeneous initial cell density $\rho(t = 0) = 1$ on one dimensional space. The parameter values in the chemotaxis model are set as $m = 3$, $C = 0.2$, $a_w = 0.05$, and $b_w = 0.05$. The other parameters are set as in Fig. 5.5.

5.3 Learned model with approximate FCD

In contrast qualitative investigation of the toy models for FCD examined above, we attempt quantitative prediction based on the learned results in Chapter 4. To this end, the model at penalty strength $\alpha = 10$ is employed since the model is expected to show the best predictive performance among the models we estimated. The model equations are written as

$$\frac{\partial x_0}{\partial t} = (k_0 + k_1 x_0)\frac{k_2 x_0 + k_3 s}{k_4 + k_5 x_1}\frac{(1 - x_0)}{(1 - x_0) + K_0^A} - (k_6 + k_7 x_0 + k_8 x_1)\frac{x_0}{x_0 + K_0^I} \tag{5.8a}$$

$$\frac{\partial x_1}{\partial t} = (k_9 + k_{10} x_1)(k_{11} x_0 + k_{12} s)\frac{(1 - x_1)}{(1 - x_1) + K_1^A} - (k_{13} + k_{14} x_1)\frac{x_1}{x_1 + K_1^I}, \tag{5.8b}$$

and intracellular cAMP concentration is estimated by

$$x_{\mathrm{cAMP}} = C_{\mathrm{cAMP}} x_0 + B_{\mathrm{cAMP}}\,[\mu\mathrm{M}]. \tag{5.9}$$

The parameter values are listed in Table 5.1. By adding dynamics of extracellular cAMP z [μM] to the learned enzymatic regulatory network model, a reaction-diffusion model is constructed as

$$\frac{\partial z}{\partial t} = k_t r(x_{\mathrm{cAMP}}) - \lambda z + D\Delta z \tag{5.10}$$

where k_t, r, λ, and D are the secretion rate, the ratio of intracellular to extracellular volumes, the degradation rate, and the diffusion constant. The values of the secretion and degradation rates of extracellular cAMP are determined as in the M&G model in which the values are from experimental works [16], and the diffusion constant is from an experiment [118], as $k_t = 0.9$ min^{-1}, $k_e = 5.4$ min^{-1}, and $D = 0.024$ mm^2 min^{-1}.

Parameters	Values	Parameters	Values
k_0	6.9	k_{11}	1
k_1	52.9	k_{12}	36.2
k_2	1	k_{13}	0.204
k_3	199	k_{14}	0.0443
k_4	1	K_0^A	2.12
k_5	25.5	K_0^I	0.217
k_6	2.88	K_1^A	0.23
k_7	6.74	K_1^I	2.21
k_8	1.15	C_{cAMP}	6.27
k_9	0.098	B_{cAMP}	0.185
k_{10}	0.145	η	0.0691

Table 5.1: Parameter values of the model at $\alpha = 10$.

By applying external perturbation to the reaction-diffusion model, we observed excitation waves and spiral-pattern formation as shown Fig.5.7. That is, the learned single-cell systems in the model can be excited by inputs from themselves, even though the learning is based only on the response dynamics to the step inputs. The self-consistency suggests that the learned system successfully generalize the input-output relation from the limited information.

Figure 5.7 (a) present the region of the parameter r where the system shows excitation wave (the shaded area), and corresponding extracellular cAMP concentration z at the resting state (the red line). The z value in this region of excitation behavior agrees with the pre-stimulus input value in the learning, that is, [0.03,0.15] (10 nM). Therefore, the observed spiral pattern is certainly a predicted feature from the data through the learning. However, as shown in Fig. 5.7 (b, c) the predicted wavelength and of spirals is larger than the experimental observation $\sim$ 2-3mm. Correspondingly, the predicted period of cAMP oscillation at each point is also overestimated, in comparison with the experimental value $\sim$ 5-6 min, as shown in Fig. 5.7 (d, e).

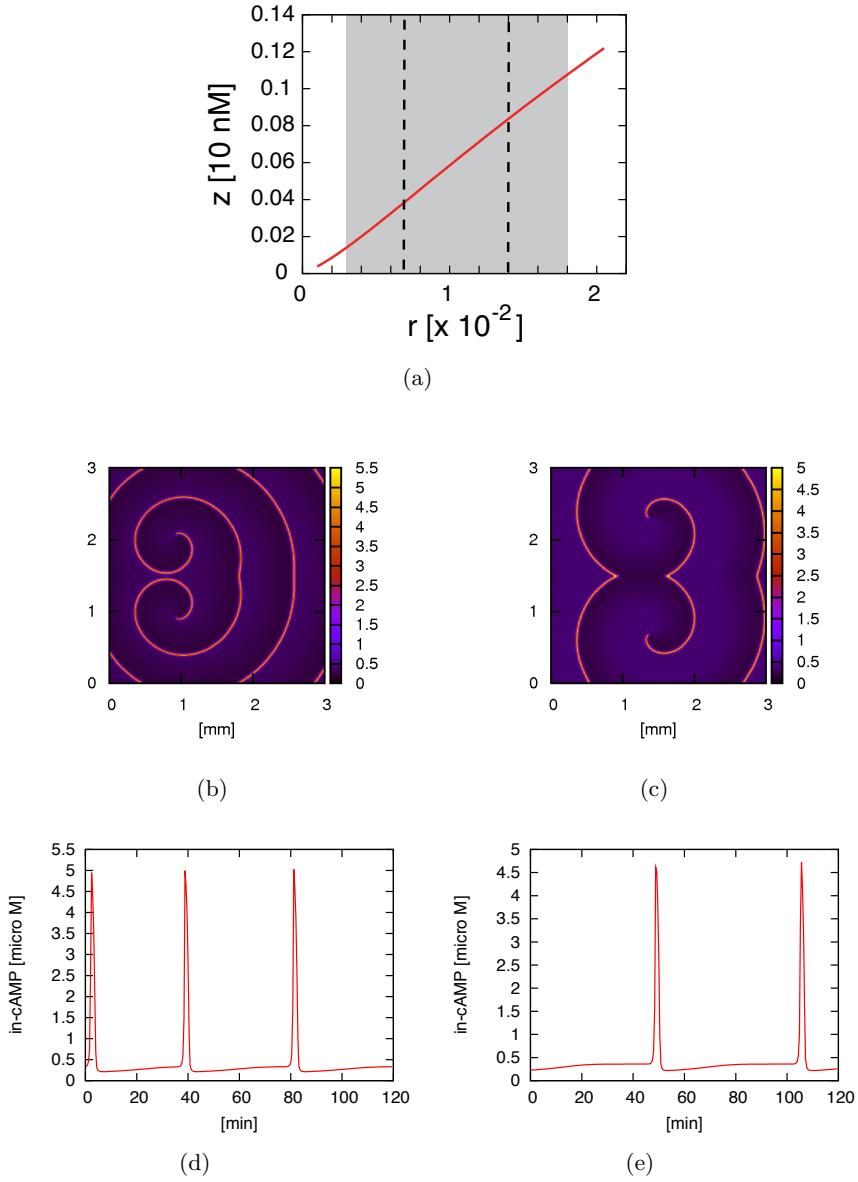


Figure 5.7: Predicted spiral patterns from the reaction-diffusion model. (a) The region in which spiral-pattern formation occurs on the r axes (the shaded area) and corresponding extracellular cAMP concentration z (red line) is presented. The broken lines indicate the values of r examined in (b-e). (b,c) Spiral patterns evoked by external perturbation are shown at $r = 0.007$ (b) and 0.014 (c). (d,e) Temporal variations of intracellular cAMP concentration are shown at $r = 0.007$ (d) and 0.014 (e).

5.3.1 Discussion

The learned system at the single-cell level reproduces spiral-pattern formation as is observed; however, the quantitative features of the pattern largely deviate from the observed ones, even though they stay in the same order. This is not surprising because, as shown in the previous section 5.2 (see Fig. 5.5), properties of excitation wave strongly depend on dynamics of inhibitor in the system, especially on the de-adaptation kinetics, which is unobservable in this case. Then, to obtain a better prediction, we require data on the cells to other inputs revealing the hidden dynamics. For example, sequential double step inputs may work effectively, because the response to the second input reflects transient state of the inhibitor depending on the interval between the inputs.

A reason why a quantitative model is important, in addition to the qualitative models with perfect FCD, is that a deviation from cell-density invariant behavior is observed in the multicellular pattern formation. The deviation is in size of the spiral waves that define aggregation territory of the cells, and seems to achieve another robustness in the asexual development. That is, the average size of the spirals negatively correlates with cell density, and then, the average number of the cells in cell aggregates is stabilized because the number is determined by the product of cell density and size of aggregation territory. Identifying origin of the deviation remains a future task.

5.4 Summary

We investigated relation of our knowledge at a single-cell level to the observed collective dynamics of *Dictyostelium* cells, by constructing multicellular models from the single-cell models. The explanation capability was evaluated from the two viewpoints, qualitative features of the toy models with perfect FCD, and quantitative features of the learned model. We found that, based on the toy models, FCD systems can reproduce key characteristics of the collective dynamics in an early phase of the asexual development. Besides, the learned model successfully reproduces an observed spatiotemporal pattern formation, though the learning algorithm and the data collection have room for improvement to increase the quantitative accuracy.

Chapter 6

Summary and Outlook

6.1 Summary

In this thesis, we studied a combined approach of statistical machine-learning and dynamical-systems analysis, aiming to extract basic characteristics of cell signaling systems that underlie various biological phenomena. Dynamical systems are employed to model cell signaling dynamics over decades. In particular, simplified description of a system by low-dimensional model leads clear explanation of the dynamics. Because of degeneracy in models to reproduce a dynamical behavior, the explanation by a simple model needs to be justified by knowledge on the biochemical mechanism. The commonly-used strategy is to reduce a complicated biochemical model to a simple one by, for example, adiabatic approximation, phase reduction method, and so forth. Nevertheless, incompleteness in data on cell signaling dynamics makes a reliable description on the biochemical mechanisms difficult. To overcome the difficulty by utilizing quantitative information in data, we propose to employ statistical techniques to estimate such simple models directly, not through mechanistic descriptions.

In Chapter 2, we have started the study by employing statistical techniques to identify bifurcation types that characterize qualitative changes in dynamics. Identification of bifurcation types has been commonly performed by rigorous, but non-statistical ways that are impractical for cell signaling dynamics because it requires unrealistically accurate measurements and precise control of systems. Our strategy drastically relaxes the strict conditions on data accuracy and experimental design. Of course, we cannot assume a priori that identifying bifurcation types is fruitful for understanding cell dynamics. Still, the successful identification of bifurcation types indicates that topological features of flow in the phase space can be captured by the estimation procedure, and then we expect the approach to be useful to extract other biologically-motivated characteristics.

In Chapter 3, we examined sparse estimation of polynomial dynamical

systems for highly nonlinear dynamics. The setting mimics a practical case where many possible terms exist in model equations for biological dynamics. For chaotic dynamics and adaptive dynamics with fold-change detection (FCD) property, the estimation procedure eliminated unnecessary terms. In addition, the results suggest a possibility to detect symmetry properties in dynamics. Further studies on this topic are highly demanded. Characterizing symmetry property and its breaking is of great importance for understanding sensory and developmental dynamics, and then further studies on this topic are highly demanded.

In Chapter 4, input-output characteristics of a cell-cell communication system in a social amoeba is investigated, with the aim to predict or explain the multicellular dynamics based on a single-cell level. Experimental studies have revealed FCD-like behaviors of the system, but determining a model at a single-cell dynamics is difficult because of the complicated biochemical mechanism and the abundance of possible models that exhibit FCD. We tackled this problem by two complementary approaches. First, we focused on consequences based only on the symmetry property that are common in the known models for FCD. The symmetry property is shown to allow robust multicellular dynamics against cell-density fluctuations. The hypothetical mechanism is similar with developmental robustness based on scaling in morphogen gradient against body-size fluctuations. We infer that our mechanism is ubiquitous. Second, we tried statistical reconstruction of the input-output characteristics in cell signaling dynamics at a single-cell level based on time-series data by the FRET measurements. The model to be estimated is designed to contain several well-known models for FCD, and we hope to select a model by the algorithm-based model reduction. Although the symmetry property is reconfirmed, the model selection so far appears to be difficult in this setting, unfortunately

In Chapter 5, we discussed on FCD models at a single-cell level and the consequences for the multicellular dynamics. FCD property and excitability are shown to be simultaneously achievable, which is prerequisite for reproducing the collective multicellular dynamics. We also introduced a model of chemotaxis with the robustness to cell-density fluctuation that is required for modeling the cell aggregation process in the asexual development of *Dictyostelium* cells. In addition, we examined quantitative predictions for the multicellular dynamics based on the estimated input-output characteristics in Chapter 4. We found large errors in prediction of spatiotemporal patterns. The results seemed to imply a need of data on dynamics of hidden components that are not revealed by the setting we adopted. The problem may be solved by experimental design techniques that are recently developed to determine optimal time-varying input for model estimation on a statistical ground.

In summary, we examined a few examples of estimation of simple dynamical systems from time-series data on artificial or experimental cell signal-

ing dynamics. Measurement technologies for cellular dynamics are rapidly evolving field. For example, two-photon microscopy enables imaging of living tissues up to an unprecedented depth. The measurements will not only contribute to improvements of known biochemically mechanistic models, but also reveal biological phenomena whose mechanisms are unknown. We hope our strategy to be an effective way to understand cell signaling dynamics underlying the newly-discovered phenomena.

6.2 Outlook

Spatial regulation on cell signaling dynamics

So far we did not estimate spatial aspects of cell signaling dynamics. However, the spatial regulations on cell signaling dynamics, at both intra- and inter-cellular levels, are known to be essential for proper functioning [1]. Data on spatiotemporal dynamics are inevitably complicated and high dimensional, and then machine-learning techniques are useful.

In addition, such techniques are also useful to deal with complicacy in models, which frequently arises by considering spatial dimension. For example, developmental biology has made a great progress to elucidate pattern formations of morphogens that are signaling chemicals providing positional information to cells in an organism [119]. In addition to the chemical signalings, "physical" quantities such as force and velocity cannot be ignored to understand developmental processes in which the body undergoes sequential deformations. In fact, recent studies indicate strong and mutual couplings between the chemical signalings and the mechanics [120,121,122]. Therefore, several attempts have been made to model the mechano-chemical dynamics, but the model complexity easily goes beyond our intuitive controls even for a phenomenological model such as cell vertex model [123]. Statistical machine-learning techniques may provide an objective way to choose parameter values and to select a better model.

Cell individuality

Another aspect we have missed is cell individuality in signaling dynamics. Although traditional models in systems biology represent signaling systems on an "average" or "typical" cell, heterogeneity within cell population is increasingly recognized. The heterogeneity is often static and is not due to temporal noise. Hence, it is not represented by the noise term in Langevin equation. In contrast, in dynamical systems modeling, the heterogeneity is represented by distributions of parameter values. Estimation techniques for static distributions of parameter values in a population have been extensively studied, for example, in ecological statistics [124]. Also, similar techniques have been developed for dynamical systems recently, and applied

to cellular dynamics [112]. At present, reliability of results by the techniques is limited because of data insufficiency. Nevertheless, since the heterogeneity is ubiquitous and biologically significant, the techniques for estimating cell individuality will be more important as more data are accumulated.

Chatacterizing biological information

In parallel with further applications, cooperations with other theoretical frameworks will be fruitful. In particular, information theory will be useful to characterize how information processing occurs through cell signaling dynamics [125]. Information-theoretic approach based on principle of entropy maximization has successfully applied to input-output characteristics of sensory neurons [126]. Furthermore, a recent attempt to apply such an approach to developmental signaling in *Drosophila* has demonstrated that a transcriptional regulation seems to achieve nearly maximal mutual information between transcriptional factor and gene expression [127]. In addition to the studies focusing on input-output characteristics, analysis of biological networks is needed to elucidate cellular dynamics, which, however, is rare except for a few recent theoretical works [128, 129]. It is partly because characterizations of probability distributions of time series from experiments are technically difficult for such dynamical systems. Statistical reconstructions of cell signaling dynamics are requested to create a bridge between experiments and theories in the field.

So far we discussed a few future directions on the approach of statistical estimation of dynamical systems in cell biology. However, application of the approach is not limited to those, but will be essential to answer various biologcal questions.

Appendix A

EM algorithm with particle smoother

A.1 State space model

We first introduce the state space model composed of the state equations and observation equations describing the system dynamics and observation process, respectively. Let us consider D-dimensional stochastic differential equations that describe a system, and d components in the system that are observed simultaneously. In the model, the state variable $x_i (i = 1, \ldots, D)$ evolves under the function $f_i(\{x_j\}, s)$, where s represents an input to the system, and the observation value y_i is obtained through the function $g_i(\{x_j\})$. The input s is considered as a bifurcation parameter in Chapter 2.

By discretizing the dynamics in time with the Euler-Maruyama scheme [54], we can write the space state model as

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

$$y_i^r = g_i(\{x_j^r\}) + \eta_i \phi_i^r, \tag{A.2}$$

where $t(\in T)$ and $r(\in R)$ are time points, Δt is an integration time, and σ_i and η_i are the noise intensities in the dynamics and observation, respectively. Both ξ_i^t and ϕ_i^r are sampled from a standard normal distribution. In general, the set of observed time points R is a subset of the entire time point set T (i.e., $R \subseteq T$) for the numerical integration. We assume that the function f_i can be expressed by the summation of linearly independent functions

$(f_i^n (n = 1, \ldots, N_i))$ as

$$f_i(\{x_j^t\}, s) = \sum_{n=1}^{N_i} k_i^n f_i^n(\{x_j^t\}, s). \tag{A.3}$$

Here, $\{k_i^n\}$ are the coefficients to be estimated.

Let us consider a data set of $Y = \{Y_a\}(a = 1, \ldots, A)$ where each Y_a is a time-series data obtained from independent measurement at input value s_a. The learning procedure estimates the parameters $\{k_i^n\}$, $\{\sigma_i\}$, $\{\eta_i\}$, and all the true states $\{x_i^t\}$ for each time-series data. In our method, the initial condition for the ith component in the ath time-series data is assumed to obey a Gaussian distribution parameterized by the mean $\mu_{i,a}$ and the variance $V_{i,a}$. Distributions at other points are automatically estimated by the particle filter algorithm explained below. Then, the parameters to be estimated are $\theta = (\{k_i^n\}, \{\sigma_i\}, \{\eta_i\}, \{\mu_{i,a}\}, \{V_{i,a}\})$. On the other hand, input values $S = \{s_a\}$ are assumed to be known.

A.2 SAEM algorithm

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

$$\log L(\theta) = \log p(Y|\theta, S) = \log \int_X p(Y|X, \theta) p(X|\theta, S). \tag{A.4}$$

Here, X $(= \{X_a\})$ denotes the entire time series of estimated states. We employ an expectation-maximization (EM) algorithm that maximizes $\log P(X, Y|\theta, S)$ (the complete-data log likelihood function), which is equivalent to maximizing the likelihood in eq. (A.4) [49]. By iterating two steps known as the E and M steps, the states X and parameters θ are estimated alternately. Since our implement of the E step includes the Monte-Carlo method as described below, stochastic approximation EM (SAEM) algorithm is adopted [53]. The SAEM procedure is described as follows.

SAEM

1. Initialize the parameter vector $\theta = \theta_0$, and set the iteration number I to zero.

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

3. (M step) Rename θ as $\theta_{\rm old}$, and update the parameter vector as

$$\theta_{\rm new} = \underset{\theta}{\rm argmax}\, Q_I(\theta) \tag{A.5}$$

$$Q_I = \begin{cases} Q(\theta, \theta_{\rm old}) & (I = 0), \\ (1 - \alpha_I) Q_{I-1}(\theta) + \alpha_I Q(\theta, \theta_{\rm old}) & (I > 0), \end{cases} \tag{A.6}$$

where

$$Q(\theta, \theta_{\rm old}) = \langle \log p(X, Y|\theta, S) \rangle_{p(X|Y, \theta_{\rm old}, S)}, \tag{A.7}$$

and $\{\alpha_I\}$ is a non-increasing sequence of positive values converging to zero.

4. Increment I by one, and iterate steps 2 and 3 until the estimation of the parameter vector converges.

The details of the E and M steps are described in the following sections.

A.2.1 E step: particle smoother

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

$$\log p(X|Y, \theta, S) = \sum_a \log p(X_a|Y_a, \theta, s_a). \tag{A.8}$$

Then, each $\log p(X_a|Y_a, \theta, s_a)$ is evaluated by using a particle filter algorithm that approximates the non-Gaussian distribution of the state x_i^t as a collection of many particles, each of which represents a sample from the distribution [50, 51]. Specifically, the algorithm required here is called a particle smoother. For the ath time series, let $x_{i,a}^{t,p}$ denote the pth particle for representing x_i^t, and let $y_{i,a}^t$ denote an observed value at time t. The procedure of the particle smoother is described as follows.

Particle smoother

For $a = 1, \ldots, A$

1. 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 $w_a^p = 1/P$

2. At each $t = 0, 1, \ldots$

   - 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.
   - If $t = r \in R$, update the weights of the particles as

$$w_a^p = \frac{w_{a,\mathrm{old}}^p l_a^{r,p}}{\sum_p w_{a,\mathrm{old}}^p l_a^{r,p}}, \tag{A.9}$$

     where

$$l_a^{r,p} = \prod_i^d p(y_{i,a}^r | \{x_{j,a}^{r,p}\}). \tag{A.10}$$

   - If $P_{eff} = 1/\sum_p (w_a^p)^2 < P_{thres}$ (i.e., if effective number of the particles fall 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.

3. Finish when all data points have passed ($t = \max(R)$), and estimate the log likelihood as

$$\log L_a(\theta) = \sum_{r \in R} \log(\frac{1}{P} \sum_p^P l_a^{r,p}). \tag{A.11}$$

Then, the posterior distribution of the time series of the state is approximated as

$$p(X_a | Y_a, \theta_{\mathrm{old}}, s_a) = \sum_{p=1}^P w_a^p \delta(X_a - X_a^p), \tag{A.12}$$

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

$$
\begin{aligned}
Q(\theta,\theta_{\rm old}) &= \langle \log p(\{X_a\},\{Y_a\}|\theta,\{s_a\})\rangle_{p(\{X_a\}|\{Y_a\},\theta_{\rm old},\{s_a\})} \\
&= \sum_{a,p} w_a^p \log p(X_a^p, Y_a|\theta, s_a) \\
&= \sum_{a,p,i} w_a^p \left(-\frac{1}{2}\log 2\pi V_{i,a} - \frac{(x_{i,a}^{0,p}-\mu_{i,a})^2}{2V_{i,a}} \right) \\
&+ \sum_{a,p,t,i} w_a^p \Bigg(-\frac{1}{2}\log 2\pi(\sigma_i)^2 \\
&- \frac{\left((x_{i,a}^{t+1,p}-x_{i,a}^{t,p}) - \Delta t \sum_{n=1}^{N_i} k_i^n f_i^n(\{x_{j,a}^{t,p}\},s_a)\right)^2}{2(\sigma_i)^2\Delta t} \Bigg) \\
&+ \sum_{a,p,r,i} w_a^p \left(-\frac{1}{2}\log 2\pi(\eta_i)^2 - \frac{(y_{i,a}^{r,p} - g_j(\{x_{j,a}^{r,p}\}))^2}{2(\eta_j)^2} \right).
\end{aligned}
\tag{A.13}
$$

A.2.2 M step

At the Ith iteration, the parameter-value update is performed by finding the θ for which $\frac{d}{d\theta}Q_I(\theta)=0$. We describe the case of $\frac{d}{d\theta}Q(\theta,\theta_{\rm old})=0$ for simplicity, although the optimization problem can be solved generally by the same method. The following example demonstrates the determination of parameters of the system dynamics (k_i^n) and the strength of the system noise (σ_i) in detail.

First, by differentiating the complete-data log likelihood with respect to k_l^m, we obtain

$$
\begin{aligned}
0 &= \frac{\partial}{\partial k_l^m} Q(\theta,\theta_{\rm old}) \\
&= \sum_{a,p,t} w_a^p \left(\frac{\left((x_{l,a}^{t+1,p}-x_{l,a}^{t,p}) - \Delta t\sum_n^{N_l} k_l^n f_l^n(\{x_{j,a}^{t,p}\},s_a)\right)\Delta t f_l^m(\{x_{j,a}^{t,p}\},s_a)}{2(\sigma_l)^2\Delta t} \right) \\
&= \sum_{a,p,t} w_a^p (\Delta x_{l,a}^{t,p} f_l^m(\{x_{j,a}^{t,p}\},s_a)) \\
&- (\sum_n^{N_l} k_l^n \sum_{a,p,t} w_a^p f_l^n(\{x_{j,a}^{t,p}\},s_a) f_l^m(\{x_{j,a}^{t,p}\},s_a)),
\end{aligned}
\tag{A.14}
$$

where $\Delta x_{l,a}^{t,p} = (x_{l,a}^{t+1,p} - x_{l,a}^{t,p})/\Delta t$. By defining the vectors b_l and k_l and a

matrix A_l as

$$(b_l)_m = \sum_{a,p,t} w_a^p (\Delta x_{l,a}^{t,p} f_l^m(\{x_{j,a}^{t,p}\}, s_a)) \quad \text{(A.15)}$$

$$(k_l)_m = k_l^m \quad \text{(A.16)}$$

$$(A_l)_{nm} = \sum_{a,p,t} w_a^p f_l^n(\{x_{j,a}^{t,p}\}, s_a) f_l^m(\{x_{j,a}^{t,p}\}, s_a), \quad \text{(A.17)}$$

the parameters for system dynamics are determined as follows.

$$k_l = (A_l)^{-1} b_l. \quad \text{(A.18)}$$

Next, using the new k_i^n calculated above, we obtain

$$\begin{aligned} 0 &= \frac{\partial}{\partial \sigma_l} Q(\theta, \theta_{\text{old}}) \\ &= \sum_{a,p,t} w_a^p \left(-\frac{1}{\sigma_l} + \frac{\Delta t \left(\Delta x_{l,a}^{t,p} - \sum_n^{N_l} k_l^n f_l^n(\{x_{j,a}^{t,p}\}, s_a) \right)^2}{(\sigma_l)^3} \right), \end{aligned} \quad \text{(A.19)}$$

and thus,

$$(\sigma_l)^2 = \frac{\Delta t}{A(|T|-1)} \sum_{a,p,t} w_a^p \left(\Delta x_{l,a}^{t,p} - \sum_n^{N_l} k_l^n f_l^n(\{x_{j,a}^{t,p}\}, s_a) \right)^2. \quad \text{(A.20)}$$

The other parameters are estimated in the same manner. Only for the variance of the initial condition $V_{i,a}$, we define the minimum value V_{min} to avoid an unnaturally small value resulting from a problem called sample impoverishment [130].

A.3 Algorithm testing

We implement the algorithm described above in C++ and MATLAB. In this section, the code is tested by using artificial data from a simple dynamical system, called the Brusselator.

A.3.1 Data generator: Brusselator

The Brusselator was proposed by Ilya Prigogine and co-workers to investigate an oscillatory behavior in a chemical system with an autocatalytic reaction [131]. We generate time-series data from the Brusselator with noise,

$$\tau \dot{x_0} = b x_1 - x_0 (x_1)^2 + \xi_0(t) \quad \text{(A.21a)}$$

$$\tau \dot{x_1} = a + x_0 (x_1)^2 - (1+b) x_1 + \xi_1(t) \quad \text{(A.21b)}$$

where the noise terms ξ_i $(i = 0, 1)$ satisfy white Gaussian statistics $\langle \xi_i(t) \rangle = 0$ and $\langle \xi_i(t)\xi_j(\tau) \rangle = 2\sigma_i^2 \delta_{i,j}\delta(t - \tau)$. As increase of value of a from 2 to 0.5, the system with $b = 2$ exhibits oscillation onset through a supercritical Hopf bifurcation. The data is generated at values of a around the bifurcation point, that is, at $a = 0.5$, 1, and 2.

The observation noise is introduced in observation equation by $y_i^r = \eta_i \phi_i^r$ $(i = 0, 1)$ where $\eta_i = 0.05$ is intensity of observation noise and ϕ_i^r is sampled from a standard normal distribution. Figure A.1 shows the artificial data to demonstrate how the dynamics fluctuates.

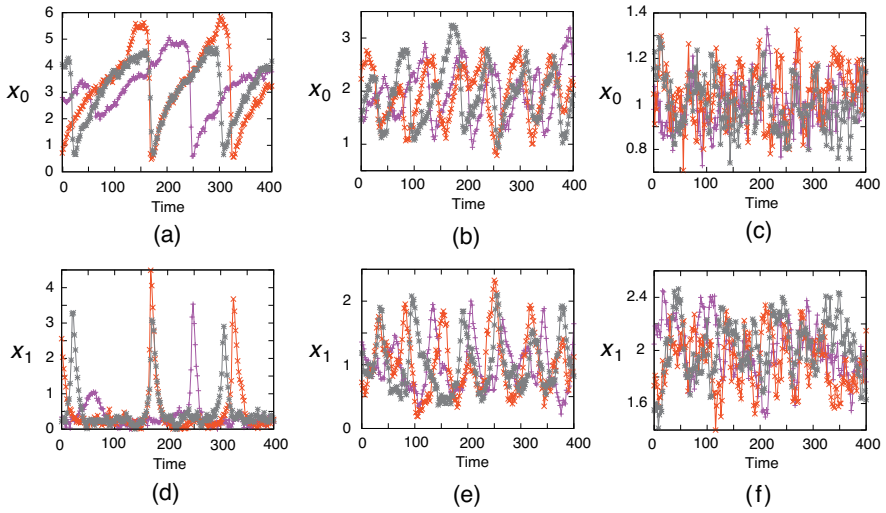


Figure A.1: Noisy time series generated from Brusselator through noisy observations of x_0 (a,b,c) and x_1 (d,e,f). The colors indicate results from independent trials. The values of the bifurcation parameter are (a,d) $a = 0.5$, (b,e) 1, (c,f) 2, respectively. The other parameter values are set as $b = 2$, $\tau = 10$, $\sigma_i/\tau = 0.05$ and $\eta_i = 0.05$ $(i = 0, 1)$. These calculations are performed by a stochastic Runge-Kutta (SRKII) algorithm [59].

A.3.2 Results

Estimation of the Brusselator

First we assume that the equations of the data generator are known. Then state equations are defined by discrete-time forms of eqs.(A.21a) and (A.21b), as

$$x_0^{t+1} = x_0^t + \Delta t(k_0 x_1 + k_1 x_0 (x_1)^2) + \sqrt{2\Delta t}\sigma_0 \xi_0^t \quad \text{(A.22a)}$$

$$x_1^{t+1} = x_1^t \Delta t(a + k_2 x_0 (x_1)^2 - k_3 x_1) + \sqrt{2\Delta t}\sigma_1 \xi_1^t. \quad \text{(A.22b)}$$

where ξ_i $(i = 0, 1)$ is sampled from a standard normal distribution. Observation equations are the same as the data generator, as $y_i^r = \eta_i \phi_i^r$ $(i = 0, 1)$.

Here, the parameters $k_{0,\ldots,3}$, $\sigma_{0,1}$, and $\eta_{0,1}$ are to be estimated.

We examined whether the parameters are correctly estimated from the data in Fig. A.1. As a result, it is found that the data is sufficient for estimation. Figure A.2(a) and (b) show the values of log likelihood and estimated parameters $\{k_i\}$ at each learning iteration. The learning quickly converges within several tens of the iterations. Besides, we confirmed that the algorithm accurately estimates the parameters in the data generator, including intensities of noise. (see Section A.4 for the detailed setting of the learning algorithm)

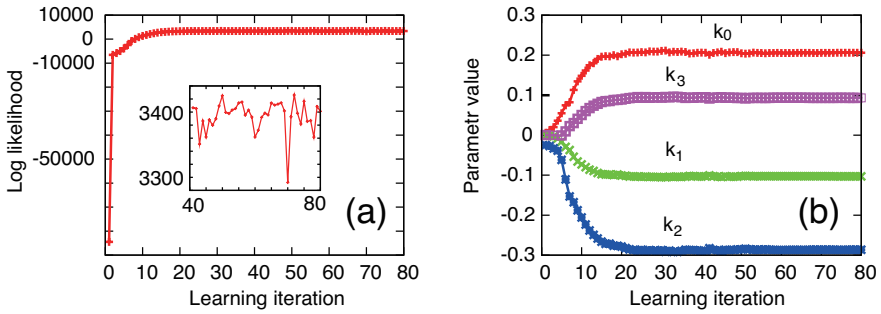


Figure A.2: The values of (a) log likelihood and (b) estimated parameters are presented at each learning iteration. The inset provides an enlarged view on log likelihood from $I = 40$ to 80, in order to demonstrate the fluctuation coming from stochastic nature of particle filter algorithm.

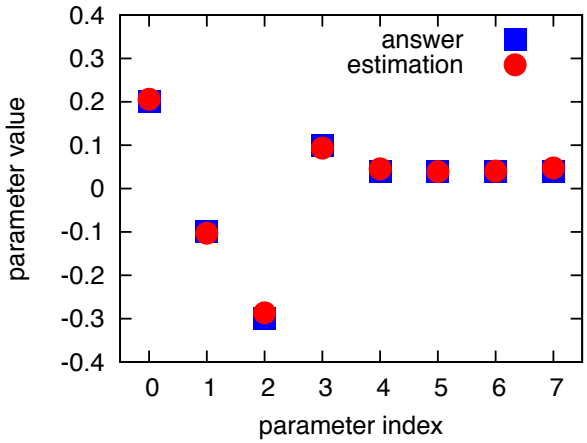


Figure A.3: The estimated parameter values (red circles) are presented in comparison with the correct values (blue squares). The parameter indices $0, \ldots, 7$ correspond to $k_{0,\ldots,3}$, $\sigma_{0,1}$, and $\eta_{0,1}$.

Estimation of a polynomial dynamical system

Next we examine the algorithm with more general state equations, that is, polynomials. By adopting polynomials of 2 variables with degree 3 as the state equations, we perform the learning based on the same data and settings as in the previous case. The state equations in continuous-time form are wrote as

$$\begin{aligned} \dot{x_0} &= k_0 + k_1 x_0 + k_2 x_1 + k_3 (x_0)^2 + k_4 x_0 x_1 + k_5 (x_1)^2 + k_6 (x_0)^3 \\ &\quad + k_7 (x_0)^2 x_1 + k_8 x_0 (x_1)^2 + k_9 (x_1)^3 + \text{noise} \qquad \text{(A.23a)} \\ \dot{x_1} &= a + k_{10} + k_{11} x_0 + \cdots + k_{18} (x_1)^3 + \text{noise}. \qquad \text{(A.23b)} \end{aligned}$$

As shown in Fig. A.4, although a bit complicated and longer transient process is observed, the learning still converges with several tens of the iterations.

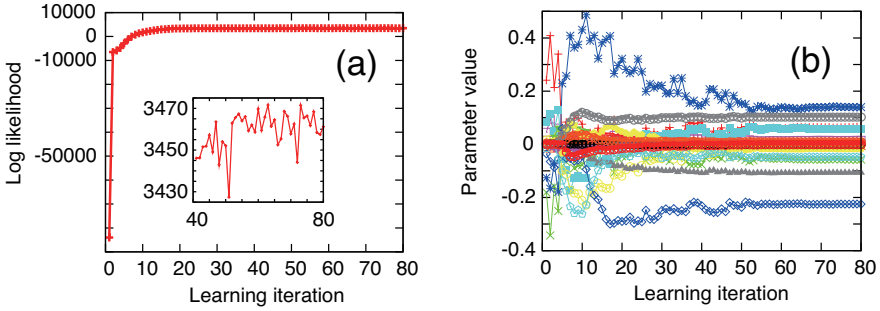


Figure A.4: The values of (a) log likelihood and (b) estimated parameters are presented at each learning iteration. The inset provides an enlarged view on log likelihood from $I = 40$ to 80, in order to demonstrate the fluctuation coming from stochastic nature of particle filter algorithm.

However, quantitative accuracy of estimation is reduced because of large number of parameters to be estimated. To quantify the accuracy, we newly generate four data sets, in addition to the data in Fig. A.1, and obtain estimation result by using each data set. The results are summarized in Fig. A.5(a) that show the mean values and the standard deviations of the estimation results. Majority of the parameters is accurately estimated, although large fluctuations remain for some parameters.

We also investigate an effect of amount of data on the estimation. To this end, 5 data sets are generated to contain 6 independent time-series data at each value of the bifurcation parameter, a. The amount of data is two times that used in Fig. A.5(a). The results are summarized in Fig. A.5(b). Although some improvements are obtained, biases and fluctuations in estimation still remain.

Finally we examine whether the algorithm can utilize information in transient dynamics, not only dynamics at attractors, as expected. In all of examples above, the data are obtained after long-time relaxation of dynamics of the data generator. Then the observed orbits certainly stay around the attractor. In contrast to those data, here we generate data sets in which initial conditions $x_0 = 0$ and $x_1 = 0$ are adopted. The data contain information on transient dynamics near the origin, unlike the previous examples. As shown in Fig. A.5(c), we can obtain the improved estimation results by using the excess information.

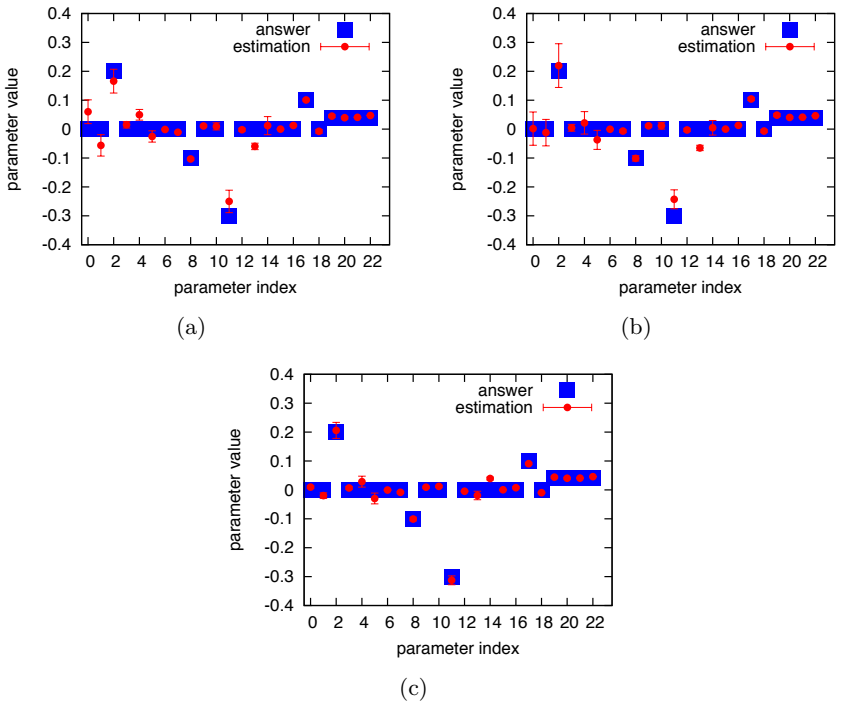


Figure A.5: The estimated parameter values (red circles) are presented in comparison with the correct values (blue squares). The mean values and the standard deviations are calculated from estimation results by 5 independent data sets. The parameter indices $0, \ldots, 22$ correspond to $k_{0,\ldots,18}$, $\sigma_{0,1}$, and $\eta_{0,1}$. (a) 3 independent time-series data at each value of the bifurcation parameter, a. (b) 6 independent data at each value of a. (c) The data is generated with initial conditions $x_0 = 0$ and $x_1 = 0$. The amount of data is the same as in (b).

Conclusion

It is confirmed that the Brusselator is estimated based on noisy time-series data. All obtained results seem to be reasonable. Then we conclude that the

algorithm described above works effectively and is properly implemented.

A.4 Settings of learning

Parameters of the learning algorithm used in Appendix A, Chapter 2, Section 3.3, and Section 3.4 are listed in Table A.1, A.2, A.3, and A.4, respectively.

Parameters in the learning algorithm	
Integration time Δt	1.0
Entire-time points T	$\{0, 1, 2, \ldots, 400\}$
Observed-time points R	$\{0, 2, 4, \ldots, 400\}$
Decreasing sequence in SAEM α_I	1 ($I \leq 50$),
	$1/(I - 50)$ (otherwise)
Particle number P	1000
Particle number threshold for resampling P_{thres}	500
Initial system dynamics parameters $\{k_i^n\}$	0
Initial system noise strength σ_i	0.04
Initial observation noise strength η_i	0.5
Minimum value of initial condition variance V_{min}	0.01

Table A.1: Pre-determined parameters of the learning algorithm used to identify Brusselator.

Parameters in the learning algorithm	
Number of time series A	9
Integration time Δt	1.0
Entire-time points T	$\{0, 1, 2, \ldots, 400\}$
Observed-time points R	$\{0, 2, 4, \ldots, 400\}$
Decreasing sequence in SAEM α_I	1 ($I \leq 30$),
	$1/\sqrt{(I - 30)}$ (otherwise)
Particle number P	1000
Particle number threshold for resampling P_{thres}	500
Initial system 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 value of initial condition variance V_{min}	0.001

Table A.2: Pre-determined parameters of the learning algorithm used to estimate bifurcations in the tyson and the ferrell models.

Parameters in the learning algorithm	
Number of time series A	18
Number of observed-time points $\|R\|$	201
Decreasing sequence in SAEM α_I	1 ($I \leq 140$), $1/(I-140)$ (otherwise)
Particle number P	400
Particle number threshold for resampling P_{thres}	300
Initial system dynamics parameters $\{k_i^n\}$	0
Initial system noise strength σ_i	0.04
Initial observation noise strength η_i	0.5
Minimum value of initial condition variance V_{min}	0.01

Table A.3: Pre-determined parameters of the learning algorithm used to identify Rössler model.

Parameters in the learning algorithm	
Number of time series A	43
Number of observed-time points $\|R\|$	43
Decreasing sequence in SAEM α_I	1 ($I \leq 450$), $1/\sqrt{(I-450)}$ (otherwise)
Particle number P	600
Particle number threshold for resampling P_{thres}	300
Initial system dynamics parameters $\{k_i^n\}$	[-0.01,0.01]
Initial system noise strength σ_i	20
Initial observation noise strength η_i	1
Minimum value of initial condition variance V_{min}	0.05

Table A.4: Pre-determined parameters of the learning algorithm used to identify a polynomial FCD model.

Appendix B

Quasi simulated annealing

B.1 Method

B.1.1 Model definition

As in the case of EM algorithm described in Appendix A, we define a state space model composed of state and observation equations. Let us consider that D-dimensional stochastic differential equations describe a system, and d-dimensional observation variables are obtained through a measurement of the system. By discretizing the dynamics in time with the Eular-Maruyam scheme [54], we write the model as

$$x_i^{t+1} = x_i^t + \Delta t f_i^1(\{x_j^t\}, \theta, s^t) + \sqrt{\Delta t} f_i^2(\{x_j^t\}, \theta)\xi_i^t \quad \text{(B.1)}$$

$$y_i^r = g_i^1(\{x_j^r\}, \theta) + g_i^2(\{x_j^r\}, \theta)\phi_i^r \quad \text{(B.2)}$$

where x_i $(i = 0, \ldots, D-1)$ and y_i $(i = 0, \ldots, d-1)$ are state and observation variables, respectively, and θ is the parameter vector, and $t(\in T)$ and $r(\in R)$ are time points, and Δt is an integration time, and s^t is the input to the system. Both ξ_i^t and ϕ_i^r are sampled from a standard normal distribution. In general, the set of observed time points R is a subset of the entire time point set T for the numerical integration (i.e., $R \subseteq T$).

We consider a data set of $Y = \{Y_a\}$ $(a = 1, \ldots, A)$ where each Y_a is a time-series data obtained from independent measurement at input values $\{s_a^t\}_{t\in T}$ that are assumed to be known. The initial condition for the ith component in the ath time-series data is assumed to obey a Gaussian distribution parameterized by the mean $\mu_{i,a}$ and the variance $V_{i,a}$. However, unlike the case in Appendix A, we do not estimate the initial condition separately from the other parameters, but calculate the conditions as described in detail below.

B.1.2 Optimization algorithm

We optimize the model using a simulated-annealing-like method. While simulated annealing (SA) algorithm optimizes a model with gradually-decreasing temperature, the quasi-SA uses just two values of temperature. However, with high initial values of intensities of system and observation noises, the noises gradually decrease and effectively work as temperature in the learning process, and then we can expect to find a good solution. The detailed procedure of the quasi-SA is described as follows.

Quasi simulated annealing

0. Define a model with a parameter vector θ and an optimization function $O(\theta)$

1. Initialize the parameter vector $\theta = \theta_0$, and set the iteration number I to zero.

2. Increment I by one, and

   (MCMC)
   If $I < I_1$, perform step 3 with $\beta = 1$,
   (Steepest descent method)
   else if $I < I_2$, perform step 3 with $\beta = \infty$,
   else if $I = I_2$, finish the optimization.

3. For each parameter $\theta_i \in \theta$,

   3a. generate a random variable as $\epsilon = \exp(\delta \times e)$ where e is sampled from a uniform distribution $[-1, 1]$, and propose a new parameter vector θ' by setting $\theta_i \to \epsilon\theta_i$.

   3b. Calculate the optimization function $O(\theta')$ and new initial conditions ($\{\mu_{i,a}\}$ and $\{V_{i,a}\}$). After that, generate a random variable u sampled from a uniform distribution $[0, 1]$, and

   If $u < \exp(\beta(O(\theta') - O(\theta)))$, accept the θ' and the new initial conditions,
   otherwise, reject the θ' and the new initial conditions.

The calculation of new initial conditions in the step 3b is performed as follows. By using tentatively estimated parameters and initial conditions, we numerically integrate the model for a time interval t_{init} in order to confine the states of the model to a region around the attractor. For each ath input, we calculate the states $\{x_i\}$ after the integration under the input value at

the initial time point, and repeat it for P times. The new estimators of initial conditions $\{\mu_{i,a}\}$ and $\{V_{i,a}\}$ are obtained as means and variances of the calculated states, respectively.

B.2 Algorithm Testing

We implement the algorithm described above in C++. In this section, the code is tested by using artificial data. As in section 3.4, a data set generated from a polynomial model for fold-change detection (eqs.(3.18a) and (3.18b)) is employed. A state space model is defined as

$$x_0^{t+1} = x_0^t + \Delta t(k_0 s^t x_1^t + k_1 x_0^t) + \sigma_0^t \xi_0^t \sqrt{\Delta t} \quad \text{(B.3a)}$$
$$x_1^{t+1} = x_1^t + \Delta t(k_2 x_1^t + k_3 x_0^t x_1^t) + \sigma_1^t \xi_1^t \sqrt{\Delta t}, \quad \text{(B.3b)}$$

and $y_i^r = x_i + \eta_i \phi_i^r$ $(i = 0, 1)$ where ξ_i^t and ϕ_i are sampled from standard normal distributions. Then, the parameter vector is defined as $\theta = (\{k_i\}, \{\sigma_i\}, \{\eta_i\})$. Below, log likelihood is adopted as the optimization function $(O(\theta) = \log L(\theta))$ and evaluated by using the particle filter algorithm described in Appendix A.

B.2.1 Estimation from the full observation

By considering the case of full observation in which both of the two variables are simultaneously observed, we perform the learning from an initial parameter set, $(k_0, k_1, k_2, k_3) = (5, -1, 1, -3)$ (other settings are listed in Table B.1). Figure B.1 shows how a behavior of the initial model deviate from the correct one.

Number of time series A	43
Number of observed-time points $\|R\|$	43
Integration time Δt	1/96
Particle number P	1000
Particle number threshold for resampling P_{thres}	500
Initial system noise strength σ_i	20
Initial observation noise strength η_i	1
Relaxation time interval t_{init}	14
Step size in parameter proposals δ	0.3
Number of MCMC iteration I_1	60
Number of total learning iteration I_2	120

Table B.1: Pre-determined parameters used in the quasi-SA.

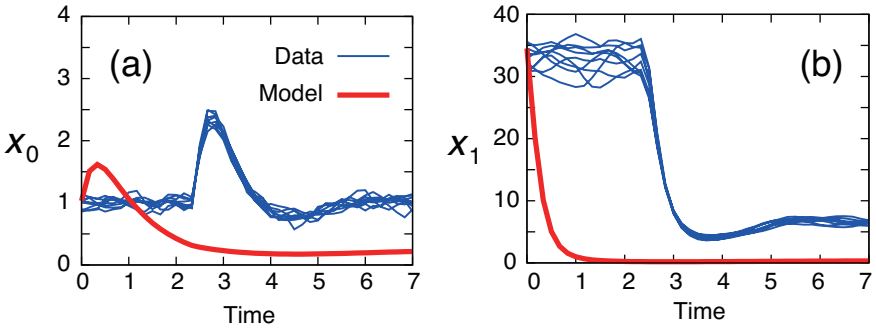


Figure B.1: Comparison between the behaviors of the initial model without noise (red) and data (blue) is presented. (a) and (b) show time series of x_0 and x_1, respectively.

As a result, we obtain correct esimations while there are cases trapped at local maxima of log likelihood. Figure B.2(a) and (b) present the values of log likelihood and estimated parameters $\{k_{0,1,2,3}\}$ at each learning iteration in a successful example. The parameters are confirmed to converge to the correct values.

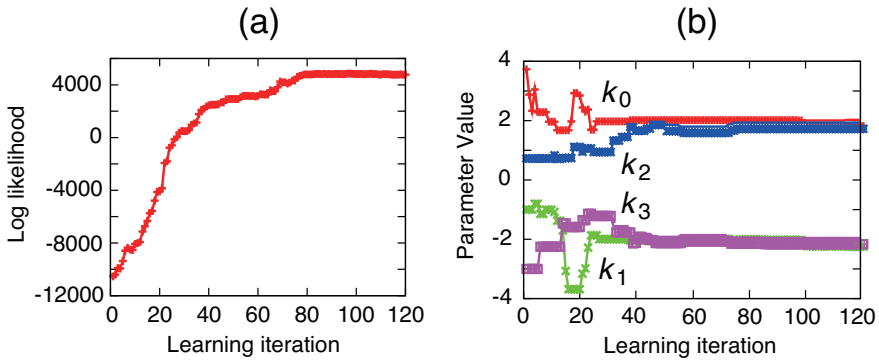


Figure B.2: The values of (a) log likelihood and (b) estimated parameters are presented at each learning iteration in the case of full observation.

B.2.2 Estimation from adaptive time series only

We also validate the quasi-SA in the case of partial observation by assuming that a variable x_1 is completely unobservable. The initial values of parameters and settings in the algorithm are the same as in the case of full observation, except for $I_1 = 100$ and $I_2 = 200$. We still obtain correct estimations, although number of cases trapped at local maxima increases compared with the full observation case. A successful example is summarized in Figs. B.3 and B.4.

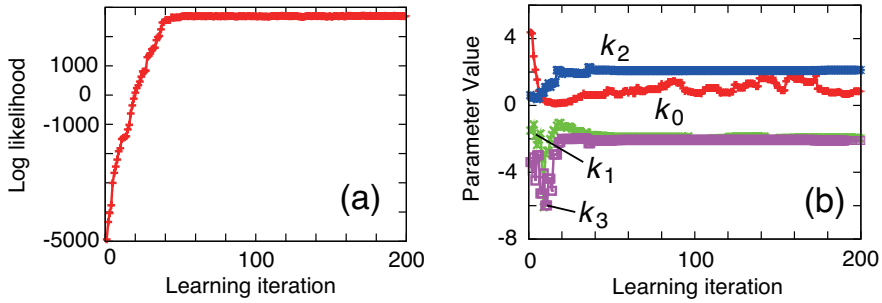


Figure B.3: The values of (a) log likelihood and (b) estimated parameters are presented at each learning iteration in the case of partial observation

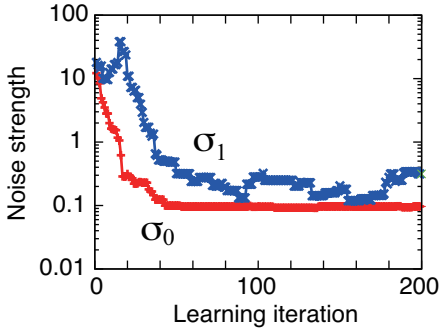


Figure B.4: The values of intensities of system noise are presented at each learning iteration for the case of partial observation.

It is found that, through the learning iterations, values of k_0 and σ_1 continue to fluctuate while other parameters and log likelihood cleary converge. We can explain this fluctuations by a symmetry property in the data generator and the state equations. That is, by variable transformations as $k_0 x_1 \to X_1$ and $\xi_1 \to k_0 \Xi_1$, we can convert the state equations from

$$\dot{x_0} = k_0 s(t) x_1 - k_1 x_0 + \xi_0(t) \quad \text{(B.4a)}$$
$$\dot{x_1} = k_2 x_1 - k_3 x_0 x_1 + \xi_1(t) \quad \text{(B.4b)}$$

to

$$\dot{x_0} = s(t) X_1 - k_1 x_0 + \xi_0(t) \quad \text{(B.5a)}$$
$$\dot{X_1} = k_2 X_1 - k_3 x_0 X_1 + \Xi_1(t). \quad \text{(B.5b)}$$

The converted equations do not include k_0, and then it is indicated that values of k_0 and σ_1 can change without affecting the observed dynamics of x_0 as long as $k_0 \times \sigma_1 = \text{const}$. This explains the main cause of the fluctuations of the values of k_0 and σ_1, although the product $k_0 \times \sigma_1$ is not

perfectly invariant because evaluation of log likelihood based the particle filter also has some fluctuation. In fact, we confirm that dynamics of the observed component x_0 looks identical to the data, in a final phase of the learning ($I > 150$).

B.2.3 Conclusion

We confirmed that the quasi-SA reasonably estimates the dynamical system based on the noisy time-series data even when only the adaptive variable is observed. Then we conclude that the algorithm described above works effectively and is properly implemented.

B.3 Settings

The settings in the quasi-SA algorithm used in Chapter 4 are listed in Table B.2.

Number of time series A	18
Number of observed-time points $\|R\|$	40
Integration time Δt	1/192
Initial observation noise strength η	1
Relaxation time interval t_{init}	26
Step size in parameter proposals δ	0.3
Number of MCMC iteration I_1	400
Number of total learning iteration I_2	700

Table B.2: Pre-determined parameters used in the quasi-SA in Chapter 4.

Bibliography

[1] B Alberts, A Johnson, J Lewis, M Raff, K Roberts, and P Walter. *Molecular Biology of the Cell*. Garland Science, 5th edition, 2008.

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

[3] S M Block, J E Segall, and H C Berg. Adaptation kinetics in bacterial chemotaxis. *Journal of Bacteriology*, Vol. 154, No. 1, pp. 312–23, 1983.

[4] J E Segall, S M Block, and H C Berg. Temporal comparisons in bacterial chemotaxis. *Proceedings of the National Academy of Sciences of the United States of America*, Vol. 83, No. 23, pp. 8987–91, 1986.

[5] D A Clark and L C Grant. The bacterial chemotactic response reflects a compromise between transient and steady-state behavior. *Proceedings of the National Academy of Sciences of the United States of America*, Vol. 102, No. 26, pp. 9150–5, 2005.

[6] S L Harmer, S Panda, and S A Kay. Molecular bases of circadian rhythms. *Annual Review of Cell and Developmental Biology*, Vol. 17, pp. 215–53, 2001.

[7] K Kaneko. *Life: An Introduction to Complex Systems Biology*. Springer, 2006.

[8] B Saltzman. *Dynamical Paleoclimatology, Volume 80: Generalized Theory of Global Climate Change*. Academic Press, 2001.

[9] S H Strogatz. *Nonlinear Dynamics And Chaos: With Applications To Physics, Biology, Chemistry, And Engineering*. Westview Press, 2001.

[10] A L Hodgkin and A F Huxley. A quantitative description of membrane current and its application to conduction and excitation in nerve. *The Journal of Physiology*, Vol. 117, No. 4, pp. 500–44, 1952.

[11] A Hoffmann, A Levchenko, M L Scott, and D Baltimore. The IκB-NF-κB signaling module: temporal control and selective gene activation. *Science*, Vol. 298, No. 5596, pp. 1241–5, 2002.

[12] K C Chen, L Calzone, A Csikasz-Nagy, F R Cross, B Novak, and J J Tyson. Integrative analysis of cell cycle control in budding yeast. *Molecular Biology of the Cell*, Vol. 15, No. 8, pp. 3841–62, 2004.

[13] R Fitzhugh. Impulses and Physiological States in Theoretical Models of Nerve Membrane. *Biophysical Journal*, Vol. 1, No. 6, pp. 445–66, 1961.

[14] J Nagumo, S Arimoto, and S Yoshizawa. An Active Pulse Transmission Line Simulating Nerve Axon. *Proceedings of the IRE*, Vol. 50, No. 10, pp. 2061–2070, 1962.

[15] G M Süel, R P Kulkarni, J Dworkin, J Garcia-Ojalvo, and M B Elowitz. Tunability and noise dependence in differentiation dynamics. *Science*, Vol. 315, No. 5819, pp. 1716–9, 2007.

[16] J Martiel. A Model Based on Receptor Desensitization for Cyclic AMP Signaling in Dictyostelium Cells. *Biophysical Journal*, Vol. 52, No. 5, pp. 807–828, 1987.

[17] B Novak and J J Tyson. Modeling the Cell Division Cycle: M-Phase Trigger, Oscillations, and Size Control. *Journal of Theoretical Biology*, Vol. 165, pp. 101–134, 1993.

[18] M Borisuk and J J Tyson. Bifurcation analysis of a model of mitotic control in frog eggs. *Journal of Theoretical Biology*, Vol. 195, No. 1, pp. 69–85, 1998.

[19] D Battogtokh and J J Tyson. Bifurcation analysis of a model of the budding yeast cell cycle. *Chaos*, Vol. 14, No. 3, pp. 653–61, 2004.

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

[21] B Pfeuty, T David-Pfeuty, and K Kaneko. Underlying principles of cell fate determination during G1 phase of the mammalian cell cycle. *Cell Cycle*, Vol. 7, No. 20, pp. 3246–57, 2008.

[22] D J Stephens and V J Allan. Light microscopy techniques for live cell imaging. *Science*, Vol. 300, No. 5616, pp. 82–6, 2003.

[23] Y Shav-Tal, R H Singer, and X Darzacq. Imaging gene expression in single living cells. *Nature Reviews Molecular Cell Biology*, Vol. 5, No. 10, pp. 855–61, 2004.

[24] M Fernández-Suárez and A Y Ting. Fluorescent probes for super-resolution imaging in living cells. *Nature Reviews Molecular Cell Biology*, Vol. 9, No. 12, pp. 929–43, 2008.

[25] R Lev Bar-Or, R Maya, L A Segel, U Alon, A J Levine, and M Oren. Generation of oscillations by the p53-Mdm2 feedback loop: a theoretical and experimental study. *Proceedings of the National Academy of Sciences of the United States of America*, Vol. 97, No. 21, pp. 11250–5, 2000.

[26] N Geva-Zatorsky, N Rosenfeld, S Itzkovitz, R Milo, A Sigal, E Dekel, T Yarnitzky, Y Liron, P Polak, G Lahav, and U Alon. Oscillations and variability in the p53 system. *Molecular Systems Biology*, Vol. 2: 2006.0033, 2006.

[27] M B Elowitz, A J Levine, E D Siggia, and P S Swain. Stochastic gene expression in a single cell. *Science*, Vol. 297, No. 5584, pp. 1183–6, 2002.

[28] J Paulsson. Summing up the noise in gene networks. *Nature*, Vol. 427, No. 6973, pp. 415–8, 2004.

[29] J M Pedraza and A van Oudenaarden. Noise propagation in gene networks. *Science*, Vol. 307, No. 5717, pp. 1965–9, 2005.

[30] C Rangel, J Angus, Z Ghahramani, M Lioumi, E Sotheran, A Gaiba, D L Wild, and F Falciani. Modeling T-cell activation using gene expression profiling and state-space models. *Bioinformatics*, Vol. 20, No. 9, pp. 1361–72, 2004.

[31] M J Beal, F Falciani, Z Ghahramani, C Rangel, and D L Wild. A Bayesian approach to reconstructing genetic regulatory networks with hidden factors. *Bioinformatics*, Vol. 21, No. 3, pp. 349–56, 2005.

[32] O Hirose, R Yoshida, S Imoto, R Yamaguchi, T Higuchi, D S Charnock-Jones, C Print, and S Miyano. Statistical inference of transcriptional module-based gene networks from time course gene expression profiles by using state space models. *Bioinformatics*, Vol. 24, No. 7, pp. 932–42, 2008.

[33] S Rogers, R Khanin, and M Girolami. Bayesian model-based inference of transcription factor activity. *BMC Bioinformatics*, Vol. 8(Suppl 2):S2, 2007.

[34] R Yoshida, M Nagasaki, R Yamaguchi, S Imoto, S Miyano, and T Higuchi. Bayesian learning of biological pathways on genomic data assimilation. *Bioinformatics*, Vol. 24, No. 22, pp. 2592–601, 2008.

[35] T Aijö and H Lähdesmäki. Learning gene regulatory networks from gene expression measurements using non-parametric molecular kinetics. *Bioinformatics*, Vol. 25, No. 22, pp. 2937–44, 2009.

[36] D J Wilkinson. Stochastic modelling for quantitative description of heterogeneous biological systems. *Nature Reviews Genetics*, Vol. 10, No. 2, pp. 122–33, 2009.

[37] D J Wilkinson. Parameter inference for stochastic kinetic models of bacterial gene regulation : a Bayesian approach to systems biology. *Bayesian Statistics*, Vol. 9, 2010.

[38] Y Taniguchi, P J Choi, G-W Li, H Chen, M Babu, J Hearn, A Emili, and X S Xie. Quantifying E. coli proteome and transcriptome with single-molecule sensitivity in single cells. *Science*, Vol. 329, No. 5991, pp. 533–8, 2010.

[39] A Eldar, V K Chary, P Xenopoulos, M E Fontes, O C Losón, J Dworkin, P J Piggot, and M B Elowitz. Partial penetrance facilitates developmental evolution in bacteria. *Nature*, Vol. 460, No. 7254, pp. 510–4, 2009.

[40] Y Kondo, K Kaneko, and S Ishihara. Identifying dynamical systems with bifurcations from noisy partial observation. *Physical Review E*, Vol. 87, 042716, 2013.

[41] J Guckenheimer and P Holmes. *Nonlinear Oscillations, Dynamical Systems, and Bifurcations of Vector Fields*. Springer, 1983.

[42] K Judd and A Mees. Modeling chaotic motions of a string from experimental data. *Physica D*, Vol. 92, pp. 221–236, 1996.

[43] H Voss, P Kolodner, M Abel, and J Kurths. Amplitude Equations from Spatiotemporal Binary-Fluid Convection Data. *Physical Review Letters*, Vol. 83, No. 17, pp. 3422–3425, 1999.

[44] T G Müller and J Timmer. Fitting parameters in partial differential equations from partially observed noisy data. *Physica D*, Vol. 171, pp. 1–7, 2002.

[45] W-X Wang, R Yang, Y-C Lai, V Kovanis, and C Grebogi. Predicting Catastrophes in Nonlinear Dynamical Systems by Compressive Sensing. *Physical Review Letters*, Vol. 106, 154101, 2011.

[46] A Sitz, J Kurths, and H Voss. Identification of nonlinear spatiotemporal systems via partitioned filtering. *Physical Review E*, Vol. 68, 016202, 2003.

[47] M Nagasaki, R Yamaguchi, R Yoshida, S Imoto, A Doi, Y Tamada, H Matsuno, S Miyano, and T Higuchi. Genomic data assimilation for estimating hybrid functional Petri net from time-course gene expression data. *Genome Informatics*, Vol. 17, No. 1, pp. 46–61, 2006.

[48] J Ohkubo. Nonparametric model reconstruction for stochastic differential equations from discretely observed time-series data. *Physical Review E*, Vol. 84, 066702, 2011.

[49] A P Dempster, N M Laird, and D B Rubin. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society: Series B*, Vol. 39, No. 1, pp. 1–38, 1977.

[50] N J Gordon, D J Salmond, and A F M Smith. Novel approach to nonlinear/non-Gaussian Bayesian state estimation. *IEE Proceedings-F*, Vol. 140, No. 2, pp. 107–113, 1993.

[51] G Kitagawa. Monte Carlo Filter and Smoother for Non-Gaussian Nonlinear State Space Models. *Journal of Computational and Graphical Statistics*, Vol. 5, No. 1, pp. 1–25, 1996.

[52] C Andrieu, A Doucet, S S Singh, and V B Tadic. Particle Methods for Change Detection, System Identification, and Control. *Proceedings of the IEEE*, Vol. 92, No. 3, pp. 423–438, 2004.

[53] B Delyon, M Lavielle, and E Moulines. Convergence of a stochastic approximation version of the EM algorithm. *Annals of Statistics*, Vol. 27, No. 1, pp. 94–128, 1999.

[54] P E Kloeden and E Platen. *Numerical Solution of Stochastic Differential Equations*. Springer, 3rd edition, 1999.

[55] B Novak and J J Tyson. Numerical analysis of a comprehensive model of M-phase control in Xenopus oocyte extracts and intact embryos. *Journal of Cell Science*, Vol. 106, pp. 1153–68, 1993.

[56] J R Pomerening, S Y Kim, and J E Jr Ferrell. Systems-level dissection of the cell-cycle oscillator: bypassing positive feedback produces damped oscillations. *Cell*, Vol. 122, No. 4, pp. 565–78, 2005.

[57] T Y-C Tsai, Y S Choi, W Ma, J R Pomerening, C Tang, and J E Jr Ferrell. Robust, tunable biological oscillations from interlinked positive and negative feedback loops. *Science*, Vol. 321, No. 5885, pp. 126–9, 2008.

[58] A W Murray and M W Kirschner. Cyclin synthesis drives the early embryonic cell cycle. *Nature*, Vol. 339, No. 6222, pp. 275–80, 1989.

[59] R Honeycutt. Stochastic Runge-Kutta algorithms. I. White noise. *Physical Review A*, Vol. 45, No. 2, pp. 600–603, 1992.

[60] H Akaike. A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, Vol. 19, No. 6, pp. 716–723, 1974.

[61] G Schwarz. Estimating the Dimension of a Model. *Annals of Statistics*, Vol. 6, No. 2, pp. 461–464, 1978.

[62] Y Kuramoto. *Chemical Oscillations, Waves, and Turbulence*. Springer, 1984.

[63] A T Winfree. *The Geometry of Biological Time*. Springer, 2nd edition, 2001.

[64] D A Rand, B V Shulgin, D Salazar, and A J Millar. Design principles underlying circadian clocks. *Journal of the Royal Society Interface*, Vol. 1, No. 1, pp. 119–30, 2004.

[65] R M Smeal, G Bard Ermentrout, and J A White. Phase-response curves and synchronized neural networks. *Philosophical transactions of the Royal Society of London Series B: Biological sciences*, Vol. 365, No. 1551, pp. 2407–22, 2010.

[66] B Pfeuty, Q Thommen, and M Lefranc. Robust entrainment of circadian oscillators requires specific phase response curves. *Biophysical Journal*, Vol. 100, No. 11, pp. 2557–65, 2011.

[67] K Nakae, Y Iba, Y Tsubo, T Fukai, and T Aoyagi. Bayesian estimation of phase response curves. *Neural Networks*, Vol. 23, No. 6, pp. 752–63, 2010.

[68] M Kyung, J Gill, M Ghosh, and G Casella. Penalized Regression, Standard Errors, and Bayesian Lassos. *Bayesian Analysis*, Vol. 5, No. 2, pp. 369–412, 2010.

[69] K Takahashi, S Tanase-Nicola, and P R ten Wolde. Spatio-temporal correlations can drastically change the response of a MAPK pathway. *Proceedings of the National Academy of Sciences of the United States of America*, Vol. 107, No. 6, pp. 2473–8, 2010.

[70] K Aoki, M Yamada, K Kunida, S Yasuda, and M Matsuda. Processive phosphorylation of ERK MAP kinase in mammalian cells. *Proceedings of the National Academy of Sciences of the United States of America*, Vol. 108, No. 31, pp. 12675–80, 2011.

[71] K T Alligood, T D Sauer, and J A Yorke. *Chaos: An Introduction to Dynamical Systems*. Springer, 2000.

[72] E N Lorenz. Deterministic Nonperiodic Flow. *Journal of the Atmospheric Sciences*, Vol. 20, No. 2, pp. 130–141, 1963.

[73] O E Rössler. An equation for continuous chaos. *Physics Letters A*, Vol. 57, No. 5, pp. 397–398, 1976.

[74] R. A. Schmitz, K. R. Graziani, and J. L. Hudson. Experimental evidence of chaotic states in the Belousov-Zhabotinskii reaction. *The Journal of Chemical Physics*, Vol. 67, No. 7, p. 3040, 1977.

[75] C Vidal, J-C Roux, S Bachelart, and A Rossi. Experimental study of the transition to turbulence in the Belousov-Zhabotinsky reaction. *Annals of the New York Academy of Sciences*, Vol. 357, No. 1, pp. 377–396, 1980.

[76] K Tomita and I Tsuda. Towards the Interpretation of Hudson's Experiment on the Belousov-Zhabotinsky Reaction. *Progress of Theoretical Physics*, Vol. 64, No. 4, pp. 1138–1160, 1980.

[77] T Matsumoto, Y Nakajima, M Saito, J Sugi, and H Hamagishi. Reconstructions and predictions of nonlinear dynamical systems: a hierarchical Bayesian approach. *IEEE Transactions on Signal Processing*, Vol. 49, No. 9, pp. 2138–2155, 2001.

[78] D E Koshland, A Goldbeter, and J B Stock. Amplification and adaptation in regulatory and sensory systems. *Science*, Vol. 217, No. 4556, pp. 220–5, 1982.

[79] Y Nakaoka, H Tokui, Y Gion, S Inoue, and F Oosawa. Behavioral Adaptation of Paramecium caudatum to Environmental Temperature. *Proceedings of the Japan Academy*, Vol. 58(B) pp. 213–217, 1982.

[80] S Asakura and H Honda. Two-state model for bacterial chemoreceptor proteins. The role of multiple methylation. *Journal of Molecular Biology*, Vol. 176, No. 3, pp. 349–67, 1984.

[81] N Barkai and S Leibler. Robustness in simple biochemical networks. *Nature*, Vol. 387, No. 6636, pp. 913–7, 1997.

[82] T S Shimizu, Y Tu, and H C Berg. A modular gradient-sensing network for chemotaxis in Escherichia coli revealed by responses to time-varying stimuli. *Molecular Systems Biology*, Vol. 6, No. 382, p. 382, 2010.

[83] T M Yi, Y Huang, M I Simon, and J Doyle. Robust perfect adaptation in bacterial chemotaxis through integral feedback control. *Proceedings of the National Academy of Sciences of the United States of America*, Vol. 97, No. 9, pp. 4649–53, 2000.

[84] D Muzzey, C A Gómez-Uribe, J T Mettetal, and A van Oudenaarden. A systems-level analysis of perfect adaptation in yeast osmoregulation. *Cell*, Vol. 138, No. 1, pp. 160–71, 2009.

[85] L Goentoro and M W Kirschner. Evidence that fold-change, and not absolute level, of beta-catenin dictates Wnt signaling. *Molecular Cell*, Vol. 36, No. 5, pp. 872–84, 2009.

[86] M D Lazova, T Ahmed, D Bellomo, R Stocker, and T S Shimizu. Response rescaling in bacterial chemotaxis. *Proceedings of the National Academy of Sciences of the United States of America*, Vol. 108, No. 33, pp. 13870–5, 2011.

[87] C Cohen-Saidon, A A Cohen, A Sigal, Y Liron, and U Alon. Dynamics and variability of ERK2 response to EGF in individual living cells. *Molecular Cell*, Vol. 36, No. 5, pp. 885–93, 2009.

[88] O Shoval, L Goentoro, Y Hart, A Mayo, E Sontag, and U Alon. Fold-change detection and scalar symmetry of sensory input fields. *Proceedings of the National Academy of Sciences of the United States of America*, Vol. 107, No. 36, pp. 15995–6000, 2010.

[89] O Shoval, U Alon, and E Sontag. Symmetry Invariance for Adapting Biological Systems. *SIAM Journal on Applied Dynamical Systems*, Vol. 10, No. 3, pp. 857–886, 2011.

[90] S Ditlevsen and P Lansky. Only through perturbation can relaxation times be estimated. *Physical Review E*, Vol. 86, No. 5, pp. 1–5, 2012.

[91] K Kamino, K Fujimoto, and S Sawai. Collective oscillations in developing cells: insights from simple systems. *Development, Growth & Differentiation*, Vol. 53, No. 4, pp. 503–17, 2011.

[92] W F Loomis and D W Smith. Molecular phylogeny of Dictyostelium discoideum by protein sequence comparison. *Proceedings of the National Academy of Sciences*, Vol. 87, No. 23, pp. 9093–9097, 1990.

[93] J D Gross, M J Peacey, and D J Trevan. Signal emission and signal propagation during early aggregation in Dictyostelium discoideum. *Journal of Cell Science*, Vol. 22, No. 3, pp. 645–656, 1976.

[94] K Tomchik and P Devreotes. Adenosine 3',5'-monophosphate waves in Dictyostelium discoideum: a demonstration by isotope dilution–fluorography. *Science*, Vol. 212, No. 4493, pp. 443–446, 1981.

[95] P Schaap. Evolutionary crossroads in developmental biology: Dictyostelium discoideum. *Development*, Vol. 138, No. 3, pp. 387–96, 2011.

[96] D Dormann, B Vasiev, and C J Weijer. Propagating waves control Dictyostelium discoideum morphogenesis. *Biophysical Chemistry*, Vol. 72, No. 1-2, pp. 21–35, 1998.

[97] T Gregor, K Fujimoto, N Masaki, and S Sawai. The onset of collective behavior in social amoebae. *Science*, Vol. 328, No. 5981, pp. 1021–5, 2010.

[98] K. Kamino et al., (in preparation).

[99] K F Swaney, C-H Huang, and P N Devreotes. Eukaryotic chemotaxis: a network of signaling pathways controls motility, directional sensing, and polarity. *Annual Review of Biophysics*, Vol. 39, pp. 265–89, 2010.

[100] P A Iglesias and P N Devreotes. Biased excitable networks: how cells direct motion in response to gradients. *Current Opinion in Cell Biology*, Vol. 24, No. 2, pp. 245–53, 2012.

[101] C Janetopoulos, T Jin, and P Devreotes. Receptor-mediated activation of heterotrimeric G-proteins in living cells. *Science*, Vol. 291, No. 5512, pp. 2408–11, 2001.

[102] Y Hashimoto, M H Cohen, and A Robertson. Cell density dependence of the aggregation characteristics of the cellular slime mould Dictyostelium discoideum. *Journal of Cell Science*, Vol. 19, No. 1, pp. 215–29, 1975.

[103] A R Gingle. Critical density for relaying in Dictyostelium discoideum and its relation to phosphodiesterase secretion into the extracellular medium. *Journal of Cell Science*, Vol. 20, No. 1, pp. 1–20, 1976.

[104] C Klein and M Darmon. The relationship of phosphodiesterase to the developmental cycle of Dictyostelium discoideum. *Biochemical and Biophysical Research Communications*, Vol. 67, No. 1, pp. 440–447, 1975.

[105] G Gerisch and D Malchow. Cyclic AMP receptors and the control of cell aggregation in Dictyostelium. *Advances in Cyclic Nucleotide Research*, Vol. 7, pp. 49–68, 1976.

[106] R P Yeh, F K Chan, and M B Coukell. Independent regulation of the extracellular cyclic AMP phosphodiesterase-inhibitor system and membrane differentiation by exogenous cyclic AMP in Dictyostelium discoideum. *Developmental Biology*, Vol. 66, No. 2, pp. 361–374, 1978.

[107] L Goentoro, O Shoval, M W Kirschner, and U Alon. The incoherent feedforward loop can provide fold-change detection in gene regulation. *Molecular Cell*, Vol. 36, No. 5, pp. 894–9, 2009.

[108] Masayo Inoue and Kunihiko Kaneko. Weber's Law for Biological Responses in Autocatalytic Networks of Chemical Reactions. *Physical Review Letters*, Vol. 107, No. 4, p. 048301, 2011.

[109] W Ma, A Trusina, H El-Samad, W A Lim, and C Tang. Defining network topologies that can achieve biochemical adaptation. *Cell*, Vol. 138, No. 4, pp. 760–73, 2009.

[110] S Börner, F Schwede, A Schlipp, F Berisha, D Calebiro, M J Lohse, and V O Nikolaev. FRET measurements of intracellular cAMP concentrations and cAMP analog permeability in intact cells. *Nature Protocols*, Vol. 6, No. 4, pp. 427–38, 2011.

[111] M Hecker, S Lambeck, S Toepfer, E van Someren, and R Guthke. Gene regulatory network inference: data integration in dynamic models-a review. *BioSystems*, Vol. 96, No. 1, pp. 86–103, 2009.

[112] J Hasenauer, S Waldherr, M Doszczak, N Radde, P Scheurich, and F Allgöwer. Identification of models of heterogeneous cell populations from population snapshot data. *BMC Bioinformatics*, Vol. 12, No. 1, p. 125, 2011.

[113] K Takeda, D Shao, M Adler, P G Charest, W F Loomis, H Levine, A Groisman, W-J Rappel, and R A Firtel. Incoherent feedforward control governs adaptation of activated ras in a eukaryotic chemotaxis pathway. *Science Signaling*, Vol. 5(205):ra2, 2012.

[114] G Gerisch and U Wick. Intracellular oscillations and release of cyclic AMP from Dictyostelium cells. *Biochemical and Biophysical Research Communications*, Vol. 65, No. 1, pp. 364–370, 1975.

[115] P A Iglesias. Chemoattractant signaling in dictyostelium: adaptation and amplification. *Science Signaling*, Vol. 5(213):pe8, 2012.

[116] E F Keller and L A Segel. Model for chemotaxis. *Journal of Theoretical Biology*, Vol. 30, No. 2, pp. 225–234, 1971.

[117] T Hofer, J A Sherratt, and P K Maini. Cellular pattern formation during Dictyostelium aggregation, *Physica D*, Vol. 85, pp. 425–44, 1995.

[118] M Dworkin and K H Keller. Solubility and diffusion coefficient of adenosine 3':5'-monophosphate. *The Journal of Biological Chemistry*, Vol. 252, No. 3, pp. 864–5, 1977.

[119] L Wolpert. One hundred years of positional information. *Trends in Genetics*, Vol. 12, No. 9, pp. 359–64, 1996.

[120] T Mammoto and D E Ingber. Mechanical control of tissue and organ development. *Development*, Vol. 137, No. 9, pp. 1407–20, 2010.

[121] S Yonemura, Y Wada, T Watanabe, A Nagafuchi, and M Shibata. α-Catenin as a tension transducer that induces adherens junction development. *Nature Cell Biology*, Vol. 12, No. 6, pp. 533–42, 2010.

[122] B Aigouy, R Farhadifar, D B Staple, A Sagner, J-C Röper, F Jülicher, and S Eaton. Cell flow reorients the axis of planar polarity in the wing epithelium of Drosophila. *Cell*, Vol. 142, No. 5, pp. 773–86, 2010.

[123] O Hamant and J Traas. The mechanics behind plant development. *The New Phytologist*, Vol. 185, No. 2, pp. 369–85, 2010.

[124] A Zuur, E N Ieno, N Walker, A A Saveliev, and G M Smith. *Mixed Effects Models and Extensions in Ecology with R.* Springer, 2009.

[125] R Cheong, A Rhee, C J Wang, I Nemenman, and A Levchenko. Information transduction capacity of noisy biochemical signaling networks. *Science*, Vol. 334, No. 6054, pp. 354–8, 2011.

[126] S Laughlin. A simple coding procedure enhances a neuron's information capacity. *Zeitschrift für Naturforschung. Section C: Biosciences*, Vol. 36c, pp. 910–2, 1981.

[127] G Tkačik, C G Callan, and W Bialek. Information flow and optimization in transcriptional regulation. *Proceedings of the National Academy of Sciences of the United States of America*, Vol. 105, No. 34, pp. 12265–70, 2008.

[128] F Tostevin and P ten Wolde. Mutual information in time-varying biochemical systems. *Physical Review E*, Vol. 81, No. 6, p. 061917, 2010.

[129] T J Kobayashi and A Kamimura. Dynamics of intracellular information decoding. *Physical Biology*, Vol. 8(5)055007, 2011.

[130] M S Arulampalam, S Maskell, N Gordon, and T Clapp. A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking. *IEEE Transactions on Signal Processing*, Vol. 50, No. 2, pp. 174–188, 2002.

[131] I Prigogine. Symmetry Breaking Instabilities in Dissipative Systems. II. *The Journal of Chemical Physics*, Vol. 48, No. 4, p. 1695, 1968.

謝辞

まず、本研究を行うにあたり多くの時間を割いていただき指導教員として終始ご指導下さった金子邦彦教授に深く感謝すると共に、模範的な院生からほど遠かった私がこれまでおかけした数々の迷惑について謝罪致します。次に、第二の指導教官とも呼びうる役割を果たして下さった、石原秀至助教に深く感謝します。また、計算機環境を整えて下さった金子研究室スーパーユーザの栗川知己、嶋岡大輔、後藤祐介、香曽我部隆裕、松原嘉哉各氏に感謝します。金子研究室院生、スタッフ、OB 諸氏、及び澤井研究室の方々からは多くの助言を頂きました。限られた紙面のために全員の名前を挙げることができないことを心苦しく思います。

最後に、共同研究者であり、院生としての五年にわたる研究生活を通して議論と相談の相手を努めてくれた神野圭太氏に感謝します。彼が同期であったのは私にとって幸運と言う他ありません。