

## Representation and inference of size control laws by neural-network-aided point processes

Atsushi Kamimura <sup>1,\*</sup> and Tetsuya J. Kobayashi <sup>1,2,3,†</sup>

<sup>1</sup>*Institute of Industrial Science, The University of Tokyo, 4-6-1 Komaba, Meguro-ku, Tokyo 153-8505, Japan*

<sup>2</sup>*Faculty of Mathematical Informatics, Graduate School of Information Science and Technology, The University of Tokyo, 7-3-1, Hongo, Bunkyo-ku, Tokyo, 113-8656, Japan*

<sup>3</sup>*Universal Biology Institute, The University of Tokyo, 7-3-1, Hongo, Bunkyo-ku, Tokyo, 113-8654, Japan*



(Received 27 January 2021; revised 24 April 2021; accepted 8 June 2021; published 8 July 2021)

The regulation and coordination of cell growth and division are long-standing problems in cell physiology. Recent single-cell measurements that use microfluidic devices have provided quantitative time-series data on various physiological parameters of cells. To clarify the regulatory laws and associated relevant parameters, such as cell size, simple mathematical models have been constructed and tested based on their capabilities to reproduce the measured data. However, the models may fail to capture some aspects of data due to presumed assumptions or simplification, especially when the data are multidimensional. Furthermore, comparing a model and data for validation is not trivial when we handle noisy multidimensional data. Thus, to extract hidden laws from data, a novel method, which can handle and integrate noisy multidimensional data more flexibly and exhaustively than the conventional ones, is necessary and helpful. By using cell size control as an example, we demonstrate that this problem can be addressed by using a neural network (NN) method, originally developed for history-dependent temporal point processes. The NN can effectively segregate history-dependent deterministic factors and unexplainable noise from given data by flexibly representing the functional forms of the deterministic relation and noise distribution. By using this method, we represent and infer the birth and division cell size distributions of bacteria and fission yeast. Known size control mechanisms, such as the adder model, are revealed as the conditional dependence of the size distributions on history. Further, we show that the inferred NN model provides a better data representation for model searching than conventional descriptive statistics. Thus, the NN method can work as a powerful tool for processing noisy data to uncover hidden dynamic laws.

DOI: [10.1103/PhysRevResearch.3.033032](https://doi.org/10.1103/PhysRevResearch.3.033032)

### I. INTRODUCTION

One major challenge in microbial physiology is determining the fundamental principles and laws underlying the regulation and coordination of cell growth and division [1]. Recent developments in microfluidic devices have enabled the tracking of microbial cells over hundreds of generations [2–5]. Various physiological parameters—particularly, cell sizes—have been measured quantitatively over time at the single-cell level [2,6–8].

To elucidate the laws of cell growth and division, a large body of research has been conducted by using physical and mathematical science techniques [1,9–14]. In particular, simple physical and mathematical models have been developed that can explain cell growth and division with a small number of relevant variables. To account for the significant cell-to-cell variety of division patterns in the measured data, the process of cell growth and division is typically formalized as a continuous-time stochastic process in which each cell

grows and divides at variable rates [10,11]. In this formulation, choosing the relevant variables for a cell's division rate is essential for defining the models. For example, the “sizer” model [10] posits that the absolute size of the cell is relevant, whereas the “adder” model [15–17] considers the added size from the birth of the cell fundamental. Although the biological mechanisms regulating cell divisions are complex and involve many biochemical reactions and signaling, such simple models have successfully reproduced the division patterns.

However, simple models may fail to capture some aspects of data due to presumed assumptions or simplification, especially when the data are high-dimensional. In addition, biological data show high variability, only part of which a simple model can explain. The unexplained part is typically represented as noise by assuming a simple distribution such as the Gaussian one. Again, the assumed shape of the distribution may not always be exact for representing the unexplained components.

This suggests that the development of a more versatile method is helpful for elucidating the laws from the data. The method should handle and integrate multidimensional noisy data and have a flexible representation power for both deterministic relations and noise distribution.

In the past decade, machine learning (ML) methods have been developed and applied to a wide variety of problems [18]. ML methods, especially deep learning (DL), have demonstrated that they can semiautomatically extract

\*kamimura@sat.t.u-tokyo.ac.jp

†tetsuya@mail.crmind.net

complicated patterns from noisy data sets by removing irrelevant dimensions and unexplainable noise factors. We can also use the representation power of DL into mathematical modeling. To this end, we should design an appropriate method to avoid generating artificial results by malfunctions of the high flexibility of DL. This implies that we should also validate the method by which we can correctly extract the known laws behind the data. Thereafter, we can use it to support searching a defined, but enormous, model space without relying on our insights and intuition.

In this study, we employ a neural network (NN) method to determine the regulation of cell size. By following a similar formulation to the previous modeling of cell size control as a stochastic process [10,11], we represent the cell size dynamics interrupted by division events as a temporal point process (TPP). Then we introduce an NN method to represent the intensity function of the process, which can depend flexibly on the history of cell size over the retrospective lineage. By training the NN, we obtain probability density functions (PDFs) of cell sizes at birth and division conditioned on their histories. The trained PDFs reproduce and confirm the previous results and presumed assumptions of size control and show that the cell division is indeed stochastic; we verify that the Markovian model is sufficient for predicting birth and division sizes, making it consistent with previous analyses that used regression models [19,20]. Adder and weak sizer principles for *Escherichia coli* (*E. coli*) and *Schizosaccharomyces pombe* (*S. pombe*), respectively, are obtained from the history dependency of the conditional PDFs. It verifies that the method can correctly extract the relations between multiple variables behind the data. The NN method also provides a better data representation than conventional data binning because it can avoid systematic biases, as observed in the conventional plot. Thus, the NN method can be used as a powerful tool for uncovering hidden dynamic laws from noisy data and for model searching better than the conventional descriptive data plotting or use of summary statistics.

The remainder of the paper is organized as follows. In Sec. II, we extend the formalism of the stochastic process of cell growth and division to the case wherein the cell size at division depends on their history. Subsequently, we explain the formulation and recent developments of intensity-based NN models for TPPs. In Sec. III, we describe the experimental data on *E. coli* and *S. pombe* cell division used in our analysis. In Sec. IV, we first present the performance of existing intensity-based models and show that a recently developed NN model is the best in its flexible expressiveness of data. With this model, we examine how the conditional PDFs of cell sizes depend on their history. We also examine the “typical” behavior by calculating the medians of the conditional PDFs and comparing them with the size control laws obtained by previous modeling. In Sec. V, we summarize and discuss our study.

## II. FORMULATION

### A. Modeling history-dependent stochastic dynamics of cell size

The stochastic dynamics of cell size can be modeled as a history-dependent TPP (hTPP) [21]. First, we characterize a

cell’s history by using a sequence of cell sizes at birth ( $l_b$ ) and division ( $l_d$ ). We refer to  $m$  as the length of the sequence with alternating  $l_b$  and  $l_d$ , which we regard as the history. We categorize sequences into two types:  $\mathbf{h}_m$  and  $\mathbf{g}_m$ . Sequence  $\mathbf{h}_m$  is the size history up to the most recent birth,  $l_b$ , whereas sequence  $\mathbf{g}_m$  is that up to the most recent division,  $l_d$ . For even  $m$ , they are represented as

$$\mathbf{h}_m = \{l_d^{i-m/2}, l_b^{i-m/2+1}, l_d^{i-m/2+1}, l_b^{i-m/2+2}, \dots, l_d^{i-1}, l_b^i\}, \quad (1)$$

$$\mathbf{g}_m = \{l_b^{i-m/2}, l_d^{i-m/2}, l_b^{i-m/2+1}, l_d^{i-m/2+1}, \dots, l_b^{i-1}, l_d^{i-1}\}. \quad (2)$$

For odd  $m$ , they are represented as

$$\mathbf{h}_m = \{l_b^{i-(m-1)/2}, l_d^{i-(m-1)/2}, l_b^{i-(m-1)/2+1}, \dots, l_d^{i-1}, l_b^i\}, \quad (3)$$

$$\mathbf{g}_m = \{l_d^{i-(m+1)/2}, l_b^{i-(m+1)/2+1}, l_d^{i-(m+1)/2+1}, \dots, l_b^{i-1}, l_d^{i-1}\}. \quad (4)$$

In these equations, the superscripts  $i$  denote the generations of the cell defined along a lineage. As each generation has two quantities,  $l_b$  and  $l_d$ , the history of  $\mathbf{h}_m$  and  $\mathbf{g}_m$  includes the sizes of approximately  $m/2$  generations.

Then we consider the next division size  $l_d^i$  of a cell with history  $\mathbf{h}_m$ . The history includes sizes up to the most recent birth size  $l_b^i$ . We define the survival probability  $\Pi(l|\mathbf{h}_m)$  that a cell with a history  $\mathbf{h}_m$  does not divide up to size  $l$ . In the following, we assume that the size dynamics are stationary; thus,  $\Pi(l|\mathbf{h}_m)$  is independent of the generation  $i$ . The survival probability satisfies the following master equation:

$$\Pi(l + dl|\mathbf{h}_m) = \Pi(l|\mathbf{h}_m)[1 - \lambda_d(l|\mathbf{h}_m)dl]. \quad (5)$$

Here,  $\lambda_d(l|\mathbf{h}_m)dl$  is the probability that division occurs within the size interval  $[l, l + dl]$  considering the size history  $\mathbf{h}_m$  and the fact that the cell does not divide up to size  $l$ . In the continuum limit, Eq. (5) can be written as follows:

$$\frac{d}{dl} \Pi(l|\mathbf{h}_m) = -\lambda_d(l|\mathbf{h}_m)\Pi(l|\mathbf{h}_m). \quad (6)$$

The formal solution of Eq. (6) is

$$\begin{aligned} \Pi(l|\mathbf{h}_m) &= \exp \left[ - \int_{l_b^i}^l \lambda_d(s|\mathbf{h}_m)ds \right] \\ &= \exp \left[ - \int_0^l \lambda_d(s|\mathbf{h}_m)ds \right], \end{aligned} \quad (7)$$

at the second equality of which the integral interval is changed by defining  $\lambda_d(s|\mathbf{h}_m) = 0$  for  $s \leq l_b^i$ . Because the probability that a cell division occurs within the size interval  $[l, l + dl]$  is

$$P(l|\mathbf{h}_m)dl = \Pi(l|\mathbf{h}_m)\lambda_d(l|\mathbf{h}_m)dl, \quad (8)$$

the PDF  $P(l|\mathbf{h}_m)$  of division size conditioned by history is written as

$$P(l|\mathbf{h}_m) = \lambda_d(l|\mathbf{h}_m) \exp \left[ - \int_0^l \lambda_d(s|\mathbf{h}_m)ds \right] = -\frac{d}{dl} \Pi(l|\mathbf{h}_m). \quad (9)$$

Similarly, we can describe the birth size PDF  $Q(l|\mathbf{g}_m)$  for  $l_b^i$  as

$$Q(l|\mathbf{g}_m) = \lambda_b(l|\mathbf{g}_m) \exp \left[ - \int_0^l \lambda_b(s|\mathbf{g}_m)ds \right], \quad (10)$$

in which the history  $\mathbf{g}_m$  includes the division size  $l_d^{i-1}$  up to generation  $i - 1$ . The stationarity for generation is also assumed. Although the rate function  $\lambda_b(l|\mathbf{g}_m)dl$  itself does not have an appropriate physical interpretation of birth size, we can use  $\lambda_b(l|\mathbf{g}_m)$  in modeling as a proxy for  $Q(l|\mathbf{g}_m)$  because  $\lambda_b(l|\mathbf{g}_m)$  has the same information as  $Q(l|\mathbf{g}_m)$ . By inferring  $\lambda_d(l|\mathbf{h}_m)$  and  $\lambda_b(l|\mathbf{g}_m)$  from data, we can represent both control laws of birth and division sizes as the dependence on the histories  $\mathbf{h}_m$  or  $\mathbf{g}_m$  and the stochasticity in size dynamics as the dependence on  $l$ .

### B. Representation and inference of the temporal point process by neural networks

To endow flexibility and high representation power to the cell size dynamics model, we employ the recent developments of neural networks for hTPPs. To briefly outline the development, we consider a given sequence of discrete events at times  $\{t_i\}_{i=1}^n \in [0, T]$ . The hTPP's behavior is modeled by defining its conditional intensity function  $\lambda(t|\mathbf{h}_t) \geq 0$ , which specifies the dependency of the next event time  $t$  on the event timing history  $\mathbf{h}_t = \{t_i : t_i < t\}$ . In the formulation, the conditional PDF of the next event time  $t_{i+1}$  is given by an equation almost identical to Eqs. (9) and (10) as follows:

$$p(t_{i+1}|\mathbf{h}_t) = \lambda(t_{i+1}|\mathbf{h}_t) \exp\left(-\int_{t_i}^{t_{i+1}} \lambda(s|\mathbf{h}_t)ds\right). \quad (11)$$

In conventional point-process modeling and inference, simple functional forms are assumed for  $\lambda(t_{i+1}|\mathbf{h}_t; \boldsymbol{\theta})$  or  $p(t_{i+1}|\mathbf{h}_t; \boldsymbol{\theta})$  to make the parameter inference of  $\boldsymbol{\theta}$  and the log-likelihood computation tractable. However, such assumptions severely restrict the model's representation power and are not sufficient, especially when  $\lambda(t|\mathbf{h}_t)$  should be a complicated function of either  $t$ , history  $\mathbf{h}_t$ , or both.

In the last couple of years, neural networks have been used in different ways to resolve this problem. Specifically, a recurrent neural network (RNN) was first used to handle history dependence [22]. The RNN has recurrent connections and loops that use its internal hidden state  $\mathbf{r}_t$  to process the sequence of the input history  $\mathbf{h}_t$ . The hidden state is sequentially updated according to state and event timing history. Thus, the RNN can convert the sequence of the input history  $\mathbf{h}_t$  nonlinearly into the hidden state  $\mathbf{r}_t$ . In this way, the conditioning of the  $\lambda(t|\mathbf{h}_t; \boldsymbol{\theta})$  by  $\mathbf{h}_t$  is replaced with  $\mathbf{r}_t$ , and the conditional intensity function is represented as  $\lambda(t|\mathbf{r}_t; \boldsymbol{\theta})$ .

Although the nonlinear history dependency was accounted for by the RNN, the representation of  $\lambda(t|\mathbf{r}_t; \boldsymbol{\theta})$  as a function of  $t$  has been restricted to simple functional forms because it is not straightforward to calculate the integral  $\Lambda(t) = \int_{t_i}^t \lambda(s|\mathbf{r}_t; \boldsymbol{\theta})ds$  in Eq. (11). To achieve the integral analytically, only a limited functional form can be used for the intensity function. For example, assume a constant intensity function  $\lambda^*(\tau = t_{i+1} - t_i) = \exp(\mathbf{v}^T \mathbf{r}_t + b)$  in which  $\mathbf{v}^T$  and  $b$  are parameters to be learned. This constant intensity function corresponds to an exponential distribution of PDF in Eq. (11), whose parameter varies depending on  $\mathbf{r}_t$ —that is, the history  $\mathbf{h}_t$ . Another simple form assumed in [22] is the exponential intensity function (exponential model)  $\lambda^*(\tau = t_{i+1} - t_i) = \exp(w\tau + \mathbf{v}^T \mathbf{r}_t + b)$ , which depends on the elapsed time  $\tau$  but can be integrated directly.

This exponential intensity function corresponds to a Gompertz distribution for the conditional PDF, where the functional form of the Gompertz distribution is  $p(\tau; \alpha, \beta) = \alpha \exp(\beta\tau - \frac{\alpha}{\beta} \exp(\beta\tau) + \frac{\alpha}{\beta})$  for  $\alpha, \beta > 0$ . As a more flexible functional form [23], a piecewise constant model can be considered in which the intensity function is discretized by piecewise constant functions as  $\lambda^*(\tau) = \text{softplus}(\mathbf{v}^T \mathbf{r}_t + b_j)$  for  $(j-1)L \leq \tau \leq jL$ , where  $j = 1, 2, \dots, \tau_{\max}/L$ , considering  $\tau_{\max}$  and  $L$ . Here,  $\text{softplus}(x) = \log(1 + \exp(x))$ .

To resolve the limited representation power of the intensity function, another NN has been introduced to flexibly model the functional form of  $\lambda(t|\mathbf{r}_t; \boldsymbol{\theta})$ . Among others, Omi *et al.* [24] proposed a fully NN model in which the integral  $\Lambda(t) = \int_{t_i}^t \lambda(s|\mathbf{r}_t; \boldsymbol{\theta})ds$ , rather than  $\lambda(s|\mathbf{r}_t; \boldsymbol{\theta})$ , was modeled by a feed-forward NN. This enabled us to efficiently compute the log-likelihood,  $\mathcal{L} = [\sum_i \log \lambda_{\boldsymbol{\theta}}(t_i) - \int_0^{t_n} \lambda_{\boldsymbol{\theta}}(s)ds]$ , by avoiding the integration of  $\lambda(s|\mathbf{r}_t)$ , regardless of the functional form of the intensity function  $\lambda(s|\mathbf{r}_t)$ . This method endows more flexibility in hTPP modeling, reduces computational complexity for parameter learning, and consequently extends its applicability.

### C. The neural network method used in this paper

In this study, we apply the fully NN model of Omi *et al.* [24] because of its flexibility owing to two neural networks: an RNN and a feed-forward NN. The fully NN model developed for hTPPs can be applied directly to the size control of cell physiology because the mathematical framework is almost identical as presented in Secs. II A and II B. Using the model, we obtain the conditional PDFs of division and birth size on their size history in the following.

## III. DATA

In this study, we use two public data sets [8,25]. Both were obtained by microfluidic single-cell measurement devices, versions of the commonly known “mother machine” [2,6]. The devices allow tracking of mother cells trapped at the bottoms of the observation channels over tens or hundreds of generations [see Fig. 1(a)]. From the data obtained, one can extract the following parameters to characterize the size dynamics and division events [see Fig. 1(b)]: the birth size  $l_b$ , the division size  $l_d$ , the size added between a consecutive pair of birth and division  $\Delta_d = l_d - l_b$ , the relative septum position  $l_{1/2}$ , and the division interval  $\tau$ .

The first data set was obtained by the measurements of *E. coli* [25]. The data were recorded every minute and contained time series of the cell length of mother cells. The measurements were conducted under three growth conditions of temperature: 37°C, 27°C, and 25°C. Figure 1(c) shows an example of a cell size time course.

The second data set was obtained by measuring *S. pombe* [8]. The data were recorded every 3 min and contain time series of the cell area. The measurements were conducted under seven culture conditions with different media and temperatures, namely, 28°C, 30°C, and 34°C in yeast extract (YE) medium and 28°C, 30°C, 32°C, and 34°C in Edinburgh minimum medium (EMM).

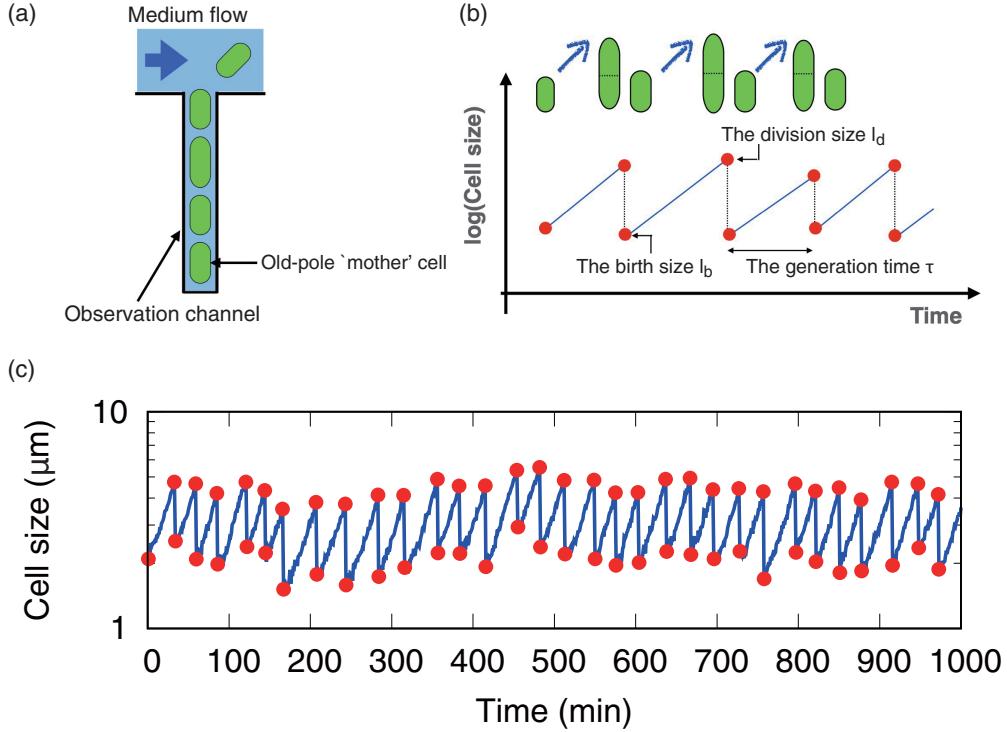


FIG. 1. (a) Schematic of the microfluidic device. The old-pole mother cell was trapped at the bottom of the observation channel. (b) Definition of the size parameters.  $\Delta_d = l_d - l_b$  is defined by the birth and division sizes of the same generation. The relative septum position  $l_{1/2}$  is defined by  $l_b^i/l_d^{i-1}$ , in which  $i-1$  and  $i$  denote the indices of consecutive generations. (c) Example of cell size trajectory from the data on *E. coli* at 37°C.

In Appendixes A and B, we describe the details of preparing and preprocessing the sequential data and procedures for NN learning.

Tables S1, S2, and S3 in the Supplemental Material [26] summarize the basic parameter statistics for *E. coli* and *S. pombe* under the YE and EMM conditions, respectively. We also present histograms of the cell sizes in Figs. S1, S2, and S3 in the Supplemental Material [26] for *E. coli* and *S. pombe* under the YE and EMM conditions, respectively.

#### IV. RESULTS

##### A. The Fully NN model achieved a better performance than other RNN-based models

In this section, we first test the performance of different RNN methods by applying them to the cell size data on *E. coli* and *S. pombe*. In a previous study [24], the fully NN model that we use in our subsequent analysis achieved a competitive or better performance compared to other RNN-based models for various synthetic and real data. Here, we show that this is also the case for cell size data. In addition to the fully NN model, we consider the three RNN models—constant, exponential, and piecewise constant—each of which has a specific form of the intensity function, as explained in Sec. II B.

Figures 2(a) and 2(b) show the trajectory examples of the actual observations and the RNN models after training for birth and division sizes of *E. coli*, respectively. For each RNN model and each generation  $i$ , the median of the PDF is calculated with the observed history of sizes before  $i$ . The medians of the fully NN and piecewise constant models agree

better with the actual observations than the other models. Significant deviations from the measured data were observed in the exponential model for birth size and in the constant model for both sizes.

In Figs. 2(c) and 2(d), we calculate the shapes of the PDFs for the four models at generation  $i = 30$  in Figs. 2(a) and 2(b). Note that because the PDFs are history dependent, their shapes change depending on the generation at which they are calculated. The conditional PDFs of the fully NN and piecewise constant models are located relatively close to the actual observation, with similar shapes. The PDFs of the exponential and constant models are much broader than the others because their functional forms are restricted to the Gompertz and exponential functions, respectively.

To measure the performances of the four models quantitatively, we calculated the negative log-likelihood  $-\log Q(l_b^* | \mathbf{g}_m)$  and  $-\log P(l_d^* | \mathbf{h}_m)$  for birth and division sizes, respectively, in which  $l_b^*$  and  $l_d^*$  denote the actual observations in the validation data. A lower negative log-likelihood indicates a better performance in prediction. Figures 2(e)–2(h) show the average performances of the four models for both *E. coli* and *S. pombe*. In almost all cases, the fully NN model achieved the superior performance, with the smallest values of the negative log-likelihood. Because the number of data was smaller than that used in Ref. [24], we changed the fraction of the training data from 80% to 95%. We found that the scores are approximately constant for the different numbers of training data, which ensures that the fully NN model is sufficiently trained. The piecewise constant model performs

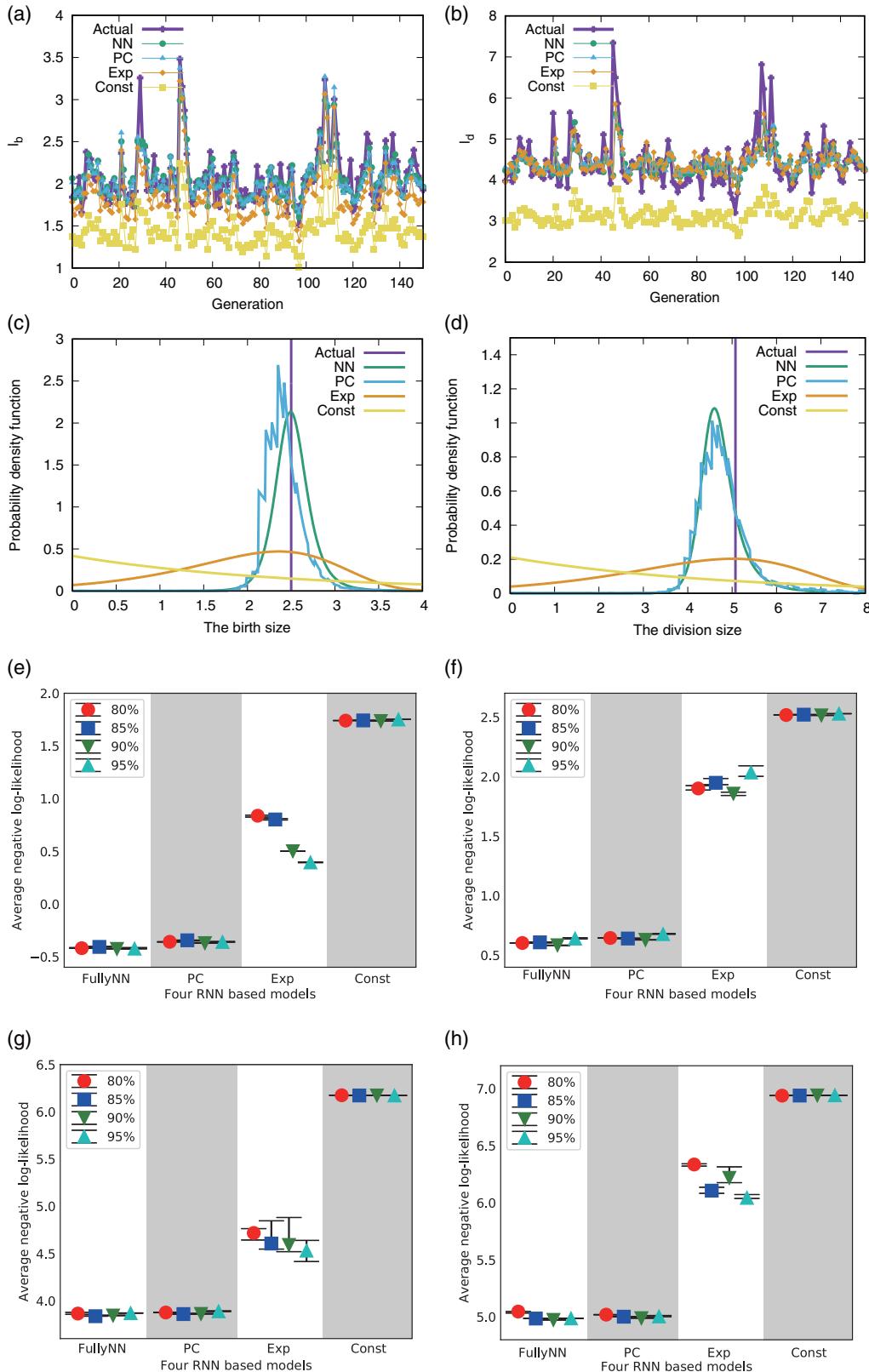


FIG. 2. (a), (b) Trajectories of the actual observations and medians of conditional probability density functions (PDFs) for *E. coli* at 37°C. The PDFs were estimated by the four RNN-based models for (a) birth size and (b) division size. (c), (d) Shapes of the PDFs of (c) birth size and (d) division size at generation  $i = 30$ . (e)–(h) Performance of the four models for (e) birth size and (f) division size of *E. coli* at 37°C and for (g) birth size and (h) division size of *S. pombe* under the YE condition at 30°C. Performance of the four RNN-based models is evaluated by the negative log-likelihood. The log-likelihood is averaged over the validation data. A lower score indicates better predictive power. The history length  $m$  was set to 10.

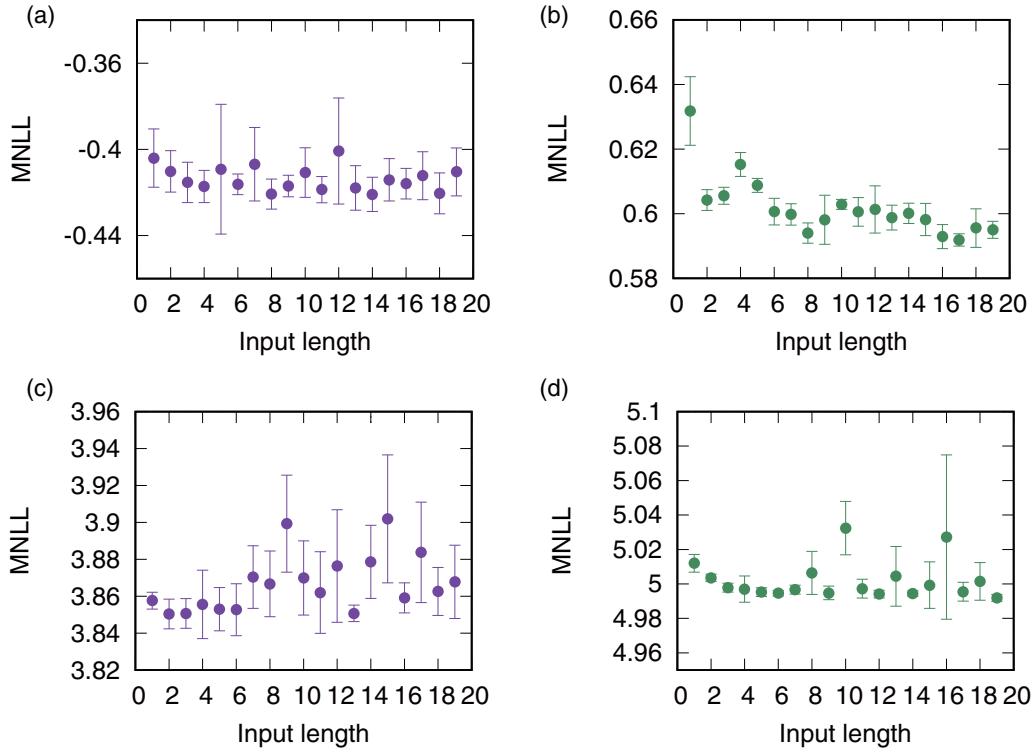


FIG. 3. Performance with different history lengths  $m$  for (a) birth size and (b) division size of *E. coli* at 37°C and for (c) birth size and (d) division size of *S. pombe* under the YE condition at 30°C. See Fig. S4 in the Supplemental Material [26] for *S. pombe* under the EMM condition. The performance was evaluated by the mean negative log-likelihood (MNLL) in which the log-likelihood was averaged over the validation data.

similarly well, but slightly worse than the fully NN model and predicts rugged conditional PDFs.

The exponential and constant models perform poorly because they have restrictions in their functional forms. The measured PDFs of cell sizes are typically single peaked and right skewed (see Sec. IV C). As a result, they cannot be approximated sufficiently by a Gompertz or exponential distribution. Further, the performances of the exponential model are unstable and its scores are more sensitive to the number of training data than the others. In any cases, the fully NN model performs better by flexibly approximating the PDFs.

### B. The most recent size in history is sufficient for birth and division size prediction

The NN method has several hyperparameters. Of particular interest is the length  $m$  of the histories  $\mathbf{h}_m$  and  $\mathbf{g}_m$ . In the previous modeling of size control, Markov models were used, and their sufficiency was tested only recently using regression models [19,20]. The dependence of the performance on the history length  $m$  can be used to systematically and quantitatively verify how far back in the history is relevant for predicting the next division or birth size.

Figure 3 shows the performance of the fully NN model for  $l_b$  and  $l_d$  as functions of the history length  $m$ . Although a slight improvement, from  $m = 1$  to  $m = 2$ , is observed for the division size  $l_d$  of *E. coli* [Fig. 3(b)], the performances do not show significant improvements with increasing  $m$  for  $m \geq 1$ . This observation indicates that the most recent elements in

history ( $l_b^i$  in  $\mathbf{h}_m$  and  $l_d^{i-1}$  in  $\mathbf{g}_m$ ) are practically sufficient, which supports the assumptions of previous size control models.

We also investigated the performance by changing other NN hyperparameters, the number of layers and the number of units in each layer (Figs. S5 and S6 in the Supplemental Material [26]). We confirm that the increase in the complexity of NN from the current parameter values does not significantly improve the results.

### C. Shapes of birth and division distributions vary depending on the history

Next, we analyze how the shapes of the birth and division size distributions depend on the history by using the estimated PDFs  $P(l_d|\mathbf{h}_m)$  and  $Q(l_b|\mathbf{g}_m)$ . Figure 4(a) shows the PDF of *E. coli* birth size  $Q(l|\mathbf{g}_m)$ . Although the shape of the PDF can, in principle, depend on all the elements in the history  $\mathbf{g}_m$ , we exclusively focus on the most recent division size because we verified it as the most influential and sufficient element of the history in Fig. 3. [See also Figs. S7 and S8 in the Supplemental Material [26] for the dependence of  $P(l|\mathbf{h}_m)$  and  $Q(l|\mathbf{g}_m)$  on previous sizes.]

As demonstrated in Fig. 4(a), the mode of the birth size PDF depends monotonously on the most recent division size, which reflects clearly that a larger cell divides into two larger daughters on average. The modes of  $Q(l|\mathbf{g}_m)$  are located approximately at the position  $l = l_d \times l_{1/2}$ , in which the relative septum position  $l_{1/2} = l_b^i/l_d^{i-1}$  is approximately 0.46 for this

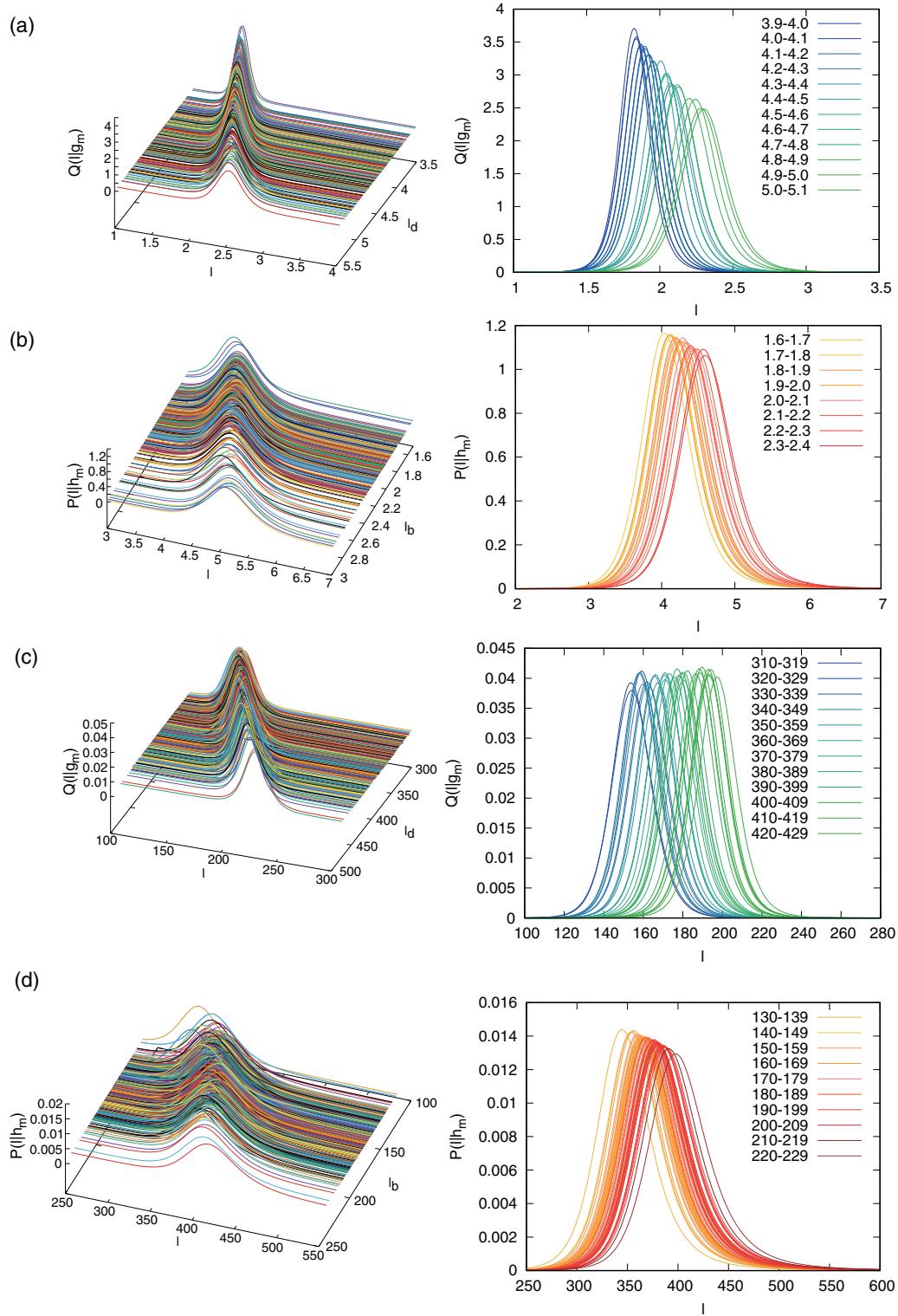


FIG. 4. The conditional probability density functions (PDFs) for (a) birth size  $Q(l|g_m)$  and (b) division size  $P(l|h_m)$  of *E. coli* at 37°C and those for (c) birth size and (d) division size of *S. pombe* under the YE condition at 30°C.  $Q(l|g_m)$  and  $P(l|h_m)$  are plotted as a function of the most recent size—that is,  $l_d$  for  $Q(l|g_m)$  and  $l_b$  for  $P(l|h_m)$ . The left panels are 3D plots of the PDFs as functions of  $l$  and the most recent size in the histories. The right panels are the slices of the PDFs for fixed values of the most recent size. In the right panels, the colors of the curves indicate the range of the most recent size. Note that the sliced PDFs shift from left to right as the most recent size increases. The history length was set to  $m = 10$ .

data set (see Table S1 in the Supplemental Material [26]). This dependency of the birth size PDF is captured even when only the most recent division size is considered as the history—that is, by setting  $m = 1$  in the model (see Fig. S9 in the Supplemental Material [26]).

A similar mode dependency is observed for the division size PDF  $P(l|\mathbf{h}_m)$  as a function of the most recent birth size [Fig. 4(b)]. However, this dependency is less prominent than that of the birth size PDF, suggesting that the division size is under tighter control than the birth size.

Figure 4(a) clarifies that the birth size of a cell with a larger parent cell shows greater variation than that with a smaller parent. This contrasts sharply with the birth size PDF of *S. pombe* [Fig. 4(c) for the YE and Fig. S10 [26] for the EMM condition], whose variation—that is, the width of the PDF—is almost independent of the previous division size. This result may indicate the difference in birth size control between *E. coli* and *S. pombe*, the latter of which has a more elaborate molecular mechanism to determine the middle of a cell [27,28]. By contrast, for the division size PDF of *S. pombe* in Fig. 4(d), we observe a dependency on the birth size similar to that of *E. coli*, which implies the existence of division size control for both *E. coli* and *S. pombe*.

All these properties are automatically extracted and clearly presented in the estimated conditional PDFs by using the fully NN model from the noisy raw data.

#### D. NN makes adder principle conspicuous by denoising data

We further elucidate the “typical” behaviors of birth and division size distributions as functions of histories by calculating the medians of the PDFs estimated by the fully NN model. We have a stable and systematic technique to obtain the median of the PDF represented implicitly by the complicated NNs (see Appendix B). Figure 5(a) exemplifies a comparison of the time series of actual birth  $l_b$  and division  $l_d$  sizes with the medians of the PDFs for *E. coli* [the same data as in Figs. 2(a) and 2(b)].

For the birth size  $l_b$ , the actual observation (orange) and medians (light blue) agree well and are distributed along  $x = y$  in the scatterplot [Fig. 5(b)]. This result demonstrates that the median of the PDF is a good predictor of  $l_b$ . For the division size  $l_d$  in Fig. 5(a), we observe more sporadic jumps in the actual observations, which the median of  $P(l|\mathbf{h}_m)$  (purple) fails to predict. Except for such outliers—for example,  $l > 10$ —the median can predict a general trend in the division size over time, which is also verified in the scatterplot in Fig. 5(c).

By using the PDF, we then attempt to detect the adder law of *E. coli* size control revealed by the previous modeling. It posits that cells add a constant size  $\Delta$  between birth and division, irrespective of birth size. A standard approach to explore size control laws is to plot the correlations among pairs of potential control parameters. Figure 5(d) shows the correlations between four parameters: division size  $l_d$ , birth size  $l_b$ , elongation  $\alpha\tau$ , and added size  $\Delta$  for *E. coli* at 37°C.

First, we focus on the correlation between the division size  $l_d$  and the birth size  $l_b$  [Fig. 5(d1)]. The plot clearly shows a positive correlation wherein a cell with a large (small) birth size divides at a large (small) division size. This trend is also consistent with the binned raw data indicated by open

circles and error bars, corresponding to the median and the interquartile range of the binned data.

The plot between the added size  $\Delta = l_d - l_b$  and the birth size  $l_b$  shows a slightly negative correlation for 37°C [Fig. 5(d11)] and approximately no correlation for 27°C and 25°C (Figs. S11 and S12 in the Supplemental Material). This negative correlation is dimmed in the binned raw data (open circles). Because of the reduced variation using the NN model, we can identify the weak sizer property in Fig. 5(d11) at 37°C. Overall, these results support the adder principle, although it is not necessarily perfect at high temperatures.

We emphasize here that our NN model does not presume specific functional forms. This is in sharp contrast to conventional statistical modeling, which assumes simple fitting functions with a theoretically tractable degree of freedom. As shown in Fig. 4, we use the neural network’s flexibility to increase the degree of freedom of fitting functions. If a linear fitting function was presumed, for example, the adder principle could be inferred and reproduced more easily. However, our NN model does not assume such a function form. Therefore, the NN model reproduced the adder principle only from the information of the data. Further, the NN model can avoid systematic biases observed in conventional binned data, as discussed in the next section.

#### E. Underlying relations may not be represented appropriately by a simple descriptive approach

To investigate whether the adder principle can account for all the other correlations, we present the expected behavior of the ideal adder model by the blue curves in Fig. 5(d) (see Sec. 1 in the Supplemental Material [26] for the ideal adder). For most plots shown in Fig. 5(d), the medians of the PDFs of the NN model agree very well with the idealized adder model (blue curves).

In contrast, in all plots—except in Figs. 5(d1), (d8), and (d11)—the binned raw data fail to capture the trend expected from the adder model. For example, if we naively assume a constant added size of the adder model, the added size  $\Delta$  should be independent of the division size and birth size. However, the binned raw data in Fig. 5(d10) show a clear positive correlation with the division size. It should be noted that the binned plots basically represent the actual distribution of the data (Fig. S13 in the Supplemental Material [26]). The deviation is not produced by the binning procedure, except for the nonmonotonic behavior between elongation and division size [Fig. 5(d2)], where the data size is relatively small for small  $\alpha\tau$ s.

This discrepancy is believed to be caused by systematic biases due to variability in the data, such as the different magnitudes of noise in the birth size and division size [1], and it does not overturn the adder hypothesis. In other words, relevant variables must be chosen for the horizontal and vertical axes to avoid such biases and to unerringly verify a model with the data. Thus, when we do not know the relevant variables *a priori*, we can obtain a correct relation only if we happen to choose the right variable, as shown in Figs. 5(d1) and (d8). Otherwise, the binned data may mislead us over the underlying relation. By using the conditional PDF and its representative value, the effect of such systematic biases is

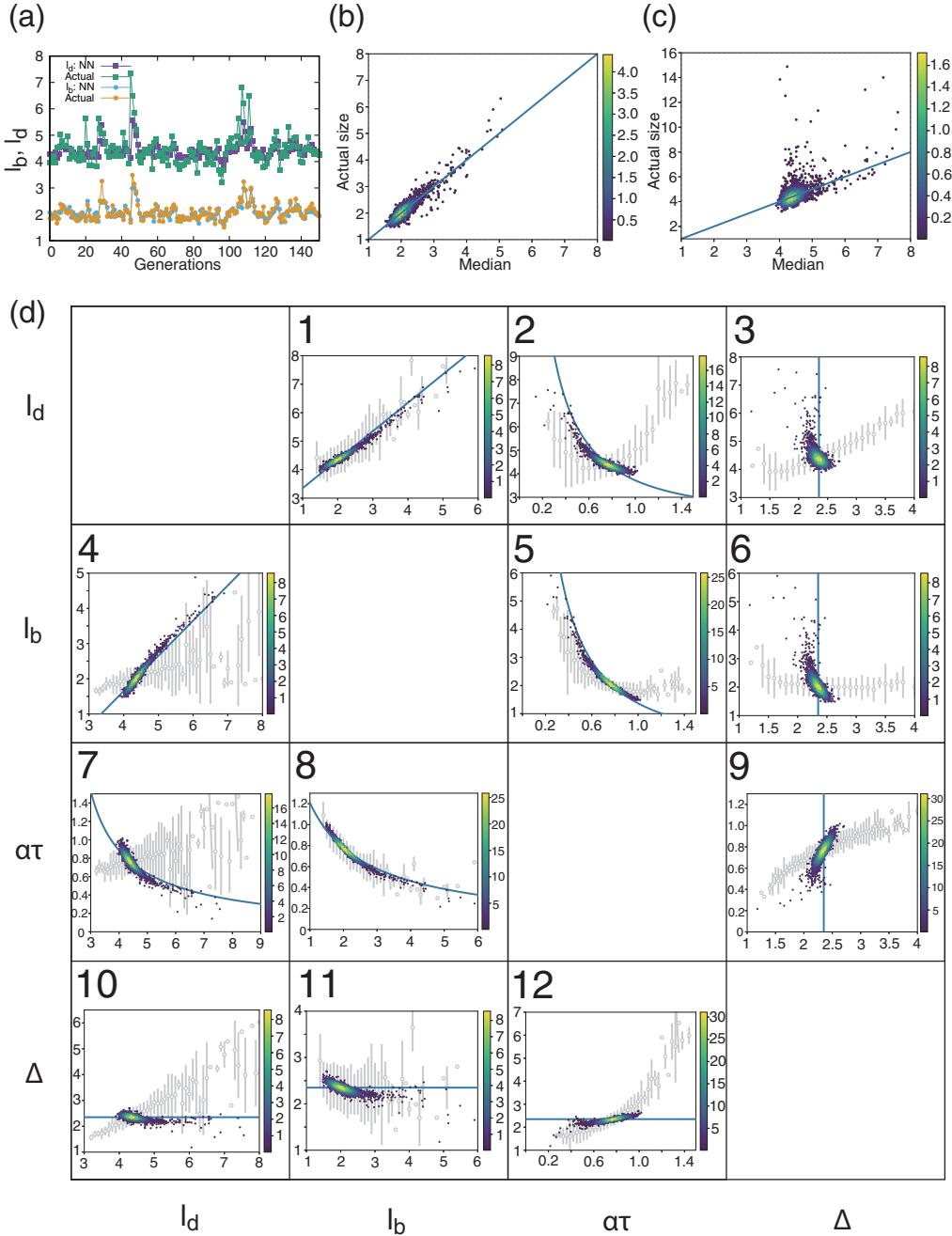


FIG. 5. Data for *E. coli* at 37°C. (a) Trajectories of the actual birth and division sizes and the medians of the PDFs obtained by the NN model. (b), (c) Density plots of the actual sizes and medians of the PDFs for (b) birth size and (c) division size. Identity lines are shown in blue. (d) All correlations between all pairs of the four parameters:  $l_d$ ,  $l_b$ ,  $\alpha\tau$ , and  $\Delta$ . Each density plot shows the correlation between each pair of parameters calculated from the medians of the PDFs. With a birth size of  $l_b$ , the PDF for the division size is used to calculate the division size ( $l_d$ ), elongation  $\alpha\tau = \log(l_d/l_b)$ , and added size  $\Delta = l_d - l_b$ . Blue curves indicate the adder model with  $\Delta = 2.35$ . Open gray circles and error bars indicate the summary statistics of the binned raw data. Circles and bars represent the medians in the binned range and their interquartile ranges, respectively.

eliminated between fluctuating variables, and a direct comparison between data and models becomes possible.

However, the discrepancy between the medians of PDFs and the binned data can still possibly be an artifact of NN modeling. In fact, the medians do not perfectly follow the idealized adder relation in some plots, as shown in Figs. 5(d3), (d6), and (d9). To exclude this possibility, we constructed synthetic data of a stochastic adder model, in which the added

size is generated from a fixed right-skewed distribution by reflecting the right skewness in the observed division distribution (see Table S4 and Sec. 2 in the Supplemental Material [26]). Then we used the synthetic data to train the NN model (Fig. 6). Here, we assume that the birth size is given by a perfect binary division  $l_b^i = l_d^{i-1}/2$  for simplicity, and the division size is given by  $l_d^i = l_b^i + z^i$ , in which the added size  $z^i$  is drawn from a log-normal distribution.

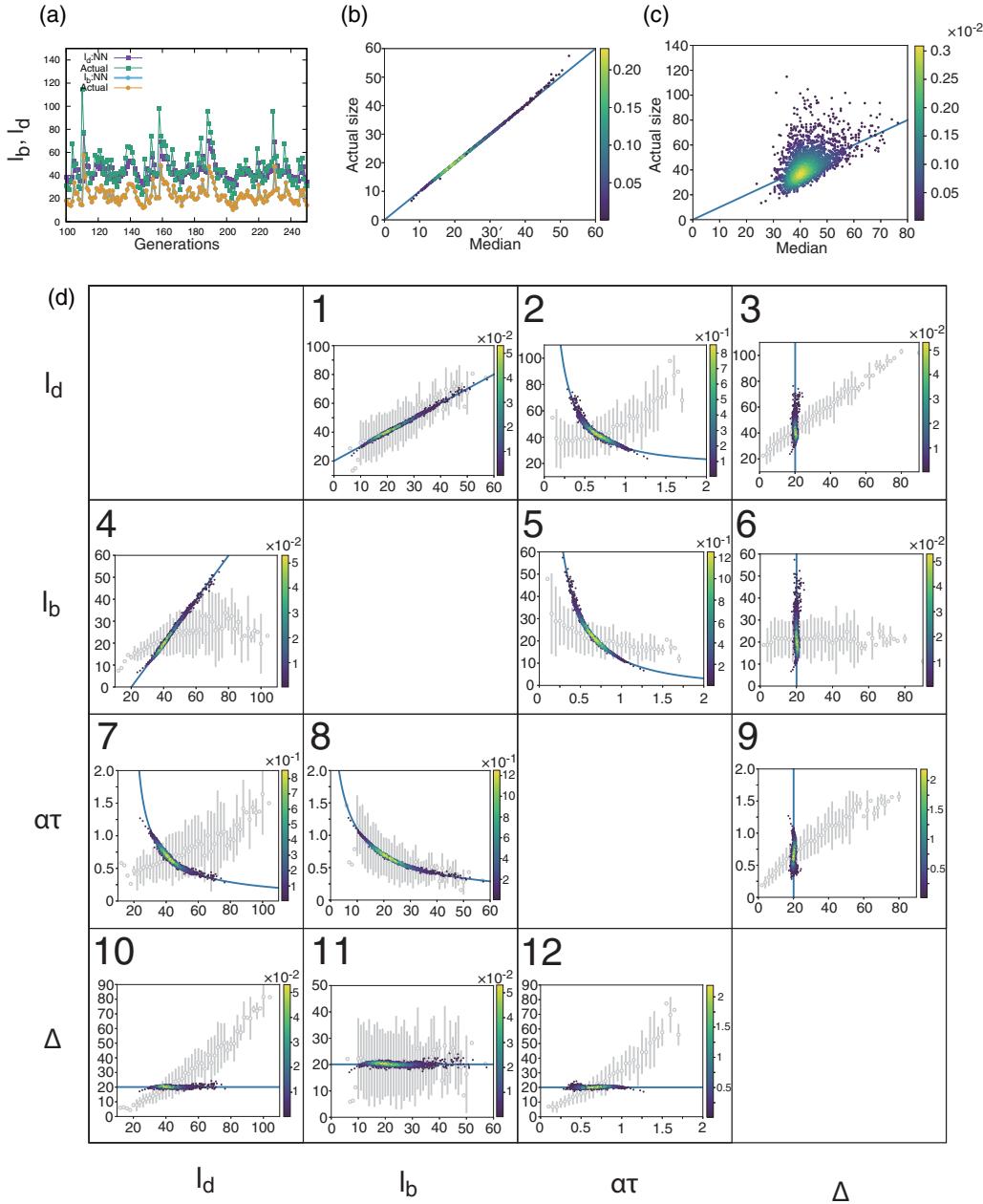


FIG. 6. Synthetic data for the adder model. The alternating sequences of  $l_b$  and  $l_d$  are synthesized as follows. For the birth size, a perfect binary division is assumed so that  $l_b^i = 0.5l_d^{i-1}$ , where the superscript denotes the generation. For the division size  $l_d^i = l_b^i + z^i$ , the added size  $z^i$  is drawn from the log-normal distribution  $L(\mu, \sigma^2)$  with  $\mu = 3$  and  $\sigma = 0.5$ . (a) Trajectories of the actual birth and division sizes and medians of the PDFs obtained by the NN model. (b), (c) Density plots of the actual sizes and medians of the PDFs for (b) birth size and (c) division size. Identity lines are shown in blue. (d) All correlations between all pairs of the four parameters:  $l_d$ ,  $l_b$ ,  $\alpha\tau$ , and  $\Delta$ . Each density plot shows the correlation between each pair of parameters calculated from the medians of the PDFs. With a birth size of  $l_b$ , the PDF for the division size is used to calculate the division size ( $l_d$ ), elongation  $\alpha\tau = \log(l_d/l_b)$ , and added size  $\Delta = l_d - l_b$ . Blue curves indicate the adder model with  $\Delta = \exp(\mu)$ . Open gray circles and error bars indicate the summary statistics of the binned raw data. Circles and bars represent the medians in the binned range and their interquartile range, respectively.

For the synthetic data, the birth size is predicted almost perfectly by the median of the trained conditional PDF [Fig. 6(b)]. Moreover, Fig. 6(c) shows that the synthetic data reproduced an asymmetry between the median and the actual size in Fig. 5(c) in a way such that the points distribute around the line  $y = x$ , but larger deviations appear more above the line. Thus, this asymmetry comes from the right skewness of the added size and, as a result, of the division size, which is not

sufficiently represented by the median of the PDF. Figure 6(d) also confirms that the PDF median can perfectly reproduce the assumed adder property, whereas the binned plots (open gray circles and error bars) fail to capture the property except in Figs. 6(d1), (d8), and (d11), similarly to the case of the *E. coli* data.

These results highlight a general advantage of the NN model over the conventional descriptive plot of raw data,

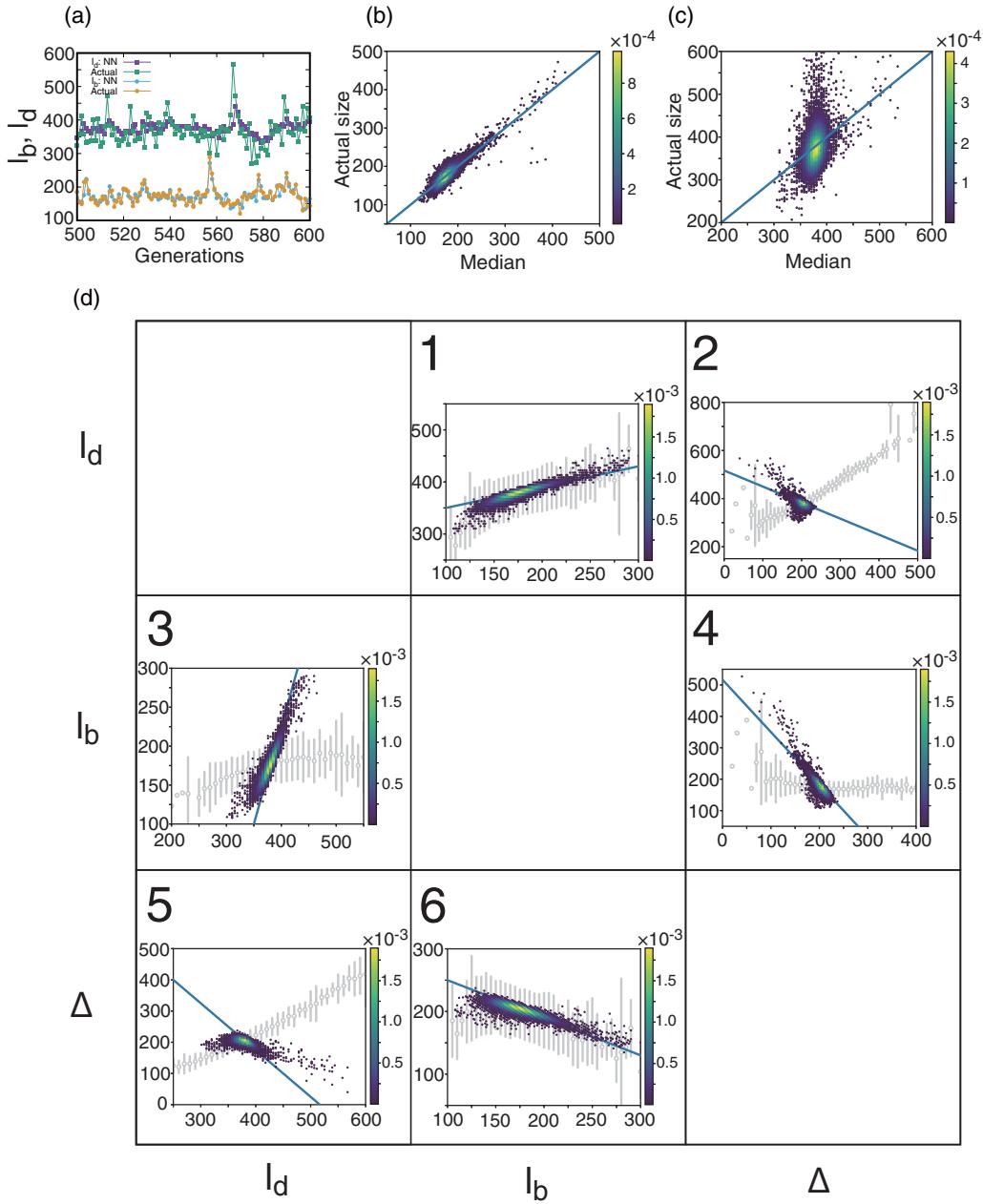


FIG. 7. Data for *S. pombe* under the YE condition at 30°C. (a) Trajectories of the actual birth and division sizes and medians of the PDFs obtained by the NN model. (b), (c) Density plots of the actual sizes and medians of the PDFs for (b) birth size and (c) division size. Identity lines are shown in blue. (d) All correlations between all pairs of the three parameters:  $l_d$ ,  $l_b$ , and  $\Delta$ . Each density plot shows the correlation between each pair of parameters calculated from the medians of the PDFs. With a birth size of  $l_b$ , the PDF for the division size is used to calculate the division size ( $l_d$ ) and the added size  $\Delta = l_d - l_b$ . Blue lines indicate  $\Delta = -0.6l_b + 310$ . Open gray circles and error bars indicate the summary statistics of the binned raw data. Circles and bars represent the medians in the binned range and their interquartile ranges, respectively.

especially when we do not know which variables should work as explanatory variables in advance. This advantage will be of particular importance when we investigate cell physiology properties other than cell size. With data binning, it will be necessary to examine whether a trend in the data is a bias derived from variation or a true relationship, similar to the case of cell size. In this sense, the NN model can benefit physical modeling by extracting relations among multiple variables from noisy data.

## F. NN works for other microbes with different size control mechanisms

Finally, we investigate *S. pombe*, whose size control law differs from that of the perfect adder [17,29]. Figure 7 shows the same analysis as in Fig. 5 for the YE condition at 30°C (see also Fig. S14 [26] for the EMM condition). The elongation is omitted because the growth of *S. pombe* may depart from the exponential trajectory and because the equation  $l_d = l_b \exp(\alpha\tau)$  used for *E. coli* may be invalid (see Fig. S15 [26]).

The NN model reproduces the birth size of *S. pombe* equally well as that of *E. coli* [Figs. 7(a) and 7(b)]. Further, Fig. 7(d6) clarifies a significant negative correlation between the birth size and the added size, whose slope is approximately  $-0.6$  (blue line). The slope of  $-1$  in this plot posits the perfect sizer principle in which a critical size exists for division. The approximated slope of  $-0.6$  indicates a size control between the perfect adder and the sizer, which is consistent with previous studies [17,29].

Similarly to Fig. 5(d), all the plots of medians in Fig. 7(d) agree with the weak sizer model of  $-0.6$  (blue curves). On the other hand, the binned data fail to capture the relations, except in Figs. 7(d1) and (d6). This result reinforces the general advantage of the NN method illustrated with the *E. coli* data.

Comparing the results for *E. coli* and *S. pombe*, a noticeable difference between Fig. 5(c) and Fig. 7(c) is found; the actual division sizes of *S. pombe* deviate from the prediction by medians. This deviation can be attributed to the sizer aspect of *S. pombe* size control. The perfect sizer principle posits that the division occurs at a certain threshold size, which makes the average division size independent of the birth size as well as the other elements in the size history. Thus, if size control approaches a sizer, the best statistical model is the history-independent i.i.d. model that faithfully reproduces the variation of division size around the threshold size.

To confirm this, we create a weak sizer model in which the added size weakly depends on the birth size and is generated in a probabilistic manner (Fig. 8 and Sec. 2 in the Supplemental Material [26]). Here, we assume that a perfect binary division  $l_b^i = l_d^{i-1}/2$  provides the birth size and that the division size is given by  $l_d^i = l_b^i + z^i$ , where the added size  $z^i$  is a realization drawn from a normal distribution with the median  $-0.6l_b + 800$ . Here, we chose the slope  $-0.6$  from Fig. 7(d6). By using these synthetic data, as shown in Fig. 8(c), the NN model reproduces the division size distribution similar to that in Fig. 7(c). Further, the plots in Fig. 8 perfectly reproduce the assumed weak sizer property. As a result, the NN model can effectively extract the relationship between variables hidden by noise and the correlation among different variables.

## V. CONCLUSION AND DISCUSSION

In this study, we have applied the NN method to the problem of cell size control. Our method describes the size dynamics as a history-dependent temporal point process and directly models the integral of its intensity function by NNs. Because it is free of the assumption of any specific functional shape for the size distribution or its dependence on the size history, our method enjoys extremely flexible expressive power, whose performance was confirmed by using the time-series data for *E. coli* and *S. pombe*.

One notable advantage of this method is that it can automatically separate two factors in size determination: a history-dependent deterministic size control and a stochastic component that cannot be explained only by the size history. With the history-dependent deterministic size control, we confirm that the size control can be well approximated as Markovian, which supports the general assumption (presumed in previous modeling) that the birth size is essential for the next division size. It is also consistent with previous

analyses that used regression models [19,20]. With the second factor, we can clarify that there still remains a large stochastic component that cannot be explained only by the information on past size, no matter how far back the history is considered. This result implies that the division is determined not only by the size but also by the interference of other processes in the self-replication of a cell [1].

A substantial portion of stochasticity, which cannot be explained by past sizes, hampers the discovery of unknown relations from data [12,30]. The NN method flexibly represents the stochasticity in data in the form of a probability distribution and its typical behavior as its representative value. This representation allows us to identify the underlying relationships between variables that are difficult to capture clearly in the conventional scatterplotting or binning of data. We verified that the size control relations obtained by the NN are consistent with those of the adder model of *E. coli* and capture the sizer aspect in *S. pombe*.

These results demonstrate that the NN method is an extremely powerful tool to semiautomatically separate deterministic and stochastic factors from measured data. It can be developed further as a complementary method to the conventional modeling method.

Because the NN model that we used was developed originally for history-dependent temporal point processes [24], the present method is also directly applicable to the cell cycle duration—that is, the division interval. The division interval is another phenotypic variable that is influenced by and interferes with cell size control. Similarly to size, the division intervals of mother and daughter cells show correlations. Although models of cell size control typically predict negative correlations [1,16], the correlation of division intervals can be positive in a subset of bacterial experiments and most observations of mammalian cells [31]. The positive correlation suggests that the division interval is a heritable quantity over generations, whose dynamics have recently been inferred as latent state dynamics from cellular lineage trees [32]. The NN method we employed may contribute to disentangling the relationship between cell size and division interval.

Further, we may extend its architecture to incorporate cell size, division interval, and other variables (such as gene expression) as a multidimensional history. The flexible expression power of the NN method is indispensable for elucidating the complicated mechanisms of cell division in which various factors are involved. However, when we incorporate multiple factors, the problem of causality must be addressed. In the formulation of the point process, a division event is the objective variable, whereas other variables, including past division events, are treated as explanatory variables. When only size is concerned, as in this study, the causal relationship between the objective and the explanatory variables is evident. However, when other factors are involved, the true causal relationship among them is not known *a priori*. For example, we should distinguish whether a high gene expression before division induces cell division with a long division interval or the long division interval leads to high gene expression. To address such problems, it is important to develop an NN method that can integrate causal inference, data from intervention experiments, and techniques such as the dual reporter system [33].

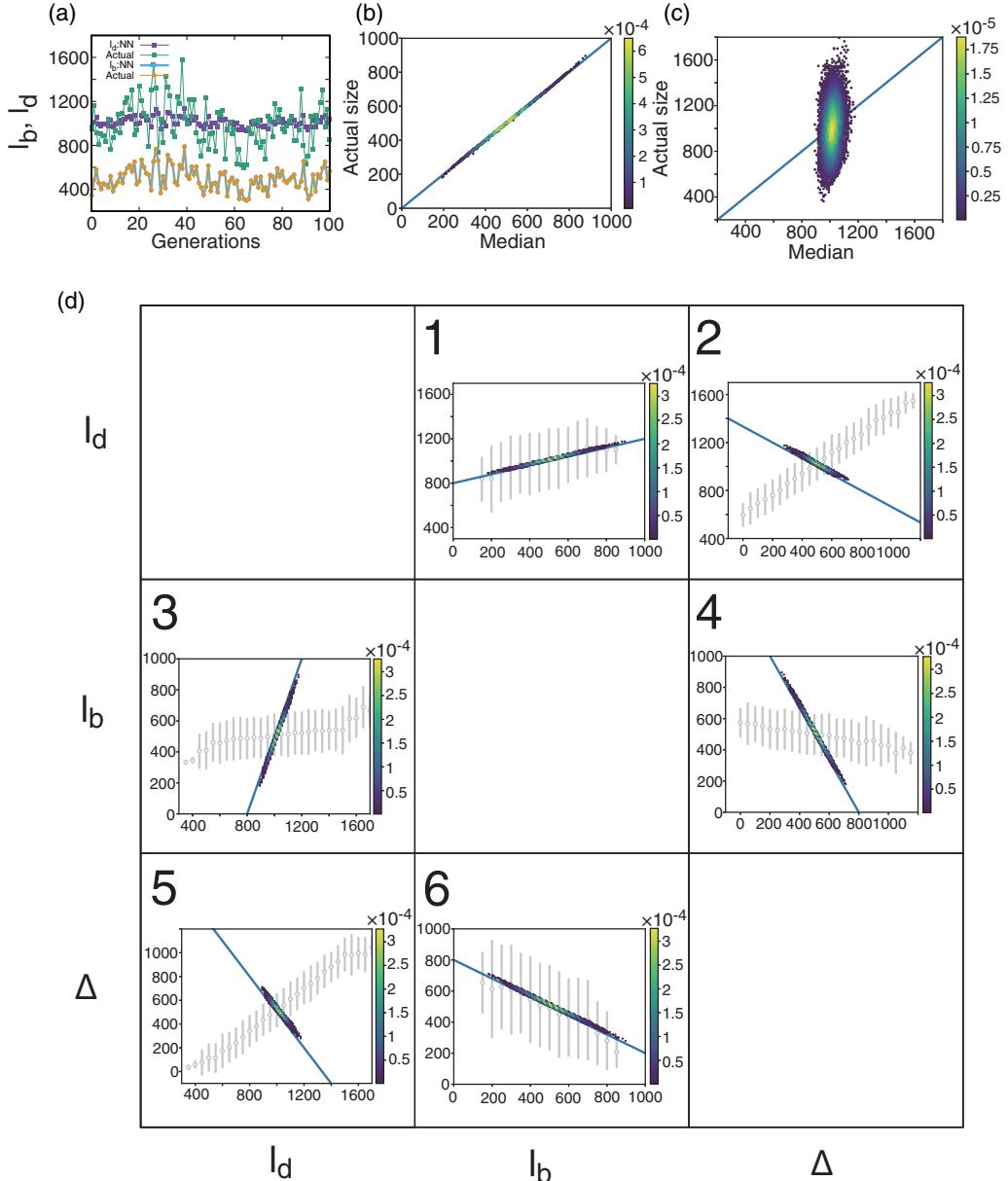


FIG. 8. Synthetic data for the weak sizer model. The alternating sequences of  $l_b$  and  $l_d$  are synthesized as follows: For the birth size, a perfect binary division is assumed such that  $l_b^i = 0.5l_d^{i-1}$ , in which the superscript denotes the generation. For the division size  $l_d^i = l_b^i + z^i$ , the added size  $z^i$  is drawn from the normal distribution  $N(\mu, \sigma^2)$  with  $\mu = -0.6l_b^i + 800$  and  $\sigma = 200$ . (a) Trajectories of the actual birth and division sizes and medians of the PDFs obtained by the NN model. (b), (c) Density plots of the actual sizes and medians of the PDFs for (b) birth size and (c) division size. Identity lines are shown in blue. (d) All correlations between all pairs of the three parameters:  $l_d$ ,  $l_b$ , and  $\Delta$ . Each density plot shows the correlation between each pair of parameters calculated from the medians of the PDFs. With a birth size of  $l_b$ , the PDF for the division size is used to calculate the division size ( $l_d$ ) and the added size  $\Delta = l_d - l_b$ . Blue lines indicate the weak sizer model, with  $\mu = -0.6l_b + 800$ . Open gray circles and error bars indicate the summary statistics of the binned raw data. Circles and bars represent the medians in the binned range and their interquartile ranges, respectively.

#### ACKNOWLEDGMENTS

We thank Yuichi Wakamoto, Yuki Sugiyama, So Nakashima, and Dimitri Loutchko for fruitful discussions. This work was supported by JST CREST Grants No. JP-MJCR1927 and No. JPMJCR2011 and JSPS KAKENHI Grants No. 19H03216 and No. 19H05799.

#### APPENDIX A: PREPARATION OF DATA

The birth size  $l_b$ , division size  $l_d$ , and division interval  $\tau$  are defined from the *E. coli* data [25] as follows: The data were recorded every minute, and the division events were indicated by “division flags” (1 if division occurred). The birth size  $l_b$  is defined by the cell length at which each division occurred, and the division size  $l_d$  is defined by the length at the measurement

immediately before division. The division interval  $\tau$  is defined as the time between consecutive division flags.

In the previous study, the measurement was conducted for each cell over 70 generations: 70  $l_b$ s, 69  $l_d$ s, and  $\tau$ s. We combine the data for all the measured cells under each of the three growth conditions ( $37^\circ\text{C}$ ,  $27^\circ\text{C}$ ,  $25^\circ\text{C}$ ) to assure sufficient training samples for learning. It contains a total of 160, 54, and 65 mother cell lineages at  $37^\circ\text{C}$ ,  $27^\circ\text{C}$ , and  $25^\circ\text{C}$ , respectively.

For *S. pombe*, we use the data published in Ref. [8]. The data were recorded every 3 min and contained millions of time slices. Here, the authors of the study introduced and defined several indices in each time slice to handle the large data as follows (see Ref. [8] for further details). The status of a cell (alive or dead) was indicated by the index “LastIndex.” If the index was 0, the cell was alive. If the index was 1, it survived at the end of the tracking in which the index “NextCell” was set to 0. If the LastIndex was set to 2 or greater, the cell died or disappeared from the channel. The division events were indicated by indices “MergeIndex” and “NewBornCell.” If the index MergeIndex was set to 1, the cell would undergo cell division by the next time slice. If the index NewBornCell was set to 1, the cell was divided between the current and the immediately preceding time slices.

The parameters for *S. pombe* are defined by monitoring the above indices as follows: The birth size  $l_b$  is defined by the cell area in which the LastIndex  $\leq 1$  and NewBornCell = 1. The division size  $l_d$  is defined by the cell area in which the LastIndex  $\leq 1$  and MergeIndex = 1. The division interval  $\tau$  is defined as the time between consecutive realizations of LastIndex  $\leq 1$  and MergeIndex = 1. Here, the time is not calculated if the end of tracking occurs (NextCell = 1) between the realizations.

For each of the seven culture conditions with different media [yeast extract medium (YE) and Edinburgh minimum medium (EMM)] and temperatures, the study introduced a time window in which stable growth was achieved [e.g.,  $t \geq 3000$  (min) under YE conditions]. To select the data in the time window, we monitor an index “Slice” in addition to the above indices. For example, the parameters are defined only if slice  $i$  satisfies  $i \geq 1000$  for YE conditions in which slice  $i$  corresponds to  $t = 3(i - 1)$  (min).

## APPENDIX B: TRAINING AND VALIDATION OF THE NEURAL NETWORK MODEL

For both data sets of *E. coli* and *S. pombe*, the sequences of the birth size  $l_b$  and the division size  $l_d$  are divided into

training and validation data. Unless otherwise mentioned, the first 80% and the last 20% of the sequences are used for training and validation, respectively. In the training phase, the model parameters are estimated by using the training data. For the optimization, the Adam optimizer with a learning rate of 0.001 is used, with the batch size set to 256 [34].

We perform the training of  $l_b$  and  $l_d$  separately as follows. For the division size  $l_d$ , the sequences of  $l_b$  and  $l_d$  are formatted in the form of  $\mathbf{h}_m$  in Eqs. (1) and (3) for a given length  $m$ . To calculate the PDF  $P(l|\mathbf{h}_m)$  of the next division size of  $l = l_d$ , the model parameters are estimated using the collection of sequences  $\mathbf{h}_m$ s. Similarly, the sequences of  $l_b$  and  $l_d$  are formatted in the form of  $\mathbf{g}_m$  in Eq. (2) and (4) to calculate the PDF  $Q(l|\mathbf{g}_m)$  of the birth size  $l_b$ .

In the validation phase, we evaluate the performance of the trained models for  $l_b$  and  $l_d$  using the validation data. For each generation time  $i$ , the PDF of the division size  $l_d$  is calculated using Eq. (9) and scored by the negative log-likelihood (NLL),  $-\log P(l_d^*|\mathbf{h}_m)$  for the actual observation  $l_d^*$  at  $i$ . Here, a smaller score indicates a better predictive performance for the validation data. Finally, the average NLL (MNLL) is calculated for  $l_d$ s for different generations within the validation data. Similarly, the performance is scored for a birth size of  $l_b$ .

The training and validation of the NN are performed by using the code provided by Omi *et al.* [24] while making necessary modifications specifically to distinguish  $\mathbf{h}_m$  and  $\mathbf{g}_m$ . The code uses TensorFlow 2.0.0.

Unless otherwise mentioned, the hyperparameters of the NN model are fixed as follows: The length of histories is  $m = 10$ . In the feed-forward network to learn the integral of the intensity function, the number of units in each layer is 64, and the number of layers is 2. The number of units in the RNN to embed the history is fixed to 64. For the piecewise model explained in Sec. II B, we fix  $L = 128$  and  $\tau_{\max} = 1.001 \times l_{\max}^{\text{data}}$ , where  $l_{\max}^{\text{data}}$  denotes the maximum size in the data.

To characterize typical behavior of trained PDFs, we calculate their medians by following Omi *et al.* [24]. The median  $\tilde{l}$  is calculated by solving  $\Lambda(\tilde{l}) = \log(2)$ , in which  $\Lambda(l)$  denotes the integral of the intensity function. We also calculate the characteristics of size control models as follows. Considering the most recent actual birth size  $l_b$  in the history  $\mathbf{h}_m$ , the added size is calculated as  $\Delta = \tilde{l}_d - l_b$ . The elongation  $\alpha\tau = \log(\tilde{l}_d/l_b)$  is also computed from the median. Here, we assume the equation  $l_d = l_b \exp(\alpha\tau)$  because the majority of *E. coli* cells elongate exponentially with time and do not display significant growth rate variations at specific cell stages (see Fig. S16 in the Supplemental Material [26]).

- 
- [1] S. Jun, F. Si, R. Pugatch, and M. Scott, Fundamental principles in bacterial physiology—history, recent progress, and the future with focus on cell size control: A review, *Rep. Prog. Phys.* **81**, 056601 (2018).
  - [2] S. Taheri-Araghi, S. D. Brown, J. T. Sauls, D. B. McIntosh, and S. Jun, Single-cell physiology, *Annu. Rev. Biophys.* **44**, 123 (2015).
  - [3] T. A. Duncombe, A. M. Tentori, and A. E. Herr, Microfluidics: Reframing biological enquiry, *Nat. Rev. Mol. Cell Biol.* **16**, 554 (2015).
  - [4] A. Grünberger, W. Wiechert, and D. Kohlheyer, Single-cell microfluidics: Opportunity for bioprocess development, *Curr. Opin. Biotechnol.* **29**, 15 (2014).
  - [5] F. S. Fritzs, C. Dusny, O. Frick, and A. Schmid, Single-cell analysis in biotechnology, systems biology, and biocatalysis, *Annu. Rev. Chem. Biomol. Eng.* **3**, 129 (2012).
  - [6] P. Wang, L. Robert, J. Pelletier, W. L. Dang, F. Taddei, A. Wright, and S. Jun, Robust growth of escherichia coli, *Curr. Biol.* **20**, 1099 (2010).

- [7] M. Hashimoto, T. Nozoe, H. Nakaoka, R. Okura, S. Akiyoshi, K. Kaneko, E. Kussell, and Y. Wakamoto, Noise-driven growth rate gain in clonal cellular populations, *Proc. Natl. Acad. Sci. USA* **113**, 3251 (2016).
- [8] H. Nakaoka and Y. Wakamoto, Aging, mortality, and the fast growth trade-off of *schizosaccharomyces pombe*, *PLoS Biol.* **15**, e2001109 (2017).
- [9] S. Cooper and C. E. Helmstetter, Chromosome replication and the division cycle of *Escherichia coli* Br, *J. Mol. Biol.* **31**, 519 (1968).
- [10] J. J. Tyson and O. Diekmann, Sloppy size control of the cell division cycle, *J. Theor. Biol.* **118**, 405 (1986).
- [11] M. Osella, E. Nugent, and M. Cosentino Lagomarsino, Coconcerted control of *Escherichia coli* cell division, *Proc. Natl. Acad. Sci. USA* **111**, 3431 (2014).
- [12] P. P. Pandey, H. Singh, and S. Jain, Exponential trajectories, cell size fluctuations, and the adder property in bacteria follow from simple chemical dynamics and division control, *Phys. Rev. E* **101**, 062406 (2020).
- [13] H. Teimouri, R. Mukherjee, and A. B. Kolomeisky, Stochastic mechanisms of cell-size regulation in bacteria, *J. Phys. Chem. Lett.* **11**, 8777 (2020).
- [14] C. Jia, A. Singh, and R. Grima, Cell size distribution of lineage data: Analytic results and parameter inference, *iScience* **24**, 102220 (2021).
- [15] W. Voorn, L. Koppes, and N. Grover, Mathematics of cell division in *Escherichia coli*: Comparison between sloppy-size and incremental-size kinetics, *Curr. Top. Mol. Gen.* **1**, 187 (1993).
- [16] A. Amir, Cell Size Regulation in Bacteria, *Phys. Rev. Lett.* **112**, 208102 (2014).
- [17] J. T. Sauls, D. Li, and S. Jun, Adder and a coarse-grained approach to cell size homeostasis in bacteria, *Curr. Opin. Cell Biol. Cell Architect.* **38**, 38 (2016).
- [18] Y. LeCun, Y. Bengio, and G. Hinton, Deep learning, *Nature* **521**, 436 (2015).
- [19] P.-Y. Ho, J. Lin, and A. Amir, Modeling cell size regulation: From single-cell-level statistics to molecular mechanisms and population-level effects, *Annu. Rev. Biophys.* **47**, 251 (2018).
- [20] M. Kohram, H. Vashistha, S. Leibler, B. Xue, and H. Salman, Bacterial growth control mechanisms inferred from multivariate statistical analysis of single-cell measurements, *Curr. Biol.* **31**, 955 (2021).
- [21] P. Brémaud, *Point Processes and Queues: Martingale Dynamics* (Springer, Berlin, 1981).
- [22] N. Du, H. Dai, R. Trivedi, U. Upadhyay, M. Gomez-Rodriguez, and L. Song, Recurrent marked temporal point processes: Embedding event history to vector, in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (ACM, San Francisco, CA, 2016), pp. 1555–1564.
- [23] H. Jing and A. J. Smola, Neural survival recommender, in *Proceedings of the Tenth ACM International Conference on Web Search and Data Mining—WSDM '17* (ACM Press, Cambridge, UK, 2017), pp. 515–524.
- [24] T. Omi, N. Ueda, and K. Aihara, Fully neural network based model for general temporal point processes, in *Advances in Neural Information Processing Systems 32 (NeurIPS 2019)*, Vancouver, BC, Canada, edited by H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett (Curran Associates, Red Hook, NY, 2019), pp. 2120–2129.
- [25] Y. Tanouchi, A. Pai, H. Park, S. Huang, N. E. Buchler, and L. You, Long-term growth data of *Escherichia coli* at a single-cell level, *Sci. Data* **4**, 170036 (2017).
- [26] See Supplemental Material at <http://link.aps.org/supplemental/10.1103/PhysRevResearch.3.033032> for supplementary figures, tables, and other data cited.
- [27] D. W. Adams and J. Errington, Bacterial cell division: Assembly, maintenance and disassembly of the z ring, *Nat. Rev. Microbiol.* **7**, 642 (2009).
- [28] T. D. Pollard and J.-Q. Wu, Understanding cytokinesis: Lessons from fission yeast, *Nat. Rev. Mol. Cell Biol.* **11**, 149 (2010).
- [29] J.-B. Nobs and S. J. Maerkl, Long-term single cell analysis of *S. pombe* on a microfluidic microchemostat array, *PLOS ONE* **9**, e93466 (2014).
- [30] F. Bertaux, J. v. Kügelgen, S. Marguerat, and V. Shahrezaei, A bacterial size law revealed by a coarse-grained model of cell physiology, *PLoS Comput. Biol.* **16**, e1008245 (2020).
- [31] T. Nozoe and E. Kussell, Cell Cycle Heritability and Localization Phase Transition in Growing Populations, *Phys. Rev. Lett.* **125**, 268103 (2020).
- [32] S. Nakashima, Y. Sugiyama, and T. J. Kobayashi, Lineage EM algorithm for inferring latent states from cellular lineage trees, *Bioinformatics* **36**, 2829 (2020).
- [33] M. B. Elowitz, A. J. Levine, E. D. Siggia, and P. S. Swain, Stochastic gene expression in a single cell, *Science* **297**, 1183 (2002).
- [34] D. P. Kingma and J. Ba, Adam: A method for stochastic optimization, [arXiv:1412.6980](https://arxiv.org/abs/1412.6980).