Stability estimation of autoregulated genes under Michaelis-Menten-type kinetics

Babak M. S. Arani[*]
Department of Aquatic Ecology and Water Quality Management, Wageningen University, P.O. Box 47, NL-6700 AA Wageningen, The Netherlands

Mahdi Mahmoudi[†]
Faculty of Mathematics, Statistics and Computer Science, Semnan University, P.O. Box 35195-363, Semnan, Iran

Leo Lahti[‡]
Department of Mathematics and Statistics, University of Turku, FI-20014 Turku, Finland

Javier González[§]
Department of Mathematics and Statistics, Lancaster University, Lancaster LA1 4YF, United Kingdom and Amazon Research Cambridge, Cambridge, United Kingdom

Ernst C. Wit[‖]
Institute of Computational Science, USI, Via G. Buffi 13, Lugano 6900, Switzerland

(Received 3 April 2017; published 11 June 2018)

Feedback loops are typical motifs appearing in gene regulatory networks. In some well-studied model organisms, including *Escherichia coli*, autoregulated genes, i.e., genes that activate or repress themselves through their protein products, are the only feedback interactions. For these types of interactions, the Michaelis-Menten (MM) formulation is a suitable and widely used approach, which always leads to stable steady-state solutions representative of homeostatic regulation. However, in many other biological phenomena, such as cell differentiation, cancer progression, and catastrophes in ecosystems, one might expect to observe bistable switchlike dynamics in the case of strong positive autoregulation. To capture this complex behavior we use the generalized family of MM kinetic models. We give a full analysis regarding the stability of autoregulated genes. We show that the autoregulation mechanism has the capability to exhibit diverse cellular dynamics including hysteresis, a typical characteristic of bistable systems, as well as irreversible transitions between bistable states. We also introduce a statistical framework to estimate the kinetics parameters and probability of different stability regimes given observational data. Empirical data for the autoregulated gene SCO3217 in the SOS system in *Streptomyces coelicolor* are analyzed. The coupling of a statistical framework and the mathematical model can give further insight into understanding the evolutionary mechanisms toward different cell fates in various systems.

DOI: 10.1103/PhysRevE.97.062407

I. INTRODUCTION

Feedback interactions are essential components in a genomic network to shape cellular functions. There are many examples of important feedback loops in every organism. Naturally occurring oscillators, such as Cdc2, have intricate feedback mechanisms that allow a sustained oscillation [1]. The $p53$-MDM2 feedback loop, in which the tumor suppressor protein $p53$ activates the gene MDM2, is negatively regulated by MDM2 [2,3]. About 40% of the known transcription factors (TFs) in *Escherichia coli* (*E. coli*) regulate their own transcription [4,5]. Often only noisy data on sparsely spaced time points are available to make sense of such systems.

In this paper, we focus on a special class of feedback loops in gene regulatory networks (GRNs), the so-called autoregulation loops. Autoregulated genes are the genes that are regulated by the TF they encode. Interestingly, in *E. coli* no transcriptional feedback cycles have been found besides autoregulation loops [6,7]. In fact, the *E. coli* transcriptional network is loosely cross connected; on average, a TF regulates three genes and any gene is regulated by two TFs [6] only. The mean network connectivity gets even less at the level of operon interactions [6]. One reason for low cross regulation is that it might be less expensive for a gene to control its regulation through its protein product than by another protein.

Both positive and negative autoregulated genes have their own biological functions. Autoinhibition, which is more common in *E. coli*, controls homeostatic regulation of the repressor gene and the genes it regulates. This stabilizes the GRN against

[*]babak.shojaeirani@wur.nl
[†]mahmoudi@semnan.ac.ir
[‡]leo.lahti@iki.fi
[§]gojav@amazon.com; also at Amazon Research Cambridge, Cambridge, United Kingdom; this work was done before joining Amazon Research Cambridge.
[‖]e.c.wit@rug.nl

cellular perturbations. Positive autoregulated genes, on the other hand, can switch between bistable states and lead to cell differentiation. This genetic switch can therefore affect other genes controlled by such a gene, especially when it has a high degree of connectivity. In *E. coli*, for example, the positive autoregulated gene cAMP receptor protein (CRP), which regulates catabolic repression, has the highest degree of connectivity despite a low mean connectivity of the entire transcriptional network [6,8]. The effect of a genetic switch can be even stronger if the activator gene jumps into an irreversible state (see Sec. II C and Fig. 5).

To model autoregulation, one common approach is to consider linear activation models (e.g., Ref. [5]), where an exact steady-state solution of an autoregulated gene can be obtained. However, more realistic generalized Michaelis-Menten (GMM) or Hill types of kinetic models produce a wider range of dynamic behavior and fit better to the available data. They are able to model a bistable reaction of autoregulated genes in response to changes in cellular conditions. Structural changes in the kinetic parameters of the system can also lead to a hysteretic reaction, when the state of the autoregulated gene depends not only on its current condition but also on its past ones. Moreover, an irreversible genetic switch is possible in some cases, when the transition between the bistable modes of the autoregulated gene is unidirectional.

In Sec. II of this paper we use a coupled deterministic system of differential equations to model over time the average quantitative behavior of gene expression levels and protein abundances in a single cell. Although autoregulation is very common, it often involves modification and other forms of cooperativity by other molecules, which are not included in our model. Nevertheless, cooperativity within an autoregulated system is possible, as shown in Sec. II C, and the model is also appropriate as a phenomenological model to describe allostery, as discussed in Sec. II D. Since our goal is to understand the stability behavior of autoregulated genes measured with noise, in Sec. III we combine our analysis with some aspects of the modern statistical inference of dynamical systems. Although our emphasis is on genomics, the phenomena of bistability and hysteresis are very common at larger scales. Ecosystems, such as lakes, coral reefs, woodlands, deserts, and oceans, can shift between alternative stable states [9].

II. GRN STABILITY DYNAMICS

A. Gene autoregulation

According to the central dogma of molecular biology, each messenger ribonucleic acid (mRNA) molecule produced in the nucleus of the cell encodes the genetic information to produce a protein. Such proteins are the building blocks of life and may have structural functions, such as enzymatic properties. Some of them, however, activate or repress the transcription of other genes. These proteins are called transcription factors and, together with the genes they regulate, form a GRN.

When a gene regulates itself, a loop appears in the GRN (see Fig. 1). By the principle of mass-action kinetics it is natural to assume that the gene expression, on average, changes according to the following ordinary differential

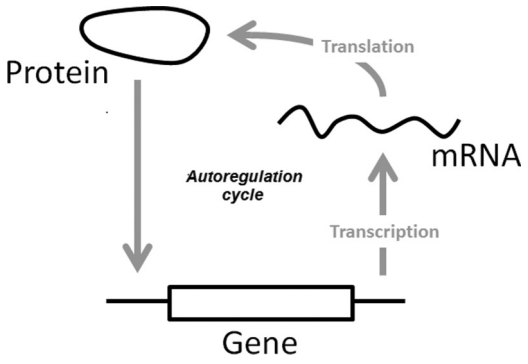

FIG. 1. Illustration of the transcription-translation cycle of an autoregulated gene.

equation (ODE),

$$\dot{x}(t) = p(t;\theta,z) - \delta x(t), \tag{1}$$

where $x(t)$ represents the mRNA concentration at time t, δ is the degradation rate of mRNA, and $p(t;\theta,z)$ is a transcription function that describes how the TF z regulates the gene given some set of parameters θ. Reversely, the TF z is encoded by the gene according to

$$\dot{z}(t) = \rho x(t) - \tau z(t), \tag{2}$$

where τ is the protein degradation rate and ρ is the translational rate of the gene.

Several models have been considered in the literature to define $p(t;z,\theta)$ in (1) ranging from linear approaches [10] to nonparametric methods [11]. In practice, experimental work suggests that the response of the mRNA abundance to the concentration of a TF follows a Hill curve [12]. This response can be well described by the family of Michaelis-Menten (MM) models. In the case of gene activation, the transcription function is assumed to satisfy

$$p^{+}(t;\theta,z) = \beta \frac{z^m}{\gamma + z^m} + \varphi,$$

for $\theta = \{\varphi,\beta,\gamma\}$ and $m \in \mathbb{N}$. In this model, the parameter φ is able to detect possible nonspecific activation. More precisely, the parameter φ is the basal transcription rate—usually zero for most *in vitro* data. The parameter β describes the maximum speed by which the TF regulates the gene [13,14]. The parameter γ represents the dissociation constant of TF from its DNA binding site [15]. Finally, the parameter m is called the Hill coefficient. This parameter exhibits a level of cooperativity, usually less than the number of DNA binding sites [16], in which a high Hill coefficient is representative of a high degree of cooperativity. Similarly, in the case of gene repression, the response can be modeled by

$$p^{-}(t;\theta,z) = \beta \frac{1}{\gamma + z^m} + \varphi.$$

In this paper, our focus will be on the case of gene activation. The system is always stable under gene repression and exhibits smooth behavior in response to changes in the parameters. See Supplemental Material for all mathematical proofs [17]. We will study the stability properties of a family of MM kinetics models.

B. Stability of MM kinetics models

Under the MM kinetics the interaction between TF and mRNA in an autoregulated gene occurs according to the following planar system of differential equations,

$$\dot{x} = \beta \frac{z}{\gamma + z} + \varphi - \delta x,$$
$$\dot{z} = \rho x - \tau z, \tag{3}$$

where we assume that all the parameters are positive and state variables x and z lie in the positive quadrant $(0,\infty)^2$. We have the following result.

Result 1. System (3) has a unique equilibrium in the positive quadrant and it is globally asymptotically stable.

Figure 2 illustrates Result 1, which shows various solutions of system (3) for $\beta = 6$, $\rho = 5$, $\delta = 0.5$, $\tau = 1$, $\gamma = 5$, and $\varphi = 0.2$ when different initial conditions $x(0)$ and $z(0)$ are considered. The equilibrium points, or values in which constant functions are solutions of the system (3) (horizontal dotted lines), are unique. Also, the solutions of the ODE for different initial points converge with t to the equilibrium point due to the global asymptotic stability of the system.

C. Hill coefficient 2: Hysteresis, bistability, and irreversible transition

In this section we deal with a more complicated family of MM kinetics models. In particular, we focus on cases where the transcription function takes the form $\beta \frac{z^2}{\gamma + z^2} + \varphi$. We show that the corresponding system of ODEs exhibits a richer class of dynamical behavior compared to the standard MM kinetics models. One "limitation" of the approach taken is that, to justify a generalized Michaelis-Menten equation with a Hill coefficient m of 2 or more, one needs cooperativity. If the regulator protein binds DNA as a dimer, which is frequently the case in bacterial signal transduction, then $m = 2$. Although it is more common in multiple species systems, this can occur in autoregulated systems, such as the recently described membrane-associated RING-CH 1 protein (MARCH1) regulator [18]. We will show that the generalized family of MM equations have the capability to represent bistability as well as hysteresis, which is a characteristic of positive feedback loops that the standard family of MM models cannot represent.

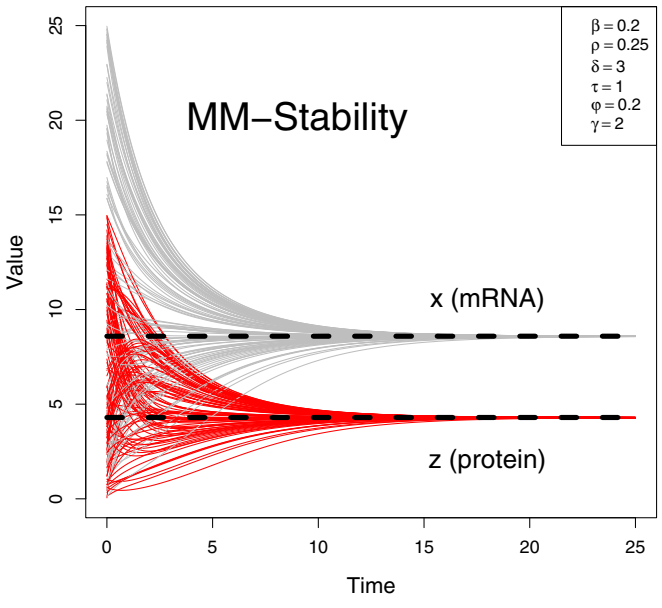

FIG. 2. Illustration of Result 1, the stability of a system with Michaelis-Menten formulation. The horizontal dotted lines represent the equilibrium point of the system. The solutions of the system converge to the equilibrium point for different initial conditions.

For the generalized MM system with Hill coefficient 2,

$$\dot{x} = \beta \frac{z^2}{\gamma + z^2} + \varphi - \delta x,$$
$$\dot{z} = \rho x - \tau z, \tag{4}$$

the following result explains its core dynamics.

Result 2. Let $A = (\beta + \varphi)\frac{\rho}{\delta\tau}$, $B = \gamma\varphi\frac{\rho}{\delta\tau}$, and

$$\Delta = 18\gamma AB - 4A^3B + \gamma^2A^2 - 4\gamma^3 - 27B^2. \tag{5}$$

Then, the equilibria of (4) lie only in the positive quadrant. Moreover:

(a) (Stability) If $\Delta < 0$, then (4) has a unique equilibrium and it is globally asymptotically stable.

(b) (Alternative stable states) If $\Delta > 0$, then (4) has three equilibria: two alternative stable equilibria separated by a saddle in between. Hence, system (4) is bistable.

(c) (Tipping point) If $\Delta = 0$, then (4) has two equilibria: a stable equilibrium and a nonhyperbolic one in which (4) undergoes a saddle-node bifurcation.

The model parameters are considered fixed in a standard analysis. However, note that the model parameters, potentially even the model structure described by the ODE, can change under changing environmental conditions. The maximum protein production rate β, the dissociation constant γ, which describes the inverse transcription efficiency, or the basal transcription rate φ could change as a function of temperature, for instance. Hence, environmental changes can therefore potentially bring the system close to a tipping point, where it is prone to abrupt shifts between alternative states [see Figs. 3(a) and 5]. It is also worth noting that factors external to our model, such as the availability of component molecules, or activity of other partially competing processes and binding targets, can affect the process *in vivo*.

Consider Fig. 3(a). Imagine that the system is rested at its upper branch and a certain parameter (here, γ) is continuously increased until at a tipping point (F_2) the system jumps down to the lower branch. If the parameter is then decreased, then the system will jump back to the upper branch at a different tipping point (F_1). In short, the system jumps to another stable branch and jumps back to its original stable branch through the so-called "hysteresis loop." Hysteresis is often associated with bistability [1], although this is not always the case for any bistable system [19], including ours [see Result 3(c)].

Here, the quantity Δ plays a central role in determining the stability dynamics of system (4). We explain how changes in Δ lead to hysteresis: In Fig. 3(a), if Δ is negative, then the system is stable and can rest in only one stable branch (for example, the upper branch). Under a saddle-node bifurcation (F_1), another stable equilibrium and an unstable saddle point bifurcate as certain parameters (for instance, γ) change and Δ crosses

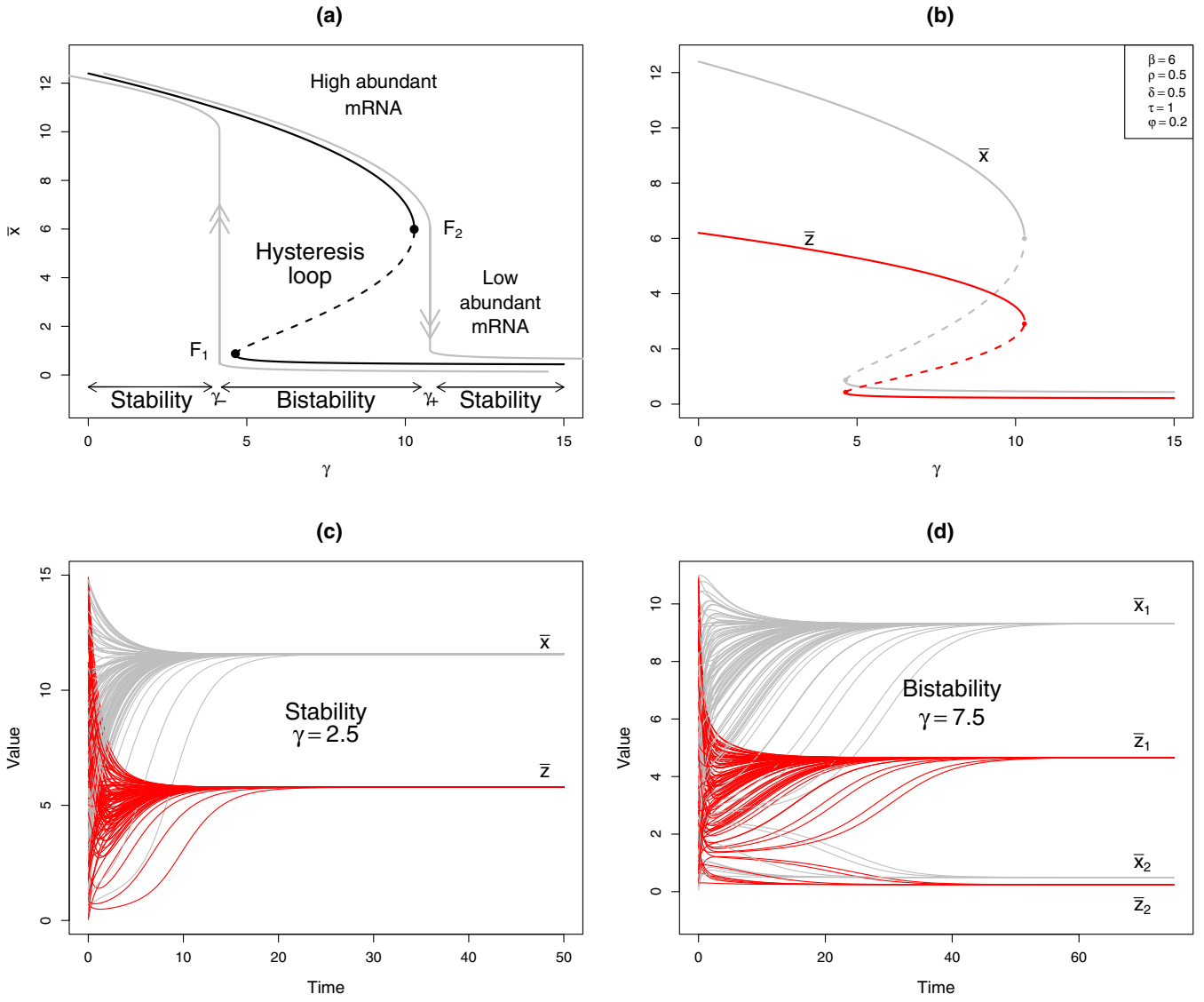

FIG. 3. Illustration of stability scenarios and hysteresis. (a) Bifurcation diagram of the mean gene expression ($\bar{x}$) as the dissociation parameter varies. It describes how the autoregulated gene drops in expression level as the TF dissociates from the promoter region. (b) The same bifurcation diagram as in (a) for both state variables. (c) ODE solution which is convergent to an equilibrium ($\gamma = 2.5$ is in the stable region). (d) The ODE solution can converge to two distinct equilibria (based on the choice of initial values) since the system is bistable ($\gamma = 7.5$ is in the bistability region).

into positive values. However, the system continues to follow the upper branch. If the parameters continue to change and at some threshold bounce the Δ back to zero again (occurrence of second bifurcation, F_2), the upper branch and an unstable dashed curve coalesce and turn into an unstable equilibrium. If further changes in the parameters make Δ negative, then this equilibrium disappears so that the system has to jump down to the lower branch. Conversely, the system jumps back to the upper branch at the first bifurcation point once the parameters change in the opposite direction. Note that the system cannot jump back to its original state at the second bifurcation point. This means that further changes of the parameters in opposite directions is necessary for the system to get back to its primary stable branch.

The bistable nature of the system (4) leads to a bimodal distribution of protein abundance and gene expression levels. This gives rise to the coexistence of two subpopulations of autoregulated genes: "low protein abundance and gene expression level" and "high protein abundance and gene expression level" (see Figs. 3 and 4).

In hysteresis, transitions between alternative stable states are reversible. However, there is an extra "cost." Biologically speaking, once the cell transits from one mode to the other one in response to changes in cellular conditions, it is able to restore its previous mode once we reverse the biological conditions. However, the same amount of changes in cellular conditions which made a cellular switch is not sufficient for the cell to regain its original mode. An extra cellular change, or cost, is necessary [see Fig. 3(a) and the first example after Result 3].

On the other hand, our model can also predict the irreversible transitions between stable states (see Fig. 5 and the second example after Result 3). This means that transitions are

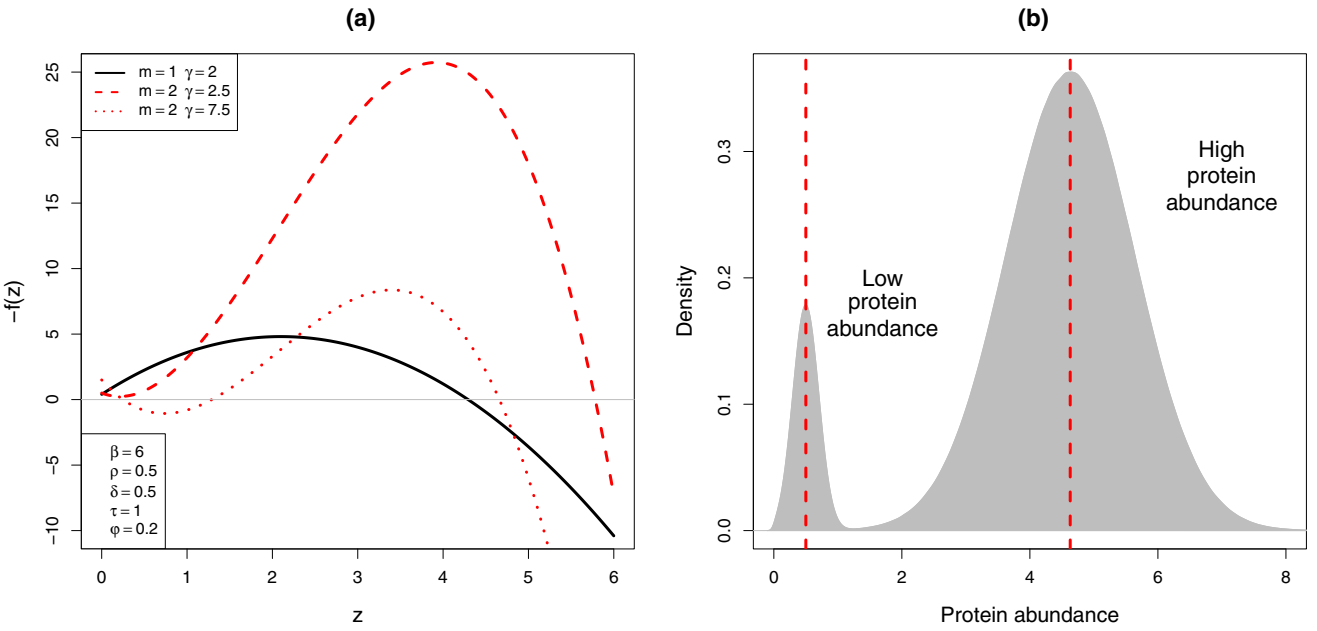

FIG. 4. (a) Solutions of the steady-state equations in the examples of Secs. II B and II C. The MM model shows a positive root whereas the two GMM shows one and three roots, respectively. (b) Bistability at a single cell level induces bimodality at the population level. In this figure we show the density plot of the abundance of a protein in a population of cells following the kinetics parameters detailed in Fig. 3(d).

only possible from one stable state to the other one and not the opposite. This may explain the existence of an interesting biological phenomenon that autoregulated genes can exhibit: Under hysteresis, the autoregulated gene is able to switch between high and low expression levels while this is not the case for transitions due to an irreversible genetic switch (transitions are only possible from low expression levels to high ones but not the opposite). Perhaps, under an irreversible genetic switch, the gene behaves in a "conservative" manner: When the cellular decision for transition is made, the gene enters a state with an everlasting fate.

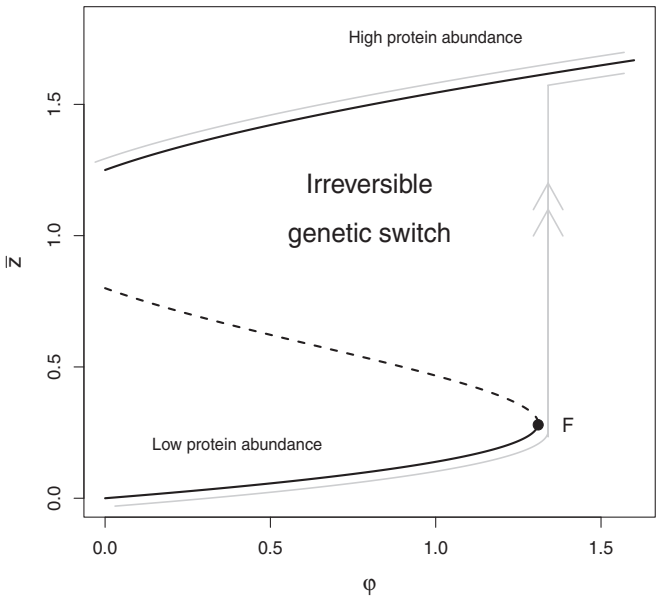

FIG. 5. Illustration of an irreversible transition in which the autoregulated gene shifts from a low expression to a high one. The system jumps up as the basal transcription rate φ passes the tipping point F, but it cannot jump down once φ decreases since the other tipping point lies in the negative range of parameter space. It describes how an autoregulated gene may be subject to an irreversible genetic switch which might, for instance, explain cell differentiation.

The following result gives us two cases under which system (4) can exhibit hysteresis as well as a case in which irreversible transition occurs.

Result 3. (a) (Hysteresis) Assume that all the parameters are fixed, except γ. Then (4) undergoes hysteresis provided that

$$\beta > 8\varphi. \tag{6}$$

(b) (Hysteresis) Assume that all the parameters are fixed, except β. Then (4) undergoes hysteresis provided that

$$\gamma > 27\left(\frac{\varphi\rho}{\delta\tau}\right)^2. \tag{7}$$

(c) (Irreversible transition) Assume that all the parameters are fixed, except φ. Then (4) exhibits an irreversible transition provided that

$$\left(\frac{\beta\rho}{\delta\tau}\right)^2 > 4\gamma. \tag{8}$$

We consider some examples. First, let $\beta = 6$, $\varphi = 0.2$, and choose the rest of the parameters in which we have $\frac{\rho}{\delta\tau} = 1$. Then condition (6) simply holds. Some calculations show that the first and second bifurcations happen at $\gamma_- \approx 4.6340$ and $\gamma_+ \approx 10.2860$ as we increase the parameter γ [see Fig. 3(a)]. Imagine that the system starts at the beginning of the upper leg of Fig. 3(a). Then, as we gradually increase the dissociation parameter γ, the system follows the upper branch. At $\gamma = \gamma_+$, the system has to jump down to the lower branch and rests there as we increase γ further. Reversely, if γ decreases, the system

jumps up to the upper branch at the first bifurcation point $\gamma = \gamma_-$. Biologically, this may mean that the autoregulated gene gradually drops in expression as the TF tends to dissociate more and more from the promoter binding site. Then it switches into a low expression regime as a certain dissociation threshold is passed.

However, if $\gamma = 1$, $\beta = 20.5$, and $\frac{\rho}{\delta\tau} = 0.1$, then condition (8) holds. By Result 3(c) an irreversible transition occurs: As we increase φ, the system jumps up at $\varphi \approx 1.3100$ to the upper branch but never jumps down if we decrease φ (see Fig. 5). Biologically speaking, as cellular perturbation φ increases, the autoregulated gene switches irreversibly. See Ref. [20] for a detailed analysis of the passage from hysteresis to irreversibility in budding yeast.

D. Generalized MM kinetics with $m > 2$

Higher Hill coefficients in natural autoregulated systems are possible. Reference [21] reports cooperativity in several *E. coli* autoregulated genes with Hill coefficients of 3. The model can also be seen as a phenomenological model for describing allosteric cooperativity, which involves the binding of exogenous ligands to the protein, which can result in fractional Hill coefficients [22,23]. Furthermore, the absence of feedback can be obtained by letting $m \to \infty$. For the cases where $m > 2$, no closed form solution exists for the analysis of the dynamical behavior of the following generalized MM formulation,

$$\begin{aligned} \dot{x} &= \beta \frac{z^m}{\gamma + z^m} + \varphi - \delta x, \\ \dot{z} &= \rho x - \tau z. \end{aligned} \tag{9}$$

Nevertheless, the same results hold as for the case $m = 2$. The following result describes the dynamic scenarios, which holds for values of m in the interval (1,2), too.

Result 4. Let $m > 2$, and consider the following polynomial,

$$f(z) = z^{m+1} - (\beta + \varphi)\frac{\rho}{\delta\tau} z^m + \gamma z - \varphi\gamma\frac{\rho}{\delta\tau},$$

which has either one, two, or three positive roots. Moreover:

(a) (Stability) If f has a unique positive root, then (9) has a unique equilibrium in $(0,\infty)^2$ and it is globally asymptotically stable.

(b) (Bistability) If f has three positive roots, then (9) has three equilibria in $(0,\infty)^2$: two alternative stable equilibria with a saddle in between. Hence, (9) is bistable.

(c) (Tipping point) If f has two positive roots, then (9) has two equilibria in $(0,\infty)^2$: a stable equilibrium and a nonhyperbolic one in which (9) undergoes a saddle-node bifurcation.

Therefore, in the case of gene activation, the generalized MM formulation also exhibits bistability and hysteresis for $m > 2$. Using more advanced mathematical methods we were able to extend some of the results of Result 3 to the case of $m > 2$ as follows.

Result 5. (a) (Irreversible versus hysteretic bistability) Assume that all the parameters are fixed except φ and let $\theta = \frac{\beta\rho}{\delta\tau}$. Then (9) exhibits an irreversible transition if

$$\gamma < \bar{\gamma} = \theta^m \frac{(m-1)^{m-1}}{m^m}. \tag{10}$$

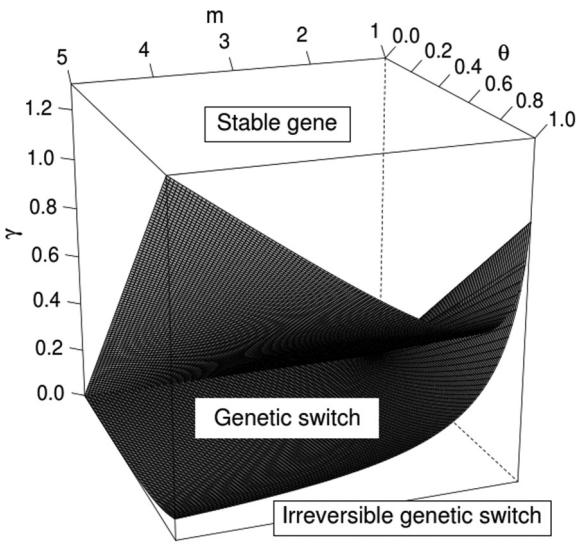

FIG. 6. Illustration of critical transitions undergone by an autoregulated gene. The upper surface [right-hand side of (11)] distinguishes transitions between stable and bistable modes of the gene while the lower surface [right-hand side of (10)] distinguishes transitions between bistable and irreversible modes.

In fact, this is the only case in which one can observe irreversible bistability. Moreover, for higher values of γ, (9) exhibits hysteresis if

$$\gamma < \frac{(m+1)^{m+1}}{4^m}\bar{\gamma}. \tag{11}$$

(b) (Hysteresis) Assume that all the parameters are fixed except β and let $\theta' = \frac{\varphi\rho}{\delta\tau}$. Then (9) undergoes hysteresis provided that

$$\gamma > \theta'^m \left(\frac{m+1}{m-1}\right)^{m+1}. \tag{12}$$

See Fig. 6 for a graphical illustration of this result. For a summary of this and other results of this paper, see Table I.

III. EXPERIMENTS

A. Simulation study

The goal of this section is to illustrate how the results obtained in Sec. II can be used to gain some knowledge about the dynamical behavior of real systems.

In this section, we will focus on the parameter Δ defined in (5), since it contains all the necessary information to know if a system is stable ($\Delta < 0$), bistable ($\Delta > 0$), or if it has a bifurcation point ($\Delta = 0$). Our goal is to infer Δ from noisy samples of gene expression and TF activity levels of autoregulated genes.

A maximum likelihood inference of Δ can be carried out through the estimation of the parameters β, ρ, δ, τ, γ, and φ. By plugging them into (5) we can obtain $\widehat{\Delta}$. To obtain such estimators we use the method proposed in Ref. [24], which has been successfully applied in the identification of both the parameters and hidden components of gene regulatory networks [14]. In the Supplemental Material [17] we have included a description of this approach for the system in (4).

TABLE I. Summary of the main results of this work. See Sec. IV for details and definitions of Δ, θ, θ' as well as further considerations on the nature of the different equilibria. CP is an abbreviation for control parameter.

Hill coefficient	Stability	Bistability	Hysteresis	Irreversible shift
$m = 1$	YES			
	Unique equilibrium	NO	NO	NO
$m = 2$	YES	YES	YES	YES
	Unique equilibrium	Alternative stable	If $\beta > 8\varphi$, CP $= \gamma$	If $\theta^2 > 4\gamma$
	If $\Delta < 0$	states if $\Delta > 0$	If $\gamma > 27\theta'^2$, CP $= \beta$	CP $= \varphi$
$m > 2$[a]			YES	YES
	YES	YES	If $\gamma < \frac{\theta}{4m}(m-1)^{\frac{m-1}{m}}(m+1)^{\frac{m+1}{m}}$	If $\gamma < \theta^m \frac{(m-1)^{m-1}}{m^m}$
	Unique equilibrium	Alternative stable	CP $= \varphi$	
		states	If $\gamma > \theta'^m(\frac{m+1}{m-1})^{m+1}$, CP $= \beta$	CP $= \varphi$

[a]Note that this result also holds for values of m in the interval (1,2).

We analyze the dynamical behavior of (4) with a simulation study where Δ is estimated for different synthetically generated data sets. Motivated by Results 2 and 3, we work with four scenarios in which we fix the values of the parameters of system (4) in such a way that stability, bistability, or bifurcation occur. See Table II for details.

Given the solutions of system (4) in the four previous scenarios, we sample x and z in 100 equally spaced points in the interval [0,4] and we perturb the resulting vector with Gaussian noise with mean zero and variance 0.01. Figures 7(a)–7(d) show the obtained samples in the four cases. Note that although the data sets look similar at first glance, they correspond to three completely different dynamic scenarios.

We repeat the data simulation procedure 500 times. In each case we estimate Δ as mentioned above. Figure 7(e) shows the estimated density functions of Δ for the 500 data sets in the four cases. The mode of the estimated distributions is always close to the true value of Δ, which indicates the ability of the approach to detect different dynamical behaviors. The noise in the data produces a certain amount of variability in the estimates of Δ, which is reflected in the shape of the distribution of $\widehat{\Delta}$. In cases of stability [Fig. 7(a)] the distribution is symmetric around the true value of Δ. The results of this experiment show how the different dynamics of an autoregulated gene can be estimated from noisy data. Notice that this is done for data collected in an interval in which the system did not necessarily converge to an equilibrium point, which shows the power of this approach in scenarios where the ability to sample in long intervals is limited.

TABLE II. Four simulated scenarios. We generate data using system (4). The value of dissociation constant γ changes in order to obtain different dynamic scenarios as illustrated in Fig. 3(a). The rest of the parameters are fixed to $\beta = 6$, $\rho = 0.5$, $\delta = 0.5$, $\tau = 1$, $\varphi = 0.2$, $x(0) = 0$, and $z(0) = 10$. The estimated stability coefficient Δ significantly diverges from 0 in the first two scenarios and it is consistent with 0 in the last two. The dynamics of the system can be inferred from noisy data.

	Dynamics	γ	Δ	$\widehat{\Delta}$	(S.D.)
Scenario A	Stability	2.5	−166.2	−164.6	(10.3)
Scenario B	Bistability	7.5	239.5	251.1	(103.7)
Scenario C	Bifurcation	4.6	0.0	2.8	(18.4)
Scenario D	Bifurcation	10.3	0.0	30.9	(80.6)

B. Autoregulation of the yeast autoregulator INO4

We analyze the stability properties of the INO4 autoregulated gene (Affymetrix probe set 1774516_at) in yeast. We use a synchronized Δbar1 strain of *S. cerevisiae* observed in duplicate across 41 time points, separated by 5 min and totally covering 200 min after synchronization [25]. This corresponds to approximately three cell cycle periods. We have only a partially observed system: We know the mRNA abundances but information about the protein abundances is not available. Therefore, we treated the protein abundance z as a latent variable in parameter estimation. This means that the formulation in (9) is potentially unidentifiable. Fortunately, we can cancel the parameter ρ by rescaling the protein concentration z in (2) without any need to rescale the mRNA concentrations x. We apply the rescaling $z = \rho\zeta$, which gives

$$\dot{x} = \beta\frac{\zeta^m}{\gamma^* + \zeta^m} + \varphi - \delta x,$$
$$\dot{\zeta} = x - \tau\zeta,$$

where $\gamma^* = \gamma/\rho^m$. For notational convenience, we will drop the asterisk in γ^*.

We fit the autoregulatory model for Hill parameters from 1 to 5 using the generalized Tikhonov regularization [26] described in Sec. II D of the Supplemental Material [17]. The minimum Akaike information criterion (AIC) is found for $m = 2$, suggesting that the best fit kinetics is more complex than a simple MM model. The estimated parameters are $\hat{\gamma} = 0.841$, $\hat{\beta} = 0.133$, $\hat{\varphi} = 0.019$, $\hat{\delta} = 0.012$, $\hat{\tau} = 3.374$. By using these values in the latent model $\rho = 1$, we obtain an estimate of the stability parameter $\widehat{\Delta} = -46.4 < 0$. This suggests that the INO4 autoregulation in the yeast system is, in fact, stable. The fit of the system for $m = 2$ is shown in Fig. 8(a).

C. Autoregulated CdaR in *Streptomyces coelicolor*

Next, we consider an experiment involving gene SCO3217 of the *Streptomyces coelicolor* bacterium. This is an au-

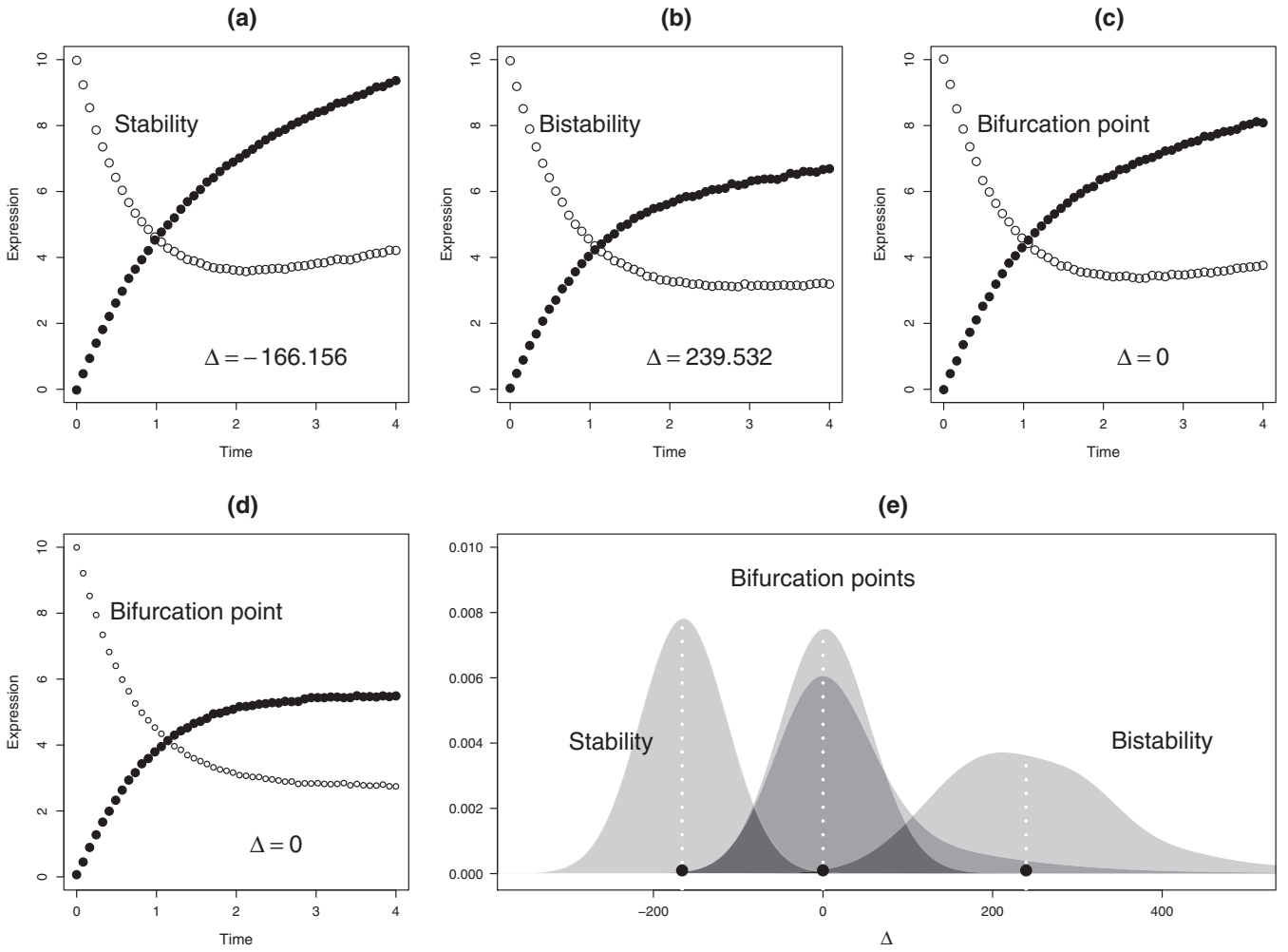

FIG. 7. Simulated data and results of the four scenarios described in Table II. (a)–(d) show the data generated from stable, bistable, and bifurcation scenarios. (e) corresponds to the estimated distributions of $\hat{\Delta}$ in the simulation study. Vertical dotted white lines and black circles represent the true value of Δ for the different scenarios.

toregulated gene that produces the transcription factor *CdaR*. This gene is an important trigger of a cascade of genes that make *Streptomyces coelicolor* produce a calcium-dependent antibiotic. The protein CdaR is an activator TF, so the generalized MM formulation in system (4) is adequate to study its autoregulatory dynamical behavior [27].

The experiment used two-channel microarrays to sample, destructively, a *Streptomyces coelicolor* wild-type strain at different times after chemical induction of the CdaR TF [13]. The *Streptomyces coelicolor* wild-type strain was grown on a solid medium. At each time point two biological replicates were collected. The original data set consists of a data set with ten measurements of the mRNA expressions and the CdaR abundances collected at 16, 18, 20, 21, 22, 23, 24, 25, 39, and 67 min after starting the experiment. At the first time point we observed an increase in both the protein and mRNA expressions, which later converge to a steady-state situation. In this analysis we study such a recovery so we only consider the seven data points collected after the first 20 min. We follow the statistical approach described in Sec. III A to estimate the parameters of the system (4) from the collected data. In order to select the best model to fit the data we use the Akaike information criterion (AIC) to choose between GMM models with Hill coefficients $m = 1,2,3,4$. A comparison between the values of the AIC for the four models shows that the model with $m = 2$ is preferred.

In Fig. 8(b) we show the data points and the smoothed functions obtained from the parameter estimation approach. The obtained parameter estimates are $\hat{\beta} = 1024.8$, $\hat{\rho} = 0.001$, $\hat{\delta} = 1.08$, $\hat{\tau} = 0.001$, $\hat{\gamma} = 976.2$, and $\hat{\varphi} = 0.001$. From these values we can infer the value of Δ and therefore gain some insight about the stability properties of the system. In particular, by plugging the parameter estimates in (5), we obtain that $\widehat{\Delta} = 8.5 \times 10^{11}$, which indicates, following Result 2, that the system is bistable. A caveat about this result is the small size of the data set used in the experiment. We think, however, that this result should be taken into account in a further analysis of this system, since variations in the initial conditions of the experiment may lead to different steady-state solutions for mRNA concentration and protein level.

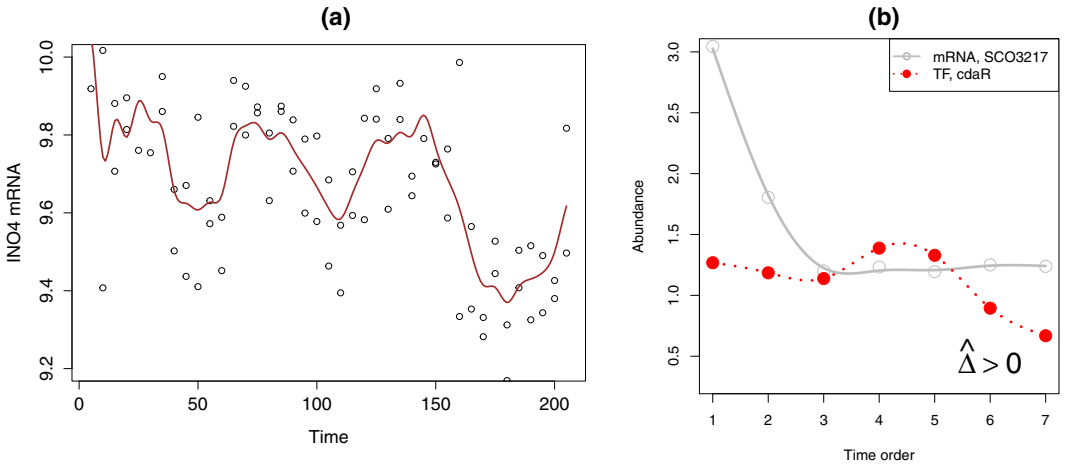

FIG. 8. (a) Observed data for the autoregulated gene INO4 in *S. serevisiae* and the obtained ODE solution after estimating the parameters of the system. The value $\hat{\Delta} < 0$ suggests that the system is stable. (b) Observed data for the autoregulated gene SCO3217 in *Streptomyces coelicolor* and obtained ODE solution after estimating the parameters of system (4). The value $\hat{\Delta} > 0$ suggests that the system is bistable, which may provoke bimodal effects at a population level.

IV. CONCLUSIONS AND FUTURE DIRECTIONS

Autoregulation is a process common in biological systems, such as GRN. For example, autoregulation is the only type of feedback loop existing in the transcriptional network of the well-characterized model organism *E. coli*. The aim of this paper has been to apply quantitative methods to unravel the implications of this mechanism. We were able to manifest diverse cellular scenarios emerging from autoregulation through a rather simple, but realistic, dynamic system model, which is analytically tractable. Although the model parameters are fixed in a standard analysis, they can change due to environmental changes (for instance, as a function of temperature).

The generalized MM kinetics model is capable of predicting the typical properties of positive and negative autoregulated genes: It leads to steady-state solutions in the case of negative autoregulation, which represents homeostatic regulation. It can exhibit bistability in the case of positive autoregulation, which represents a developmental differentiation. In the latter case, the gene can shift between alternative stable states (low and high gene expression levels). Furthermore, in response to gradual changes in the cellular conditions, the generalized MM is able to show a discontinuous switchlike response, which is common in biological systems, including cell cycle regulation and cell differentiation.

We applied the model to two typical noisy time-course expression data involving the autoregulated genes INO4 in *S. cerevisiae* and SCO3217 in *S. coelicolor*. In the first case we find that the overall system is stable, whereas in the second example the situation is more complicated. Our statistical analysis about the dynamical behavior of this autoregulated gene reveals that its kinetic parameters lie well in the bistability region of the generalized MM model. Furthermore, we found a correlation between bistable behavior and the bimodal distribution of gene expressions using simulated data. The model is also capable of exhibiting irreversible shifts between bistable states. Such a phenomenon is representative of an irreversible genetic switch. This can, perhaps, lead to so-called "conservative" cellular decision making since the cell cannot restore its primary state. However, more research is necessary to verify this through experimental data.

Each cell is always subject to cellular noise or perturbations, which can alter cellular activities. Under high noise levels, the cell might alternate between bistable modes with kinetic parameters far from the bifurcation points. It is an intriguing question what the maximum cellular noise can be that the gene can absorb without being tipped into an alternative state. Such a question could perhaps be answered through the Freidlin-Wentzell theory of random perturbations [28] using a suitable potential function, which for gradient systems always exists. However, many systems including GRN are not gradient systems, so therefore it would be interesting if a quasipotential landscape with meaningful biological interpretations could be constructed [29]. An alternative approach would be to extend our deterministic model to the following simple Langevin system with additive noise,

$$\begin{aligned} dx &= \left(\beta \frac{z^m}{\gamma + z^m} + \varphi - \delta x\right) dt + \sigma_1 dW_1, \\ dz &= (\rho x - \tau z) dt + \sigma_2 dW_2, \end{aligned} \tag{13}$$

where W_1, W_2 are Wiener processes and σ_1, σ_2 are the corresponding noise intensities. Then, one can consider the corresponding backward Kolmogorov [30] or backward Fokker-Planck [31] equation of (13) and calculate the mean first exit time from the attraction basins.

ACKNOWLEDGMENTS

J.G. and E.W. are thankful for the support of Project No. SBC-EMA-435065. L.L. was supported by the Academy of Finland (Grants No. 295741 and No. 307127). E.W. acknowledges support from the H2020 COST Action COSTNET (CA15109).

B.M.S.A. and M.M. contributed equally to this work.

[1] J. R. Pomerening, E. D. Sontag, and J. E. Ferrell, Nat. Cell Biol. **5**, 346 (2003).

[2] G. Lahav, N. Rosenfeld, A. Sigal, G. N. Zatorsky, A. J. Levine, M. B. Elowitz, and U. Alon, Nat. Genet. **36**, 147 (2004).

[3] E. Batchelor, A. Loewer, and G. Lahav, Nat. Rev. Cancer **9**, 371 (2009).

[4] N. Rosenfeld, M. B. Elowitz, and U. Alon, J. Mol. Biol. **323**, 785 (2002).

[5] J. E. M. Hornos, D. Schultz, G. C. P. Innocentini, J. Wang, A. M. Walczak, J. N. Onuchic, and P. G. Wolynes, Phys. Rev. E **72**, 051907 (2005).

[6] D. Thieffry, A. M. Huerta, E. Pérez-Rueda, and J. Collado-Vides, Bioessays **20**, 433 (1998).

[7] R. Milo, S. Shen-Orr, S. Itzkovitz, N. Kashtan, D. Chklovskii, and U. Alon, Science **298**, 824 (2002).

[8] H. Ishizuka, A. Hanamura, T. Inada, and H. Aiba, EMBO J. **13**, 3077 (1994).

[9] M. Scheffer, S. R. Carpenter, J. A. Foley, C. Folke, and B. Walker, Nature (London) **413**, 591 (2001).

[10] T. Chen, H. L. He, and G. M. Church, Biocomput. **4**, 29 (1999).

[11] T. Äijö and H. Lähdesmäki, Bioinformatics **25**, 2937 (2009).

[12] H. D. Jong, J. Comput. Biol. **9**, 67 (2002).

[13] R. Khanin, V. Vinciotti, V. Mersinias, C. Smith, and E. Wit, Biometrics **63**, 816 (2007).

[14] J. Gonzalez, I. Vujacic, and E. Wit, Stat. Appl. Genet. Mol. Biol. **12**, 109 (2013).

[15] S. Goutelle, M. Maurin, F. Rougier, X. Barbaut, L. Bourguignon, M. Ducher, and P. Maire, Fundam. Clin. Pharmacol. **22**, 633 (2008).

[16] J. N. Weiss, FASEB J. **11**, 835 (1997).

[17] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevE.97.062407 for the 5 Results described in the main paper and an overview of the inference methods for ODE systems.

[18] M.-C. Bourgeois-Daigneault and J. Thibodeau, J. Immunol. **188**, 4959 (2012).

[19] G. M. Guidi and A. Goldbeter, J. Phys. Chem. A **101**, 9367 (1997).

[20] G. Charvin, C. Oikonomou, E. D. Siggia, and F. R. Cross, PLoS Biol. **8**, e1000284 (2010).

[21] A. Y. Mitrophanov and E. A. Groisman, Bioessays **30**, 542 (2008).

[22] A. Whitty, Nat. Chem. Biol. **4**, 435 (2008).

[23] S. J. Edelstein and N. Le Novère, J. Mol. Biol. **425**, 1424 (2013).

[24] J. González, I. Vujačić, and E. Wit, Pattern Recognit. Lett. **45**, 26 (2014).

[25] P. Eser, C. Demel, K. C. Maier, B. Schwalb, N. Pirkl, D. E. Martin, P. Cramer, and A. Tresch, Mol. Syst. Biol. **10**, 717 (2014).

[26] I. Vujačić, S. M. Mahmoudi, and E. Wit, Stat **5**, 132 (2016).

[27] R. Khanin, V. Vinciotti, and E. Wit, Proc. Natl. Acad. Sci. U.S.A. **103**, 18592 (2006).

[28] M. Freidlin and A. Wentzell, *Random Perturbations of Dynamical Systems* (Springer, Berlin, 1984).

[29] J. X. Zhou, M. D. S. Aliyu, E. Aurell, and S. Huang, J. R. Soc., Interface **9**, 3539 (2012).

[30] A. Kolmogorov, Math. Ann. **104**, 415 (1931).

[31] C. W. Gardiner, J. Opt. Soc. Amer. B: Opt. Phys. **1**, 409 (1984).