

# Probability Metrics to Calibrate Stochastic Chemical Kinetics

Heinz Koepl<sup>\*</sup>, Gianluca Setti<sup>†</sup>, Serge Pelet<sup>‡</sup>, Mauro Mangia<sup>†</sup>, Tatjana Petrov<sup>\*</sup> and Matthias Peter<sup>‡</sup>

<sup>\*</sup>School of Communication and Computer Sciences

Ecole Polytechnique Federale de Lausanne (EPFL), Switzerland

<sup>†</sup> Department of Electronics, Computer Sciences and Systems

University of Bologna, Italy

<sup>‡</sup>Institute of Biochemistry

Swiss Federal Institute of Technology Zurich (ETHZ), Switzerland

**Abstract**—Calibration or model parameter estimation from measured data is an ubiquitous problem in engineering. In systems biology this problem turns out to be particularly challenging due to very short data-records, low signal-to-noise ratio of data acquisition, large intrinsic process noise and limited measurement access to only a few, of sometimes several hundreds, state variables. We review state-of-the-art model calibration techniques and also discuss their relation to the general reverse-engineering problem in systems biology. For biomolecular circuits involving low-copy-number molecules we adopt a Markov process setup and discuss a calibration approach based on suitable metrics between probability measures and propose the metrics computation for the multivariate case. In particular, we use Kantorovich’s distance and devise an algorithm, for the case when FACS (fluorescence-activated cell sorting) measurements are given. We discuss a case study involving FACS data for the high-osmolarity glycerol (HOG) pathway in budding yeast.

## I. INTRODUCTION

New experimental measurement techniques allow us to sneak a peek on the biomolecular processes within a single cell. Time-lapse fluorescence microscopy provides relative protein abundances over time and can even access the spatial arrangement of those inside the cell [1]. Flow cytometry and fluorescence-activated cell sorting (FACS) [2] outputs adequate statistics on the distribution of protein abundances of a large population of cells. Below the sensitivity of fluorescent-proteins-based techniques, single molecules approaches, such as mRNA detection using the MS2 bacteriophage, or fluorescence *in vivo* Hybridization (FIVH) have been developed to actually count single mRNA copies within a cell [3].

The cell-to-cell variability observed with such single-cell recordings is partially attributed to *molecular noise*, i.e., the intrinsic stochasticity of a chemical kinetics. Another part comes from the extrinsic part [4] that is due to fluctuations in the intracellular environment of the cell, i.e. concentration variation, spatial position of regulatory molecules and cell volume. However, due to noise amplification and bistability in many biomolecular circuits, molecular noise can give rise to such extrinsic noise.

The availability of such measurements exposing the inherent stochasticity of the molecular mechanisms calls for a probabilistic framework to be able to give them a suitable mathematical model. Starting with the work of Delbrück a Markov process theory of chemical kinetics [5], well founded on thermodynamic principles [6] has been developed. Recent

years have witnessed the limitation of this approach to capture *in vivo* dynamics of intracellular processes. In particular, the assumption of well-stirredness and Brownian-type diffusion is shown to be inappropriate in some scenarios and spatial simulation algorithms that also account for macromolecular crowding [7] were proposed [8].

Recently, effort has been made to account for and exploit the significant stochasticity of single-cell recordings in model calibration. Markov-chain Monte Carlo methods have been deployed to sample from the complicated posterior distribution in Bayesian estimation schemes in the case of sampled data [9]. Reinker *et al.* fitted a hidden Markov model to noisy data of single-mRNA measurement, whereas Golightly *et al.* proposed to infer the parameters of the approximate chemical Langevin equation to cope with the state-space-explosion of the underlying exact probabilistic transition system [10]. Rutter *et al.* [11] applied approximate variational inference on the transition system to keep the computational complexity manageable. Approximate Bayesian computation is used by Toni *et al.* [12] to infer kinetic rates and perform model selection. The first use of probability metrics for systems biology is reported in [13] and it serves as the basis for the approach taken here. To this end, it is important to stress that calibration is different from, and less ill-posed, than the general inverse problem of reverse-engineering regulatory networks, where network topology and kinetic parameters are inferred simultaneously [14].

The remaining part of the paper is organized as follows. In Section II we introduce the standard mathematical framework to describe stochastic chemical kinetics. Section III lays out the foundation for Section IV, where the considered calibration approach and the particular problem statement is discussed in detail. This section contains two case studies – a simple synthetic one and one that involves FACS measurements. Section V draws conclusions and indicates directions for further research.

## II. STOCHASTIC CHEMICAL KINETICS

We call a *reaction system* a tuple  $\mathcal{S} \equiv (\mathcal{A}, m_0, \mathcal{R}, \mathbf{c})$  with  $\mathcal{A} \equiv \{A_1, A_2, \dots, A_N\}$  the set of chemical species,  $\mathcal{R} \equiv \{R_1, R_2, \dots, R_M\}$  the set of reactions,  $m_0 : \mathcal{A} \mapsto \mathbb{N}_0$  the initial species’ multiplicity and  $\mathbf{c} \in \mathbb{R}_+^M$  the vector of kinetic rate constants. A reaction  $R_k$  among species is thus defined

by the construct

$$R_k : S_{1k}A_1 + S_{2k}A_2 + \cdots + S_{Nk}A_N \\ \xrightarrow{c_k} P_{1k}A_1 + P_{2k}A_2 + \cdots + P_{Nk}A_N$$

where here the symbol “+” indicates chemical composition of species and not algebraic addition. The constant  $c_k$  is determined by thermodynamic quantities, such as the difference in free energy. We denote  $S_{ik} \in \mathbb{N}_0$  and  $P_{ik} \in \mathbb{N}_0$  as the stoichiometric coefficients for the substrates (left-hand-side) and the products (right-hand-side) of reaction  $R_k$ , respectively. Moreover, we denote the matrix  $\mathbf{N} \in \mathbb{Z}^{N \times M}$  as the stoichiometric matrix of  $\mathcal{S}$ , the elements of which are defined as  $N_{ik} \equiv P_{ik} - S_{ik}$ . Thus for a given reaction  $R_k$  the  $k$ -th column of  $\mathbf{N}$  corresponds to the net consumption or production of the involved molecular species.

The *macroscopic state* of a reaction system  $\mathcal{S}$  is a function  $\mathbf{X} : \mathbb{R}_0 \mapsto \mathcal{X} \subseteq \mathbb{N}_0^N$  with elements  $X_i(t) = m_t(A_i)$ ,  $t \in \mathbb{R}_0$ ,  $0 < i \leq N$  and multiplicity  $m_t : \mathcal{A} \times \mathbb{R}_0 \mapsto \mathbb{N}_0$ . We call  $\mathbf{X}$  the macroscopic state because it may serve as a summary of the microscopic states, involving for instance position and velocity of every single species in a reaction volume  $V$ .

A reaction system  $\mathcal{S}$  is considered *well-stirred* if the probability of finding  $\mathcal{S}$  in state  $\mathbf{x} \equiv (x_1, \dots, x_N)^T$  at any time  $t$ ,  $P(\mathbf{X}(t) = \mathbf{x})$  is expressible solely as a function of  $\mathbf{x}$  at previous time points  $t' < t$ . Examples of not well-stirred systems, where the dynamics cannot be expressed self-consistently by the macroscopic state are diffusion limited reaction systems [8].

We call a well-stirred  $\mathcal{S}$  *Markovian* if its macroscopic state  $\mathbf{X}$  evolves according to a first-order Markov process, that is if  $\Pr[\mathbf{X}(t_n) = \mathbf{X}_n | \mathbf{X}(t_{n-1}) = \mathbf{x}_{n-1}, \dots, \mathbf{X}(t_0) = \mathbf{x}_0] = \Pr[\mathbf{X}(t_n) = \mathbf{x}_n | \mathbf{X}(t_{n-1}) = \mathbf{x}_{n-1}]$ . Simple examples for well-stirred reaction system that exhibit non-Markovian dynamics are systems with delays [15].

Consider a well-stirred Markovian  $\mathcal{S}$ . The *hazard*  $a_k(\mathbf{x})$  of reaction  $R_k$  is the probability per time that  $R_k$  fires, i.e.

$$\Pr[\mathbf{X}(t + dt) = \mathbf{x} + \mathbf{N}\mathbf{e}_k | \mathbf{X}(t) = \mathbf{x}] = a_k(\mathbf{x})dt$$

where  $\mathbf{e}_k$  is the  $k$ -th basis vector of a  $M$ -dimensional vector space.  $\mathcal{S}$  is said to obey *mass-action* if

$$a_k(\mathbf{x}) \equiv c_k \prod_{j=1}^N \binom{x_j}{S_{jk}}. \quad (1)$$

Throughout this work, we consider well-stirred, Markovian reaction systems with mass-action kinetics.

### III. METRICS ON STOCHASTIC PROCESSES

Let  $\Omega$  be a metric space with metric  $\rho : \Omega \times \Omega \mapsto \mathbb{R}_0$  and  $\mathcal{F}$  be the  $\sigma$ -algebra of Borel subsets of  $\Omega$ . Take two probability measures  $P$  and  $Q$  on  $(\Omega, \mathcal{F})$  and respective random variables  $Z_i : \Omega \mapsto \Xi$ ,  $i \in \{1, 2\}$ , then the Kantorovich distance  $\kappa$  between them is defined as

$$\kappa(P, Q) \equiv \inf_{P \in \mathcal{P}(Z_1, Z_2)} \mathbb{E}_P [\rho(Z_1, Z_2)], \quad (2)$$

where  $\mathcal{P}(Z_1, Z_2)$  denotes the set of all joint distributions  $P : \mathcal{F} \times \mathcal{F} \mapsto [0, 1]$  having marginals  $P$  and  $Q$  and  $\mathbb{E}_P$  computes the expectation with respect to one such joint density  $P \in \mathcal{P}$ .

For  $\rho$  being a  $L_p$  metric, the Kantorovich metric is sometimes referred to as *the minimal  $L_p$  metric*,  $l_p$ . The following dual representation of  $l_1$  is important for its application [16].

*Theorem 1 (Kantorovich-Rubinstein [16]):* The minimal metric  $l_1(P, Q)$  between two probability measures  $P$  and  $Q$  is identical to the Lipschitz metric  $\mu_L$ , the definition of which is

$$\mu_L(P, Q) \equiv \sup_{\|f\|_L \leq 1} \left\{ \int_{\Omega} f(\omega) [dP(\omega) - dQ(\omega)] \right\}$$

with  $\omega \in \Omega$  and  $f : \Omega \mapsto \mathbb{R}$ . The supremum is taken over all functions  $f$  that have a *Lipschitz semi-norm* below one

$$\|f\|_L \equiv \sup_{\omega_1 \neq \omega_2} \frac{|f(\omega_1) - f(\omega_2)|}{\rho(\omega_1, \omega_2)}.$$

For the case of  $\Xi = \mathbb{R}^1$  and  $\rho(Z_1, Z_2) = |Z_1 - Z_2|$  the minimal metric thus becomes

$$l_1(P, Q) = \int_0^1 |F_1^{-1}(u) - F_2^{-1}(u)| du \\ = \int_{-\infty}^{\infty} |F_1(z) - F_2(z)| dz, \quad (3)$$

where we made use of the cumulative distribution function  $F_1$  and  $F_2$  for the random variables  $Z_1$  and  $Z_2$ , respectively. This original result for the one-dimensional case is shown to be extensible for any  $L_p$  norm [16] not just  $p = 1$ . With the supremum norm  $p \rightarrow \infty$ , we recover the Kolmogorov-Smirnov distance and for  $p = 2$ , the Cramer-von-Mises distance as special cases.

### IV. MODEL CALIBRATION

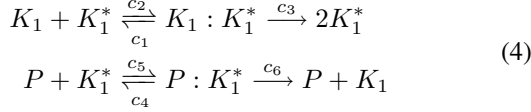
The dual representation of  $\kappa$  for the one-dimensional case (3) is applied to calibration and model reduction in [13] by defining one-dimensional “reporter variables” of the underlying multivariate process. Unfortunately, explicit expression for the dual representation of  $\kappa$  as for  $\Xi = \mathbb{R}^1$  or  $\Xi = \mathbb{Z}^1$  in (3) in the multivariate case does not exist, with the exception of the Gaussian distribution [17]. For multivariate discrete processes  $\Xi = \mathbb{Z}^N$ , we propose however to resort to the primal representation (2) of  $\kappa$  that defines a finite linear program and can thus be computed efficiently. With this, one can exploit the dependencies between species abundances as experimentally provided by, for instance, multi-channel time-lapse fluorescence microscopy.

In the following, however, we focus on the important practical situation of FACS measurement, where the intracellular fluorescence intensity of tagged proteins of one type is recorded for a large population ( $\approx 10^4$ ) of cells at one time instant  $t_0$ . In particular, we address the question whether model parameters can uniquely be determined by one such large-cell-population snapshot.

#### A. Case study – bistable Lisman switch

To illustrate our approach we consider the simple bi-stable reaction system due to Lisman [18]. The system is proposed as a generic mechanism for long-term modification of synaptic efficiency and memory storage within neuronal cells. The original [18] reaction system  $\mathcal{S} = (\mathcal{A}, m_0, \mathcal{R}, \mathbf{c})$  contains five

species  $\mathcal{A} = \{K_1, K_1^*, P, K_1 : K_1^*, P : K_1^*\}$  and the following set  $\mathcal{R}$  of six reactions.



Subsequently, we write  $\mathcal{S}(\mathbf{c})$  for the model and make explicit its parameter-dependence. The data-generating system is assumed to be system  $\mathcal{S}(\hat{\mathbf{c}})$  with some constant  $\hat{\mathbf{c}}$ . For small copy numbers  $m_0(A)$ ,  $\forall A \in \mathcal{A}$ , the chemical master equation [5]

$$\begin{aligned} \frac{\partial p(\mathbf{x}, \mathbf{c}, t)}{\partial t} &= \sum_{j=1}^M a_j(\mathbf{x} - \mathbf{N}e_j, c_j) p(\mathbf{x} - \mathbf{N}e_j, \mathbf{c}, t) \\ &\quad - \sum_{j=1}^M a_j(\mathbf{x}, c_j) p(\mathbf{x}, \mathbf{c}, t), \end{aligned} \quad (5)$$

for system  $\mathcal{S}(\mathbf{c})$  can be solved numerically, where we introduced the probability density  $p : \mathcal{X} \times \mathbb{R}_+^M \times \mathbb{R}_0 \mapsto [0, 1]$ . For this case study we assume that we only have measurement access to one species of  $\mathcal{S}(\hat{\mathbf{c}})$ , chosen to be  $K_1^*$ . Assigning a linear index to any element of the state space by bijection  $n : \mathcal{X} \mapsto \{1, 2, \dots, |\mathcal{X}|\}$  we can write (5) in a standard state-space form

$$\dot{\mathbf{p}}(\mathbf{c}, t) = \mathbf{Q}^T(\mathbf{c})\mathbf{p}(\mathbf{c}, t) \quad (6)$$

$$\boldsymbol{\pi}(\mathbf{c}, t) = \mathbf{M}\mathbf{p}(\mathbf{c}, t) \quad \text{with} \quad \mathbf{p}(\mathbf{c}, 0) = \mathbf{p}_0, \quad (7)$$

where in (6)  $p_j(\mathbf{c}, t) \equiv p(\tilde{\mathbf{x}}, \mathbf{c}, t)$ ,  $j = n(\tilde{\mathbf{x}})$  and where  $\mathbf{Q}(\mathbf{c})$  is the infinitesimal generator of the Markov process [19]. In (7), we also introduced an observation equation where  $\mathbf{M}$  is the marginalization matrix for  $\mathbf{p}(\mathbf{c}, t)$  and  $\boldsymbol{\pi}(\mathbf{c}, t)$  is the marginal probability over the copy numbers of  $K_1^*$ . The observable distribution of the data-generating system is thus  $\hat{\boldsymbol{\pi}}(t_0) \equiv \boldsymbol{\pi}(\hat{\mathbf{c}}, t_0)$ . The Kantorovich distance (3) then reads

$$\kappa(\mathbf{c}) \equiv \kappa(\hat{\boldsymbol{\pi}}(t_0), \boldsymbol{\pi}(\mathbf{c}, t_0)) = \|\mathbf{F}[\hat{\boldsymbol{\pi}}(t_0) - \boldsymbol{\pi}(\mathbf{c}, t_0)]\|_p^p, \quad (8)$$

where matrix  $\mathbf{F}$  computes the cumulative distributions from their respective densities. Calibration of  $\mathcal{S}(\mathbf{c})$  to  $\mathcal{S}(\hat{\mathbf{c}})$  corresponds to the problem

$$\min_{\mathbf{c}} \kappa(\mathbf{c}) \quad \text{subject to: } c_k \geq 0 \quad \forall k \in \{1, 2, \dots, M\}, \quad (9)$$

that is solved for  $p = 2$  using an iterative second-order scheme by supplying gradient and Hessian of  $\kappa$  with respect to  $\mathbf{c}$

$$\begin{aligned} \frac{\partial \kappa(\mathbf{c})}{\partial \mathbf{c}} &= -2 [\hat{\boldsymbol{\pi}}(t_0) - \boldsymbol{\pi}(\mathbf{c}, t_0)]^T \mathbf{F}^T \mathbf{F} \mathbf{M} \frac{\partial \mathbf{p}(\mathbf{c}, t_0)}{\partial \mathbf{c}}. \\ \frac{\partial^2 \kappa(\mathbf{c})}{\partial c_k \partial c_l} &= -2 [\hat{\boldsymbol{\pi}}(t_0) - \boldsymbol{\pi}(\mathbf{c}, t_0)]^T \mathbf{F}^T \mathbf{F} \mathbf{M} \frac{\partial^2 \mathbf{p}(\mathbf{c}, t_0)}{\partial c_k \partial c_l} \\ &\quad + 2 \left( \frac{\partial \mathbf{p}(\mathbf{c}, t_0)}{\partial c_k} \right)^T \mathbf{M}^T \mathbf{F}^T \mathbf{F} \mathbf{M} \left( \frac{\partial \mathbf{p}(\mathbf{c}, t_0)}{\partial c_l} \right), \end{aligned}$$

where in the former we applied the notational convention that the derivative with respect to a column vector is a row vector. The involved first and second-order sensitivities of  $\mathbf{p}(\mathbf{c}, t_0)$

are best obtained from solving the corresponding variational equations

$$\begin{aligned} \frac{d}{dt} \left( \frac{\partial \mathbf{p}(\mathbf{c}, t)}{\partial c_k} \right) &= \mathbf{Q}(\mathbf{c}) \left( \frac{\partial \mathbf{p}(\mathbf{c}, t)}{\partial c_k} \right) + \frac{\partial \mathbf{Q}(\mathbf{c})}{\partial c_k} \mathbf{p}(\mathbf{c}, t) \\ \frac{d}{dt} \left( \frac{\partial^2 \mathbf{p}(\mathbf{c}, t)}{\partial c_k \partial c_l} \right) &= \mathbf{Q}(\mathbf{c}) \left( \frac{\partial^2 \mathbf{p}(\mathbf{c}, t)}{\partial c_k \partial c_l} \right) \\ &\quad + \frac{\partial \mathbf{Q}(\mathbf{c})}{\partial c_l} \left( \frac{\partial \mathbf{p}(\mathbf{c}, t)}{\partial c_k} \right) + \frac{\partial \mathbf{Q}(\mathbf{c})}{\partial c_k} \left( \frac{\partial \mathbf{p}(\mathbf{c}, t)}{\partial c_l} \right) \end{aligned}$$

jointly with (6) and (7) for  $t \in [0, t_0]$ . The initial conditions are zero matrices of appropriate dimension for both cases. Alternatively to the variational equations, the convolutional representation for the two sensitivities, involving the matrix exponential of  $\mathbf{Q}(\mathbf{c})$ , can be computed. The explicit expressions of gradient and Hessian of (8), together with an initial parameter set, fully specifies a Newton algorithm to solve (9) in the general case.

Moreover, one can use the Hessian to determine local identifiability. For the case study, we randomly sampled many  $\hat{\mathbf{c}}$  for the data-generating  $\mathcal{S}(\hat{\mathbf{c}})$  and computed the Hessian at the optimal point  $\hat{\mathbf{c}}$  and in its vicinity. The Hessian turned out to be singular. We can thus conclude that  $\mathcal{S}(\mathbf{c})$  can be calibrated to the data of  $\mathcal{S}(\hat{\mathbf{c}})$  but the parameters cannot be uniquely determined.

### B. Case study – Osmo-Stress Response in Yeast

The HOG pathway in yeast *saccharomyces cerevisiae* is activated by osmo-sensors at the cell membrane and results in the activation of the Hog1p MAP (mitogen activated protein) kinase, which drives the adaptation of the cell to its new environment and causes an up-regulation of roughly 300 genes [20]. To study this process at the single cell level, a fluorescence expression reporter under the control of a stress responsive promoter (pSTL1) was integrated in the genome of the yeast and the fluorescence of the cell was quantified by flow cytometry. If the cells are challenged by high osmotic stress (0.4M NaCl), a twenty fold increase in fluorescence signal is measured compared to the basal auto-fluorescence of the cells. Interestingly, at intermediate NaCl concentration (0.1 M) we observe a bimodal response in the expression response with only half of the cells expressing the fluorescent reporter (P in Fig. 1) [21]. We show that a chromatin remodeling event driven by the RSC complex [22] can be used to model this behavior. The transcriptional model is depicted in Fig. 1.

The complete model involves 13 mass-action reactions, including a simple activation model for Hog1p through NaCl. Furthermore, we assume a linear scaling relation  $y = \alpha x$  between the fluorescence signal  $y$  and the modeled reporter abundance  $x = m_{t_0}(P)$ . All 13 kinetic rates with the scaling constant  $\alpha$  are calibrated based on (8). For the model with realistic abundance levels, (5) cannot be solved numerically anymore and we resort to Monte Carlo stochastic simulation [6], [23] of the system. However, with an empirical estimate for  $\boldsymbol{\pi}(\mathbf{c}, t_0)$ , gradient-type algorithms solving (9) show bad convergence due to the variance in the estimate. Supported by theoretical considerations [24] and FACS data analysis [2], we deploy a log-normal mixture model as a kernel

density estimator to reduce the variance. In the case of Monte Carlo simulations gradient and Hessian cannot be computed analytically, and we resort to a gradient-type finite-differencing scheme. Every iteration comprises  $10^4$  stochastic simulations, estimation of the mixture model parameters and computation of (8) between the mixtures. The results are shown in Fig. 2.

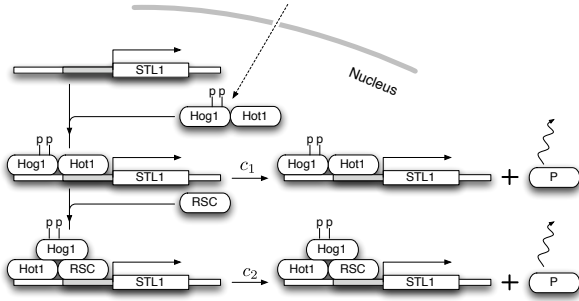


Fig. 1. Proposed minimal gene regulation model for the response to osmo-stress that explains bimodality in the expression profile (for P); efficient transcription (with rate  $c_2$ ) only takes place if the chromatin remodeling complex (RSC) is recruited by active Hog1 to the STL1 promoter site; the general transcription factors and the RNA polymerase II are not modeled as they appear to be unregulated with respect to active Hog1; curly arrows denote protein degradation.

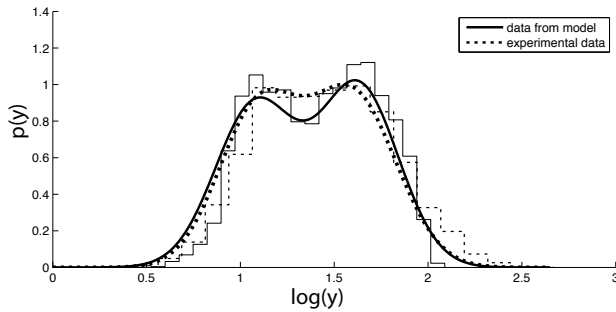


Fig. 2. Histogram and log-normal fit of the bimodal expression profile for experimental fluorescence intensity data  $y$  (dotted) and corresponding calibrated model intensity using  $\hat{y} = \hat{\alpha}^{-1}x$  (solid), with estimated  $\hat{\alpha}$ ; local identifiability cannot be proved due to the lack of the Hessian expression.

## V. CONCLUSIONS

We review the state-of-the-art in calibrating stochastic models of chemical kinetics. We discuss in detail the mathematical framework of probability metrics to compare two stochastic processes. Focus is put on the particular problem of calibrating a model to FACS measurements. For the example of the bistable Lisman switch, we show by spectral analysis of the Hessian that such data is insufficient to uniquely determine the parameters of the model. In a second example we discuss a novel molecular mechanism to explain bimodality, observed in the expression profile of osmo-stress genes in yeast. We calibrate a stochastic model to FACS measurements of this system. Future work includes the development of a general framework to determine how many FACS snapshots are necessary to guarantee identifiability of a given model.

## ACKNOWLEDGMENT

HK acknowledges the support from the Swiss National Science Foundation, grant no. 200020-117975/1. SP and MP acknowledge the support from E.U. project UNICELLSYS (contract No. 201142). MM acknowledges the support from the Italian MIUR, internationalization project ACASTI. TP acknowledges support from SystemsX.ch, the Swiss Initiative in Systems Biology.

## REFERENCES

- [1] T. Yanagida and Y. Ishii, Eds., *Single-molecule dynamics in life science*, Wiley-VCH, Weinheim, 2009.
- [2] J. V. Watson, *Flow Cytometry Data Analysis: Basic Concepts and Statistics*, Cambridge University Press, UK, 1992.
- [3] A. Raj and A. van Oudenaarden, "Single-molecule approaches to stochastic gene expression," *Annu. Rev. Biophys.*, vol. 38, pp. 255–270, 2009.
- [4] M. B. Elowitz, A. J. Levine, E. D. Siggia, and P. S. Swain, "Stochastic gene expression in a single cell," *Science*, vol. 297, pp. 1183–1186, 2002.
- [5] D. A. McQuarrie, "Stochastic approach to chemical kinetics," *J. Appl. Probab.*, vol. 4, no. 3, pp. 413–478, 1967.
- [6] D. T. Gillespie, "A rigorous derivation of the chemical master equation," *Physica A*, vol. 188, pp. 404–425, 1992.
- [7] A. P. Minton, "The influence of macromolecular crowding and macromolecular confinement on biochemical reactions in physiological media," *J. Biol. Chem.*, vol. 276, pp. 10577–10580, 2001.
- [8] S. S. Andrews and D. Bray, "Stochastic simulation of chemical reactions with spatial resolution and single molecule detail," *Phys. Biol.*, vol. 1, pp. 137–151, 2004.
- [9] R. J. Boys, D. J. Wilkinson, and T. B. L. Kirkwood, "Bayesian inference for a discretely observed stochastic kinetic model," *Stat. Comput.*, vol. 18, no. 2, pp. 125–135, 2008.
- [10] A. Golightly and D. J. Wilkinson, "Bayesian inference for nonlinear multivariate diffusion models observed with error," *Comput. Stat. Data An.*, vol. 83, pp. 1891–1901, 2008.
- [11] A. Rutter, G. Sanguinetti, and M. Opper, "Approximate inference for stochastic reaction processes," in *Learning and Inference in Computational Systems Biology*, pp. 189–205. The MIT Press, 2009.
- [12] T. Toni, D. Welch, N. Strelkowa, A. Ipsen, and M. P. H. Stumpf, "Approximate Bayesian computation scheme for parameter inference and model selection in dynamical systems," *J. Roy. Soc. Interface*, vol. 6, no. 31, pp. 187–202, 2009.
- [13] D. Thorsley and E. Klavins, "Model reduction of stochastic processes using Wasserstein pseudometrics," in *Amer Contr Conf*, June 11–13 2008, pp. 1374–1381.
- [14] R. Bonneau, "Learning biological networks: from modules to dynamics," *Nat. Chem. Biol.*, vol. 4, no. 11, pp. 658–664, 2008.
- [15] D. Bratsun, D. Volfson, L. S. Tsimring, and J. Hasty, "Delay-induced stochastic oscillations in gene regulation," *P. Natl. Acad. Sci. USA*, vol. 102, no. 41, pp. 14593–14598, 2005.
- [16] S. T. Rachev, *Probability metrics and the stability of stochastic models*, Probability and Mathematical Statistics. Wiley, New York, 1991.
- [17] D. C. Dowson and B. V. Landau, "The Fréchet distance between multivariate normal distributions," *J. Multivariate Anal.*, vol. 12, pp. 450–455, 1982.
- [18] J. E. Lisman, "A mechanism for memory storage insensitive to molecular turnover: A bistable autophosphorylating kinase," *P. Natl. Acad. Sci. USA*, vol. 82, pp. 3055–3057, 1985.
- [19] W. J. Stewart, *Introduction to the numerical solution of Markov chains*, Princeton University Press, Princeton, New Jersey, 1994.
- [20] S. Hohmann, M. Krantz, and B. Nordlander, "Yeast osmoregulation," *Methods Enzymol.*, vol. 428, pp. 29–45, 2007.
- [21] S. Pelet, F. Rudolf, and M. Peter, "manuscript in preparation."
- [22] G. Mas *et al.*, "Recruitment of a chromatin remodelling complex by the Hog1 MAP kinase to stress genes," *EMBO J.*, vol. 28, pp. 326–336, 2009.
- [23] D. J. Wilkinson, *Stochastic Modelling for Systems Biology*, Chapman & Hall, 2006.
- [24] C. Furusawa, T. Suzuki, A. Kashiwagi, T. Yomo, and K. Kaneko, "Ubiquity of log-normal distributions in intra-cellular reaction dynamics," *Biophys.*, vol. 1, pp. 25–31, 2005.