

Moment-based inference predicts bimodality in transient gene expression

Christoph Zechner^{a,1}, Jakob Ruess^{a,1}, Peter Krenn^a, Serge Pelet^b, Matthias Peter^b, John Lygeros^a, and Heinz Koeppl^{a,2}

^aAutomatic Control Laboratory, and ^bInstitute of Biochemistry, ETH Zurich, CH-8092 Zurich, Switzerland

Edited by Peter J. Bickel, University of California, Berkeley, CA, and approved April 3, 2012 (received for review January 5, 2012)

Recent computational studies indicate that the molecular noise of a cellular process may be a rich source of information about process dynamics and parameters. However, accessing this source requires stochastic models that are usually difficult to analyze. Therefore, parameter estimation for stochastic systems using distribution measurements, as provided for instance by flow cytometry, currently remains limited to very small and simple systems. Here we propose a new method that makes use of low-order moments of the measured distribution and thereby keeps the essential parts of the provided information, while still staying applicable to systems of realistic size. We demonstrate how cell-to-cell variability can be incorporated into the analysis obviating the need for the ubiquitous assumption that the measurements stem from a homogeneous cell population. We demonstrate the method for a simple example of gene expression using synthetic data generated by stochastic simulation. Subsequently, we use time-lapsed flow cytometry data for the osmo-stress induced transcriptional response in budding yeast to calibrate a stochastic model, which is then used as a basis for predictions. Our results show that measurements of the mean and the variance can be enough to determine the model parameters, even if the measured distributions are not well-characterized by low-order moments only—e.g., if they are bimodal.

extrinsic variability | high-osmolarity glycerol pathway | moment dynamics | parameter inference | stochastic kinetic models

Building predictive computational models of intracellular reaction kinetics is still a dauntingly ill-posed task (1), characterized by low-dimensional experimental readouts of the hypothesized high-dimensional process. Single-cell technologies hold promise to partly alleviate this ill-posedness by exploiting the observed variability for the calibration of stochastic kinetic models (2, 3). The same technologies, however, also reveal that isogenic cells in a single population exhibit large cell-to-cell variability (4, 5). The variation can be shown to be a convolution of two sources, namely the intrinsic molecular noise and extrinsic factors that render single cells different even in the absence of molecular noise; in many cases, the latter was reported to dominate the former (4, 5). Extrinsic factors comprise difference in cell size, cell-cycle stage, expression capacity, local growth conditions—to name but a few (6, 7). Thus, although single-cell technology offers a way out of the predicament of ill-posedness, it requires new methods to deal properly with intrinsic and extrinsic variability. The effect of extrinsic variability on the dynamics of stochastic models is studied in refs. 7 and 8, whereas first attempts have been made to address the inverse problem of quantifying the extrinsic (9) and any additional intrinsic (10) components from measurements. Because the latter is based on path sampling, its applicability remains limited to small systems. Naturally, extrinsic variability is bypassed when calibrating a stochastic model to one single cell (3, 11), for instance through live-cell imaging data. However, the extent to which such a single path observation of the hypothesized stochastic process—given the notoriously sparse acquisition times—can sufficiently confine unknown process parameters remains questionable.

Stochastic kinetic models, able to capture the intrinsic noise, were proposed for modeling single-cell data and its variability (2,

12, 13). Models that track probabilities over the integer-valued state-space of molecule-counts suffer from the curse of dimensionality and are computationally prohibitive for all but the simplest systems. Similar limitations apply to approximations thereof that retain the discreteness of the state-space (14, 15). While extracting sample paths of such processes is straightforward (16), acquiring their statistics—often necessary for calibration—is hampered by the slow convergence of empirical estimates for high-dimensional models (17). This is particularly challenging because most calibration or inference methods rely on iterative schemes, making it necessary to recompute statistics. Alternative methods set out to reduce the computational burden by tracking only low-order moments instead of the whole probability distribution. A standard scheme in this class is moment closure, which provides a means to capture the stochasticity of reactions while leveraging the scalability of ordinary differential equation models (18–20).

Here we introduce a moment-based inference scheme for calibrating stochastic models with heterogeneous single-cell measurements. We show how by extending the method of moment closure by conditional moment equations one can properly account for extrinsic factors. The proposed method requires no Monte Carlo simulations over extrinsic factors, making this approach very scalable. Moreover, besides parameter estimates and their confidence bounds, the method allows one to quantitatively characterize the cell-to-cell variability, ultimately dissecting the unspecific conglomerate of extrinsic factors (6). Every additionally accounted moment of the stochastic process can make the calibration less ill-posed; in the same way as the mean of the process contains information, so does its variance. Importantly, we show that this also holds true if the process is poorly captured by the accounted moments, for instance, if we just consider first and second order moments of a multimodal process distribution.

We instantiate this computational framework by addressing a widely discussed—and we believe ubiquitous—process motif, namely the transiently induced gene expression (21, 22). Often signaling pathways are activated for a short time window, in which the activated signaling output—such as a mitogen-activated protein kinase (MAPK)—needs to initiate transcription by translocation and interaction with possibly several intermediates. If many intermediates need to be in place, some cells do not manage to transcribe at all, ultimately giving rise to bimodal protein expression profiles. The particular case study we consider is the high-osmolarity glycerol (HOG) pathway in budding yeast (23), where for intermediate induction levels a bimodality in the induced stress proteins was observed (21). We perform time-lapsed flow cytometry measurements to calibrate a stochastic kinetic

Author contributions: C.Z., J.R., M.P., J.L., and H.K. designed research; C.Z., J.R., P.K., S.P., and H.K. performed research; P.K., S.P., and M.P. contributed new reagents/analytic tools; C.Z., J.R., and H.K. analyzed data; and C.Z., J.R., J.L., and H.K. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

¹C.Z. and J.R. contributed equally to this work.

²To whom correspondence should be addressed. E-mail: koeppl@ethz.ch.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1200161109/-DCSupplemental.

model and derive population statistics using the proposed framework. This allows us to perform an *in silico* homogenization of the cell population and to start quantifying and dissecting the different sources of variability.

Results

Population Dynamics and Extrinsic Variability. The probability distribution of stochastic models is governed by the chemical master equation (CME) whose moments can be approximated using moment closure techniques (see *Materials and Methods*). When considering a heterogeneous cell population, the state of each cell follows an individual CME that depends on a certain extrinsic condition. Consequently, a single cell's dynamics can be described by a vector of approximate conditional moments $\tilde{\mu}_{|z}$ up to some order n with z as a realization of some extrinsic variable Z , i.e.,

$$\frac{d}{dt}\tilde{\mu}_{|z} = A(\theta, z)\tilde{\mu}_{|z} + B(\theta, z)f(\tilde{\mu}_{|z}), \quad [1]$$

where θ denotes the set of intrinsic parameters that are shared among cells, matrices $A(\theta, z)$ and $B(\theta, z)$ are determined by the model structure and function f is obtained by moment closure. In this work we assume a constant extrinsic condition or that it varies on time-scales much larger than the duration of the experiment. Accordingly, Z is modeled as a random vector, governed by a multidimensional probability distribution P_Z , which is in turn described by a set of parameters α that we refer to as the extrinsic statistics. For instance, one could assume α to be the moments or a parametrization of P_Z . This gives rise to a hierarchical Bayesian model (24), as illustrated in Fig. 1A.

Single-cell population data is characterized by a convolution of intrinsic and extrinsic variability. Determining the moments of a heterogeneous population requires computing the expectation of $\tilde{\mu}_z$ with respect to P_Z (see Fig. 1B). One way to achieve this is to sample values z from P_Z , solve system Eq. 1 and then average over the obtained solutions, which naturally comes with a large computational effort. However, our analysis in *SI Appendix, section S.1* shows that it is also possible to directly derive a population-based system of moment equations. The resulting approximate population moments can be written as

$$\begin{aligned} \frac{d}{dt}\tilde{\mu} &= A(\theta)\tilde{\mu} + B(\theta)\tilde{\mu}_{X,Z} + C(\theta)\tilde{\alpha} + D(\theta)g(\tilde{\mu}, \tilde{\mu}_{X,Z}, \tilde{\alpha}) \\ \frac{d}{dt}\tilde{\mu}_{X,Z} &= E(\theta)\tilde{\mu}_{X,Z} + F(\theta)\tilde{\alpha} + G(\theta)h(\tilde{\mu}, \tilde{\mu}_{X,Z}, \tilde{\alpha}), \end{aligned} \quad [2]$$

where $\tilde{\mu}_{X,Z}$ contains the cross moments of the species and the extrinsic variable of order up to n , matrices $A(\theta)$, $B(\theta)$, $C(\theta)$,

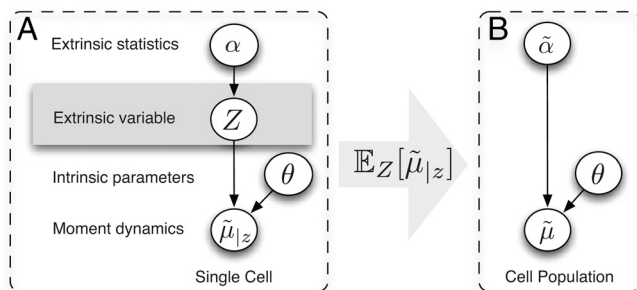


Fig. 1. Hierarchical Bayesian models for the single cell (left) and population moments (right). (A) The approximate moment dynamics of a single cell $\tilde{\mu}_{|z}$ depend on the intrinsic parameters as well as the extrinsic condition z . We assume the extrinsic variable Z to be time-invariant and to stem from a distribution described by a set of extrinsic statistics α . (B) Marginalization with respect to the extrinsic condition gives rise to the graphical model of the approximate population dynamics $\tilde{\mu}$, depending on the intrinsic parameters θ and the moments of Z of order up to n , which are denoted $\tilde{\alpha}$.

$D(\theta)$, $E(\theta)$, $F(\theta)$, and $G(\theta)$ are determined by the model structure and functions g and h are obtained by moment closure. Note that Eq. 2 only depends on lower-order moments of P_Z , denoted here by $\tilde{\alpha}$. In order to compute approximations of the population moments, Eq. 2 is solved with $\tilde{\alpha}$ as the extrinsic statistics—and θ as the intrinsic parameters.

Moment-Based Inference. In practical scenarios, both the intrinsic parameters as well as the extrinsic statistics have to be inferred from the measurements. Although an extension to the general case is straightforward, we assume—for the sake of clarity—that only a single species is measured from a cell population at time points t_l , $l \in \{1, \dots, L\}$. We define $\gamma = \{\tilde{\alpha}, \theta\}$ and denote the approximate time evolution of the measured species' k -th order moment, computed from Eq. 2, as $t \mapsto \tilde{\mu}_k(t, \gamma)$, $k \in \{1, \dots, n\}$. With the k -th order experimental moments $\hat{\mu}_k = \{\hat{\mu}_k(t_l)\}$, $l \in \{1, \dots, L\}$ and their corresponding estimated variances $\sigma_k^2(t_l)$ with $k \in \{1, \dots, n\}$, the posterior distribution over γ is given by

$$p(\gamma|\hat{\mu}_1, \dots, \hat{\mu}_n) = \frac{1}{K} \prod_{k=1}^n \prod_{l=1}^L p(\hat{\mu}_k(t_l)|\gamma)p(\gamma), \quad [3]$$

with $p(\gamma)$ as the parameter prior and K as a normalizing constant independent of γ . For the large-sample case encountered in flow cytometry we can make use of the central limit theorem and assume that $p(\hat{\mu}_k(t_l)|\gamma) = \mathcal{N}(\tilde{\mu}_k(t_l, \gamma), \sigma_k^2(t_l))$ (for more details see *Materials and Methods* and *SI Appendix, sections S.3.2 and S.4.5*). A common strategy to obtain Bayesian point estimates is to maximize Eq. 3 with respect to the parameter γ (24) (see also *Materials and Methods*).

Whether the moments of the measured distributions carry enough information to jointly determine the intrinsic parameters and the extrinsic statistics, in general has to be answered by performing an identifiability analysis of the closed moment system (1, 25). Using a simple birth-death process, we analytically demonstrate that, in principle, this is possible (see *SI Appendix, section S.2*).

A Simple Model of Transient Gene Activation. To test whether the moment-based inference scheme can identify parameters even in the case of multimodal process distributions, we studied the four-species model depicted in Fig. 2A, which can be thought of as a simple model of transiently induced gene expression. Degradation of A serves as a simplistic mechanism to model a temporal window of transcription factor activity. During this temporal window, the gene B manages to switch into a state, where protein C is produced only in a fraction of the cells (parameter configuration and initial conditions are given in *SI Appendix, Table S.1* and section S.3.2). To generate protein distributions at 10 different time points, we used Gillespie's stochastic simulation algorithm (16). We then computed empirical means and variances of these distributions, treated them as experimental measurements, and performed a parameter search to maximize the parameter posterior using a Metropolis-Hastings (M-H) Markov chain Monte Carlo (MCMC) (26) sampler (see *Materials and Methods*). The inferred moments are depicted in Fig. 2C. Even though in this case, mean and variance do not paint a full picture of the underlying multimodal protein distribution, all parameters—and thus the protein distributions—were estimated accurately up to a small approximation error (see Fig. 2B and *SI Appendix, section S.3.2*).

Hog1-Induced Gene Expression in Yeast. The moment-based inference scheme allowed us to study gene expression, activated by the HOG signaling pathway in budding yeast (23). Upon hyper-osmotic shock yeast cells induce the MAPK Hog1 signaling cascade.

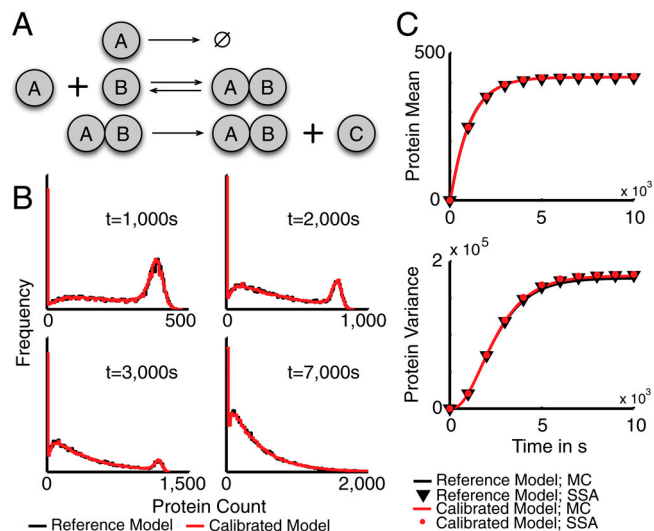


Fig. 2. Moment-based inference using synthetic data. (A) A simple model of transient gene activation: The binding of A to the target gene B aggregates all necessary steps involved in gene activation such as the binding of additional transcription factors, polymerase binding, or chromatin remodeling. Also protein synthesis is reduced to the simplest possible model—i.e., a first order production, abstracting messenger RNA (mRNA) transcription and degradation, translation, and protein folding. (B) The protein distributions predicted by the calibrated model (red) compared to the distributions generated from the reference model (black) at four representative time points (see *SI Appendix, section S.3.2* for further time points). Estimates of the distributions were obtained by stochastic simulation (20,000 runs). (C) The time evolutions of the approximate protein mean and variance obtained from moment closure (MC) differ only little from approximations computed by stochastic simulation (SSA). Therefore, the model parameters can be inferred up to negligibly small deviations.

The role of this kinase is twofold. In the cytoplasm, Hog1 phosphorylates its substrate to increase the internal concentration of glycerol in the cell. In parallel, a large fraction of the active Hog1 translocates to the nucleus where it triggers the activation of a transcriptional program leading to the upregulation of roughly 300 genes (27). Once the internal glycerol concentration allows to balance the external osmotic pressure, the HOG pathway is deactivated, leading to loss of active MAPK and a rapid termination of the transcriptional process.

To quantify the amount of transcription induced by this pathway a fluorescent expression reporter was generated using the promoter pSTL1 (promoter of the sugar transporter-like protein 1), a well-characterized Hog1 expression target driving the expression of a fluorescent protein construct (quadrupleVenus-qV). It was recently shown in ref. 21 that the transient activation of the MAPK Hog1 in conjunction with a slow step in the transcription activation process of the promoter results in a bimodality in the expression profiles of this fluorescent expression reporter.

Nuclear enrichment of Hog1 was measured by microscopy and the pSTL1-qV reporter abundance was quantified by flow cytometry at nine different time points for NaCl concentrations of 0 M, 0.1 M, 0.12 M, and 0.2 M.

The Model. Components involved in activation and translocation of Hog1 are present in high abundance (e.g., around 6800 Hog1 molecules per cell) (28). Consequently, intrinsic fluctuations of active Hog1 are relatively small. Experimental results in ref. 29 and our own data support this and also that Hog1 signaling is robust against cell-to-cell variations. Motivated by this, we assume Hog1 signaling to be deterministic rather than stochastic and that the mean dynamics reflect well the signaling behavior (30, 31). Continuous-time functions of nuclear Hog1 were ob-

tained from the experimental data by linear regression with radial basis functions (24) across different NaCl concentrations (see *SI Appendix, section S.4.2*).

Several transcription factors such as Sko1 or Hot1 are under control of active Hog1 once it translocates to the nucleus as shown in ref. 32. This and the experimentally observed switch-like induction of fluorescent reporter expression suggest a high cooperativity in the pSTL1 promoter dynamics. In a purely stochastic mass-action model, one way to model cooperativity is to require multiple Hog1 copies to bind to the promoter before messenger RNAs (mRNA) can be transcribed. However, the previous high copy-number considerations allow us to simplify this step into transforming the fitted Hog1 abundance curves using a Hill-function with tunable parameters (see *SI Appendix, section S.4.2*). The output of this function is then treated as a time-varying kinetic parameter modulating the gene activation rate. Efficient transcription of mRNA requires interaction of the active gene with chromatin remodeling complexes (generic remodeler denoted as CR) (21). Translation is modeled as a one-step linear production event, depending on the number of ribosomes. We assume that extrinsic variability enters the system in the chromatin remodeling (variability in the number of CR) (33) and the translation efficiency (variability in the number of ribosomes) (5). A graphical representation of the model is given in Fig. 3A.

pSTL1-qV Mean and Variance Predict Transient Bimodality. The parameters of the model from Fig. 3A were inferred from the time courses of the experimental means and variances (see Fig. 3B) using NaCl concentrations 0 M, 0.12 M, and 0.2 M. We then validated the model by comparing the distributions predicted by the model for 0.1 M NaCl with the experimental results. The pSTL1-qV expression profiles for each measurement time point and NaCl concentration were computed from the calibrated model using stochastic simulation. A comparison between the experimental and the predicted distributions is shown in Fig. 4A. Even though only means and variances were used in the inference, the bimodal distributions are accurately predicted by the model (see also *SI Appendix, section S.4.5*).

We further validated the model using an additional snapshot dataset from ref. 21, where the pSTL1-qV reporter abundance was measured for several other NaCl concentrations between 0 M and 0.3 M, 45 min upon osmotic shock. From the model predictions and the measured distributions, we computed the coefficient of variation (CV) and a dose-response as functions of the NaCl concentration (Fig. 4B). The area around 0.1 M NaCl, where the CV is large and the dose-response curve is rising, indicates the NaCl concentration interval where the expression is in a bimodal regime. Note that also at 0.3 M NaCl, a concentration much larger than the concentrations that were used in the inference, the CV is predicted accurately.

To study the stochastic pSTL1-qV induction, we simulated the model to estimate the average number of cells that (i) never activate the pSTL1 promoter, (ii) activate the promoter at least once, and (iii) induce transcription. Our model predicts that for all NaCl concentrations except 0 M all cells manage to activate the promoter and, therefore, that the bimodality has to be caused by the subsequent—and comparably slow—chromatin remodeling step (see *SI Appendix, section S.4.7*). Further, we performed an *in silico* knock-down of CR by rescaling each cell's amount of CR by a hand-tuned factor, such that the percentage of responding cells saturated around 60% as measured in the experiment (see Fig. 4B). We found that the transition between the non- and all-responding domain is shifted to higher NaCl values and that the slope of the transition edge is decreased.

In Silico Homogenization of the Cell Population. After calibrating the model, we switched off extrinsic variability by setting each cell's extrinsic condition to the inferred mean value. We then recom-

demonstrates that averaged population data may contain too little information to identify the reaction rates. Contrary to that, additionally measuring the variance in the example of Fig. 2A allows one to uniquely identify all the parameters, even though the measured distributions are bimodal. This implies that the question of whether the measured distributions are well-characterized by low-order moments only is not necessarily of importance. In ref. 2 the authors presented a method that makes use of the information provided by the whole distributions. However, for larger systems, approximation of the probability distribution becomes computationally cumbersome. Focusing the analysis on lower-order moments, as proposed in this paper, means discarding a part of the information but makes the parameter identification feasible for larger systems.

The moment-based inference scheme allowed us to estimate the parameters of a stochastic model of the osmo-stress induced transcriptional activation in budding yeast using distribution measurements of a heterogeneous cell population and thereby enabled us to explain and predict experimental data and to provide computational support for existing biological hypotheses. The inferred model characteristics and parameters agree well with state-of-the-art literature. For instance the predictions, obtained for the *in silico* knock-down of CR, agree well with results in ref. 21, where the authors performed knock-down experiments for different components of the SAGA complex, which is recruited during chromatin remodeling (see Fig. 4B). The pSTL1-qV half-life was estimated to be around 90 min; this value agrees well with ref. 34, where the authors report a high stability of similar fluorescent reporters. The Hill coefficient of the pSTL1 promoter dynamics was estimated to be $n_H \approx 6$, indicating high cooperativity in the binding of active Hog1 to the target gene. This seems to be crucial for the cell to achieve the strong switch-like behavior observed in the experimental data and agrees well with previously reported results, where Hog1 dependent transcription factors were shown to have multiple binding sites (35).

The inferred model parameters predict large cell-to-cell variations in the chromatin remodeling as well as in the translation efficiency (CVs around 0.3–0.4). This is in agreement with the experimental results from ref. 33, where the authors found that several chromatin remodeling factors show large variations—e.g., a CV around 0.3 for the transcription adapter 2 (Ada2).

According to our model, variability in the chromatin remodeling is widely independent of the FSC gating radius, indicating that extrinsic noise is suppressed only in the translation efficiency by applying gates on morphological features. In conjunction with the observation that an *in silico* homogenization of the cell population leads to different CVs of the pSTL1-qV distributions (Fig. 4B), this suggests that studies that solely rely on cell gating to eliminate cell-to-cell variability may lead to biased results. Explicitly including extrinsic variability may resolve a systematic mismatch and, additionally, allows to quantify effects of variability on the system. For instance the *in silico* homogenization of the cell population (Fig. 4B) indicates that the dose response is widely insensitive against cell-to-cell variability. This provides computational support for the hypothesis in ref. 21 that the partitioning between reporter inducing and noninducing cells is primarily caused by intrinsic stochasticity.

Materials and Methods

Flow Cytometry Measurements. The yeast cells bearing the integrated pSTL1-quadrupleVenus (21) were cultivated in synthetic (SD) medium (using a yeast nitrogen base w/o folic acid, w/o riboflavin). An overnight saturated culture was diluted and grown in log phase for 24 h (OD600 nm kept below 0.2 by several dilutions). Hog1 driven gene expression was induced by adding 5 mL (3x) salt solution (SD medium + NaCl) to 10 mL culture containing flasks. At each time point a 0.5 mL aliquot was taken and protein translation was stopped by adding cycloheximide (final CHX concentration: 100 μ g/mL). After maturation of the fluorescent reporters (about 2 h), 350 μ L cells were added to 400 μ L PBS, briefly sonicated and filtered. Finally the fluorescence

was measured using a BD LSR II (excitation: 488 nm, emission: 525/50 nm). If not explicitly stated, we applied cell-gates of log-diameter 0.5 with respect to the forward—and log—diameter 2.5 to the side scatter channel, respectively. Each flow cytometry distribution was obtained from around 55,000 cells, leading to an effective number of cells around 20,000 after applying the gate.

Moment Closure. Equations that describe the time evolution of all the moments of the distribution can be derived from the CME (36). Extracting the equations for the moments up to some order n leads to a finite open system that possibly depends on higher order moments. Approximating the higher order moments by some nonlinear function f of the lower-order moments (see ref. 19 and *SI Appendix, sections S.3.1 and S.4.1*) leads to a closed system of the form

$$\frac{d}{dt}\tilde{\mu} = A(\theta)\tilde{\mu} + B(\theta)f(\tilde{\mu}),$$

where $\tilde{\mu}$ is a vector containing the moments of order up to n and $A(\theta)$ and $B(\theta)$ are determined by the model structure.

Moment Uncertainties and Data Modeling. Asymptotically unbiased estimates for central moments of order k at time t_i were computed from M samples as

$$\hat{\mu}_k(t_i) = \begin{cases} \frac{1}{M} \sum_{i=1}^M x_i(t_i) & k = 1 \\ \frac{1}{M} \sum_{i=1}^M (x_i(t_i) - \hat{\mu}_1(t_i))^k & k > 1. \end{cases}$$

The central limit theorem implies that for large M (i.e., around 20,000 within our experiments) the moment estimates are approximately normally distributed—i.e., $\hat{\mu}_k(t_i) \sim \mathcal{N}(\mu_k(t_i), \sigma_k^2(t_i))$ —with $\mu_k(t_i)$ as the true k -th moment. We further validated this assumption for both case studies by comparing bootstrapped distributions of the empirical moments to normal distributions using probability-probability (P-P) and quantile-quantile (Q-Q) plots. Additionally, we used a Kolmogorov–Smirnov–Lilliefors test to assess normality (see *SI Appendix, sections S.3.2 and S.4.5*). For $k = 1$ and $k = 2$, the estimators variance can be estimated as

$$\sigma_1^2(t_i) = \frac{1}{M} \hat{\mu}_2^2(t_i) \quad \text{and} \quad \sigma_2^2(t_i) = \frac{1}{M} \left(\hat{\mu}_4(t_i) - \frac{M-3}{M-1} \hat{\mu}_2^2(t_i) \right),$$

respectively.

Modeling Fluorescence Intensities. We assumed that the measured fluorescence intensity for a given cell is proportional to the number of fluorescent proteins (33)—i.e., $I_{\text{tot}}(t_i) = \epsilon I_{\text{tot}}(t_i)$ with scaling parameter ϵ . Due to the nonidentifiability in the translation step (see *SI Appendix, section S.4.5*), only the product of the translation rate and ϵ can be determined. Furthermore, we assumed that the reporter abundance $I_R(t_i)$ is corrupted by autofluorescence and measurement artifacts, modeled as an additive random variable $I_{AF}(t_i)$, independent of the reporter abundance—i.e., $I_{\text{tot}}(t_i) = I_R(t_i) + I_{AF}(t_i)$. Mean and variance of $I_{AF}(t_i)$ were estimated from the flow cytometry data for 0 M NaCl, collected over the measurement time points. As this allows very accurate estimates (M in the order of hundreds of thousands), the uncertainty of those estimates can be well neglected. The experimental means and variances of pSTL1-qV abundance at a given measurement time point were calculated as $\hat{\mu}_R^k(t_i) = \hat{\mu}_{\text{tot}}^k(t_i) - \mu_{AF}^k(t_i)$ for $k \in \{1, 2\}$. Note that moment-based inference and analysis of the model can be carried out without any assumptions on the autofluorescence distribution. In order to compare protein distributions from the model with experimentally obtained distributions, we sampled autofluorescence values from the measured flow cytometry distribution for 0 M NaCl.

Model Calibration. For all experiments, we assumed flat prior distributions over parameters $\gamma_j \in \gamma$ (with zero probability for negative values). In the M-H MCMC scheme, for each of the J parameters in γ , we used independent log-normal proposal distributions such that $q(\gamma^{\text{new}}|\gamma^{\text{old}}) = \prod_{j=1}^J q(\gamma_j^{\text{new}}|\gamma_j^{\text{old}})$ with $q(\gamma_j^{\text{new}}|\gamma_j^{\text{old}}) = \mathcal{L}(\ln \gamma_j^{\text{old}}, v_j^2)$. A detailed configuration can be found in *SI Appendix, sections S.3.2 and S.4.5*. Proposed parameter samples are accepted with probability

$$a = \min \left\{ 1, \frac{p(\gamma^{\text{new}}|\hat{\mu}_1, \dots, \hat{\mu}_n)q(\gamma^{\text{old}}|\gamma^{\text{new}})}{p(\gamma^{\text{old}}|\hat{\mu}_1, \dots, \hat{\mu}_n)q(\gamma^{\text{new}}|\gamma^{\text{old}})} \right\}.$$

From the resulting Markov chain we extracted the parameter configuration which maximized the posterior density. The distance between predicted and measured protein distributions was quantified using the Kolmogorov metric (see *SI Appendix, section S.4.5*).

ACKNOWLEDGMENTS. We are grateful to J. Paulsson, A. Hilfinger, and J. Hasenauer for interesting discussions. C.Z. and H.K. acknowledge the sup-

port from the Swiss National Science Foundation, Grant PP00P2 128503. The work of J.R. and J.L. was supported in part by SystemsX.ch under the project YeastX and by the European Commission under the project MoVeS. S.P. and M.P. acknowledge the support from the European project UNICELLSYS, the European Research Council, the SystemsX.ch organization (LiverX), the Competence Centre for Systems Physiology and Metabolic Disease, the Swiss National Science Foundation, and the ETH Zurich.

- Hengl S, Kreutz C, Timmer J, Maiwald T (2007) Data-based identifiability analysis of non-linear dynamical models. *Bioinformatics* 23:2612–2618.
- Munsky B, Trinh B, Khammash M (2009) Listening to the noise: Random fluctuations reveal gene network parameters. *Mol Syst Biol* 5:318.
- Boys RJ, Wilkinson DJ, Kirkwood TBL (2007) Bayesian inference for a discretely observed stochastic kinetic model. *Stat Comput* 18:125–135.
- Elowitz MB, Levine AJ, Siggia ED, Swain PS (2002) Stochastic gene expression in a single cell. *Science* 297:1183–1186.
- Colman-Lerner A, et al. (2005) Regulated cell-to-cell variation in a cell-fate decision system. *Nature* 437:699–706.
- Snijder B, Pelkmans L (2011) Origins of regulated cell-to-cell variability. *Nat Rev Mol Cell Biol* 12:119–125.
- Shahrezaei V, Ollivier JF, Swain PS (2008) Colored extrinsic fluctuations and stochastic gene expression. *Mol Syst Biol* 4:196.
- Hilfinger A, Paulsson J (2011) Separating intrinsic from extrinsic fluctuations in dynamic biological systems. *Proc Natl Acad Sci USA* 108:12167–12172.
- Hasenauer J, et al. (2011) Identification of models of heterogeneous cell populations from population snapshot data. *BMC Bioinformatics* 12:125.
- Koepl H, Zechner C, Ganguly A, Pelet S, Peter M (2012) Accounting for extrinsic variability in the estimation of stochastic rate constants. *Int J Robust Nonlin*, 10.1002/rnc.2804.
- Golightly A, Wilkinson DJ (2005) Bayesian inference for stochastic kinetic models using a diffusion approximation. *Biometrics* 61:781–788.
- Mettetal JT, Muzzey D, Pedraza JM, Ozbudak EM, van Oudenaarden A (2006) Predicting stochastic gene expression dynamics in single cells. *Proc Natl Acad Sci USA* 103:7304–7309.
- Guido NJ, et al. (2006) A bottom-up approach to gene regulation. *Nature* 439:856–860.
- Munsky B, Khammash M (2006) The finite state projection algorithm for the solution of the chemical master equation. *J Chem Phys* 124:044104.
- Wolf V, Goel R, Mateescu M, Henzinger TA (2010) Solving the chemical master equation using sliding windows. *BMC Syst Biol* 4:42.
- Gillespie D (1976) A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. *J Comput Phys* 22:403–434.
- Fishman GS (1996) Monte Carlo: Concepts, algorithms, and applications. (Springer, Berlin).
- Whittle P (1957) On the use of the normal approximation in the treatment of stochastic processes. *J Roy Stat Soc B Met* 19:268–281.
- Hespanha J (2008) Moment closure for biochemical networks. *Proceedings of the 3rd International Symposium on Communications, Control and Signal Processing (IEEE)* pp 142–147.
- Ruess J, Miliadis-Argeitis A, Summers S, Lygeros J (2011) Moment estimation for chemically reacting systems by extended Kalman filtering. *J Chem Phys* 135:165102.
- Pelet S, et al. (2011) Transient activation of the HOG MAPK pathway regulates bimodal gene expression. *Science* 332:732–735.
- Shalem O (2008) Transient transcriptional responses to stress are generated by opposing effects of mRNA production and degradation. *Mol Syst Biol* 4:4.
- Hohmann S (2002) Osmotic stress signaling and osmoadaptation in yeasts. *Microbiol Mol Biol Rev* 66:300–372.
- Bishop CM (2007) Pattern Recognition and Machine Learning. New York (Springer).
- Bellman R, Astrom K (1970) On structural identifiability. *Math Biosci* 7:329–339.
- Gilks WR, Richardson S, Spiegelhalter DJ (1996) Markov chain Monte Carlo in practice. (Chapman & Hall, New York).
- Gasch AP, et al. (2000) Genomic expression programs in the response of yeast cells to environmental changes. *Mol Biol Cell* 11:4241–4257.
- Ghaemmaghami S, et al. (2003) Global analysis of protein expression in yeast. *Nature* 425:737–741.
- Muzzey D, Gómez-Urbe Ca, Mettetal JT, van Oudenaarden A (2009) A systems-level analysis of perfect adaptation in yeast osmoregulation. *Cell* 138:160–171.
- Zi Z, Liebermeister W, Klipp E (2010) A quantitative study of the Hog1 MAPK response to fluctuating osmotic stress in *Saccharomyces cerevisiae*. *PLoS One* 5:e9522.
- Macia J, et al. (2009) Dynamic signaling in the Hog1 MAPK pathway relies on high basal signal transduction. *Sci Signal* 2:ra13.
- Capaldi AP, et al. (2008) Structure and function of a transcriptional network activated by the MAPK Hog1. *Nat Genet* 40:1300–1306.
- Newman JRS, et al. (2006) Single-cell proteomic analysis of *S. cerevisiae* reveals the architecture of biological noise. *Nature* 441:840–846.
- Mateus C, Avery SV (2000) Destabilized green fluorescent protein for monitoring dynamic changes in yeast gene expression with flow cytometry. *Yeast* 16:1313–1323.
- Proft M, Gibbons F, Copeland M, Roth F, Struhl K (2005) Genomewide identification of Sko1 target promoters reveals a regulatory network that operates in response to osmotic stress in *Saccharomyces cerevisiae*. *Eukaryot Cell* 4:1343–1352.
- Hespanha J (2006) Modeling and analysis of stochastic hybrid systems. *IEE Proceedings Control Theory And Applications* 153:520–535.