Using a Single Fluorescent Reporter Gene to Infer Half-Life of Extrinsic Noise and Other Parameters of Gene Expression

Michał Komorowski,[†*] Bärbel Finkenstädt,[†] and David Rand[‡]
[†]Department of Statistics, and [‡]Systems Biology Centre and Mathematics Institute, University of Warwick, Coventry, United Kingdom

ABSTRACT Fluorescent and luminescent proteins are often used as reporters of transcriptional activity. Given the prevalence of noise in biochemical systems, the time-series data arising from these is of significant interest in efforts to calibrate stochastic models of gene expression and obtain information about sources of nongenetic variability. We present a statistical inference framework that can be used to estimate kinetic parameters of gene expression, as well as the strength and half-life of extrinsic noise from single fluorescent-reporter-gene time-series data. The method takes into account stochastic variability in a fluorescent signal resulting from intrinsic noise of gene expression, kinetics of fluorescent protein maturation, and extrinsic noise, which is assumed to arise at transcriptional level. We use the linear noise approximation and derive an explicit formula for the likelihood of observed fluorescent data. The method is embedded in a Bayesian paradigm, so that certain parameters can be informed from other experiments allowing portability of results across different studies. Inference is performed using Markov chain Monte Carlo. Fluorescent reporters are primary tools to observe dynamics of gene expression and the correct interpretation of fluorescent data is crucial to investigating these fundamental processes of cellular life. As both magnitude and frequency of the noise may have a dramatic effect on the cell fitness, the quantification of stochastic fluctuation is essential to the understanding of how genes are regulated. Our method provides a framework that addresses this important question.

INTRODUCTION

Fluorescent and luminescent proteins are among the most commonly used reporters of gene expression (1). In particular, they are used to quantify changes in protein concentration over time (2) and as reporters of transcriptional activity (3) in single cells and tissue. Hence an abundance of data is becoming available that is useful for the estimation of kinetic parameters of expression of many different genes.

The significance of single gene expression dynamics has resulted in numerous theoretical models (4–7) and experimental studies (8–11) that revealed aspects of the stochastic nature of this process (see (12,13) for reviews). Usually the systems being considered are far from thermodynamic equilibrium (14) and they may involve small copy numbers of reacting macromolecules (15). Determining the origins and the magnitude of the stochastic effects is of interest because of the implications for cell fate decisions, development, and nongenetic individuality (see (12,13,16) for reviews). One of the important advances in the studies of noise in gene expression is the development of experimental methods based on using two equivalent reporters in the same cell. This allows the determination of extrinsic and intrinsic components of the total gene expression noise (11,17). Intrinsic noise is defined as a source of variability creating differences between the expression of two identical genes placed in the same cell. By contrast, extrinsic noise refers to the sources that affect the two genes equally in any given cell.

A basic assumption behind using fluorescent or luminescent proteins as reporters of dynamical gene expression, particularly in experiments investigating noise in gene expression, is that the observed fluorescence intensity is proportional to the number of proteins being expressed in the cell (8,9,11,18). There is a reasonable basis to the assumption that such proportionality exists for molecules that are actively fluorescent (19). Nevertheless, before the expressed protein becomes visible to fluorescent detection techniques, it must undergo a maturation process that can last from a few minutes to greater than a day (20,21). This process comprises three major steps: folding; cyclization of the tripeptide motif; and oxidation of the cyclized motif (22). The dynamics of this process significantly contributes to the observed variability of a fluorescent signal and has the potential to impact both estimates of the number of proteins present and estimates of the variability in gene expression (21,23). Even though the maturation process has been recognized, it is most often neglected in the quantitative analysis of fluorescent data (e.g., (9,11,18,24–26)).

The presence of extrinsic and intrinsic noises and stochastic effects of protein maturation indicate that extracting information from the fluorescent signal is not straightforward. Stochastic fluctuations arising at each level of gene expression are masked by subsequent steps of this process, so that the observed variability is a filtered mixture of multiple noise sources. In particular, the fluctuations in transcription rate, which is of great importance to the understanding of gene regulation, are masked by random events that occur between the release of mRNA molecules and the occurrence of fluorescent proteins. Therefore, a precise interpretation of the fluorescent signal requires a mathematical

Submitted July 9, 2009, and accepted for publication March 4, 2010.

*Correspondence: m.komorowski@imperial.ac.uk

Michał Komorowski's present address is Centre for Bioinformatics, Imperial College London, London SW7 2AZ, UK.

Editor: Herbert Levine.

0006-3495/10/06/2759/11 $2.00

doi: 10.1016/j.bpj.2010.03.032

model and a statistical method for its calibration. Various approaches have been proposed to address this problem (7,18,27–30). Nevertheless, none of the currently available inference methods takes into account the stochasticity of the protein maturation kinetics or infers strength of extrinsic fluctuations from commonly used single reporter gene data.

In this article, we calculate protein distributions that account for the variability that originates from the fluorescent protein maturation, transcriptional extrinsic noise, and the intrinsic noise of gene expression. The calculated distributions are used to generate predictions of fluctuating protein levels in steady state as well as away from steady state. We combine the model with an efficient statistical inference framework to fit a time course of fluorescence. The method allows for the estimation of translation rate, the decay rate (half-life), and magnitude of transcriptional extrinsic fluctuations from data of a single reporter gene experiment.

The quantification of fluctuations in protein abundance is important to the understanding of how genes are regulated. For example, it has been demonstrated that both magnitude and frequency of the noise may determine cell fitness (3). Small changes in protein concentration may have a significant effect if they last for long enough, whereas large fluctuations in concentration may not have any effect if they occur too frequently to influence cellular processes (12). This observation stimulated studies of protein level dynamics (31,32) and reveals the need for a method to quantify the stochastic characteristics of the expression of different genes.

Our approach constitutes a general framework for the interpretation of fluorescent time-lapse steady-state and out-of-steady-state data because it simultaneously addresses two important problems: it infers the strength of transcriptional noise in a way that often allows quantification of the transcriptional extrinsic variability using only a single fluorescent reporter gene rather than the dual reporters used previously; and it accounts for stochasticity of the fluorescent protein maturation.

The article is organized as follows. First, we introduce the mathematical model of gene expression that incorporates stochasticity of protein maturation kinetics and extrinsic noise and calculate matured protein distributions out of steady state. We briefly analyze the influence of kinetic parameters on stochastic properties of the fluorescent signal. Finally, we present the statistical method to fit a time course of fluorescence and quantify observed stochasticity in fluorescent signal. We demonstrate applicability of the framework using examples of a gene that is expressed both in a steady state and out of steady state. We explain why all the model components are necessary to reliably interpret the fluorescent signal.

METHODS

In this section, we extend the standard model of single gene expression by adding the protein maturation process and a model for extrinsic noise. Subsequently we analyze stationary fluorescence fluctuations predicted by the model using the autocorrelation function and the power spectral density. Finally, we use the linear noise approximation (30,33,34) to construct a statistical method for estimation of model parameters from fluorescent-reporter-gene time series.

Model of fluorescent gene expression

Although gene expression involves numerous biochemical reactions, the current common consensus is to model it in terms of only three biochemical species (DNA, mRNA, and protein) and four reaction channels (transcription, mRNA degradation, translation, and protein degradation) (4,7,35). Such a simple model has been successfully used in a variety of applications and can generate data with the same statistical behavior as more complicated models (36,37).

We assume what are now standard simplifications employed in this model. We assume that the process begins with production of mRNA molecules (R) at time-dependent rate $k_r(t)$. Each mRNA molecule may be independently translated into protein molecules (P) at rate k_p. Both mRNA and protein molecules are degraded at rates γ_r and γ_p, respectively. To model the expression of a fluorescent protein, we extend the standard model in a similar way to that seen in (21,23). After translation, proteins are folded at a rate k_f and subsequently matured (oxidated) at a rate k_m. The number of unmatured folded proteins and matured proteins are denoted by P_f and P_m. Matured proteins are capable of emitting a fluorescent signal when illuminated. Here, we neglect the cyclization, because it is much faster than the other two folded proteins that constitute the maturation process (see 22). We also assume that both folded and matured proteins degrade at rate γ_p. The reactions in this model can thus be summarized as the following stoichiometric equations:

$$
\begin{aligned}
&R_1 : \mathrm{DNA} \xrightarrow{k_r(t)} \mathrm{DNA} + R && R_5 : P \xrightarrow{k_f} P_f\\
&R_2 : R \xrightarrow{\gamma_r} \varnothing && R_6 : P_f \xrightarrow{\gamma_p} \varnothing\\
&R_3 : R \xrightarrow{k_p} R + P && R_7 : P_f \xrightarrow{k_m} P_m\\
&R_4 : P \xrightarrow{\gamma_p} \varnothing && R_8 : P_m \xrightarrow{\gamma_p} \varnothing.
\end{aligned}
$$

We model biochemical reactions as Poisson birth and death processes. Precisely, we assume that the probability for each reaction to occur in a small time interval is proportional to the product of the length of that interval, the rate of the reaction, and the number of molecules that may undergo the reaction. The probability that more than one event will take place in a small time interval is of the higher order, with respect to the length of the interval. Finally, we assume that events taking place in disjoint time intervals are independent when conditioned on events in the previous interval. This specification leads to the Chemical Master Equation (see Supporting Material). Unfortunately, for many tasks such as inference, the Chemical Master Equation is not a convenient mathematical tool and hence various types of approximations have been developed. As shown in Komorowski et al. (30), the linear noise approximation provides a useful and reliable inference framework. The linear noise approximation models biochemical reactions through a stochastic dynamic model that essentially approximates a Poisson process by an ordinary differential equation model with an appropriately defined noise process. Using the linear noise approximation, our model equations are (see Supporting Material for derivation)

$$
dr = (k_r(t) - \gamma_r r)dt + \sqrt{\tau(t) + \gamma_r \phi_r(t)}dW_1, \qquad (1)
$$

$$
\begin{aligned}
dp = {} & \big(k_p r - (\gamma_p + k_f)p\big)dt + \sqrt{k_p \phi_r(t) + \gamma_p \phi_p(t)}dW_2\\
& - \sqrt{k_f \phi_p(t)}dW_3, \qquad (2)
\end{aligned}
$$

$$dp_f = \left(k_f p - \left(\gamma_p + k_m\right)p_f\right)dt + \sqrt{k_f\phi_p(t)}dW_3 + \sqrt{\gamma_p\phi_{p_f}(t)}dW_4 - \sqrt{k_m\phi_{p_f}(t)}dW_5, \quad (3)$$

$$dp_m = \left(k_m p_f - \gamma_p p_m\right)dt + \sqrt{k_m\phi_{p_f}(t)}dW_5 + \sqrt{\gamma_p\phi_{p_m}(t)}dW_6, \quad (4)$$

where r, p, p_f, and p_m are the concentrations of mRNA, unfolded protein, folded protein, and mature protein, respectively; $\{dW_i\}_{(i=1,\ldots,6)}$ expressions denote increments of independent Wiener processes; $\tau(t)$ is the mean transcription rate at time t; and variables ϕ_r, ϕ_p, ϕ_{Pf}, and ϕ_{Pm} are macroscopic concentrations of mRNA, unfolded protein, folded protein, and mature protein, respectively, described by the following ordinary differential equations (see Supporting Material for derivation):

$$\dot{\phi}_r = \tau(t) - \gamma_r\phi_r, \quad (5)$$

$$\dot{\phi}_p = k_p\phi_p - \left(\gamma_p + k_f\right)\phi_p, \quad (6)$$

$$\dot{\phi}_{p_f} = k_f\phi_p - \left(\gamma_p + k_m\right)\phi_{p_f}, \quad (7)$$

$$\dot{\phi}_{p_m} = k_m\phi_{p_f} - \gamma_p\phi_{p_m}. \quad (8)$$

The macroscopic variables describe the behavior of the system in the thermodynamic limit. This is the limit of an infinitely large number of reacting molecules, where fluctuations average out, leading to a deterministic behavior (34).

Extending the standard model by extrinsic noise

Genetically identical cells exhibit significant diversity even when exposed to the same environmental conditions. Recent studies concluded that this noise has intrinsic and extrinsic sources that could be distinguished by placing two independent gene reporters in the same cell to partition observed variability into these two categories (11,17). Noise sources that create differences between the two reporters within the same cell are called intrinsic noise. Extrinsic noise, on the other hand, refers to sources that affect the two reporters equally in any given cell but create differences between two cells. Noise arising from the stochastic events of births and deaths of mRNA and proteins molecules can be identified as intrinsic. Differences between cells, either in environment or in the concentration of any factor that affects gene expression, will result in extrinsic noise (see (12) for more details).

This definition of the two sources of variability implies that in the derived model (Eqs. 1–8), intrinsic noise due to the birth and death events is modeled by diffusion terms (terms that include dW_i).

The sources of extrinsic variability are defined less clearly. Here we focus on the stochasticity arising from fluctuations in the overall transcription rate, as it is argued in the literature (9,25,38), that it dominates over other sources of extrinsic noise. As proposed by Chabot et al. (9) and Shahrezaei et al. (38), transcriptional extrinsic noise can arise from multiplicative factors in the transcription rate. In this case

$$k_r(t) = D(t)\tau(t)(1 + \zeta(t)), \quad (9)$$

where $\tau(t)$ is a macroscopic transcription term (deterministic function which typically varies smoothly with time) and $\zeta(t)$ is a stochastic perturbation representing the extrinsic noise. The random process $D(t)$ expresses the changing transcriptional environment due to binding and unbinding of transcription factors to the regulatory region of the gene and changes in activity due to chromatin modification. In many situations, the former process is highly dynamic, with fast on- and off-rates. In this case, it follows from Eq. 10 in Rausenberger and Kollmann (32) that the fluctuations are small and can be ignored. On the other hand, changes in transcription due to chromatin modification tend to be on a much larger timescale. It is therefore reasonable to ignore the fluctuations in D, and replace it by a constant. In this case, we obtain

$$k_r(t) = D_0\tau(t)(1 + \zeta(t)) \quad (10)$$

If these assumptions do not hold, then, in the linear approximation

$$\zeta(t) = \zeta_1(t) + \zeta_2(t),$$

where $\zeta_1(t)$ is the extrinsic noise and $\zeta_2(t)$ is due to the fluctuations in $D(t)$. In this case, we cannot separate the extrinsic noise and that in $\zeta_2(t)$. Nevertheless, measurement of the combined noise is extremely interesting. Moreover, it is likely that further experiments could be used to separate the effects. For example, it is possible to reduce or eliminate chromatin modification.

To allow for a potential memory of the extrinsic factor, $\zeta(t)$ is modeled as an Ornstein-Uhlenbeck (OU) process:

$$d\zeta = \left(-\gamma_\zeta\zeta\right)dt + \sigma_\zeta dW_7. \quad (11)$$

This form of transcriptional extrinsic noise has been indicated by experimental data (9). The OU process has an exponentially decaying autocorrelation function (ACF) of the form (39)

$$ACF_\zeta(t) = \frac{\sigma_\zeta^2}{2\gamma_\zeta}\exp\left(-\gamma_\zeta t\right). \quad (12)$$

The parameter γ_ζ can be thus interpreted as a decay rate of the extrinsic fluctuations and $log(2)/\gamma_\zeta$ constitutes the half-life of the extrinsic noise in the rate k_r . Small values of γ_ζ correspond to slow transcriptional fluctuations and a slowly decaying ACF. In this case, we say that transcription has long memory. The stationary variance of the OU process is given by $\sigma_\zeta^2/2\gamma_\zeta$ (39) and this quantity describes the strength of the extrinsic fluctuations. The model that incorporates protein maturation dynamics and extrinsic noise and for which we construct an inference method is given by Eqs. 1–11.

Analysis of the fluorescent protein fluctuations

Before we present our inference method, we examine how the model parameters determine the memory of fluorescence fluctuations and how they affect the filtering of the stochasticity arising from the different reactions constituting the expression process. We are particularly interested in how transcriptional memory and the strength of transcriptional fluctuations are masked by translation and protein maturation processes.

To understand how memory is determined by model parameters we analytically calculate the autocorrelation function for the fluctuations of matured proteins p_m in the stationary state. We assure existence of the steady state by assuming that the macroscopic component of transcription is constant $\tau(t) = b$ and obtain (see Supporting Material for derivation)

$$ACF_{p_m}(t) = a_1\exp\left(-\gamma_\zeta t\right) + a_2\exp\left(-\gamma_r t\right) + a_3\exp\left(-\gamma_p t\right) + a_4\exp\left(-\left(\gamma_p + k_f\right)t\right) + a_5\exp\left(-\left(\gamma_p + k_m\right)t\right), \quad (13)$$

where a_1, …, a_5 are time-independent functions of model parameters. We say that the observed fluctuations have long memory (are slow) if the ACF is a slowly decreasing function of time when compared to the timescale of an experiment. Equation 13 shows that there are five main parameters that determine how the ACF depends on time and therefore jointly determine the total memory of the observed fluctuations. These parameters are: decay rate of transcriptional fluctuations γ_ζ, mRNA rate γ_r, protein degradation rate γ_p, kinetic parameter of protein folding k_f, and kinetic parameters of protein maturation k_m.

Therefore, estimates of all these parameters are necessary to understand the origins of the observed fluorescence fluctuations.

The Fourier transform of the ACF (13) gives the power spectrum of the fluorescent protein fluctuations. Analysis of the spectrum (see Supporting Material) reveals that the variability generated at the transcriptional level undergoes low pass filtering. Therefore, fast transcriptional fluctuations (large γ_ξ) will be filtered out. The strength of the filtering depends on γ_r, γ_p, k_f, and k_m. For large values of these parameters, high frequencies have a smaller contribution to the observed variability.

The above analysis, similar to the more detailed studies (21,23,40), is important from the point of view of inference. It shows that the filtering effect influences the identifiability of model parameters. Fast transcriptional fluctuations will not be present in the fluorescent signal and therefore the precision of estimates for γ_ξ and ${\sigma_\xi}^2$ will be limited. In further sections, we demonstrate that our inference framework can detect this effect and account for it so that estimates of other model parameters are not affected.

INFERENCE FROM FLUORESCENT MICROSCOPY EXPERIMENTAL DATA

In this section, we present a method for estimating parameters of the model (Eqs. 1–11) from sequences of single cell fluorescent microscopy measurements

$$\mathbf{u} = (u_{t_0}, \ldots, u_{t_n}). \tag{14}$$

Let $\mathbf{y}$ denote values of the process p_m evaluated at times $t_0, \ldots, t_n$,

$$\mathbf{y} = \left(p_{mt_0}, \ldots, p_{mt_n}\right). \tag{15}$$

Because the linear noise approximation implies Gaussian distribution, it can be shown (see Supporting Material) that

$$P(\mathbf{y}|\Theta) = \psi(\mathbf{y}|\mu(\Theta), \Sigma(\Theta)), \tag{16}$$

where Θ is a vector of all unknown parameters from Eqs. 1–11, and $\psi(\cdot|\mu(\Theta), \Sigma(\Theta))$ is a multivariate Gaussian density with mean vector $\mu(\Theta)$ and covariance matrix $\Sigma(\Theta)$ whose elements can be calculated numerically in a straightforward way (see Supporting Material).

To find the distribution of the measurements $\mathbf{u}$ we define the relation between the time series of protein concentration $\mathbf{y}$ and the measurements $\mathbf{u}$, assuming that the fluorescent signal is proportional to the number of fluorescent molecules with additional measurement error as

$$u_{t_i} = \lambda p_{mt_i} + \epsilon_{t_i}, \tag{17}$$

where λ is an unknown proportionality constant and ϵ_{t_i} is a measurement error. For mathematical convenience, we assume that the joint distribution of the measurement error is normal with mean 0 and known covariance matrix Σ_ϵ, i.e.,

$$(\epsilon_{t_0}, \ldots, \epsilon_{t_n}) \sim N(0, \Sigma_\epsilon).$$

If measurement errors are independent with a constant variance ${\sigma_\epsilon}^2$, then $\Sigma_\epsilon = {\sigma_\epsilon}^2 I$.

Equations 16 and 17 and the normality of the measurement error imply that the likelihood of the vector $\mathbf{u}$ is Gaussian:

$$P(\mathbf{u}|\Theta) = \psi\left(\mathbf{u}|\lambda\mu(\Theta), \lambda^2\Sigma(\Theta) + \Sigma_\epsilon\right). \tag{18}$$

Henceforth λ is an element of vector Θ and will be estimated from experimental data. Equation 18 provides the joint distribution of a single time series. Often not only single but also many isogenic cells are simultaneously observed under a fluorescent microscope. In this case, the data matrix comprises l time series

$$\mathbf{U} = \left(\mathbf{u}^{(1)}, \ldots, \mathbf{u}^{(l)}\right). \tag{19}$$

Because the time series corresponding to different cells is independent, the likelihood function takes the form

$$P(\mathbf{U}|\Theta) = \prod_{i=1}^{l} \psi\left(\mathbf{u}^{(i)}\middle|\lambda\mu(\Theta), \lambda^2\Sigma(\Theta) + \Sigma_\epsilon\right). \tag{20}$$

Because the likelihood is given explicitly, both maximum likelihood and a Bayesian approach can be used in a straightforward way. To account for prior information on parameters, our methodology is embedded in the Bayesian paradigm where the posterior distribution $P(\Theta|\mathbf{U})$ satisfies (41)

$$P(\Theta|\mathbf{U}) \propto P(\mathbf{U}|\Theta)\pi(\Theta). \tag{21}$$

Equations 20 and 21 allow us to use the standard Metropolis-Hastings algorithm (41) to generate samples from the posterior $P(\Theta|\mathbf{U})$.

RESULTS

In this section, we show that parameters of extrinsic noise can be inferred from single-reporter fluorescent microscopy time series, in contrast to currently available methods that require double-reporter gene experimental data (3,9). In addition, we estimate the kinetic parameters of gene expression such as the transcription profile and the translation rate. Also, the scaling factor λ that relates the fluorescent signal to the number of matured fluorescent proteins can be inferred from data.

The estimation of the model parameters is possible under the assumption that informative prior distributions for degradation rates γ_r and γ_p are obtained in additional experiments. These experiments are often not difficult to conduct (42). Similarly, we use informative prior distributions for the parameters of the protein maturation process. These values are not gene- or promoter-dependent but characterize the fluorescent reporter. They can either be found in the literature (22) or estimated in experiments similar to those used to obtain degradation rates (42).

Because the transcription and translation rates and the parameters of extrinsic noise (decay rate and variance) provide the insightful explanation of the observed fluorescent variability, our method can be seen as a quantification of different types of stochastic behaviors. To demonstrate its applicability we consider two examples—the first is an

inference from steady-state fluctuations, and the second is based on oscillatory, out-of-steady-state expression.

Stationary fluctuations

In this section, we consider a gene that is expressed at steady state by assuming that the deterministic component of the transcription rate is constant (i.e., $\tau(t) = b$). Using a modified version of Gillespie's algorithm (38) that allows for fluctuation in reaction rates (see Supporting Material for details), we generated 50 time series for parameter values that give rise to four different types of stochastic fluctuations. The parameters values are given in Table 1 and the corresponding fluorescence signals are plotted in Fig. 1.

Type *A* represents fast transcriptional fluctuations (half-life 8 min) that, due to the low-pass filtering effect, have relatively small impact on the observed signal. In addition, the mRNA and protein degradation rates γ_r and γ_p are relatively large so that the observed variability demonstrates rather homogeneous, short memory behavior.

Types *B* and *C* demonstrate the effect of long (half-life 83 min) and very long (half-life 69 h) transcriptional memory. The degradation rates of mRNA and protein γ_r and γ_p are large (similarly to type *A*) so that the observed long-term memory behavior at the fluorescent protein level is a result of the slow transcriptional fluctuations.

As the ACF in Eq. 13 indicates, slow fluorescence fluctuations may appear which are not necessarily due to long memory in transcription but are, for instance, due to a low mRNA degradation rate. This regime of behavior is reflected in type D where long-term memory of fluorescence appears despite short-term memory of the transcriptional fluctuations (half-life 8 min).

The results of the inference are presented in Table 2, Figs. 2 and 3. All kinetic parameters of gene expression, particularly the transcription and translation (b, k_p) rates as well as the proportionality constant λ, can be estimated with reasonable precision. For the cases with slow extrinsic fluctuations (*B* and *C*), the parameters of the extrinsic noise γ_ζ and σ^2_ζ have been estimated from data. In cases *A* and *D* where extrinsic fluctuations are fast the obtained posterior distribution are not much different from the uninformative prior distributions (Fig. 3). This is due to a lack of information about these parameters in the data, which results from low-pass filtering predicted by the analysis of the power spectral density (Supporting Material). Although we cannot precisely estimate the values of γ_ζ and σ^2_ζ, we can detect the filtering effect that is revealed by the similarity of the prior and posterior distributions. This is presented in Fig. 3, where prior and posterior distributions for these parameters are plotted. We used uninformative exponential priors (see Table 1). In contrast to cases *A* and *D*, the posteriors and prior distributions are significantly different for cases *B* and *C* as the slow extrinsic fluctuations are displayed by the data.

TABLE 1 Parameter values that correspond to the four different noise characteristics

Parameter	*A*	*B*	*C*	*D*	Prior distributions
γ_r	0.44	0.44	0.44	0.1	Γ(0.44, 0.01)
γ_p	0.52	0.52	0.52	0.52	Γ(0.52, 0.01)
b	100	200	200	0.5	Exp(1000)
k_p	1	0.5	0.5	30	Exp(1000)
γ_ζ	5	0.5	0.01	5	Exp(10)
σ_ζ	1	0.1	0.002	1.25	Exp(10)
λ	1	1	1	1	Exp(10)
k_m	4.16	4.16	4.16	4.16	Γ(4.16, 0.01)
k_f	0.74	0.74	0.74	0.74	Γ(0.74, 0.01)

All rates given are per hour. These values give rise to the four different types of stochastic behavior (Fig. 1) and have been used to generate data to obtain the estimates presented in Table 2 and Fig. 2. Last column contains prior distributions used for estimation.

This example demonstrates that our method can detect the influence of extrinsic fluctuations on observed variability, and that if enough information is present in the data, the half-life and variance of the extrinsic fluctuations can be accurately estimated.

The separation of slow and fast fluctuations can be achieved by fitting a two-component autocorrelation function as shown in Rosenfeld et al. (3). Nevertheless, such an ad hoc procedure will not provide information about the kinetic parameters of gene expression and cannot distinguish between the sources of fast and slow fluctuations. Moreover, Eq. 13 shows that fluorescent fluctuations can contain more than two timescales. Therefore, our method provides a more insightful quantification method. However, its application requires prior knowledge about degradation and maturation rates.

An oscillatory gene

Most often, experimental data exhibit nonequilibrium behavior (9,31). Theoretical models of gene expression have focused on analysis of steady-state distributions (4,6,7) with relatively little work done to analyze nonequilibrium protein fluorescent trajectories (9,32). In this section we demonstrate that our method can be applied to a system that never reaches a steady state. Although we draw similar conclusions to those in the previous section, this study demonstrates that the method can be applied to a variety of biologically relevant experiments (9,31). We use oscillatory dynamics (similarly as in (9)) as an example of nonequilibrium expression. In this case the deterministic component $\tau(t)$ of the transcription process $k_r(t)$ is modeled as

$$\tau(t) = b_0 \sin\left(\frac{2\pi}{24}(b_1 t + b_2)\right) + b_3. \qquad (22)$$

Both slow (half-life 3.5 h) and fast (half-life 21 min) regimes of transcriptional fluctuations are considered (see Table S1 in the Supporting Material for all parameter

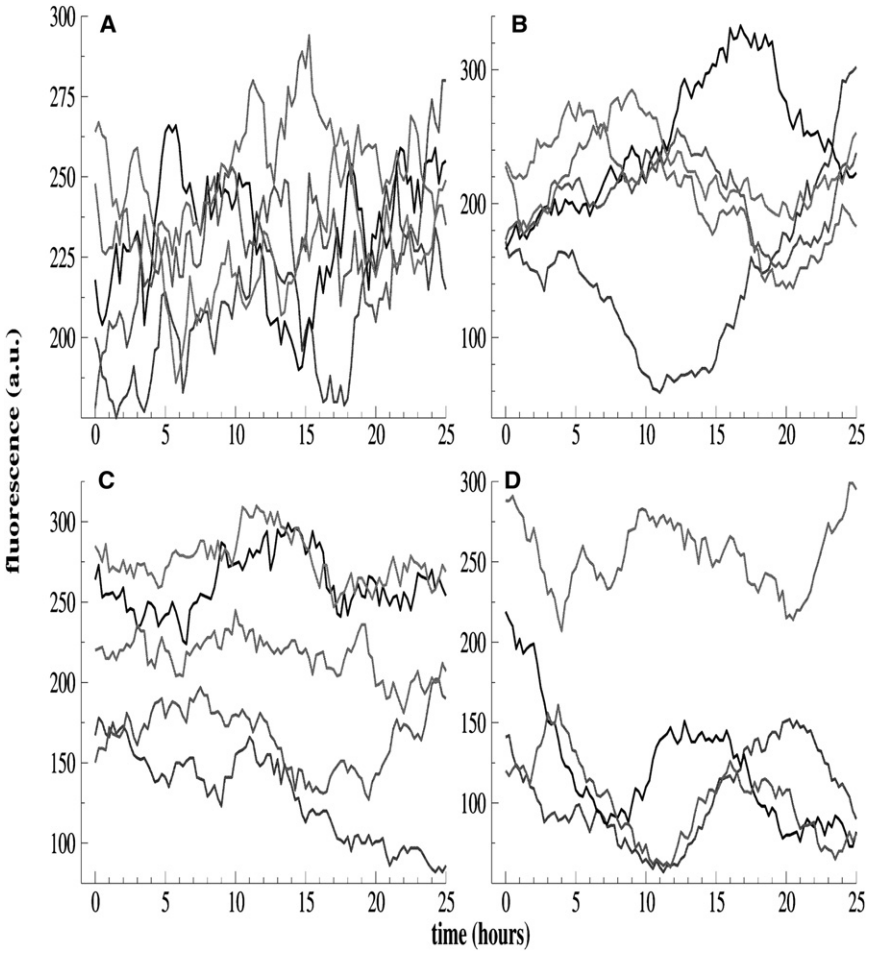

FIGURE 1 Different noise characteristics exhibited in the fluctuations of fluorescent signal. (*A*) Fast extrinsic fluctuations. (*B*) Medium extrinsic fluctuations. (*C*) Slow extrinsic fluctuations. (*D*) Fast extrinsic fluctuations and long mRNA half-life. Data has been generated using the Gillespie algorithm (see Supporting Material) with parameters presented in Table 1.

values). Fig. 4 shows data generated using Gillespie's algorithm (see Supporting Material). As presented in Table S1 and Fig. 5, the parameters of transcription and translation processes are estimated with accurate precision. For the case of slow extrinsic fluctuations, the parameters γ_ζ and $\sigma^2{}_\zeta$ are also inferred precisely. In the case of fast extrinsic fluctuations, the inferred posterior distributions of γ_ζ and $\sigma^2{}_\zeta$ are not much different from the uninformative prior distributions, which demonstrate the detection of the filtering effect.

Necessity of all model components

We find that all the components of the model (1–11,22) are necessary to ensure reliable interpretation of the fluorescent

TABLE 2 Posterior medians and 95% credibility intervals

Parameter	Estimate *A*	Estimate *B*	Estimate *C*	Estimate *D*
γ_r	0.46 (0.34–0.58)	0.38 (0.27–0.51)	0.44 (0.27–0.6)	0.1 (0.07–0.11)
γ_p	0.49 (0.36–0.61)	0.54 (0.37–0.7)	0.54 (0.42–0.68)	0.5 (0.38–0.61)
b	95.61 (32.90–599.35)	223 (24.18–1433)	336 (92–1255)	0.44 (0.28–0.91)
k_p	0.93 (0.07–2.94)	0.46 (0.04–2.09)	0.38 (0.04–1.3)	26.20 (10.37–43.8)
γ_ζ	14.34 (4.23–30.24)	0.61 (0.36–1.23)	0.01 (0.006–0.014)	6.36 (0.9–25.44)
$\sigma^2{}_\zeta$	5.17 (0.21–19.11)	0.15 (0.05–0.59)	0.002 (0.001–0.003)	7.016 (0.30–25.3)
λ	1.04 (0.78–1.29)	0.95 (0.69–1.25)	0.99 (0.73–1.20)	1.07 (0.82–1.32)
k_m	4.16 (3.98–4.31)	4.16 (3.97–4.31)	4.16 (3.96–4.31)	4.16 (3.97–4.305)
k_f	0.75 (0.57–0.90)	0.69 (0.54–0.85)	0.73 (0.55–0.87)	0.71 (0.54–0.85)

Estimates *A–D* each corresponds to inference from 100 independent time series generated using Gillespie's algorithm with parameters given in Table 1. Data were extracted every 15 min and 101 point per trajectory were collected. Independent and normally distributed error with variance $\sigma^2{}_\epsilon = 1$ was added to each data point. For estimation, variance of the measurement error was assumed to be known. Rates given are per hour. The estimates are based on the final 20,000 iterations of a run of 30,000 MCMC iterations. To ensure identifiability of all model parameters, we assumed that, for both degradation rates and protein maturation parameters, k_f and k_m informative prior distributions are available (see Table 1). Prior distributions for all other parameters were specified to be noninformative.

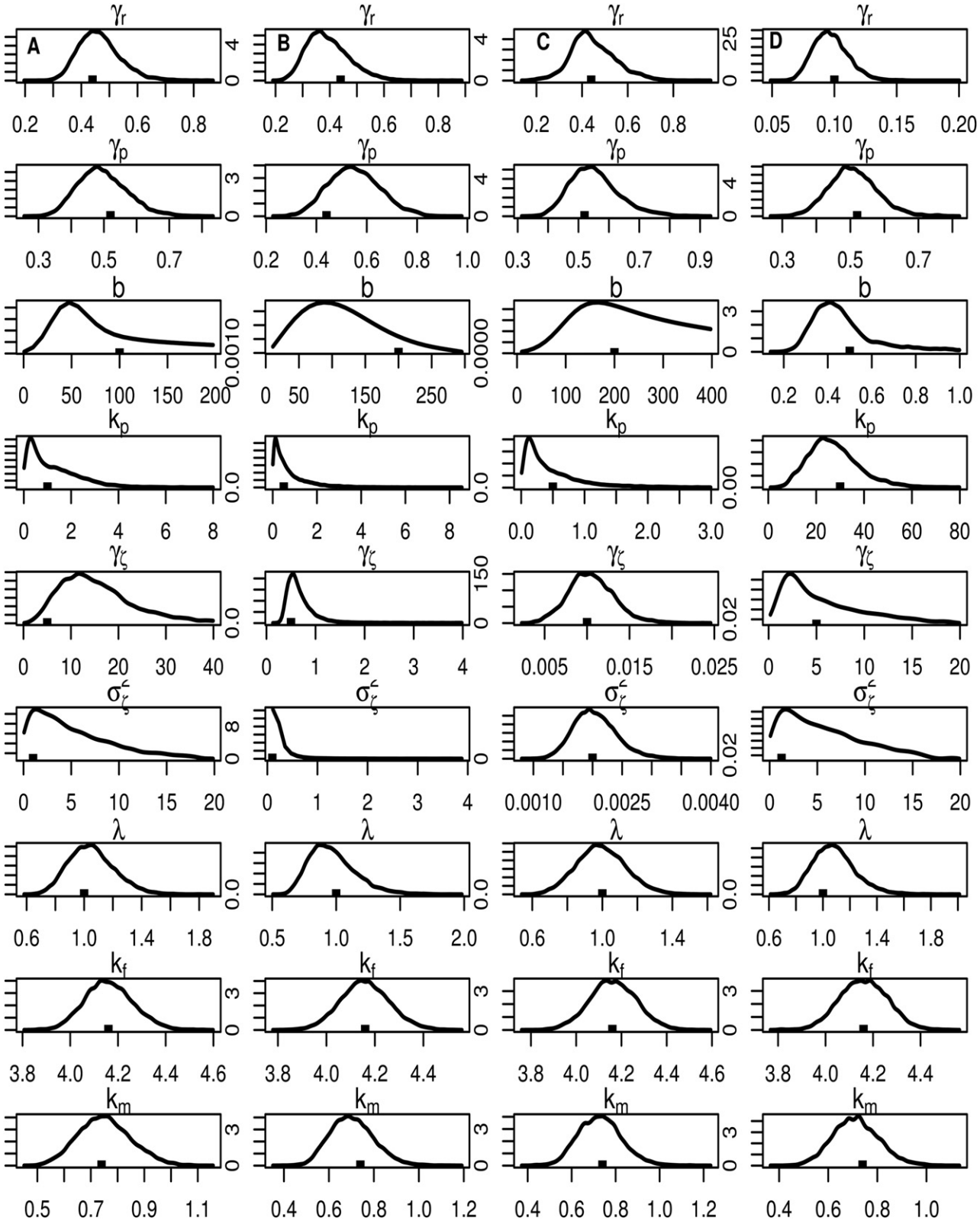

FIGURE 2 Posterior distributions corresponding to estimates presented in Table 2. (*Solid lines*) Kernel density estimators of the posterior distributions obtained from MCMC samples. (*Solid points*) True value of the parameters.

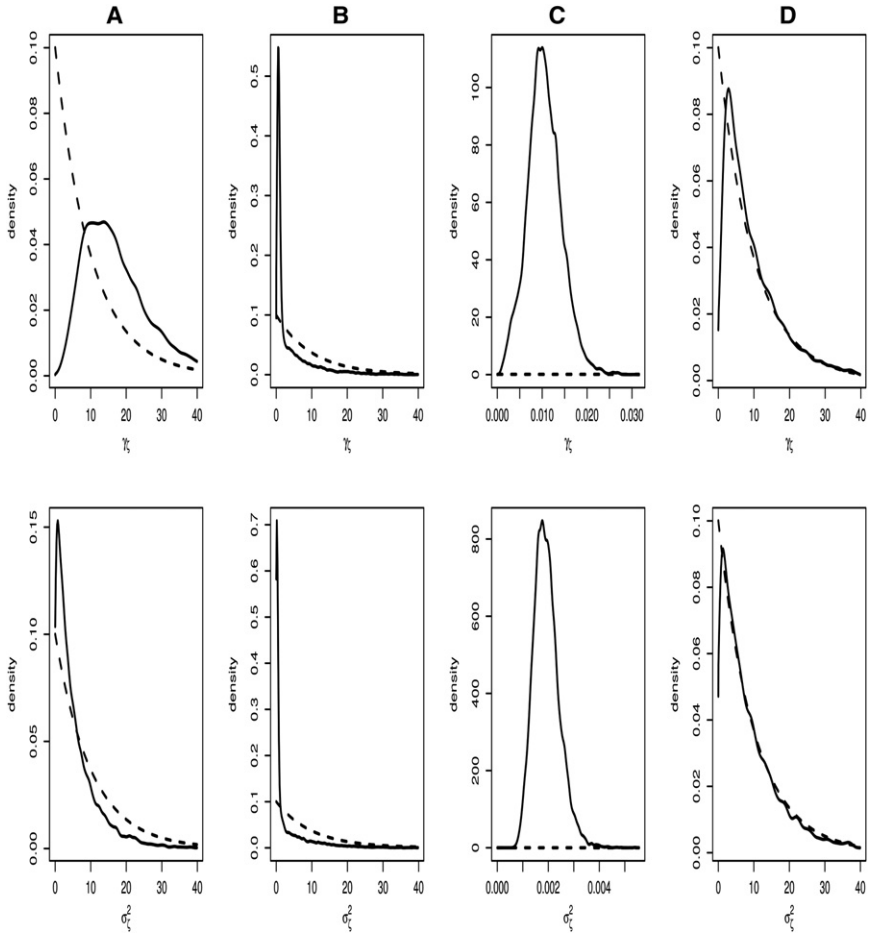

FIGURE 3 Detection of extrinsic noise in steady-state data. Prior distributions (*red line*) and posterior distributions (*black line*) of parameters γ_ζ (*top row*) and $\sigma^2{}_\zeta$ (*bottom row*). Posterior distributions correspond to estimates given in Table 2. For fast extrinsic fluctuations (*A* and *D*), prior and posterior distribution are similar, demonstrating that extrinsic fluctuations have been filtered out. In contrast, posterior distributions for slow extrinsic fluctuation (*B* and *C*) are significantly different from prior distributions and represent information about extrinsic fluctuations contained in the data.

signal. To show this, we consider two submodels of model (1–11,22). The first submodel assumes immediate maturation, i.e., we assume that we observe

$$u_{t_i} = \lambda p_{t_i} + \epsilon_{t_i} \text{ and } k_f = k_m = 0.$$

The second submodel assumes immediate maturation and lack of extrinsic noise, i.e.,

$$\gamma_\zeta = \sigma^2_\zeta = 0$$

We have generated 400 independent trajectories from the full model using Gillespie's algorithm (see Supporting Material), assuming that the deterministic part of transcription is oscillatory, as given by Eq. 22. We intentionally simulated a large data set in this example to minimize uncertainty about the model parameters arising from any shortage of data. Then we used the full model (1–11,22) and both submodels to perform inference from the generated data. The results are presented in Table 2 in the Supporting Material. As already demonstrated, estimation using models from Eqs. 1–11 and 22 provides accurate values. Because a large data set has been used, this demonstrates that application of the linear noise approximation does not result in any significant estimation bias. Inference using submodel 1 results in substantial bias in the estimates of the translation rate k_p and of the phase shift parameter b_2. This demonstrates that the incorporation of the protein maturation process is necessary to obtain the underlying transcription profile.

Estimates of all model parameters were subject to substantial bias if submodel 2 was used. As intuitively expected, this bias decreases as both protein maturation process and extrinsic fluctuations become fast enough (data not shown). Nevertheless, fast maturation and fast extrinsic fluctuations are not common (3,20,22,38) and therefore our method provides a much needed and convenient tool to interpret a fluorescent signal in the presence of slow extrinsic noise and slow maturation.

DISCUSSION

The aim of this article is to suggest a reliable framework for the interpretation of fluorescent reporter gene, single-cell, steady-state, and out-of-steady-state data. We have developed a model that shows how the observed variability depends on the kinetic parameters of a fluorescent reporter expression. The model is combined with a statistical inference framework that allows us to explain the behavior observed in an experiment in terms of the underlying parameter values. Apart from stochasticity resulting from randomness of transcription, translation, and degradation events, our

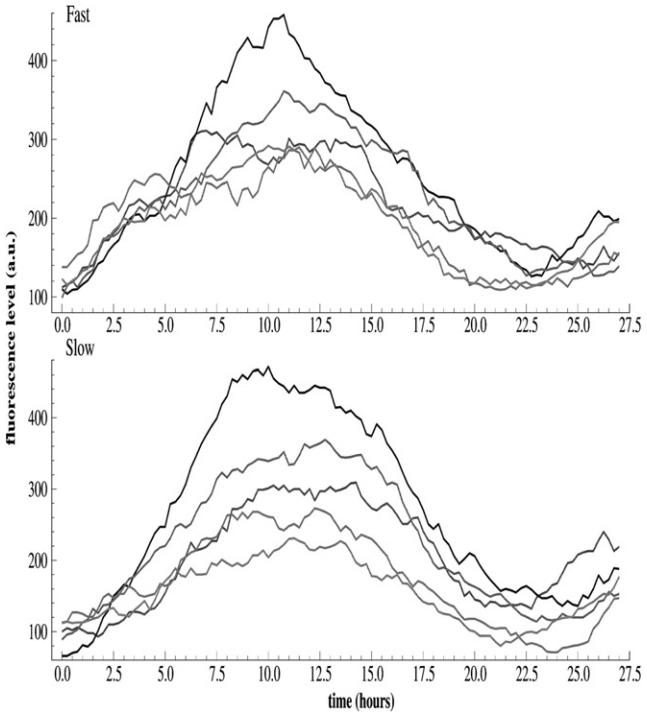

FIGURE 4 Different noise characteristics exhibited in the fluctuations of the fluorescence level for out-of-steady-state expression. (*Top*) Fast extrinsic fluctuations. (*Bottom*) Slow extrinsic fluctuations. Data generated using Gillespie's algorithm using parameters presented in Table S1 of Supporting Material.

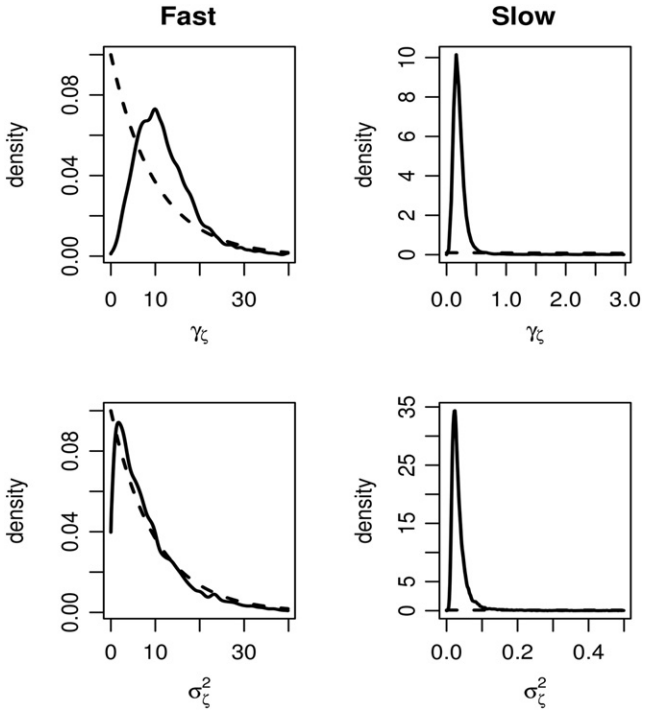

FIGURE 5 Detection of extrinsic noise in out-of-steady-state data. Prior distributions (*red line*) and posterior distributions (*black line*) of parameters γ_ζ (*top row*) and σ^2_ζ (*bottom row*). Distributions correspond to the estimates for an oscillatory gene given in Table S1 in Supporting Material. Fast extrinsic fluctuations are not exhibited in the data, therefore prior and posterior distributions are similar. In case of slow extrinsic fluctuations, posterior distribution is significantly distinct from prior distribution and contains information about extrinsic noise present in the data. Prior distributions used in both cases are the same, but look merely different due to the different *y*-axis scales.

approach accounts for variability arising from the kinetics of fluorescent protein maturation as well as extrinsic noise. Because the sources of extrinsic variability are currently unknown, we modeled it as fluctuations in transcription. Although this assumption may be or may not be true for any particular experimental system, the methodology presented here may be used to build analogous models with different extrinsic noise sources and can be combined with the statistical model selection framework (43) to investigate origins of extrinsic variability. In the context of this article, the method allows us to infer properties of extrinsic noise such as strength and half-life from single reporter-gene time-lapse data, whereas other established methods require double reporter-gene experiments.

To perform parameter inference we used the linear noise approximation to derive an explicit formula for the likelihood of fluorescent reporter gene data measured with error. The procedure suggested here is implemented in a Bayesian framework using Markov chain Monte Carlo (MCMC) simulation to generate posterior distributions. We assure identifiability of model parameters by assuming that informative prior distributions for mRNA and protein degradation rates as well as maturation parameters of fluorescent reporter are available and also that the variance of measurement error is known. Therefore, the disadvantage of this approach is that it requires additional prior experiments to determine these parameters; nevertheless, they can be measured in a relatively straightforward way described in Gordon et al. (42). For some fluorescent proteins such as GFP, maturation rates can be found in Tsien (22). We have successfully tested our approach using data simulated with Gillespie's algorithm and demonstrated that protein maturation and extrinsic noise must be taken into account to reliably interpret the fluorescent signal.

We also investigated how the maturation process and transcriptional extrinsic noise influence the dynamic properties of the fluorescence fluctuations as characterized by the ACF and the power spectral density. These investigations revealed that both processes significantly affect the rate at which the ACF decays. Furthermore, they showed that the maturation process works as a low-pass filter that filters out fast fluctuations in the transcription rate.

In the field of quantitative gene expression, promoter-fluorescent-protein fusions are commonly used as reporters of transcriptional activity. This technique is used to address many important questions, particularly to investigate the ability of a living cell to grow, divide, sense, and respond to its environment in the presence of spontaneous

fluctuations in their biochemical machinery. Experiments focused on establishing the origins of variability in gene expression observed from isogenic cell populations have influenced the view of how genes are regulated and how variability between cells arises (3,11,24). Recent investigations draw attention to the assumption in the current studies that the fluorescent protein expression reflects the endogenous protein expression (21,23,44), potentially leading to errors in interpretation. Here we confirm these findings indicating that to accurately explain the magnitude, origins, and temporal dynamics of variability in gene expression from fluorescence measurements, a mathematical model is required that accounts for the properties of the reporter protein. Our novel inference framework accounts for this factor and therefore allows us to reliably obtain a dynamical, detailed picture of the noise in terms of the model parameters.

SUPPORTING MATERIAL

Supporting Material containing derivation of theoretical results and details about algorithm implementation is available at http://www.biophysj.org/biophysj/supplemental/S0006-3495(10)00365-6.

M.K. thanks B.F. and D.R. for supervising his PhD research and for three years of fruitful collaboration.

D.R. is funded by the Engineering and Physical Sciences Research Council (Senior Fellowship grants No. EP-C544587-1 and No. GR-S29256-01) to M.K. by the University of Warwick, and M.K. and D.R. were funded by the European Union Biomedical Simulations and Imaging Laboratory Network contract No. 005137.

REFERENCES

1. Chalfie, M., Y. Tu, …, D. C. Prasher. 1994. Green fluorescent protein as a marker for gene expression. *Science*. 263:802–805.
2. Nelson, D. E., A. E. C. Ihekwaba, …, M. R. White. 2004. Oscillations in NF-κB signaling control the dynamics of gene expression. *Science*. 306:704–708.
3. Rosenfeld, N., J. W. Young, …, M. B. Elowitz. 2005. Gene regulation at the single-cell level. *Science*. 307:1962–1965.
4. Thattai, M., and A. van Oudenaarden. 2001. Intrinsic noise in gene regulatory networks. *Proc. Natl. Acad. Sci. USA*. 151:588–598.
5. Kepler, T. B., and T. C. Elston. 2001. Stochasticity in transcriptional regulation: origins, consequences, and mathematical representations. *Biophys. J.* 81:3116–3136.
6. Paulsson, J. 2006. Summing up the noise in gene networks. *Nature*. 427:415–418.
7. Friedman, N., L. Cai, and X. S. Xie. 2006. Linking stochastic dynamics to population distribution: an analytical framework of gene expression. *Phys. Rev. Lett*. 97:168302.
8. Ozbudak, E. M., M. Thattai, …, A. van Oudenaarden. 2002. Regulation of noise in the expression of a single gene. *Nat. Genet*. 31:69–73.
9. Chabot, J. R., J. M. Pedraza, …, A. van Oudenaarden. 2007. Stochastic gene expression out-of-steady-state in the cyanobacterial circadian clock. *Nature*. 450:1249–1252.
10. Xie, X. S., P. J. Choi, …, G. Lia. 2008. Single-molecule approach to molecular biology in living bacterial cells. *Annu. Rev. Biophys*. 37:417–444.
11. Elowitz, M. B., A. J. Levine, …, P. S. Swain. 2002. Stochastic gene expression in a single cell. *Science*. 297:1183–1186.
12. Raser, J. M., and E. K. O'Shea. 2005. Noise in gene expression: origins, consequences, and control. *Science*. 309:2010–2013.
13. Raj, A., and A. van Oudenaarden. 2008. Nature, nurture, or chance: stochastic gene expression and its consequences. *Cell*. 135:216–226.
14. Keizer, J. 1987. Statistical Thermodynamics of Nonequilibrium Processes. Springer, New York.
15. Guptasarma, P. 1995. Does replication-induced transcription regulate synthesis of the myriad low copy number proteins of *Escherichia coli*? *Bioessays*. 17:987–997.
16. Arias, A. M., and P. Hayward. 2006. Filtering transcriptional noise during development: concepts and mechanisms. *Nat. Rev. Genet*. 7:34–44.
17. Swain, P. S., M. B. Elowitz, and E. D. Siggia. 2002. Intrinsic and extrinsic contributions to stochasticity in gene expression. *Proc. Natl. Acad. Sci. USA*. 99:12795–12800.
18. Finkenstädt, B., E. A. Heron, …, D. A. Rand. 2008. Reconstruction of transcriptional dynamics from gene reporter data using differential equations. *Bioinformatics*. 24:2901–2907.
19. Wu, J. Q., and T. D. Pollard. 2005. Counting cytokinesis proteins globally and locally in fission yeast. *Science*. 310:310–314.
20. Nagai, T., K. Ibata, …, A. Miyawaki. 2002. A variant of yellow fluorescent protein with fast and efficient maturation for cell-biological applications. *Nat. Biotechnol*. 20:87–90.
21. Dong, G., and D. McMillen. 2008. Effects of protein maturation on the noise in gene expression. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys*. 77:21908.
22. Tsien, R. Y. 1998. The green fluorescent protein. *Annu. Rev. Biochem*. 67:509–544.
23. Wang, X., B. Errede, and T. C. Elston. 2008. Mathematical analysis and quantification of fluorescent proteins as transcriptional reporters. *Biophys. J*. 94:2017–2026.
24. Pedraza, J., and A. van Oudenaarden. 2005. Noise propagation in gene networks. *Science*. 307:1965–1969.
25. Blake, W. J., M. Kaern, …, J. J. Collins. 2003. Noise in eukaryotic gene expression. *Nature*. 422:633–637.
26. Elerian, O., S. Chib, and N. Shephard. 2001. Likelihood inference for discretely observed nonlinear diffusions. *Econometrica*. 69:959–993.
27. Reinker, S., R. Altman, and J. Timmer. 2006. Parameter estimation in stochastic biochemical reactions. *Systems Biol. IEE Proc*. 153: 168–178.
28. Golightly, A., and D. J. Wilkinson. 2005. Bayesian inference for stochastic kinetic models using a diffusion approximation. *Biometrics*. 61:781–788.
29. Heron, E. A., B. Finkenstädt, and D. A. Rand. 2007. Bayesian inference for dynamic transcriptional regulation; the Hes1 system as a case study. *Bioinformatics*. 23:2596–2603.
30. Komorowski, M., B. Finkenstädt, …, D. A. Rand. 2009. Bayesian inference of biochemical kinetic parameters using the linear noise approximation. *BMC Bioinformatics*. 10:343.
31. Sigal, A., R. Milo, …, U. Alon. 2006. Variability and memory of protein levels in human cells. *Nature*. 444:643–646.
32. Rausenberger, J., and M. Kollmann. 2008. Quantifying origins of cell-to-cell variations in gene expression. *Biophys. J*. 95:4523–4528.
33. Elf, J., and M. Ehrenberg. 2003. Fast evaluation of fluctuations in biochemical networks with the linear noise approximation. *Genome Res*. 13:2475–2484.
34. Van Kampen, N. 2006. Stochastic Processes in Physics and Chemistry. North-Holland, Amsterdam, The Netherlands.
35. Komorowski, M., J. Miekisz, and A. Kierzek. 2009. Translational repression contributes greater noise to gene expression than transcriptional repression. *Biophys. J*. 96:2064–2081.

36. Dong, C. G., L. Jakobowski, and D. R. McMillen. 2006. Systematic reduction of a stochastic signaling cascade model. *J. Biol. Phys.* 32:173–176.

37. Iafolla, M. A., and D. R. McMillen. 2006. Extracting biochemical parameters for cellular modeling: a mean-field approach. *J. Phys. Chem. B*. 110:22019–22028.

38. Shahrezaei, V., J. Ollivier, and P. Swain. 2008. Colored extrinsic fluctuations and stochastic gene expression. *Mol. Syst. Biol.* DOI:10.1038/msb.2008.31.

39. Gardiner, C. 1985. Handbook of Stochastic Methods. Springer, New York.

40. Austin, D. W., M. S. Allen, …, M. L. Simpson. 2006. Gene network shaping of inherent noise spectra. *Nature*. 439:608–611.

41. Gamerman, D., and H. F. Lopes. 2006. Markov Chain Monte Carlo Stochastic Simulation for Bayesian Inference, 2nd Ed. Chapman & Hall/CRC, Boca Raton, FL.

42. Gordon, A., A. Colman-Lerner, …, R. Brent. 2007. Single-cell quantification of molecules and rates using open-source microscope-based cytometry. *Nat. Methods*. 4:175–181.

43. Gelman, A., J. Carlin, …, D. Rubin. 2004. Bayesian Data Analysis. Texts in Statistical Science Series. Chapman & Hall/CRC, Boca Raton, FL.

44. Chubb, J. 2008. Faculty of 1000 Biology: evaluations for G. Q. Dong and D. R. McMillen. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* http://www.f1000biology.com/article/id/1119120/evaluation.