The Annals of Applied Statistics
2013, Vol. 7, No. 4, 1960–1982
DOI: 10.1214/13-AOAS669

QUANTIFYING INTRINSIC AND EXTRINSIC NOISE IN GENE TRANSCRIPTION USING THE LINEAR NOISE APPROXIMATION: AN APPLICATION TO SINGLE CELL DATA[1]

BY BÄRBEL FINKENSTÄDT[2,*], DAN J. WOODCOCK[2,*], MICHAL KOMOROWSKI[3,†], CLAIRE V. HARPER[4,‡], JULIAN R. E. DAVIS[5,‡], MIKE R. H. WHITE[5,‡] AND DAVID A. RAND[2,6,*]

University of Warwick, Polish Academy of Sciences† and University of Manchester‡*

A central challenge in computational modeling of dynamic biological systems is parameter inference from experimental time course measurements. However, one would not only like to infer kinetic parameters but also study their variability from cell to cell. Here we focus on the case where single-cell fluorescent protein imaging time series data are available for a population of cells. Based on van Kampen's linear noise approximation, we derive a dynamic state space model for molecular populations which is then extended to a hierarchical model. This model has potential to address the sources of variability relevant to single-cell data, namely, intrinsic noise due to the stochastic nature of the birth and death processes involved in reactions and extrinsic noise arising from the cell-to-cell variation of kinetic parameters. In order to infer such a model from experimental data, one must also quantify the measurement process where one has to allow for nonmeasurable molecular species as well as measurement noise of unknown level and variance. The availability of multiple single-cell time series data here provides a unique testbed to fit such a model and quantify these different sources of variation from experimental data.

1. Introduction. The effect of stochasticity ("*noise*") on linear and nonlinear dynamical systems has been studied for some time, yet significant new aspects continue to be discovered. Here we consider population dynamical systems, that is, systems which model the dynamics of species of populations stochastically by

Received January 2013; revised June 2013.

[1]The Centre for Cell Imaging has been supported through BBSRC REI Grant BBE0129651.

[2]Supported by BBSRC and EPSRC (GR/S29256/01, BB/F005814/1) and EU BIOSIM Network Contract 005137.

[3]Supported from University of Warwick, Department of Statistics, and the Foundation for Polish Science under the program Homing Plus HOMING 2011-3/4.

[4]Supported by The Professor John Glover Memorial Postdoctoral Fellowship.

[5]Supported by a Wellcome Trust Programme Grant 67252.

[6]Supported by an EPSRC Senior Fellowship (EP/C544587/1).

Key words and phrases. Linear noise approximation, kinetic parameter estimation, intrinsic and extrinsic noise, state space model and Kalman filter, Bayesian hierarchical modeling.

birth and death processes. This modeling framework has been widely applied in many scientific fields, including molecular biology, ecology, epidemiology and chemistry. Examples include predator-prey population dynamics [McKane and Newman (2005)], SIR-type epidemic modeling [Simoes, Telo da Gama and Nunes (2008)], genetic networks [Scott, Ingalls and Kaern (2006)], molecular clocks [Gonze, Halloy and Goldbeter (2002)] and biochemical networks [Wilkinson (2011)].

Basic cellular processes such as gene expression are fundamentally stochastic with randomness in molecular machinery and interactions leading to cell-to-cell variations in mRNA and protein levels. This stochasticity has important consequences for cellular function and it is therefore important to quantify it [Thattai and van Oudenaarden (2001), Paulsson (2004, 2005), Swain, Elowitz and Siggia (2002)]. Elowitz et al. (2002) defined *extrinsic noise* in gene expression in terms of fluctuations in the amount or activity of molecules such as regulatory proteins and polymerases which in turn cause corresponding fluctuations in the output of the gene. They pointed out that such fluctuations represent sources of extrinsic noise that are global to a single cell but vary from one cell to another. On the other hand, *intrinsic noise* for a given gene was defined in terms of the extent to which the activities of two identical copies of that gene in the same intracellular environment fail to correlate because of the random microscopic events that govern the timing and order of reactions. We can therefore consider that much of what makes up extrinsic noise can be expressed mathematically in terms of the stochastic variation between kinetic parameters across a population of cells.

The availability of replicate single cell data provides us with a unique opportunity to estimate and explicitly quantify such between-cell variation. Recent developments in fluorescent microscopy technology allow for levels of reporter proteins such as green fluorescent protein (GFP) and luciferase to be measured in vivo in individual cells [Harper et al. (2011)]. Here, an important issue is to relate the unobserved dynamics of expression of the gene under consideration to the observed fluorescence levels of the reporter protein [Finkenstädt et al. (2008)]. This is facilitated by knowledge of the kinetic parameters associated with the translational and degradational processes of the reporter protein and mRNA. In this study we present a methodology for estimating these rates and their cell-to-cell variation. The approach can be seen as an example of a general modeling framework which has the potential to explicitly quantify and decouple both intrinsic and extrinsic noise in population dynamical systems.

The structure of our study is as follows: starting with a single cell stochastic model, we introduce the modeling approach and inference methodology. This is then extended toward a population of cells by introducing a Bayesian hierarchical structure where the cell-to-cell variation in certain parameters such as degradation rates is quantified by a probability distribution. We also introduce the basic idea of the linear noise approximation (LNA) which is a key ingredient to rendering inference computationally efficient given the complexity of a hierarchical model and

the amount of data. The performance of the Markov chain Monte Carlo (MCMC) estimation algorithms is first tested on simulated data from the model and results are presented for both simulated and real data.

2. Model of gene expression and data. Our stochastic model for a single cell follows the general model of gene expression [Paulsson (2005)] shown in Figure 1. The active gene transcribes mRNA which is then translated into protein. During this process both mRNA and protein are degraded. We wish to infer this model from multiple single cell protein imaging time series of the sort shown in Figure 2 resulting from two types of experiments (see Appendix A.1 for technical details). In the first experiment (Figure 2, left panel) the synthesis of protein was inhibited by adding the translational inhibitor Cycloheximide. In the second experiment (Figure 2, right panel) the transcriptional inhibitor Actinomycin D was added to block the synthesis of mRNA. In both cases only the fluorescent form of the protein is observable. Single cell images of green fluorescent protein (GFP) molecules were collected every 5 min at discrete time points $t_i; i = 1, \ldots, T$, and quantified to give a time series $Y(t_i)$. The images were collected simultaneously for a collection of cells. Allowing for an appropriately formulated measurement process, our aim is to provide a statistical methodology for estimating the kinetic parameters and for quantifying their variability between cells. The model of gene expression in Figure 1 constitutes a system of two different molecular subpopulations, namely, mRNA and protein, with state vector $X(t) = (X_1(t), X_2(t))^T$ where $X_i(t), i = 1, 2$, denotes the number of molecules of each species, respectively. There are $m = 4$ possible reactions (transcription, degradation of the mRNA, translation, degradation of the protein), where a reaction of type j changes $X(t)$ to

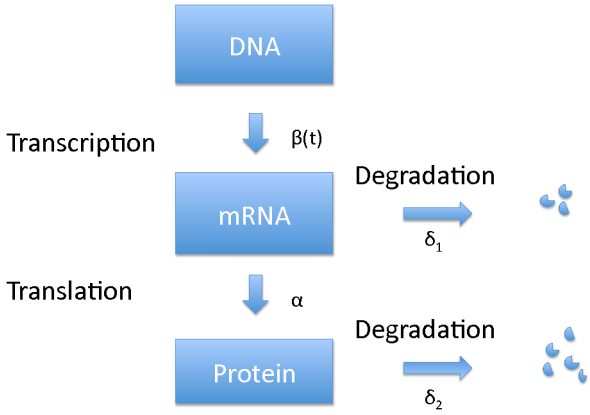

FIG. 1. *The expression of a gene into its protein is determined by four processes*: *transcription, translation, mRNA degradation and protein degradation.*

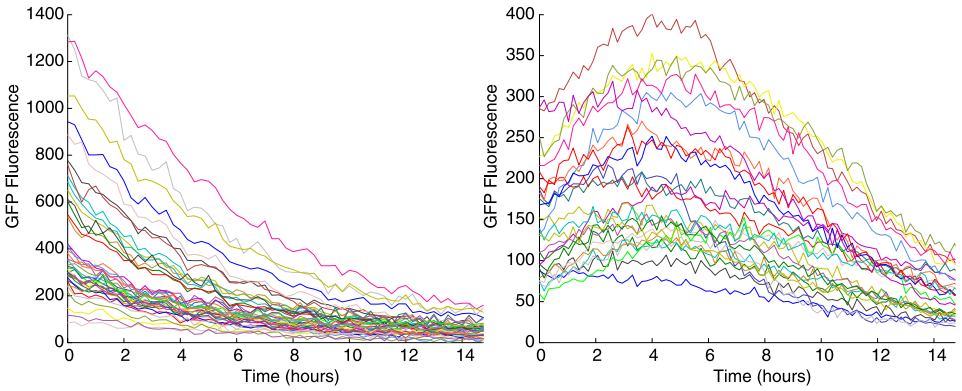

FIG. 2. Left: *observed fluorescence level from translation inhibition experiment* (40 *cells*, 59 *observations per cell*). Right: *observed fluorescence level from transcription inhibition experiment* (25 *cells*, 88 *observations per cell*). *For both experiments measurements were taken simultaneously in cells every* 5 *minutes*.

$X(t) + v_j$ with

$$v_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \qquad v_2 = \begin{pmatrix} -1 \\ 0 \end{pmatrix}, \qquad v_3 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \qquad v_4 = \begin{pmatrix} 0 \\ -1 \end{pmatrix},$$

called stoichiometric vectors. Each reaction occurs at a rate $w_j(X(t))$ as summarized in Table 1.

3. Inference for reaction networks. Reaction networks such as the one introduced above constitute continuous time Markov jump processes and thus satisfy the Chapman–Kolmogorov equation for which one can obtain the forward form known as the master equation (ME) describing the evolution of the probability $P(X_1 = n_1, X_2 = n_2; t)$. Although an exact numerical simulation algorithm is provided [Gillespie (1977)], the ME is rarely tractable and, hence, an explicit formula for the exact likelihood is not available for parameter inference. Additionally, longitudinal data from such systems are usually discrete in time and only partially observed, that is, not all molecular species are measurable during the experiment. This poses further challenges to the estimation problem.

TABLE 1
Summary of reactions in the standard model of gene expression

Event	**Effect**	**Transition rate**
Transcription	$(X_1, X_2) \to (X_1 + 1, X_2)$	$w_1 = \beta(t)$
Degradation of mRNA	$(X_1, X_2) \to (X_1 - 1, X_2)$	$w_2 = \delta_1 X_1(t)$
Translation	$(X_1, X_2) \to (X_1, X_2 + 1)$	$w_3 = \alpha X_1(t)$
Degradation of protein	$(X_1, X_2) \to (X_1, X_2 - 1)$	$w_4 = \delta_2 X_2(t)$

One way forward is to consider suitable approximations of the likelihood function. In particular, the diffusion approximation describing the process by a set of Itô stochastic differential equations (SDEs), also called chemical Langevin equations, has been of use in this context. This approach is rigorously modeling the intrinsic noise of the stochastic dynamics of the kinetic processes provided that the assumptions of the SDE approximation itself are valid. However, the associated likelihood is also usually intractable and approximations have to be considered using numerical simulations [Golightly and Wilkinson (2005, 2008), Heron, Finkenstädt and Rand (2007)]. Here, the basic idea, which has earlier been considered in econometric applications of SDEs to discretely observed data [Durham and Gallant (2002), Elerian, Chib and Shephard (2001)], is to assume a Gaussian approximation to the transition density [Kloeden and Platen (1999)] for a sufficiently small time interval and to augment the discretely observed data, along with any other unobserved variables, by introducing a number of latent data points, thus creating a fine virtual discrete time grid for which the normal approximation is valid. In particular, for partially observed systems the resulting estimation algorithms are challenging to implement and computationally intensive even for single time series because the dimension of the resulting posterior density becomes very large. Heron, Finkenstädt and Rand (2007) and Golightly and Wilkinson (2008) consider additional measurement error in a partially observed system but assume that the observed data is measured at the correct population level and that the variance of the measurement error is known. Attempts to incorporate realistic assumptions about the measurement process have been extremely limited within this framework and applications have only been on artificially simulated discretized time series data from chemical networks.

Another approach that has received attention in theoretical studies of chemical networks [Elf and Ehrenberg (2003), Paulsson (2004)] is the linear noise approximation (LNA) [Van Kampen (1976, 1997)]. This decomposes the system into a set of ordinary differential equations (ODEs) for the mean and a set of linear SDEs with Gaussian transition densities for the fluctuations around the mean. While the validity of the LNA is the subject of ongoing research [Wallace et al. (2012)], its use for Bayesian inference in chemical networks was first suggested and studied for some example networks in Komorowski et al. (2009) who also provide an application to single cell time series experimental data. The LNA has the advantage that the approximated likelihood is multivariate Gaussian. Hence, it is straightforward to incorporate both Gaussian measurement error (with unknown variance) and a model relating the observed imaging data to the unobserved molecular populations, some of which may not be measurable.

4. Linear noise approximation. The LNA as formulated by Kurtz (1971, 1981) is derived directly from the underlying Markov jump process and is valid for any time interval of fixed length. However, here we follow a simplified derivation of the LNA by Wallace (2010) [see also Wilkinson (2011)] which is adequate

for our purposes. Readers should consult Kurtz (1981) for the precise assumptions and validity of the approximation.

Consider an approximation for a fixed time interval of length τ where for each reaction j we define K_j to be the number of events of type j that occur within the interval of length τ. K_j depends on the rate w_j and on τ. The system state vector can then be updated with the stoichiometric vectors v_j as

$$X(t+\tau) = X(t) + \sum_{j=1}^{m} v_j K_j\big(w_j\big(X(t)\big), \tau\big). \tag{4.1}$$

Under the assumption that τ is small enough so that the rate function $w_j(X(t))$ can be considered constant over $[t, t+\tau)$ for all j, known as the first leap condition [Gillespie (2001)], the distribution of each K_j is Poisson with mean and variance $\mathrm{E}[K_j] = \mathrm{Var}[K_j] = w_j(X(t))\tau$. If, furthermore, τ is large enough so that $w_j(X(t))\tau$ is large for all j (second leap condition), then K_j is Gaussian $K_j \sim N(w_j(X(t))\tau, w_j(X(t))\tau)$. With the Gaussian assumption the updating rule in equation (4.1) becomes

$$X(t+\tau) = X(t) + \tau \sum_{j=1}^{m} v_j w_j + \sqrt{\tau} \sum_{j=1}^{m} v_j \sqrt{w_j}\varepsilon_j, \tag{4.2}$$

where $\varepsilon_j \sim N(0,1)$, $j = 1, \ldots, m$, are independent standard normal random variables. For simplicity of notation we use w_j instead of $w_j(X(t))$. The LNA makes the ansatz

$$X(t) = \Omega\phi(t) + \sqrt{\Omega}\zeta(t), \tag{4.3}$$

that is, the process $X(t)$ can be written as the deterministic solution of the macroscopic equations $\phi(t) = (\phi_1(t), \phi_2(t))^T$ of the concentrations plus a residual stochastic process $\zeta(t) = (\zeta_1(t), \zeta_2(t))^T$ where both components are scaled appropriately by the volume Ω of the system. Van Kampen (1997) derives the LNA from a system size expansion of the ME where terms of first order give the macroscopic rate equations or ODEs, and terms of second order give a set of SDEs for the stochastic process $\zeta(t)$ (see Appendix A.2). A simplified derivation (see Appendix A.2) shows that the LNA also provides a first order Taylor approximation to equation (4.2). The macroscopic solutions for our model are

$$\begin{aligned} \frac{d\phi_1(t)}{dt} &= \beta(t) - \delta_1\phi_1(t), \\ \frac{d\phi_2(t)}{dt} &= \alpha\phi_1(t) - \delta_2\phi_2(t). \end{aligned} \tag{4.4}$$

The residual stochastic process is characterized by

$$d\zeta(t) = J\zeta(t)\,dt + B(t)\,dW(t), \tag{4.5}$$

where $W = (W_1, W_2)^T$, W_1, W_2 are independent Wiener processes,

$$B(t) = \begin{pmatrix} \sqrt{\beta(t) + \delta_1\phi_1(t)} & 0 \\ 0 & \sqrt{\alpha\phi_1(t) + \delta_2\phi_2(t)} \end{pmatrix}$$

and

$$J = \begin{pmatrix} -\delta_1 & 0 \\ \alpha & -\delta_2 \end{pmatrix}.$$

Due to the linearity of (4.5), the transition densities $P(\zeta(t_{i+1})|\zeta(t_i))$ for arbitrary time length are Gaussian. For our model J is time-independent and, hence, the mean is given by

$$\mu(t_i) = e^{J\Delta_i}\zeta(t_i),$$

that is, the solution for the deterministic part of (4.5) from a starting value $\zeta(t_i)$, where $\Delta_i = t_{i+1} - t_i$ is the length of the interval. The covariance matrix is

$$\Sigma(t_i) = \int_{t_i}^{t_{i+1}} e^{J(t_{i+1}-s)} B(s)B(s)^T \left(e^{J(t_{i+1}-s)}\right)^T ds. \tag{4.6}$$

The approach does not require equidistant measurements. In general, the integrals needed for the mean and covariance of the transition densities arising from the LNA can be either determined explicitly or computed numerically. Since $P(\zeta(t_{i+1})|\zeta(t_i))$ is $N(\mu(t_i), \Sigma(t_i))$, we have that $P(X(t_{i+1})|X(t_i))$ is $N(\Omega\phi(t) + \sqrt{\Omega}\mu(t_i), \Omega\Sigma(t_i))$. Thus, the LNA estimates the variances of the species abundances and the covariances between them and a transition from $X(t_i)$ to $X(t_{i+1})$ follows the state space equation

$$X(t_{i+1}) = F(t_i)X(t_i) + c(t_i) + \varepsilon_{t_i}, \qquad \varepsilon_{t_i} \sim N\big(\mathbf{0}, \Sigma_\varepsilon(t_i)\big), \tag{4.7}$$

where for our model $F(t_i) = e^{J\Delta_i}$, $c(t_i) = \Omega[\phi(t_{i+1}) - e^{J\Delta_i}\phi(t_i)]$ and $\Sigma_\varepsilon(t_i) = \Omega\Sigma(t_i)$.

In general, we assume a measurement equation of the form

$$Y(t_i) = \kappa X(t_i) + u(t_i), \tag{4.8}$$

where κ is a matrix, $Y(t_i)$, $i = 1, \ldots, T$, are the observed data for a single cell, and the $u(t_i) \sim N(\mathbf{0}, \Sigma_u(t_i))$ represent measurement errors with covariance matrix $\Sigma_u(t_i)$. Note that κ accounts for the assumption that imaging data may be proportional to molecular population size. State variables that cannot be measured reduce the rank of the matrix. As in all our applications below, only the protein is imaged, κ will be a scalar and $u(t_i)$ will be one-dimensional.

Let $\mathbf{X} = (X(t_1), \ldots, X(t_T))^T$ and $\mathbf{Y} = (Y(t_1), \ldots, Y(t_T))^T$. The joint density of $\mathbf{Y}$ or likelihood $L(\mathbf{Y}|\theta)$ can be obtained by writing the system of equations (4.7) for all discrete time points t_i, $i = 1, \ldots, T$ as

$$\mathbf{BX} = \mathbf{Z} + \mathbf{C} + \boldsymbol{\varepsilon}, \tag{4.9}$$

where

$$\mathbf{B} = \begin{pmatrix} \mathbf{I} & \mathbf{0} & \cdots & \cdots & \cdots & \mathbf{0} \\ -e^{J\Delta_1} & \mathbf{I} & \ddots & & & \vdots \\ \mathbf{0} & \ddots & \ddots & \ddots & & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & & \ddots & \ddots & \ddots & \mathbf{0} \\ \mathbf{0} & \cdots & \cdots & \mathbf{0} & -e^{J\Delta_{T-1}} & \mathbf{I} \end{pmatrix},$$

$$\mathbf{C} = \big(c(t_0), \ldots, c(t_{T-1})\big)^T, \qquad \mathbf{Z} = \big(e^{-J\Delta_0} X(0), \mathbf{0} \ldots \mathbf{0}\big)^T$$

and $\boldsymbol{\varepsilon} \sim N(\mathbf{0}, \Sigma_{\boldsymbol{\varepsilon}})$ with covariance matrix $\Sigma_{\boldsymbol{\varepsilon}} = \operatorname{diag}(\Sigma_\varepsilon(t_0), \ldots, \Sigma_\varepsilon(t_{T-1}))$.

It follows that

$$\mathbf{X} \sim N\big(\mathbf{B}^{-1}(\mathbf{C} + \mathbf{Z}), \mathbf{B}^{-1}\big(\Sigma_{\boldsymbol{\varepsilon}} + \operatorname{Var}(\mathbf{Z})\big)\big(\mathbf{B}^{-1}\big)^T\big). \tag{4.10}$$

The matrix formulation of the observational equation (4.8) for all time points is

$$\mathbf{Y} = \boldsymbol{\kappa}\mathbf{X} + \mathbf{u}, \tag{4.11}$$

where $\boldsymbol{\kappa}$ is the matrix that has κ along the diagonal and is zero otherwise, and $\mathbf{u} = (u(t_1), \ldots, u(t_T))^T$ with $\mathbf{u} \sim N(\mathbf{0}, \Sigma_{\mathbf{u}})$. Hence, the likelihood $L(\mathbf{Y}|\theta)$ is [see also Komorowski et al. (2009)]

$$\mathbf{Y} \sim N\big(\boldsymbol{\kappa}\mathbf{B}^{-1}(\mathbf{C} + \mathbf{Z}), \boldsymbol{\kappa}\mathbf{B}^{-1}\big(\Sigma_{\boldsymbol{\varepsilon}} + \operatorname{Var}(\mathbf{Z})\big)\big(\mathbf{B}^{-1}\big)^T\boldsymbol{\kappa}^T + \Sigma_{\mathbf{u}}\big). \tag{4.12}$$

We assume that measurement errors $u(t_i)$ are i.i.d. normal so that $\Sigma_{\mathbf{u}} = \sigma_u^2\mathbf{I}$, but the approach can be adapted to other specifications of the error covariance matrix.

In situations where the sample size and/or the dimension of $\mathbf{Y}$ is large, the density corresponding to (4.12) may be expensive to compute, as it requires inversion of large matrices. Since the model for $Y(t_i)$ can be written in state space form defined by equations (4.7) and (4.8), an equivalent form of (4.12) that is easier to handle is obtained using the prediction error decomposition with log likelihood

$$\begin{aligned} \log L(\mathbf{Y}; \theta) &= \sum_{i=1}^{T} \log f(Y_{t_i} | Y_{t_1}, \ldots, Y_{t_{i-1}}; \theta) \\ &= \sum_{i=1}^{T} \left[-\frac{\dim(Y_{t_i})}{2} \log 2\pi - \frac{1}{2} \log |R_{t_i}| - \frac{1}{2} e_{t_i}^T R_{t_i}^{-1} e_{t_i} \right], \end{aligned} \tag{4.13}$$

where $e_{t_i} = Y_{t_i} - \hat{Y}_{t_i|t_{i-1}}$ is the prediction error, $\hat{Y}_{t_i|t_{i-1}} = E(Y_{t_i}|Y_{t_1}, \ldots, Y_{t_{i-1}}; \theta)$ is the optimal predictor of Y_{t_i} given the information up to time t_{i-1} and R_{t_i} is the variance matrix of the prediction error. These quantities can be computed as part of the Kalman filter recursions (see Appendix A.3).

5. A hierarchical model for multiple cells. The full data matrix of an experiment contains N multiple imaging time series

$$\tilde{\mathbf{Y}} = \big(\mathbf{Y}^{(1)}, \mathbf{Y}^{(2)}, \ldots, \mathbf{Y}^{(N)}\big),$$

where $\mathbf{Y}^{(i)}$ are the imaging data for a cell now indexed by i; $i = 1, \ldots, N$. Bayesian hierarchical modeling [Gamerman and Lopes (2006)] provides a natural framework to account for the cell-to-cell variability of kinetic parameters. Assuming that replicates are independent, the full likelihood for all cells in the experiment is

$$L(\tilde{\mathbf{Y}}; \boldsymbol{\theta}) = \prod_{i=1}^{N} L\big(\mathbf{Y}^{(i)} | \theta^{(i)}\big), \tag{5.1}$$

where $\theta^{(i)}$ denotes the vector containing all parameters for cell i and $L(\mathbf{Y}^{(i)}|\theta^{(i)})$ is the likelihood for a single cell as derived above. In contrast to assuming that a reaction j in all cells is described by exactly the same value of the associated kinetic parameter θ_j, in a hierarchical model it is a sample from a population distribution $p(\theta_j|\Theta_j)$ which is governed by an unknown parameter vector Θ_j quantifying the mean and variance of each hierarchical parameter across the population of cells. Let $\boldsymbol{\theta} = (\theta^{(1)}, \theta^{(2)}, \ldots, \theta^{(N)})$ denote the matrix of parameter vectors and let $p(\boldsymbol{\theta}|\Theta)$ denote the joint distribution of $\boldsymbol{\theta}$ assuming

$$p(\boldsymbol{\theta}|\Theta) = \prod_{j} p(\theta_j|\Theta_j),$$

where Θ is the vector of all hyperparameters. In the hierarchical model we wish to infer upon the posterior $p(\Theta|\tilde{\mathbf{Y}})$,

$$p(\Theta|\tilde{\mathbf{Y}}) \propto L(\tilde{\mathbf{Y}}; \boldsymbol{\theta}) p(\boldsymbol{\theta}|\Theta) p(\Theta), \tag{5.2}$$

where $p(\Theta)$ denotes the prior distribution of the hyperparameters. This is achieved by formulating an appropriate MCMC algorithm that samples from $p(\Theta|\tilde{\mathbf{Y}})$.

5.1. *Translation inhibitor experiment.* We start with the translation inhibition experiment, as it allows us to estimate the protein half life which is then used as prior information for the other experiment. We assume that under the influence of the translational inhibitor the level of protein synthesis drops down to zero or possibly a small basal level τ_2 in case inhibition is not fully achieved while the initial protein degrades at per capita rate δ_2 [Gordon et al. (2007)]. The resulting model thus does not depend on the mRNA process and simplifies to the univariate case where $X(t) = X_2(t)$ with a single macroscopic equation

$$\frac{d\phi_2(t)}{dt} = \tau_2 - \delta_2 \phi_2(t).$$

The LNA noise process is thus one-dimensional $\eta = (\eta_2)$, where

$$B(t) = \sqrt{\tau_2 + \delta_2 \phi_2(t)}$$

and $J = -\delta_2$. The following parameters are assumed to be hierarchical $\delta_2^{(i)}$, $\tau_2^{(i)}$, $\sigma_u^{(i)}, i = 1, \ldots, N$. We reparameterize $\tilde{\tau}_2^{(i)} = \kappa \tau_2^{(i)}$, which significantly improves convergence of the estimation algorithm. Details on the specification of the distributions of the parameters are given in Appendix A.4.1. As the initial conditions $\phi_2^{(i)}(0)$ may be very different across cells, in particular, in experiments where the cell behavior is not synchronized, we estimate them independently for each cell rather than assuming a hierarchical structure. For simplicity, the scaling parameter κ is assumed to be constant for all cells.

5.2. *Transcription inhibitor experiment*. Similarly to the previous experiment, we assume that under the influence of a transcriptional inhibitor mRNA synthesis drops to some small basal level τ_1 while the initial amount of mRNA degrades and is also translated into protein which then degrades. The model is thus given by the full two-species model with $\beta(t) = \tau_1$. As only the protein is imaged, the measurement equation is formulated in the same way as for the previous experiment. We specify $\tau_1^{(i)}$, $\delta_1^{(i)}$, $\alpha^{(i)}$ and $\sigma_u^{(i)}$ as hierarchical parameters (see Appendix A.4.2 for details) and use the estimation results for the protein degradation parameter from the translation inhibitor experiment as the informative prior. Our simulation studies show that this is a crucial step, as without this prior the other parameters were not identifiable due to the mRNA not being observable.

6. Results.

6.1. *Simulation studies*. In order to develop the MCMC estimation algorithms, we generated artificial data of similar sample size and sampling frequency as the observed data for chosen parameter values as displayed in Table 2. Such simulation studies are vital to developing the estimation algorithm, assessing its performance, checking for bias and gaining an understanding of the precision with which parameters can be estimated. We also study the simpler case of the translation inhibition experiment to compare estimation for two scenarios, namely, a system with a large and a small number of molecules where the set parameter κ was adjusted to give values in a similar range so that the measurement error had a similar impact in both scenarios. Artificial data was generated with exact intrinsic stochasticity using the stochastic simulation algorithm [Gillespie (1977)] and normal measurement error.

Table 2 summarizes estimation results for the simulation study confirming that estimation based on the LNA is successfully reproducing posterior estimates with reasonable precision. Retrieving both κ and measurement error variance is a significant achievement, not least because it gives us an idea of the size of the molecular populations. The simulation study shows that κ is estimated with more precision for the smaller molecular system. A possible reason for this is that since the intrinsic noise scales with factor $\sqrt{\kappa}$ between the measurement and population level, the

TABLE 2
Results for simulated data

Parameter	True mean	Estimated mean	True variance	Estimated variance
		Translation inhibitor experiment		
τ_2	$3.675 \cdot 10^4$	182 (54, 1408)	$6.345 \cdot 10^8$	$25 \cdot 10^3$ $(2 \cdot 10^3, 1.5 \cdot 10^6)$
$\tilde{\tau}_2$	3.675	3.61 (2.76, 4.55)	6.345	9.48 (4.84, 18.70)
δ_2	0.576	0.56 (0.54, 0.57)	0.005	0.004 (0.002, 0.006)
σ_u^2	12	11.86 (11.02, 12.54)	3	3.89 (1.34, 7.06)
κ	10^{-4}	0.02 (0.00, 0.05)	–	–
τ_2	3.675	3.40 (2.23, 4.69)	6.345	6.54 (1.57, 17.14)
$\tilde{\tau}_2$	3.675	3.43 (2.42, 4.49)	6.345	6.67 (1.82, 17.93)
δ_2	0.576	0.56 (0.53, 0.58)	0.005	0.004 (0.001, 0.009)
σ_u^2	12	12.27 (11.05, 13.25)	3	5.12 (1.61, 11.31)
κ	1	1.01 (0.82, 1.17)	–	–
		Transcription inhibitor experiment		
τ_1	40	37.40 (30.56, 43.39)	2	5.172 (1.576, 9.84)
δ_1	0.2	0.193 (0.183, 0.208)	0.005	0.0080 (0.00013, 0.0173)
α	3.5	3.731 (2.557, 4.940)	2	1.483 (0.684, 4.056)
σ_u^2	10	9.124 (8.275, 10.144)	2	1.615 (0.363, 4.602)
κ	0.25	0.239 (0.221, 0.254)	–	–

True values and posterior estimates of the mean and variance of the distribution of hierarchical parameters for simulated data. κ is not hierarchical. For the translation inhibitor model two cases are considered: large number of molecules with $\phi_2^{(i)}(0) = 2 \times 10^6, \kappa = 10^{-4}$ (top), and small number of molecules with $\phi_2^{(i)}(0) = 500, \kappa = 1$ (bottom). For the transcription inhibition model, one case is considered with $\phi_1^{(i)}(0) = 500, \phi_2^{(i)}(0) = 2000$. Estimates are medians (with 95% interval in brackets) of the posterior chains from 40 K iterations after convergence is achieved. All rates are per hour. The choice of rate parameters was motivated by values that generate artificial data with approximately similar dynamics as the real data and using preliminary results from fitting ODEs to aggregate data.

information about the intrinsic noise is essential in identifying κ. For larger population sizes the trajectories become smoother and the intrinsic noise will be less informative about κ. We thus conjecture that while it is the stochastic approximation which facilitates calibration of the model in terms of population size, it will be less successful in doing so if the molecular population is large and a simpler ODE approximation may be assumed to be adequate.

We also developed and studied performance of our estimation algorithm for the transcription inhibitor experiment via a simulation study as reported in Table 2. Inference is more challenging for the two-dimensional model when only one vari-

able is measurable. The parameter traces tend to be more correlated and more time is needed on fine-tuning the algorithm. To enhance efficiency over the conventional Metropolis–Hastings algorithm, we implemented a modified MCMC algorithm based on the Metropolis–Hastings method. In particular, we used block sampling [Gamerman and Lopes (2006)] in combination with the multiple-try Metropolis (MTM) algorithm with antithetic sampling as in Craiu and Lemieux (2007). The original MTM algorithm [Liu, Liang and Wong (2000)] generates a number of proposals for each parameter and selects one with a probability that is proportional to its likelihood. We then construct a backward step from the chosen proposal so that the detailed balance condition is satisfied and then accept or reject this proposal in the conventional Metropolis manner. The antithetic multiple correlated try Metroplis (MCTM) method incorporates negative correlation into this framework to maximize the Euclidean distance between these proposals. Together with the reparameterization it was found to improve convergence of the estimation algorithm.

6.2. *Results for experimental data.* The results for the real data are given in Table 3 and are plotted in Figures 3 and 4 for the translation and transcription in-

TABLE 3
Results for experimental data

Parameter	Estimated mean	Estimated variance
	Translation inhibitor experiment	
τ_2	50.67 (24.12, 114.40)	1086.40 (206.41, 6155.13)
$\tilde{\tau}_2$	3.51 (2.84, 4.22)	5.07 (2.35, 9.59)
δ_2	0.57 (0.54, 0.59)	0.004 (0.002, 0.007)
σ_u^2	6.36 (5.11, 7.65)	20.07 (10.25, 35.47)
κ	0.07 (0.02, 0.12)	–
	Transcription inhibitor experiment	
τ_1	3.53 (2.64, 5.02)	9.38 (3.79, 23.90)
$\tilde{\tau}_1$	1.53 (1.10, 2.55)	4.29 (0.93, 21.12)
δ_1	0.13 (0.12, 0.15)	0.004 (0.001, 0.013)
α	3.93 (3.27, 4.80)	6.09 (3.35, 11.11)
$\tilde{\alpha}$	0.44 (0.34, 0.62)	0.08 (0.04, 0.17)
σ_u^2	5.14 (4.82, 5.90)	1.09 (0.55, 5.94)
κ	0.11 (0.09, 0.14)	–

Posterior estimates of the mean and variance of the distribution of hierarchical parameters for experimental data from translation and transcription inhibitor experiments. κ is not hierarchical. All rates are per hour. Estimates were computed from posterior chains as described in Table 2.

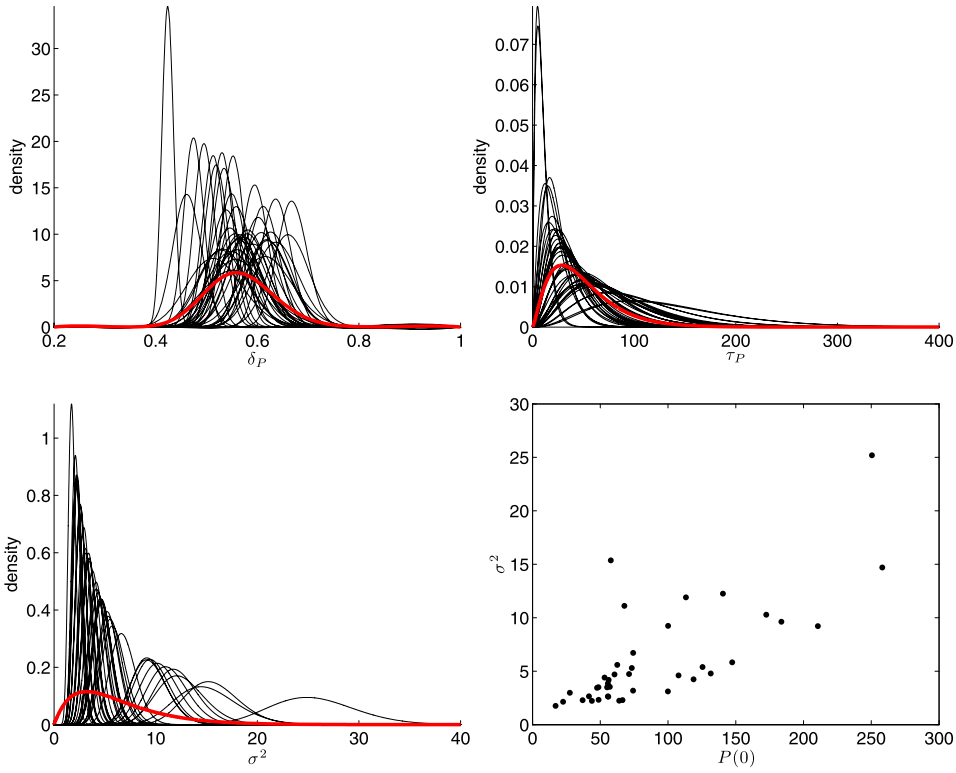

FIG. 3. Translation inhibitor experiment. *Estimated posterior densities of parameters* δ_2 (*top left*), τ_2 (*top right*) *and* σ_u^2 (*bottom left*) *from experimental data of the translation inhibitor experiment. Black solid lines give estimated posterior densities for the single cells using kernel estimation. Red solid curve gives their estimated joint density as specified by the hierarchical model. The scatterplot* (*bottom right*) *gives the estimated standard deviation* σ_u *of the measurement error against estimated initial condition of the macroscopic solution* $\phi_2(0)$ *for each cell. The empirical Spearman correlation coefficient is* 0.77.

hibition, respectively. Diagnostic tests applied to the standardized prediction error (A.13) computed for the mean posterior parameter estimates indicate that residuals do not exhibit significant autocorrelation and their distributions are compatible with normality. We use the coefficient of variation CV (ratio of standard deviation to mean) to compare between-cell variability of different parameters.

Degradation rates. Out of all estimated rates it seems the degradation rates for protein and mRNA exhibit the least cell-to-cell variation. The mean protein degradation rate was estimated to be around 0.576, which corresponds to a half-life of approximately 1.2 h. The estimated cell-to-cell variation in the degradation rate is 0.063 (standard deviation) and the CV is 0.1, indicating that there is only small variation between cells. The 2-sigma band for the estimated density of the

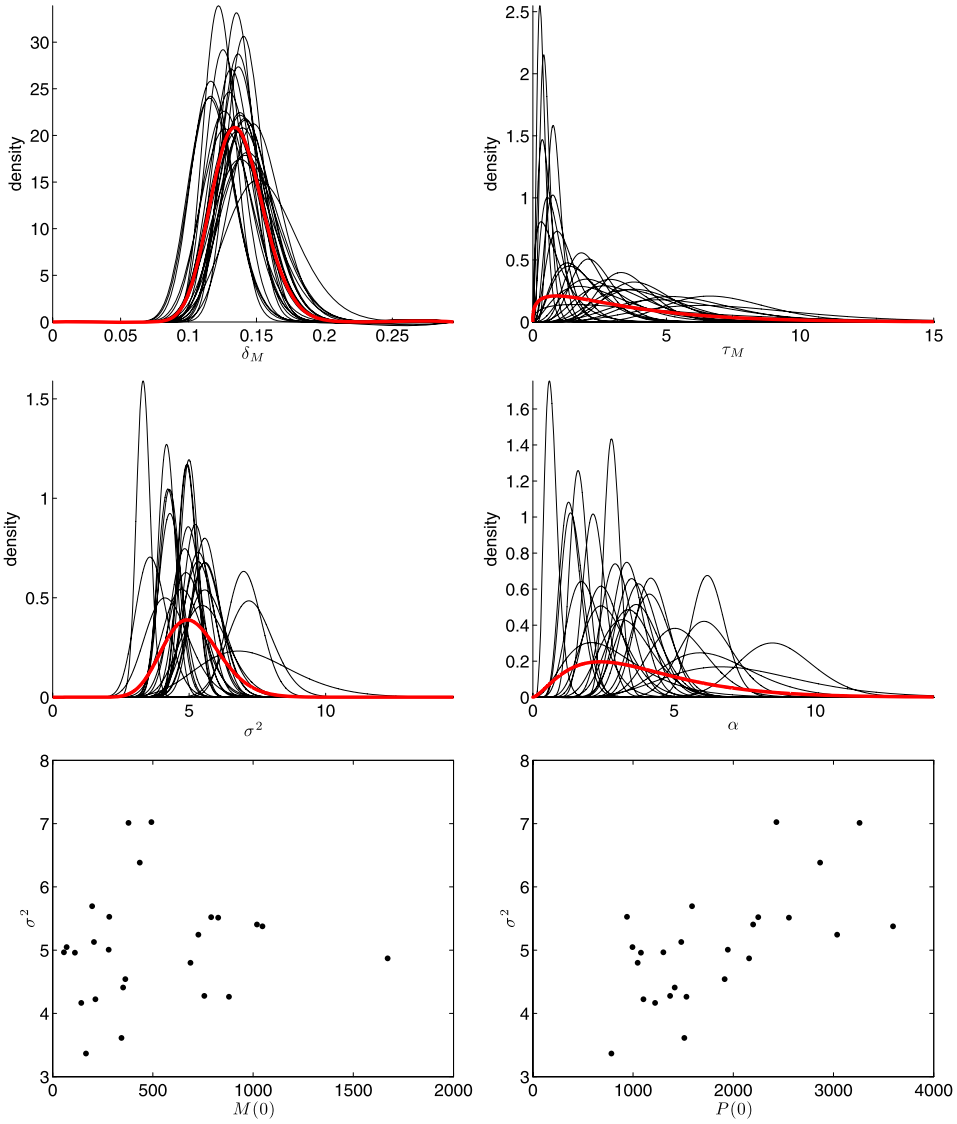

FIG. 4. Transcription inhibitor experiment. *Estimated posterior densities of parameters* δ_1 (*top left*), τ_1 (*top right*), σ_u^2 (*middle left*) *and* α (*middle right*) *from experimental data of the transcription inhibitor experiment. Black solid lines give estimated posterior densities for the single cells using kernel estimation. Red solid curve gives their estimated joint density as specified by the hierarchical model. The scatterplots give the estimated standard deviation* σ_u *of the measurement error against estimated initial conditions of the macroscopic equations* $\phi_1(0)$ (*bottom left*) *and* $\phi_2(0)$ (*bottom right*). *The empirical Spearman correlation coefficients are* 0.11 *and* 0.63, *respectively.*

protein degradation rate of the hierarchical model (red line in Figure 3 top left) is (0.45–0.70). Thus, almost all cells in the sample exhibit a mean protein half-life between 1 h and 1.5 h. The mRNA degradation rate was estimated around 0.137 (Figure 4 top left), corresponding to a half-life of 5 h. The cell-to-cell variation is about 0.02 (standard deviation) and is also very small with a CV of 0.05. From the 2-sigma band we find that almost all cells have an mRNA half-life between 4 h and 7 h.

Transcription and translation rates. The rate τ_1 under transcriptional inhibition is smaller than the rate τ_2 under translational inhibition. This is reasonable, as population size for protein can be expected to be larger than for mRNA. The CVs are 0.67 and 0.87 for τ_2 and τ_1, respectively, indicating considerably more between-cell variability than for the degradation rates. This may also be due to the fact that τ_1 and τ_2 will vary with cell size and abundance. The translation rate α is estimated around 4 with similar cell-to-cell variation as transcription rates with a CV of 0.62. We note that neither τ_2 nor τ_1 appear to be close to zero, indicating that neither treatment achieved complete inhibition. Our estimation shows that degradation rates show less variability than transcription rates. It is interesting to speculate that this is because of the extra variability that would arise from a combination of the bursting structure that has been observed in genes [Harper et al. (2011), Suter et al. (2011)] and the related effect of chromatin remodeling and transcription factor variability.

Scaling factor κ, molecular population size and measurement error. The scaling factor κ was estimated to be around 0.07 and 0.10 in the translation and transcription inhibitor experiment, respectively. Both values overlap largely in their posterior distribution, indicating that the scaling is similar between the two experiments, which is reasonable as the experimental protocol with respect to taking the images was similar. Simple plug-in estimates of the initial molecular population sizes can be obtained from the Markov chain traces. For the translation inhibitor experiment the average (over the cells) initial protein abundance is estimated to be around 1260 (median) with 95 % central range (660–3660) and similar for the transcription inhibitor experiment [1530 (median) with 95 % central range (880–3290)]. The average initial mRNA abundance is smaller [370 (median) with 95% central range (70–1150)]. There is considerable variation between cells in the initial conditions and it becomes clear that the variation seen in the data of Figure 2 is predominantly due to the cell-to-cell variation in initial population size. The estimated variance σ_u^2 of the measurement error is of similar size in both experiments. We find that the individual cell estimates of the measurement error variance correlate strongly with the initial protein abundance of the cell in both experiments (Figures 3 and 4, bottom right). The correlation between the measurement error variance and the initial mRNA population (Figure 4, bottom left) is weaker, which is also reasonable as the imaging data are more directly related to protein rather than mRNA abundance.

7. Summary and discussion. In this study we have introduced a Bayesian hierarchical model based on the LNA that can be used in the context of stochastic compartmental population models. The model has the potential to tease out different sources of variability that are relevant to inference about transcriptional, translational and degradational processes at the molecular level, namely:

- *intrinsic stochasticity* due to the natural stochastic nature of the birth and death processes involved in chemical reactions,
- *extrinsic variability*, that is, arising from the cell-to-cell variation of kinetic parameters associated with these processes, and
- *measurement noise* which is additive and not part of the dynamic process.

We focus on drawing inference about these sources of stochasticity from real experimental time series data in molecular systems biology. The two experiments considered here are an example of a scenario where the use of a rich stochastic model in combination with a Bayesian approach to inference can deal with shortcomings such as unobservable population species. It demonstrates that, provided the data are indeed of the kind to display intrinsic stochasticity, that is, come from a single cell, assuming a stochastic model is more informative than an ODE approach even if the latter fits the mean well. Here it allowed us to decouple the different sources of noise and to obtain an estimate of the scaling factor κ, leading to inference about the size of the underlying molecular species. Pilot simulation studies are an essential tool not only to develop the estimation algorithm but also to study parameter identifiability. The posterior information about the kinetic parameters obtained here will allow us further to study nonlinear transcription functions and spatial characteristics of the gene under consideration, that is, Prolactin, in spatio-temporal experiments using the same reporter construct.

Our simulation studies confirm that the LNA works well for inference in our model, also for smaller molecular population sizes. This is in line with results in Komorowski et al. (2009). Recently, Stathopoulos and Girolami (2013) combined the LNA with their Riemann manifold MCMC sampling methods and found that this combined approach is both statistically and computationally efficient for reaction networks also in the case of smaller populations. A study by Wallace et al. (2012) provides new insight on the range of validity of the LNA and supports its applicability to systems such as studied in this paper. The rigorous presentation of the LNA by Kurtz (1971, 1981) clearly states the underlying assumptions, in particular, the need to restrict to a finite time horizon before taking the limit $\Omega \to \infty$ except in the case when the ODE has a single stable stationary solution. Therefore, this method cannot be expected to work when molecular numbers are too small. The threshold population size above which the method gives good results will depend on the nature of the dynamical system and the time horizon set. For example, a system near to a bifurcation or with multiple attractors is expected to have a significantly larger threshold than one with a single globally attracting attractor. Similarly, the LNA is likely to struggle when the ODE solution in the LNA is not

equal to the mean value of the stochastic process. An obvious remedy is to reset the LNA by allowing for the initial conditions of the ODE to evolve conditional on the observed data and recent work by Fearnhead, Giagos and Sherlock (2013) shows that this improves estimation in a nonlinear predator-prey interaction model. We did not encounter this problem because in our model the Jacobian does not depend on the deterministic solution. We are also currently studying the use of the LNA for inference in systems with small populations and with nonlinear switch-type transcription functions and so far have found that the LNA copes well with the nonlinearity of the switch function and only begins to break down for extremely small population sizes.

The hierarchical model provides a useful and natural approach to study and quantify the cell-to-cell variability between kinetic parameters. The resulting model is complex and the use of the LNA has been crucial to facilitate its inference. Computational feasibility and ability to link between the model and experimental data through a realistically modeled measurement process are essential for studying problems in systems biology where in the future many more single cell data sets from complex systems will become available.

APPENDIX

A.1. Experiment. GH3-DP1 cells containing a stably integrated 5kb human prolactin-destabilised EGFP reporter gene [Harper et al. (2011)] were grown in Dulbecco's minimal essential medium plus 10% FCS and maintained at 37°C 5% CO2. Cells were seeded onto 35-mm glass coverslip-based dishes (IWAKI, Japan) and cultured for 20 h before imaging. The dish was transferred to the stage of a Zeiss Axiovert 200 microscope equipped with an XL incubator (maintained at 37°C, 5% CO2, in humid conditions). Fluorescence images were obtained using a Fluar x20, 0.75 numerical aperture (Zeiss) dry objective. Stimulus (5 μM forskolin and 0.5 μM BayK-8644) to induce an increase in prolactin gene expression was added directly to the dish for 6 h, followed by treatment with 10 μg/ml cycloheximide (for protein degradation rate) or 3 μg/ml actinomycin D (for mRNA degradation rate) and imaged for at least a further 15 h.

A.2. The linear noise approximation. The LNA approximates transitions densities by a Gaussian distribution with an appropriately-defined covariance matrix. It is usually derived as an approximation to the master equation by van Kampen's system-size expansion [Van Kampen (1976, 1997)]. However, here we give a simplified derivation of the LNA by Wallace (2010) [see also Wilkinson (2011)] which is less general than Kurtz (1971) but more intuitive, as it treats births and deaths separately. Start by re-expressing the rates explicitly as functions of the system size Ω,

$$\Omega w_j\left(\frac{X(t)}{\Omega}\right).$$

Let w_j^+ and w_j^- denote reaction rates associated with birth and death, respectively, and let $\varepsilon_j(t) \sim N(0,1)$. Then equation (4.2) can be expressed as

$$X(t+\tau) = X(t) + \tau\Omega \sum_{j=1}^{m} \left[w_j^+\left(\frac{X(t)}{\Omega}\right) - w_j^-\left(\frac{X(t)}{\Omega}\right)\right] v_j + \sqrt{\tau\Omega} \sum_{j=1}^{m} \sqrt{w_j^+\left(\frac{X(t)}{\Omega}\right) + w_j^-\left(\frac{X(t)}{\Omega}\right)}\, v_j \varepsilon_j(t)$$

or, equivalently,

$$X(t+\tau) = X(t) + \tau\Omega \sum_{j=1}^{m} A_j\left(\frac{X(t)}{\Omega}\right) v_j + \sqrt{\tau\Omega} \sum_{j=1}^{m} \sqrt{B_j\left(\frac{X(t)}{\Omega}\right)}\, v_j \varepsilon_j(t), \tag{A.1}$$

where

$$A_j\left(\frac{X(t)}{\Omega}\right) = w_j^+\left(\frac{X(t)}{\Omega}\right) - w_j^-\left(\frac{X(t)}{\Omega}\right),$$
$$B_j\left(\frac{X(t)}{\Omega}\right) = w_j^+\left(\frac{X(t)}{\Omega}\right) + w_j^-\left(\frac{X(t)}{\Omega}\right).$$

We now make the Ansatz that $X(t)$ can be decomposed into a deterministic solution with a stochastic perturbation the variance of which scales with $\sqrt{\Omega}$,

$$X(t) = \Omega\phi(t) + \sqrt{\Omega}\zeta(t), \tag{A.2}$$

where $\phi(t)$ is the macroscopic or ODE solution for the concentration and $\zeta(t)$ is a stochastic process. Inserting this into (A.1) and dividing by Ω gives

$$\phi(t+\tau) + \frac{1}{\sqrt{\Omega}}\zeta(t+\tau) = \phi(t) + \frac{1}{\sqrt{\Omega}}\zeta(t) + \tau \sum_{j=1}^{m} A_j\left(\phi + \frac{1}{\sqrt{\Omega}}\zeta\right) v_j + \sqrt{\frac{\tau}{\Omega}} \sum_{j=1}^{m} \sqrt{B_j\left(\phi + \frac{1}{\sqrt{\Omega}}\zeta\right)}\, v_j \varepsilon_j(t). \tag{A.3}$$

Applying a Taylor expansion to $A_j(\frac{X(t)}{\Omega})$ and $B_j(\frac{X(t)}{\Omega})$ about the deterministic term ϕ gives

$$A_j\left(\phi + \frac{1}{\sqrt{\Omega}}\zeta\right) \approx A_j(\phi) + \frac{1}{\sqrt{\Omega}}\zeta D_\phi(A_j(\phi)) + \cdots$$

and

$$B_j\left(\phi + \frac{1}{\sqrt{\Omega}}\zeta\right) \approx B_j(\phi) + \frac{1}{\sqrt{\Omega}}\zeta D_\phi\big(A_j(\phi)\big) + \cdots.$$

Inserting these into (A.3) and collecting terms of order Ω^0 give

$$\phi(t+\tau) = \phi(t) + \tau \sum_j A_j(\phi) v_j$$

or, expressing the sum in matrix form,

$$\phi(t+\tau) = \phi(t) + \tau A(\phi),$$

which translates into the macroscopic ODE model

$$\frac{d\phi(t)}{dt} = A(\phi). \tag{A.4}$$

Next, collecting terms of order $\Omega^{-1/2}$ gives an equation for the noise process

$$\zeta(t+\tau) = \zeta(t) + \tau \sum_j D_\phi\big(A_j(\phi)\big) v_j \zeta(t) + \sqrt{\tau} \sum_j \sqrt{B_j(\phi)} v_j \varepsilon_j(t).$$

The corresponding SDE or Langevin form is

$$\delta\zeta(t) = J(t)\zeta(t)\,dt + \sum_j \sqrt{B_j(\phi)}\,dW_j(t), \tag{A.5}$$

where $W_j(t)$ is a Wiener process, one for each population, and $J(t)$ is the Jacobian matrix of the macroscopic equations

$$J_{ij}(t) = \frac{\partial A_j(\phi)}{\partial \phi_i} = \frac{\partial[w_j^+(\phi(t)) - w_j^-(\phi(t))]}{\partial \phi_i}.$$

Equations (A.2), (A.4) and (A.5) together specify the Linear Noise Approximation (LNA) derived by Van Kampen (1997).

A.3. Kalman filter and prediction error. For ease of notation we use t instead of t_i and $t \pm 1$ instead of $t_{i\pm1}$. The state space model defined by (4.7) and (4.8) has the form

$$X_{t+1} = F_t X_t + c_t + \varepsilon_t, \qquad \varepsilon_t \sim N(\mathbf{0}, \Sigma_{\varepsilon,t}), \tag{A.6}$$

$$Y_t = \kappa_t X_t + u_t, \qquad u_t \sim N(\mathbf{0}, \Sigma_{u,t}), \tag{A.7}$$

where the structure of the matrices $F_t, c_t, \Sigma_{\varepsilon,t}$ in the state equation (A.6) is invoked by the LNA. The measurement equation (A.7) is general, allowing $\kappa = \kappa_t$ and $\Sigma_u = \Sigma_{u,t}$ to vary over time. The Kalman filter is defined by the recursions given in the following prediction and updating equations. Let $\hat{X}_t$ denote the optimal linear estimator of X_t based on the information available at time t and let P_t denote

its variance matrix. Based on information up to time $t-1$, we get the prediction equations

$$\hat{X}_{t|t-1} = F_{t-1}\hat{X}_{t-1} + c_{t-1}, \tag{A.8}$$

$$P_{t|t-1} = F_{t-1}P_{t-1}F_{t-1}^T + \Sigma_{\varepsilon,t-1}. \tag{A.9}$$

The updating equations when Y_t becomes available are

$$\hat{X}_t = \hat{X}_{t|t-1} + P_{t|t-1}\kappa_t^T R_t^{-1}(Y_t - \kappa_t \hat{X}_{t|t-1}), \tag{A.10}$$

$$P_t = P_{t|t-1} - P_{t|t-1}\kappa_t^T R_t^{-1}\kappa_t P_{t|t-1}, \tag{A.11}$$

where

$$R_t = \kappa_t P_{t|t-1}\kappa_t^T + \Sigma_{u,t} \tag{A.12}$$

is the variance matrix of the prediction error

$$e_t = Y_t - \kappa_t \hat{X}_{t|t-1}. \tag{A.13}$$

The prediction error in (A.13) and its variance matrix (A.12) are used for the likelihood (4.13).

A.4. Further details on specification of parameter distributions.

A.4.1. *Translation inhibition experiment.* The hierarchical parameters are assumed to have distributions

$$\delta_2^{(i)} \sim \Gamma\big(\mu_{\delta_2}, \sigma_{\delta_2}^2\big), \qquad \tilde{\tau}_2^{(i)} \sim \Gamma\big(\mu_{\tilde{\tau}_P}, \sigma_{\tilde{\tau}_P}^2\big), \qquad \sigma_u^{2,(i)} \sim \Gamma\big(\mu_{\sigma_u}, \sigma_{\sigma_u}^2\big),$$

where $\Gamma(\mu_{(\cdot)}, \sigma_{(\cdot)}^2)$ denotes a gamma density parameterized to have mean $\mu_{(\cdot)}$ and variance $\sigma_{(\cdot)}^2$. The hyperparameters are $\Theta_H = (\mu_{\delta_2}, \sigma_{\delta_2}^2, \mu_{\tilde{\tau}_2}, \sigma_{\tilde{\tau}_2}^2, \mu_{\sigma_u}, \sigma_{\sigma_u}^2)$. For the prior distribution of Θ_H we assume a product of vague exponential densities with parameter 10^4 for each element of Θ_H. We also reparameterize $\tilde{\phi}_2^{(i)}(0) = \kappa\phi_P^{(i)}(0)$ assuming prior distributions $\tilde{\phi}_2^{(i)}(0) \sim \mathrm{Exp}(10^4)$ and $\kappa \sim \mathrm{Exp}(10^4)$. Let $\Theta = (\Theta_H, \tilde{\phi}_2(0), \kappa)$, where $\tilde{\phi}_2(0)$ denotes the vector of initial conditions for cells $i = 1, \ldots, N$, then we wish to estimate Θ via their posterior distribution as given in (5.2). We use a standard Metropolis–Hastings algorithm [Chib and Greenberg (1995), Gamerman and Lopes (2006)] to generate a sample from the posterior distribution.

A.4.2. *Transcription inhibition experiment.* To improve convergence of the estimation algorithm for the two-dimensional model, we reparameterized $\tilde{\tau}_1^{(i)} = \kappa\alpha^{(i)}\tau_1^{(i)}$ and $\tilde{\alpha}^{(i)} = \kappa\alpha^{(i)}$, as well as the initial conditions $\tilde{\phi}_1^{(i)}(0) = \kappa\alpha^{(i)}\phi_1^{(i)}(0)$

and $\tilde{\phi}_2^{(i)}(0) = \kappa\phi_2^{(i)}(0)$. The hierarchical parameters are assumed to have the following distributions:

$$\delta_1^{(i)} \sim \Gamma\big(\mu_{\delta_1}, \sigma_{\delta_1}^2\big), \qquad \tilde{\tau}_1^{(i)} \sim \Gamma\big(\mu_{\tilde{\tau}_1}, \sigma_{\tilde{\tau}_1}^2\big),$$
$$\tilde{\alpha}^{(i)} \sim \Gamma\big(\mu_{\tilde{\alpha}}, \sigma_{\tilde{\alpha}}^2\big), \qquad \sigma_u^{2,(i)} \sim \Gamma\big(\mu_{\sigma_u}, \sigma_{\sigma_u}^2\big).$$

We have used a vague prior for $\sigma_u^{2,(i)}$ rather than importing a prior informed by the translation inhibitor experiment, as it is possible that the setting of the experiment and the use of a camera might have resulted in a very different variance of the measurement error. The vector of hyperparameters is $\Theta_H = (\mu_{\delta_1}, \sigma_{\delta_1}^2, \mu_{\tilde{\tau}_1}, \sigma_{\tilde{\tau}_1}^2, \mu_{\tilde{\alpha}}, \sigma_{\tilde{\alpha}}^2, \mu_{\sigma_u}, \sigma_{\sigma_u}^2)$ and the full parameter vector is $\Theta = (\Theta_H, \tilde{\phi}_1(0), \tilde{\phi}_2(0), \kappa)$, where $\tilde{\phi}_1(0)$ and $\tilde{\phi}_2(0)$ denote the vectors of unknown initial conditions for mRNA and protein, respectively. Similar to the previous experiment, the prior for each element of Θ was $\mathrm{Exp}(10^4)$ except we set $\mu_{\delta_2} = 0.57$ and $\sigma_{\delta_2}^2 = 0.004$, importing the posterior results on protein degradation from the translation inhibitor experiment.

Acknowledgments. The authors wish to thank Kirsty Hey (Department of Statistics, University of Warwick) and an anonymous referee for valuable suggestions. MK was funded during Ph.D. by a scholarship from University of Warwick. Hamamatsu Photonics and Carl Zeiss Limited provided technical support. The Endocrinology Group, University of Manchester, is supported by the Manchester Academic Health Sciences Centre (MAHSC) and the NIHR Manchester Biomedical Research Centre.

Author contribution. DJW and MK performed numerical estimations supervised by BF. DJW constructed the MCMC sampler for the transcription inhibition model and MK for the translation inhibition. BF proposed the use of hierarchical modeling and wrote the manuscript with input from DJW, MK and DAR. CVH devised the experimental single cell degradation approach and performed and analyzed the experiments supervised by JRED and MRHW. DAR provided help on the mathematical modeling.

REFERENCES

CHIB, S. and GREENBERG, E. (1995). Understanding the Metropolis–Hastings algorithm. *Amer. Statist*. **49** 327–336.

CRAIU, R. V. and LEMIEUX, C. (2007). Acceleration of the multiple-try Metropolis algorithm using antithetic and stratified sampling. *Stat. Comput*. **17** 109–120. MR2380640

DURHAM, G. B. and GALLANT, A. R. (2002). Numerical techniques for maximum likelihood estimation of continuous-time diffusion processes. *J. Bus. Econom. Statist*. **20** 297–316.

ELERIAN, O., CHIB, S. and SHEPHARD, N. (2001). Likelihood inference for discretely observed nonlinear diffusions. *Econometrica* **69** 959–993. MR1839375

ELF, J. and EHRENBERG, M. (2003). Fast evaluation of fluctuations in biochemical networks with the linear noise approximation. *Genome Res*. **13** 2475–2484.

ELOWITZ, M. B., LEVINE, A. J., SIGGIA, E. D. and SWAIN, P. S. (2002). Stochastic gene expression in a single cell. *Science* **297** 1183–1186.

FEARNHEAD, P., GIAGOS, V. and SHERLOCK, C. (2013). Inference for reaction networks using the linear noise approximation. Available at arXiv:1205.6920.

FINKENSTÄDT, B., HERON, E. A., KOMOROWSKI, M., EDWARDS, K., TANG, S., HARPER, C. V., DAVIS, J. R. E., WHITE, M. R. H., MILLAR, A. J. and RAND, D. A. (2008). Reconstruction of transcriptional dynamics from gene reporter data using differential equations. *Bioinformatics* **24** 2901–2907.

GAMERMAN, D. and LOPES, H. F. (2006). *Markov Chain Monte Carlo*: *Stochastic Simulation for Bayesian Inference*, 2nd ed. Chapman & Hall/CRC, Boca Raton, FL. MR2260716

GILLESPIE, D. T. (1977). Exact stochastic simulation of coupled chemical reactions. *J. Phys. Chem*. **81** 2340–2361.

GILLESPIE, D. T. (2001). Approximate accelerated stochastic simulation of chemically reacting systems. *J. Chem. Phys*. **115** 1716.

GOLIGHTLY, A. and WILKINSON, D. J. (2005). Bayesian inference for stochastic kinetic models using a diffusion approximation. *Biometrics* **61** 781–788. MR2196166

GOLIGHTLY, A. and WILKINSON, D. J. (2008). Bayesian inference for nonlinear multivariate diffusion models observed with error. *Comput. Statist. Data Anal*. **52** 1674–1693. MR2422763

GONZE, D., HALLOY, J. and GOLDBETER, A. (2002). Robustness of circadian rhythms with respect to molecular noise. *Proc. Natl. Acad. Sci. USA* **99** 673–678.

GORDON, A., COLMAN-LERNER, A., CHIN, T. E., BENJAMIN, K. R., YU, R. C. and BRENT, R. (2007). Single-cell quantification of molecules and rates using open-source microscope-based cytometry. *Nature Methods* **4** 175–182.

HARPER, C. V., FINKENSTÄDT, B., FRIEDRICHSEN, S., WOODCOCK, D., SEMPRINI, S., SPILLER, D., MULLINS, J. J., RAND, D. A., DAVIS, J. R. E. and WHITE, M. R. H. (2011). Dynamic analysis of stochastic transcription cycles. *PLoS Biology* **9** 1000607.

HERON, E. A., FINKENSTÄDT, B. and RAND, D. A. (2007). Bayesian inference for dynamic transcriptional regulation; the Hes1 system as a case study. *Bioinformatics* **23** 2589–2595.

KLOEDEN, P. E. and PLATEN, E. (1999). *Numerical Solution of Stochastic Differential Equations*, 3rd ed. *Appl. Math*. **22**. Springer, Berlin.

KOMOROWSKI, M., FINKENSTÄDT, B., HARPER, C. V. and RAND, D. A. (2009). Bayesian inference of biochemical kinetic parameters using the linear noise approximation. *BMC Bioinformatics* **10** 343.

KURTZ, T. G. (1971). Limit theorems for sequences of jump Markov processes approximating ordinary differential processes. *J. Appl. Probab*. **8** 344–356. MR0287609

KURTZ, T. G. (1981). *Approximation of Population Processes*. *CBMS-NSF Regional Conference Series in Applied Mathematics* **36**. SIAM, Philadelphia, PA. MR0610982

LIU, J. S., LIANG, F. and WONG, W. H. (2000). The multiple-try method and local optimization in Metropolis sampling. *J. Amer. Statist. Assoc*. **95** 121–134. MR1803145

MCKANE, A. J. and NEWMAN, T. J. (2005). Predator-prey cycles from resonant amplification of demographic stochasticity. *Phys. Rev. Lett*. **94** 218102.

PAULSSON, J. (2004). Summing up the noise in gene networks. *Nature* **427** 415–418.

PAULSSON, J. (2005). Models of stochastic gene expression. *Physics of Life Reviews* **2** 157–175.

SCOTT, M., INGALLS, B. and KAERN, M. (2006). Estimations of intrinsic and extrinsic noise in models of nonlinear genetic networks. *Chaos* **16** 026107.

SIMOES, M., TELO DA GAMA, M. M. and NUNES, A. (2008). Stochastic fluctuations in epidemics on networks. *Journal of the Royal Society Interface* **5** 555–566.

STATHOPOULOS, V. and GIROLAMI, M. A. (2013). Markov chain Monte Carlo inference for Markov jump processes via the linear noise approximation. *Philos. Trans. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci.* **371** 20110541, 21. MR3005674

SUTER, D. M., MOLINO, N., GATFIELD, D., SCHNEIDER, K., SCHIBLER, U. and NAEF, F. (2011). Mammalian genes are transcribed with widely different bursting kinetics. *Science* **332** 472–474.

SWAIN, P. S., ELOWITZ, M. B. and SIGGIA, E. D. (2002). Intrinsic and extrinsic contributions to stochasticity in gene expression. *Proc. Natl. Acad. Sci. USA* **99** 12795–12800.

THATTAI, M. and VAN OUDENAARDEN, A. (2001). Intrinsic noise in gene regulatory networks. *Proc. Natl. Acad. Sci. USA* **98** 8614–8619.

VAN KAMPEN, N. G. (1976). The expansion of the master equation. *Adv. Chem. Phys.* **34** 245–309.

VAN KAMPEN, N. G. (1997). *Stochastic Processes in Physics and Chemistry*. North-Holland, Amsterdam.

WALLACE, E. W. J. (2010). A simplified derivation of the linear noise approximation. Preprint. Available at arXiv:1004.4280.

WALLACE, E. W. J., GILLESPIE, D. T., SANFT, K. R. and PETZOLD, L. R. (2012). Linear noise approximation is valid over limited times for any chemical system that is sufficiently large. *Systems Biology* **IET 6** 102–115.

WILKINSON, D. J. (2011). *Stochastic Modelling for Systems Biology*, 2nd ed. Chapman & Hall/CRC, Boca Raton, FL.

B. FINKENSTÄDT
DEPARTMENT OF STATISTICS
UNIVERSITY OF WARWICK
COVENTRY, CV47AL
UNITED KINGDOM
E-MAIL: B.F.Finkenstadt@warwick.ac.uk

D. J. WOODCOCK
D. A. RAND
SYSTEMS BIOLOGY CENTRE
UNIVERSITY OF WARWICK
COVENTRY, CV47AL
UNITED KINGDOM
E-MAIL: D.J.Woodcock@warwick.ac.uk
D.A.Rand@warwick.ac.uk

M. KOMOROWSKI
POLISH ACADEMY OF SCIENCES
DIVISION OF MODELLING IN BIOLOGY
AND MEDICINE
02-106 WARSZAWA
POLAND
E-MAIL: mkomor@ippt.gov.pl

C. V. HARPER
M. R. H. WHITE
FACULTY OF LIFE SCIENCES
UNIVERSITY OF MANCHESTER
MANCHESTER, M13 9PT
UNITED KINGDOM
E-MAIL: Claire.Harper@manchester.ac.uk
Mike.White@manchester.ac.uk

J. R. E. DAVIS
FACULTY OF MEDICAL AND HUMAN SCIENCES
UNIVERSITY OF MANCHESTER
MANCHESTER, M13 9PT
UNITED KINGDOM
E-MAIL: Julian.Davis@manchester.ac.uk