Parameter inference with analytical propagators for stochastic models of autoregulated gene expression

Frits Veerman, Nikola Popović[*] and Carsten Marr[*]

14th June 2018

Abstract

Stochastic gene expression in regulatory networks is conventionally modelled via the Chemical Master Equation (CME) (van Kampen 1981). As explicit solutions to the CME, in the form of so-called propagators, are not readily available, various approximations have been proposed (Zechner et al. 2013, Feigelman et al. 2016, Popović, Marr and Swain 2016). A recently developed analytical method (Veerman, Marr and Popović 2017) is based on a scale separation that assumes significant differences in the lifetimes of mRNA and protein in the network, allowing for the efficient approximation of propagators from asymptotic expansions for the corresponding generating functions. Here, we showcase the applicability of that method to a 'telegraph' model for gene expression that is extended with an autoregulatory mechanism. We demonstrate that the resulting approximate propagators can be successfully applied for Bayesian parameter inference in the non-regulated model with synthetic data; moreover, we show that in the extended autoregulated model, autoactivation or autorepression may be refuted under certain assumptions on the model parameters. Our results indicate that the method showcased here may allow for successful parameter inference and model identification from longitudinal single cell data.

1 Introduction and background

Gene expression in regulatory networks is an inherently stochastic process (Elowitz et al. 2002). Mathematical models typically take the form of a Chemical Master Equation (CME), which describes the temporal evolution of the probabilities of observing specific states in the network (van Kampen 1981). Recent advances in single-cell fluorescence microscopy (Crane et al. 2014, Filipczyk et al. 2015, Hoppe et al. 2016, Suter et al. 2011, Zenklusen, Larson and Singer 2008) have allowed for the generation of experimental longitudinal data, whereby the fluorescence intensity of mRNA or protein abundances in single cells is measured. While most common models assume the availability of protein abundance data, the abundance of mRNA may equally be of interest, depending on the model (Janicki et al. 2004). Here, we focus exclusively on protein abundances, which we assume to be measured at regular sampling intervals Δt, for the sake of simplicity. Given the resulting data set D, parameter inference is performed on the basis of the log-likelihood

$$L(\Theta|D) = \sum_i \log\, P_{n_{i+1}|n_i}(\Delta t, \Theta), \quad \text{for } i = 0, \ldots, N, \tag{1.1}$$

that can be calculated over a range of values for the model parameters to yield a 'log-likelihood landscape', the maximum of which corresponds to the most likely parameter set Θ subject to D. Here, the propagator $P_{n_{i+1}|n_i}(\Delta t, \Theta)$ encodes the probability for the transition $n_i \to n_{i+1}$ between

[*]Joint corresponding authors.

protein numbers n_i and n_{i+1} to occur after time Δt, given Θ. Due to the complex nature of the underlying gene networks, explicit expressions for $P_{n_{i+1}|n_i}$ are difficult to obtain in general. Hence, a variety of approximations have been developed, which can be either numerical (Zechner et al. 2013, Feigelman et al. 2016) or analytical (Schnoerr, Sanguinetti and Grima 2017, Popović, Marr and Swain 2016), to name but a few. Here, we apply the analytical method recently developed by the current authors (Veerman, Marr and Popović 2017), which was based on ideas presented by Popović, Marr and Swain 2016, to obtain fully time-dependent approximate propagators; an outline of the method is given in Section 2.

Our aim in the present article is to demonstrate the applicability of these propagators, as well as to evaluate their performance in the context of Bayesian parameter inference for synthetic data. Specifically, we showcase the resulting inference procedure for a family of models for stochastic gene expression. First, in Section 3, we consider a model that incorporates DNA on/off states ('telegraph model'); see also the work of Raj et al. 2006 and Shahrezaei and Swain 2008. Subsequently, in Section 4, that model is extended with an autoregulatory mechanism, where the protein influences its own production through an autocatalytic reaction. In Section 5, we summarise our results and present an outlook to future research; finally, in Appendix A, we collate the analytical formulae that underly our inference procedure for the family of models showcased here.

2 Method

2.1 Calculation of propagators

Our method (Veerman, Marr and Popović 2017) is based on an analytical approximation of the probability generating function that is introduced for analysing the CME corresponding to the given gene expression model. Propagators can be calculated from the generating function via the Cauchy integral formula, which implies

$$P_{n_{i+1}|n_i}(\Delta t, \Theta) = \frac{1}{2\pi i} \oint_\gamma \frac{F(z; \Delta t, n_i, \Theta)}{z^{n_{i+1}+1}} \, \mathrm{d}z; \tag{2.1}$$

here, $F(z; t, n_i, \Theta)$ is the generating function of the (complex) variable z, which additionally depends on time t, the protein number n_i, and the model parameter set Θ. The integration contour γ is a closed contour in the complex plane around $z = 0$. The choice of contour is arbitrary; however, it can have a significant effect on computation times and the accuracy of the resulting integrals; see the work by Bornemann 2011. Here, we choose γ to be a regular 150-sided polygon approximating a circle of radius 0.8 that is centred at the origin of the complex plane, which results in a 'hybrid analytical-numerical' procedure for the evaluation of $P_{n_{i+1}|n_i}$.

2.2 Parameter inference

The parameter inference procedure proposed here can be divided into the following steps:

1) Data binning. The simulated data D is presented as a time series $\{n_i\}$, $0 \leq i \leq N$, which yields N transitions $n_i \to n_{i+1}$. Generically, some of these transitions occur more than once. We bin the data accordingly to create a binned data set $D_0 = \{(k_j, (n_0 \to n)_j)\}$, with $0 \leq j \leq N_0$ for $N_0 \leq N$, where k_j denotes the frequency of the transition $(n_0 \to n)_j$; see also Figure 1.

2) Marginalisation. Frequently, some of the involved species in a model are not observed, and hence have to be marginalised over. In the models discussed in Sections 3 and 4, this is the case for mRNA. Marginalisation over unobserved species is usually carried out on the transition

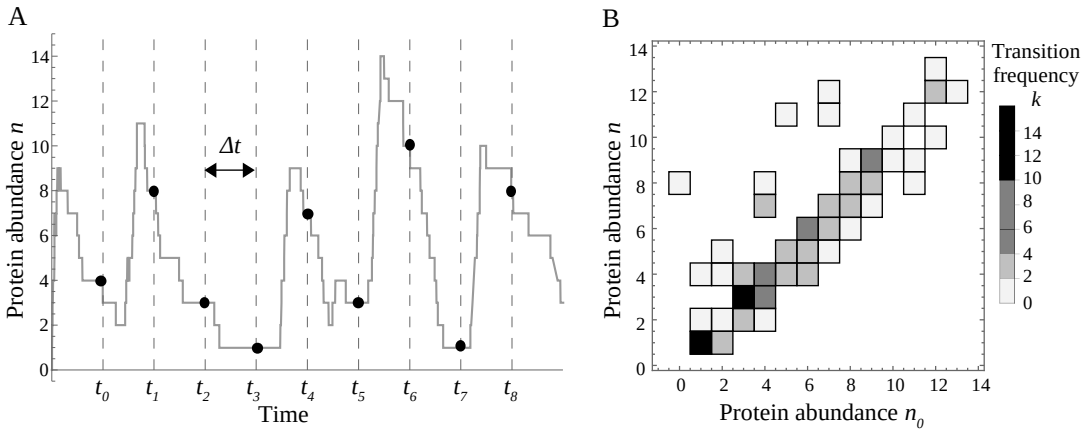

Figure 1: (A) Simulated time series of protein abundance n, with measurements at times t_i and sampling interval Δt. (B) Histogram of the frequency k_j of transitions $(n_0 \to n)_j$, from a longer time series with 100 transitions.

probabilities in (2.1). However, since the marginalisation procedure is linear, it commutes with the Cauchy integral. Introducing the linear 'marginalisation operator' $\mathbb{M}$, we may write

$$\mathbb{M}\, P_{n_{i+1}|n_i}(\Delta t, \Theta) = \frac{1}{2\pi i} \oint_\gamma \frac{\mathbb{M}\, F(z; \Delta t, n_i, \Theta)}{z^{n_{i+1}+1}}\, \mathrm{d}z, \tag{2.2}$$

where $\mathbb{M}$ now acts on the generating function F. Therefore, given the analytical approximation for F resulting from our method (Veerman, Marr and Popović 2017), we define

$$\widehat{F}(z; \Delta t, n_i, \widehat{\Theta}) = \mathbb{M}\, F(z; \Delta t, n_i, \Theta), \tag{2.3}$$

where $\widehat{\Theta} \subset \Theta$ is the subset of parameters that remain after the marginalisation procedure has been applied. Note that $\widehat{F}$ is still a fully analytical, general expression which depends on the as yet unspecified values of its arguments.

3) Evaluation. We choose a set $\widehat{\Theta}_0$ of numerical values for the parameters in $\widehat{\Theta}$. Moreover, we specify the integration contour γ, which we discretise as described in 2) to approximate the Cauchy integral in (2.1) by a finite sum. Suppose that the contour γ is discretised as $\{\zeta(l)\}$, with $0 \le l \le L$ and $\zeta(0) = \zeta(L)$; then, the integral of a function G along γ is approximated as

$$\oint_\gamma G(z)\, \mathrm{d}z \approx \sum_{l=0}^{L-1} G(\zeta(l))\, \Delta\zeta(l), \quad \text{with } \Delta\zeta(l) = \zeta(l+1) - \zeta(l). \tag{2.4}$$

Now, for every element $(k, n_0 \to n)_j$ of the binned data set D_0, we evaluate $\widehat{F}$, as given in (2.3), for the chosen parameter values $\widehat{\Theta}_0$ along the discretised contour. We hence obtain the array

$$\left\{ \frac{1}{2\pi i} \frac{\widehat{F}(\zeta(l); \Delta t, (n_0)_j, \widehat{\Theta}_0)}{\zeta(l)^{(n)_j+1}} \Delta\zeta(l) \right\} \quad \text{for } 0 \le l \le L-1 \text{ and } 0 \le j \le N_0, \tag{2.5}$$

which we sum over l to find

$$p_j(\widehat{\Theta}_0, \Delta t) = \sum_{l=0}^{L-1} \frac{1}{2\pi i} \frac{\widehat{F}(\zeta(l); \Delta t, (n_0)_j, \widehat{\Theta}_0)}{\zeta(l)^{(n)_j+1}} \Delta\zeta(l) \tag{2.6}$$

as the approximate value of the propagator for the transition $(n_0 \to n)_j$.

4) Calculation of the log-likelihood To calculate the log-likelihood of the parameter subset $\widehat{\Theta}_0$, we substitute the approximate propagators p_j, as defined in (2.6), into (1.1) to obtain

$$L(\widehat{\Theta}_0|D) = \sum_{j=0}^{N_0} k_j \log\ p_j(\widehat{\Theta}, \Delta t). \tag{2.7}$$

3 Showcase 1: The telegraph model

To demonstrate our parameter inference procedure, we consider a stochastic gene expression model that incorporates DNA on/off states ('telegraph model') (Raj et al. 2006, Shahrezaei and Swain 2008):

$$\begin{array}{rcll} D & \underset{k_1}{\overset{k_0}{\rightleftharpoons}} & D^* & \text{(DNA activation/deactivation)}, \\ D^* & \xrightarrow{\nu_0} & D^* + M & \text{(transcription of DNA to mRNA)}, \\ M & \xrightarrow{\nu_1} & M + P & \text{(translation of mRNA to protein)}, \\ M & \xrightarrow{d_0} & \emptyset & \text{(decay of mRNA)}, \\ P & \xrightarrow{d_1} & \emptyset & \text{(decay of protein)}. \end{array} \tag{3.1}$$

In recent work (Veerman, Marr and Popović 2017), we presented an analytical method for obtaining explicit, general, time-dependent expressions for the generating function associated to the CME that arises from the model in (3.1). A pivotal element of the application of that method to (3.1) is the assumption that the protein decay rate d_1 is notably smaller than the mRNA decay rate d_0, which implies that the parameter $\varepsilon := \frac{d_1}{d_0}$ is small; hence, the associated generating function is *approximated* to a certain order $O = k$, corresponding with a theoretical accuracy that is proportional to ε^k. For more details on the resulting approximation, we refer to Appendix A.

We simulate the model in (3.1) using Gillespie's algorithm (Gillespie 1977), for fixed values of the (rescaled) parameters

$$\kappa_0 := \frac{k_0}{d_1} = 1.3,\ \kappa_1 := \frac{k_1}{d_1} = 1.2,\ \lambda := \frac{\nu_0}{d_1} = 3.3,\ \mu := \frac{\nu_1}{d_0} = 2.85,\ \varepsilon := \frac{d_1}{d_0} = 0.1,\ \text{and } d_1 = 1 \tag{3.2}$$

on the time interval $0 \leq t \leq 10$, and measure the protein abundance n with a fixed sampling interval Δt. As our method assumes that Δt is of order ε, cf. again Appendix A, we set $\Delta t = \varepsilon = 0.1$, which yields $N = 100$ transitions. Based on the simulated measurement data, we perform the parameter inference procedure described in Section 2. As the data consists of protein abundances only, and as propagators for the model in (3.1) depend on both protein and mRNA abundances, we marginalise over mRNA abundance assuming a Poisson distribution with parameter $\lambda\varepsilon\frac{\kappa_0}{\kappa_0+\kappa_1} = \frac{\nu_0}{d_0}\frac{k_0}{k_0+k_1}$. We assume that the values of κ_0, κ_1, ε, and d_1 are known, and calculate the log-likelihood in (2.7) for varying λ and μ. We scan these two parameters in $\{10^{-3} \leq \lambda \leq 10^3,\ 10^{-3} \leq \mu \leq 10^2\}$, using a logarithmically spaced grid of 50×40 grid points. Figure 2 shows the resulting log-likelihood landscapes and, in particular, a comparison of the performance of the leading (zeroth) order approximation for the generating function, see Figure 2(A), with that of the first order approximation in Figure 2(B).

To quantify the performance of the method developed by Veerman, Marr and Popović 2017 for parameter inference, we compare four different scenarios:

(a) Parameter values as in (3.2), with sampling interval $\Delta t = \varepsilon = 0.1$ on the time interval $0 \leq t \leq 10$, corresponding to $N = 100$ transitions, which is the original setup that yields the results shown in Figure 2.

(b) As in (a), with the time interval increased to $0 \leq t \leq 100$, which yields $N = 1000$ transitions.

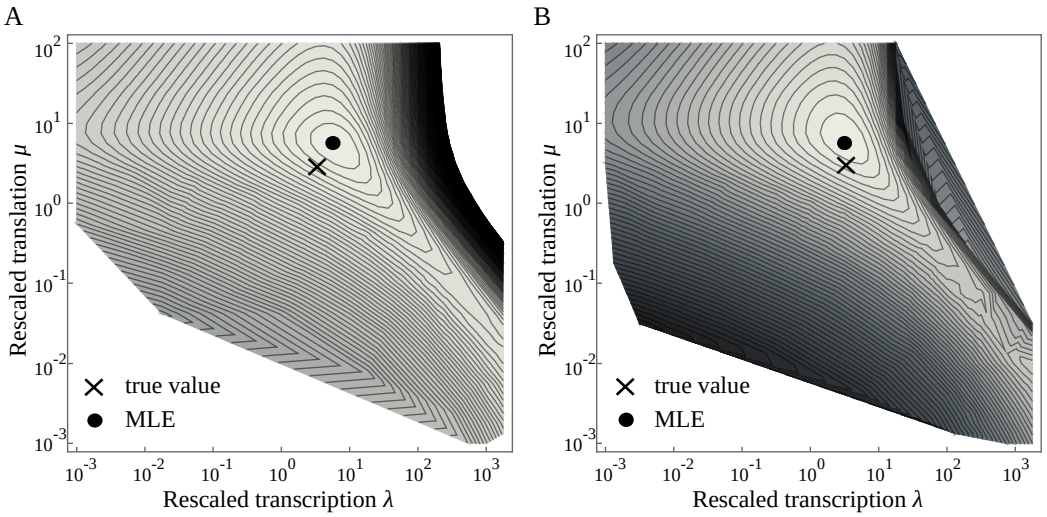

Figure 2: Log-likelihood landscapes inferred from a simulation of the telegraph model in (3.1) with $N = 100$ transitions and parameter values as in (3.2); (A): leading (zeroth) order approximation, (B): first order approximation; true value (cross) versus maximum likelihood estimate (MLE; dot).

(c) As in (a), with $\varepsilon = 0.01$; the sampling interval is decreased accordingly to $\Delta t = \varepsilon = 0.01$; measurements are taken on the time interval $0 \leq t \leq 1$, which yields $N = 100$ transitions.

(d) As in (a), with $\mu = 28.5$.

For each scenario, we infer the most likely values of the parameters λ and μ, for increasing approximation order O. The inferred values of λ and μ are compared to the 'true' values λ_{true} and μ_{true}, where we consider relative errors to quantify the performance of our inference procedure. The results of that comparison are shown in Figure 3.

4 Showcase 2: An autoregulated telegraph model

We extend the telegraph model in (3.1) with an autoregulatory mechanism, where the DNA activation rates are influenced by the presence of protein. Autoregulation is modelled in a catalytic manner, via one of the two following reactions:

$$D + P \xrightarrow{a_P} D^* + P \qquad \text{(autoactivation through protein)}, \tag{4.1a}$$

$$D^* + P \xrightarrow{r_P} D + P \qquad \text{(autorepression through protein)}. \tag{4.1b}$$

The above pair of autoregulation mechanisms was introduced by Hornos et al. 2005, and implemented e.g. by Iyer-Biswas and Jayaprakash 2014; see Section 5 for a discussion of the physical validity of these mechanisms.

To assess the performance of our parameter inference procedure, we fix the parameter values as in (3.2). We generate six data sets, as follows:

(A) Simulate the model in (3.1) without autoregulation ('null model'; $a_P = r_P = 0$) on the time interval $0 \leq t \leq 10$, which yields $N = 100$ transitions.

(B) As in (A), with the time interval increased to $0 \leq t \leq 100$, which yields $N = 1000$ transitions.

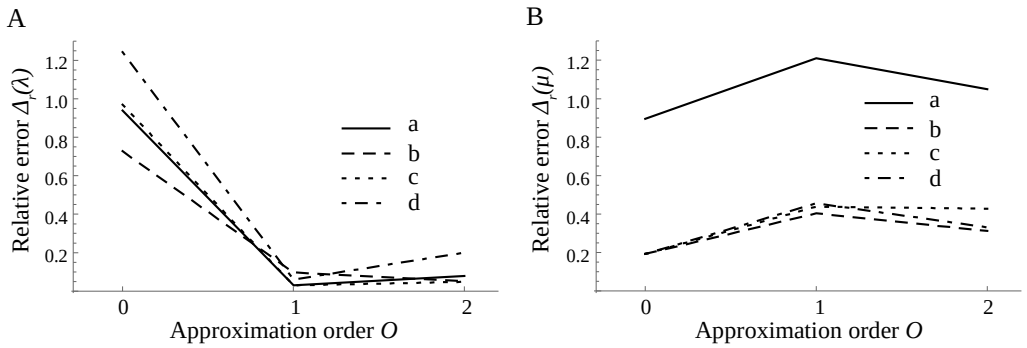

Figure 3: Relative error $\Delta_r(x) := \frac{x - x_{\text{true}}}{x_{\text{true}}}$ of the inferred parameters λ (A) and μ (B), for increasing approximation order O; for $O = k$, the propagators $P_{n_{i+1}|n_i}$ are approximated up to and including terms of order ε^k. (a) $N = 100$ transitions, $\varepsilon = 0.1$, $\lambda_{\text{true}} = 3.3$, and $\mu_{\text{true}} = 2.85$. (b) $N = 1000$ transitions, other parameters as in (a). (c) $\varepsilon = 0.01$, number of transitions N and other parameters as in (a). (d) $\mu_{\text{true}} = 28.5$, number of transitions N and other parameters as in (a). The accuracy of inference for λ clearly increases when the approximation order O is increased from 0 to 1; the increase in accuracy from $O = 1$ to $O = 2$ is obfuscated by grid size effects. A ten-fold increase in the number of transitions (b) increases the accuracy of the leading order approximation, while a ten-fold increase in the value of μ_{true} (d) decreases the accuracy of the leading order approximation. For μ, there is no noticeable increase in accuracy with the approximation order O, within the parameter grid used. However, the accuracy of inference for μ increases overall when the number of transitions is increased (b), the small parameter ε is decreased (c), or the value of μ_{true} is increased (d).

(C) Simulate the extended model {(3.1),(4.1a)} with autoactivation for $a_P\delta = 0.3$ on the time interval $0 \leq t \leq 10$, which yields $N = 100$ transitions.

(D) As in (C), with the time interval increased to $0 \leq t \leq 100$, which yields $N = 1000$ transitions.

(E) Simulate the extended model {(3.1),(4.1b)} with autorepression for $r_P\delta = 0.3$ on the time interval $0 \leq t \leq 10$, which yields $N = 100$ transitions.

(F) As in (E), with the time interval increased to $0 \leq t \leq 100$, which yields $N = 1000$ transitions.

Every data set consists of 10 runs of equal length.

Generating functions for the autoregulated extension (4.1) of the telegraph model in (3.1) have been derived in the theoretical companion article (Veerman, Marr and Popović 2017) to the current work, under the assumption that the autoregulation rate a_P or r_P is small compared to the protein decay rate d_1. That assumption implies that the ratios $\frac{a_P}{d_1} := \alpha_P\delta$ and $\frac{r_P}{d_1} := \rho_P\delta$ are small.

Parameter inference now proceeds as follows. We fix a data set, and take a single run from that data set. For that run, we determine the likelihood of the autoactivated model in {(3.1),(4.1a)}, varying $0 \leq \alpha_P\delta \leq 1$; likewise, we determine the likelihood of the autorepressed model in {(3.1),(4.1b)}, varying $0 \leq \rho_P\delta \leq 1$. The likelihood of the non-regulated model in (3.1) is then used to determine the model score according to the Bayesian information criterion (BIC) (Schwarz 1978), where

$$\Delta\text{BIC} = 2[\log(L) - \log(L_0)]. \tag{4.2}$$

Here, L is the likelihood of an autoregulated extension, with autoregulation as in (4.1), of the model in (3.1), while L_0 denotes the likelihood of the non-regulated model in (3.1). The information difference ΔBIC – which is also known as the log Bayes factor – quantifies the evidence for the model in question. Typically, a ΔBIC-value above 3 is considered strong evidence (Kass and Raftery 1995). We repeat the above procedure for all 10 runs in the data set, and determine the mean and standard deviation; the outcome is illustrated in Figure 4.

5 Discussion

In the present article, we showcase a parameter inference procedure that is based on a recently developed analytical method (Veerman, Marr and Popović 2017) which allows for the efficient numerical approximation of propagators via the Cauchy integral formula on the basis of asymptotic series for the underlying generating functions. The resulting hybrid analytical-numerical approach reduces the need for computationally expensive simulations; moreover, due to its perturbative nature, it is highly applicable over relatively short time scales, such as occur naturally in the calculation of the log-likelihood in (1.1).

We present results for synthetic data in a family of models for stochastic gene expression from the literature under the assumption that lifetimes of protein are significantly longer than those of mRNA, which introduces a small parameter ε and, hence, a separation of scales. For an extensive discussion of the validity of our assumption that ε is small, we refer to (Veerman, Marr and Popović 2017).

In Section 3, we discuss a simple ('telegraph') gene expression model without autoregulation, showing that our approach can successfully infer relevant model parameters. Unlike in previous work by Feigelman, Popović and Marr 2015, the underlying implementation avoids potential bias due to zero propagator values and large initial protein numbers through the use of 'implicit' series expansions in ε; see Appendix A for an in-depth argument.

In Section 4, we perform a model comparison in an autoregulated extension of the standard telegraph model. We consider three types of gene regulation: autoactivation, autorepression, and no regulation of DNA activity (null model). For each type, we simulate data with 100 and 1000 protein transitions, respectively. Throughout, we find that 100 data points are insufficient to reject model hypotheses with our approach. With 1000 data points, however, we can successfully reject the non-regulated and the autorepressed model for simulated data from an autoactivated model, in which case we can even infer the correct order of the autoactivation parameter. For simulated autorepression, we can reject the model with autoactivation, but not the non-regulated model. Our approach fails to identify the correct model for data from a non-regulated model for 1000 transitions, where the autoactivated model is clearly, but wrongly favoured. We believe that more research is needed into the sources of these discrepancies in dependence on model parameters and the order of our approximation.

In both showcases, we observe a trade-off between the accuracy of inference versus the required computation time. Computation times seem to increase exponentially with the approximation order, at least for the setup realised in this article. For practical purposes, we hence propose an algorithm whereby the fastest, leading order approximation is used to obtain a first estimate for the underlying model parameters; that estimate can then be improved by including higher order corrections, resulting in a much more computationally efficient procedure.

It is insightful to compare our results with other recent work on parameter inference in regulated gene expression models. In work by Feigelman et al. 2016, three models for regulated gene expression with a slightly different structure compared to the models studied in the present article were simulated and inferred via a stochastic particle filtering-based inference procedure that employs genealogical information of dividing cells. Interestingly, positive and negative autoregulation could be successfully rejected there for data that was simulated from a no-feedback model. However, the no-feedback model could not be rejected for data originating from the

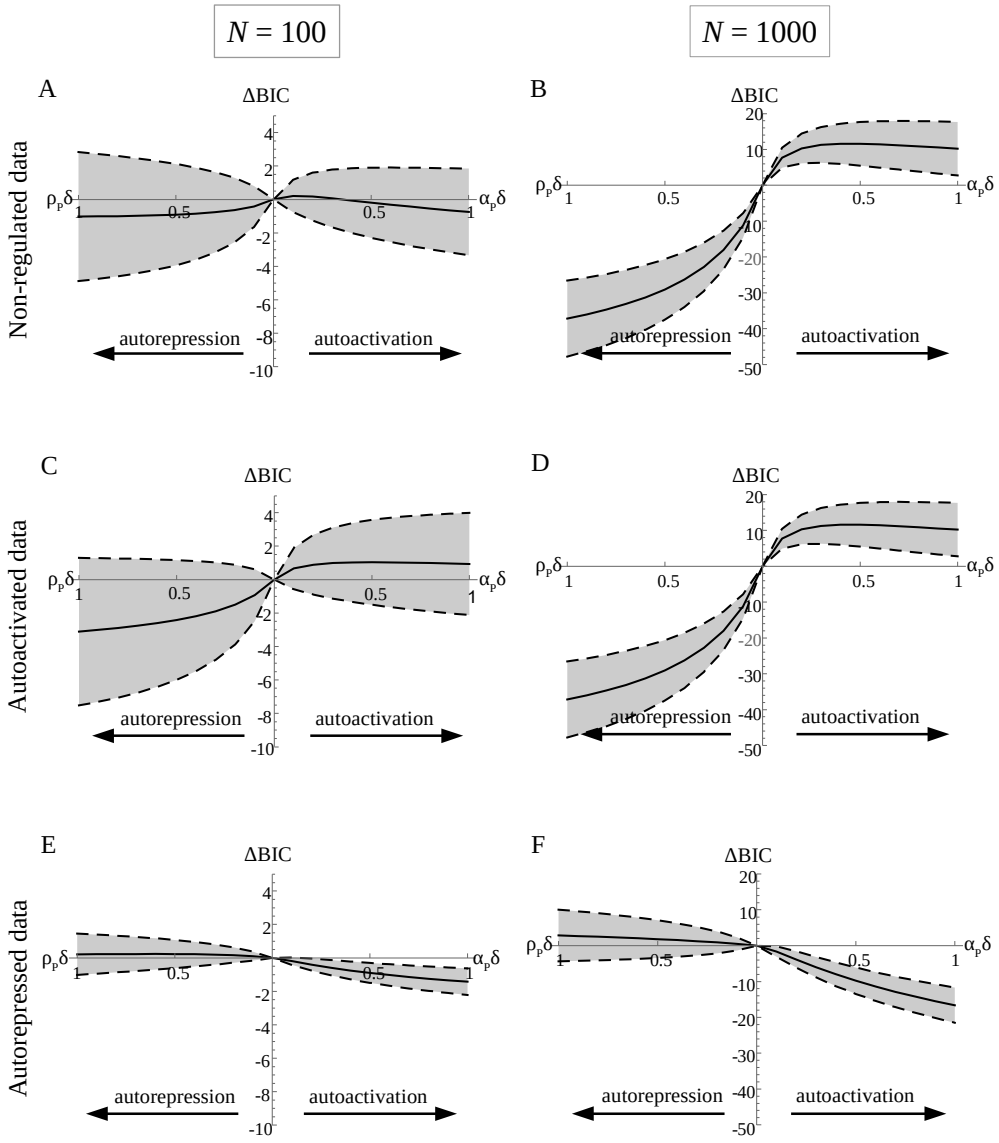

Figure 4: Results on parameter inference for the extended autoregulated model {(3.1),(4.1)} on the basis of various types of synthetic data, where the performance of the model is quantified via the Bayesian information criterion (BIC). On the vertical axis, ΔBIC, as defined in (4.2), indicates the difference between the BIC score of the autoactivated or autorepressed model and the score of the non-regulated model in (3.1); the higher the ΔBIC score, the more likely the associated model is. In each panel, the solid curve indicates the mean values based on 10 model runs; dashed curves indicate the uncertainty (one standard deviation). On the horizontal axis, the strength of autoregulation is measured by $\alpha_P\delta$ (increasing to the right) or $\rho_P\delta$ (increasing to the left). (A) Data generated from the null (non-regulated) model in (3.1), with $N = 100$ transitions; (C) data generated with the model in (3.1) with autoactivation as in (4.1a), for $\alpha_P\delta = 0.3$ and $N = 100$ transitions; (E) data generated with the model in (3.1) with autorepression as in (4.1b), for $\rho_P\delta = 0.3$ and $N = 100$ transitions. (B,D,F) as (A,C,E), but with $N = 1000$ transitions; note that the vertical axis has a different scaling. All other model parameters were assumed to be known. We see that 1000 transitions suffice to correctly refute autorepression in (B,D), and to correctly refute autoactivation in (F). In the case of 100 transitions, no conclusion can be drawn from (A) and (C); (E) correctly refutes autoactivation; however, the low ΔBIC score, with $|\Delta\text{BIC}| \leq 2$, indicates low significance.

corresponding models with positive or negative feedback. From this comparison with Feigelman et al. 2016, we conclude that the structure of the data, the intensity of regulatory feedback, and the chosen inference procedure together will influence the extent of insight which can be obtained from a inference approach that is based on stochastic models of gene expression.

We emphasise that the application of the analytical method showcased here is not restricted to specific models; the goal of the present article is to demonstrate the applicability of that method, and to investigate its performance, rather than to assess the biological validity of a given model. It is important to note that our approach can equally be extended to recent, physiologically more relevant modifications of the telegraph model with autoregulation (Hornos et al. 2005) by Grima, Schmidt and Newman 2012 and Congxin et al. 2018; another feasible alternative model can be obtained by introduction of a refractory state (Zoller et al. 2015).

The input for our propagator-based approach is the abundance of the involved species, viz. of protein. Such abundances are challenging to obtain experimentally due to the unknown relation with the fluorescence that is observed under a microscope. A linear relation is regularly assumed (Suter et al. 2011, Zechner et al. 2013); an improvement over that assumption may be achievable through recent work by Bakker and Swain 2018.

Finally, the showcases presented in this article are based on synthetic data that was generated *in silico*; in future work, we plan to consider experimental data, such as can be found in work by Suter et al. 2011.

Acknowledgements

The authors thank Ramon Grima and Peter Swain (both University of Edinburgh) for valuable comments and suggestions. This work has been supported by the Leverhulme Trust, through Research Project Grant RPG-2015-017 ('A geometric analysis of multiple-scale models for stochastic gene expression').

References

Bakker, E. and P.S. Swain (2018). "Estimating numbers of fluorescent molecules in single cells by analysing fluctuations in photobleaching". In: *bioRxiv*. DOI: 10.1101/272310.

Bornemann, F. (2011). "Accuracy and Stability of Computing High-order Derivatives of Analytic Functions by Cauchy Integrals". In: *Foundations of Computational Mathematics* 11.1, pp. 1–63. DOI: 10.1007/s10208-010-9075-z.

Congxin, L. et al. (2018). "Frequency modulation of transcriptional bursting enables sensitive and rapid gene regulation". In: *Cell Systems* 6.4, pp. 409–423. DOI: 10.1016/j.cels.2018.01.012.

Crane, M.M. et al. (2014). "A microfluidic system for studying ageing and dynamic single-cell responses in budding yeast". In: *PLoS One* 9.6, e100042. DOI: 10.1371/journal.pone.0100042.

Elowitz, M.B. et al. (2002). "Stochastic Gene Expression in a Single Cell". In: *Science* 297.5584, pp. 1183–1186. DOI: 10.1126/science.1070919.

Feigelman, J., N. Popović and C. Marr (2015). "A case study on the use of scale separation-based analytical propagators for parameter inference in models of stochastic gene regulation". In: *Journal of Coupled Systems and Multiscale Dynamics* 3.2, pp. 164–173. DOI: 10.1166/jcsmd.2015.1074.

Feigelman, J. et al. (2016). "Analysis of cell lineage trees by exact bayesian inference identifies negative autoregulation of nanog in mouse embryonic stem cells". In: *Cell Systems* 3.5, pp. 480–490. DOI: 10.1016/j.cels.2016.11.001.

Filipczyk, A. et al. (2015). "Network plasticity of pluripotency transcription factors in embryonic stem cells". In: *Nature Cell Biology* 17, pp. 1235–1246. DOI: 10.1038/ncb3237.

Gillespie, D.T. (1977). "Exact stochastic simulation of coupled chemical reactions". In: *The Journal of Physical Chemistry* 81.25, pp. 2340–2361. DOI: 10.1021/j100540a008.

Grima, R., D.R. Schmidt and T.J. Newman (2012). "Steady-state fluctuations of a genetic feedback loop: An exact solution". In: *Journal of Chemical Physics* 190, p. 035104. DOI: 10.1063/1.4736721.

Hoppe, P.S. et al. (2016). "Early myeloid lineage choice is not initiated by random PU.1 to GATA1 protein ratios". In: *Nature* 535, pp. 299–302. DOI: 10.1038/nature18320.

Hornos, J.E.M. et al. (2005). "Self-regulating gene: An exact solution". In: *Physical Review E: statistical, nonlinear, biological, and soft matter physics* 72, p. 051907. DOI: 10.1103/PhysRevE.90.051907.

Iyer-Biswas, S. and C. Jayaprakash (2014). "Mixed Poisson distributions in exact solutions of stochastic autoregulation models". In: *Physical Review E: statistical, nonlinear, biological, and soft matter physics* 90, p. 052712. DOI: 10.1103/PhysRevE.90.052712.

Janicki, S.M. et al. (2004). "From silencing to gene expression: real-time analysis in single cells". In: *Cell* 116.5, pp. 683–698. DOI: 10.1016/S0092-8674(04)00171-0.

Kass, R.E. and A.F. Raftery (1995). "Bayes factors". In: *Journal of the Americal Statistical Association* 90.430, pp. 773–795. DOI: 10.1080/01621459.1995.10476572.

Popović, N., C. Marr and P.S. Swain (2016). "A geometric analysis of fast-slow models for stochastic gene expression". In: *Journal of Mathematical Biology* 72.1, pp. 87–122. DOI: 10.1007/s00285-015-0876-1.

Raj, A. et al. (2006). "Stochastic mRNA synthesis in mammalian cells". In: *PLoS Biology* 4.10, pp. 1707–1719. DOI: 10.1371/journal.pbio.0040309.

Schnoerr, D., G. Sanguinetti and R. Grima (2017). "Approximation and inference methods for stochastic biochemical kinetics—a tutorial review". In: *J. Phys. A* 50.9, pp. 093001, 60. ISSN: 1751-8113. DOI: 10.1088/1751-8121/aa54d9.

Schwarz, G. (1978). "Estimating the dimension of a model". In: *Ann. Statist.* 6.2, pp. 461–464. ISSN: 0090-5364. DOI: 10.1214/aos/1176344136.

Shahrezaei, V. and P.S. Swain (2008). "Analytical distributions for stochastic gene expression". In: *Proceedings of the National Academy of Sciences of the United States of America* 105.45, pp. 17256–17261. DOI: 10.1073/pnas.0803850105.

Suter, D.M. et al. (2011). "Mammalian genes are transcribed with widely different bursting kinetics". In: *Science* 332.6028, pp. 472–474. DOI: 10.1126/science.1198817.

van Kampen, N. G. (1981). *Stochastic processes in physics and chemistry*. Vol. 888. Lecture Notes in Mathematics. North-Holland Publishing Co., Amsterdam-New York, pp. xiv+419. ISBN: 0-444-86200-5. DOI: 10.1002/bbpc.19830870424.

Veerman, F., C. Marr and N. Popović (2017). "Time-dependent propagators for stochastic models of gene expression: an analytical method". In: *Journal of Mathematical Biology*. DOI: 10.1007/s00285-017-1196-4.

Zechner, C. et al. (2013). "Scalable inference of heterogeneous reaction kinetics from pooled single-cell recordings". In: *Nature Methods* 11, pp. 197–202. DOI: doi:10.1038/nmeth.2794.

Zenklusen, D., D.R. Larson and R.H. Singer (2008). "Single-RNA counting reveals alternative modes of gene expression in yeast". In: *Nature Structural & Molecular Biology* 15, pp. 1263–1271. DOI: 10.1038/nsmb.1514.

Zoller, B. et al. (2015). "Structure of silent transcription intervals and noise characteristics of mammalian genes". In: *Molecular Systems Biology* 11.7, p. 823. DOI: 10.15252/msb.20156257.

A Analytical details

A.1 Approximate generating functions

The generating functions used to approximate propagators in the present article, cf. Section 2, are derived via the analytical method presented by Veerman, Marr and Popović 2017. For the telegraph model in Section 3, the leading order generating function $\widehat{F}_0$ is given by

$$\widehat{F}_0(z;n_0,\Delta t,\varepsilon,\lambda,\mu,\chi) = \left[1-(1-z)\mathrm{e}^{-\Delta t}\right]^{n_0} \times \exp\left(-\varepsilon(1-\chi)\lambda\mu(1-z)\mathrm{e}^{-\Delta t}\frac{1-\mathrm{e}^{-[1+\mu(1-z)\mathrm{e}^{-\Delta t}]\frac{\Delta t}{\varepsilon}}}{1+\mu(1-z)\mathrm{e}^{-\Delta t}}\right); \quad \text{(A.1)}$$

here, $\chi = \frac{\kappa_1}{\kappa_0+\kappa_1}$. All parameters have been rescaled according to (3.2). The generating function has been marginalised over mRNA abundance, using a Poisson distribution with parameter $\lambda\varepsilon\frac{\kappa_0}{\kappa_0+\kappa_1}$. Analogously, the first order approximation $\widehat{F}_1$ of the generating function is given by

$$\begin{aligned}\widehat{F}_1(z;n_0,\Delta t,\varepsilon,\lambda,\mu,\chi) = {}& \left[1-(1-z)\mathrm{e}^{-\Delta t}\right]^{n_0} \\ & \times\left[1+\frac{\varepsilon(1-\chi)\lambda\mu(1-z)\mathrm{e}^{-\Delta t}}{1+\mu(1-z)\mathrm{e}^{-\Delta t}}\left(\frac{1-\mathrm{e}^{-[1+\mu(1-z)\mathrm{e}^{-\Delta t}]\frac{\Delta t}{\varepsilon}}}{1+\mu(1-z)\mathrm{e}^{-\Delta t}}-\frac{\Delta t}{\varepsilon}\right)\right] \\ & \times\exp\left\{-\varepsilon(1-\chi)\lambda\mu(1-z)\mathrm{e}^{-\Delta t}\frac{1-\mathrm{e}^{-[1+\mu(1-z)\mathrm{e}^{-\Delta t}]\frac{\Delta t}{\varepsilon}}}{1+\mu(1-z)\mathrm{e}^{-\Delta t}}\left[1+\frac{\varepsilon}{\mathrm{e}^{[1+\mu(1-z)\mathrm{e}^{-\Delta t}]\frac{\Delta t}{\varepsilon}}-1}\right.\right. \\ & \left.\left.\times\left(\frac{\mathrm{e}^{[1+\mu(1-z)\mathrm{e}^{-\Delta t}]\frac{\Delta t}{\varepsilon}}-1}{[1+\mu(1-z)\mathrm{e}^{-\Delta t}]^2}-\frac{\frac{\Delta t}{\varepsilon}}{1+\mu(1-z)\mathrm{e}^{-\Delta t}}+\mu(1-z)\mathrm{e}^{-\Delta t}\left(\frac{1}{2}\frac{\Delta t}{\varepsilon}\right)^2\right)\right]\right\}. \quad \text{(A.2)}\end{aligned}$$

For the autoregulated model discussed in Section 4, the same expressions for the generating functions are used; however, χ now depends on the autoregulatory mechanism according to

$$\chi = \begin{cases} \dfrac{\kappa_1}{\kappa_0+\kappa_1} & \text{(no autoregulation)}, \\ \dfrac{\kappa_1}{\kappa_0+\alpha_P\delta\, n_0+\kappa_1} & \text{(autoactivation)}, \\ \dfrac{\kappa_1+\rho_P\delta\, n_0}{\kappa_0+\kappa_1+\rho_P\delta\, n_0} & \text{(autorepression)}. \end{cases} \quad \text{(A.3)}$$

A.2 'Implicit' expansions

It is important to note that neither the leading order generating function in (A.1) nor the first order approximation given by (A.2) are expressed as asymptotic series in powers of ε, as would be expected on the basis of the perturbative approach taken by Veerman, Marr and Popović 2017. The underlying reasoning can be summarised as follows.

First, in the derivation of the generating functions, it was assumed that the sampling time Δt was small, i.e. of order ε; note that this assumption is satisfied in all numerical simulations shown in the current article, where $\Delta t = \varepsilon$ throughout. Thus, we can write

$$\Delta t = \varepsilon\Delta s. \quad \text{(A.4)}$$

With the above scaling for Δt, an expansion of $\widehat{F}_0$ and $\widehat{F}_1$, as defined in (A.1) and (A.2), respectively, into asymptotic series in ε to the appropriate order yields

$$\widehat{F}_0 = z^{n_0}, \quad \text{(A.5)}$$

$$\widehat{F}_1 = z^{n_0}\left\{1+\varepsilon(1-z)\left[\frac{n_0\Delta s}{z}-\frac{(1-\chi)\lambda\mu}{1+\mu(1-z)}\left(\frac{\mu(1-z)\left\{1-\mathrm{e}^{[-(1+\mu(1-z)]\Delta s}\right\}}{1+\mu(1-z)}+\Delta s\right)\right]\right\}. \quad \text{(A.6)}$$

From (A.5), one can immediately conclude that

$$\oint_\gamma \widehat{F_0} = \delta_{n,n_0}. \tag{A.7}$$

From the series for $\widehat{F_1}$ in (A.6), we see that we can write

$$F_0(z) = z^{n_0} \sum_{k=-1}^{\infty} f_{1,k} z^k; \tag{A.8}$$

hence, it follows that

$$\oint_\gamma \widehat{F_1} = 0 \quad \text{if } n_0 > n + 1. \tag{A.9}$$

More generally, an expansion of the generating function to order k in ε will yield

$$\oint_\gamma \widehat{F_k} = 0 \quad \text{if } n_0 > n + k. \tag{A.10}$$

From these observations, we conclude that decreasing transitions ($n_i \to n_{i+1}$), where $n_i > n_{i+1} + k$, will be assigned a probability that is identically zero. Hence, if such transitions *do* occur in the data, the model is ruled out immediately, as our perturbative approach excludes the possibility that such transitions can occur. One can understand this phenomenon by considering the definition of the small parameter ε, which is defined as the ratio of the protein decay rate d_1 over the mRNA decay rate d_0. A leading order approximation of $\varepsilon = 0$ is thus equivalent to taking the protein decay rate $d_1 \to 0$ which, in turn, implies that protein does not decay at all, since (natural) protein decay is the only reaction in (3.1) that can decrease protein abundance. By the same reasoning, a straightforward expansion of the generating function to order ε^k will restrict the model to transitions ($n_i \to n_{i+1}$), where $n_{i+1} - n_i \geq -k$. It would follow that either the order O of the method would be limited from below by the data, leading to high-order expansions in ε and, hence, to long computation times, or that the method could only be applied to a subset of the data, which would introduce a bias.

Lastly, an asymptotic expansion such as (A.6) implicitly assumes that all parameters and variables in the model are of order 1 in ε. For the series expansion of $\widehat{F_1}$ in (A.6), that assumption would significantly restrict the range of λ; in comparison, in Figure 2, likelihood values for λ up to order ε^{-3} are calculated. More importantly, the above assumption would restrict the range of n_0, implying that only a subset of the data – with sufficiently low protein numbers – could be used as input for parameter inference.

We emphasise that none of these difficulties occur with the expressions in (A.1) and (A.2), where the expansion order in ε is expressed 'implicitly' in the respective functional forms of $\widehat{F_0}$ and $\widehat{F_1}$.