Multivariate moment closure techniques for stochastic kinetic models

Eszter Lakatos,[a] Angelique Ale, Paul D. W. Kirk, and Michael P. H. Stumpf[b]
Department of Life Sciences, Centre for Integrative Systems Biology and Bioinformatics, Imperial College London, London SW7 2AZ, United Kingdom

(Received 27 April 2015; accepted 2 August 2015; published online 4 September 2015)

Stochastic effects dominate many chemical and biochemical processes. Their analysis, however, can be computationally prohibitively expensive and a range of approximation schemes have been proposed to lighten the computational burden. These, notably the increasingly popular linear noise approximation and the more general moment expansion methods, perform well for many dynamical regimes, especially linear systems. At higher levels of nonlinearity, it comes to an interplay between the nonlinearities and the stochastic dynamics, which is much harder to capture correctly by such approximations to the true stochastic processes. Moment-closure approaches promise to address this problem by capturing higher-order terms of the temporally evolving probability distribution. Here, we develop a set of multivariate moment-closures that allows us to describe the stochastic dynamics of nonlinear systems. Multivariate closure captures the way that correlations between different molecular species, induced by the reaction dynamics, interact with stochastic effects. We use multivariate Gaussian, gamma, and lognormal closure and illustrate their use in the context of two models that have proved challenging to the previous attempts at approximating stochastic dynamics: oscillations in p53 and Hes1. In addition, we consider a larger system, Erk-mediated mitogen-activated protein kinases signalling, where conventional stochastic simulation approaches incur unacceptably high computational costs. © *2015 AIP Publishing LLC.* [http://dx.doi.org/10.1063/1.4929837]

I. INTRODUCTION

At the macroscopic scale, chemical reaction systems are typically described using mass-action kinetics. These capture how the concentrations of chemical species evolve over time. But only at a scale where all chemical species involved in a reaction system are available in sufficiently high abundances is it possible to use that the behaviour is governed by the speed of reactions, expressed by reaction rates. However, in biochemical systems, the amount of most reactants is often in the regime of a few dozen or hundred molecules, which are distributed over relatively large volumes and hence it is not guaranteed that reactant molecules of a specific reaction can be found in sufficient proximity for the reaction to occur.[1] Furthermore, thermal noise also increases inherent stochasticity of production and degradation reactions and results in further variation in the copy numbers of reactants. All of these factors profoundly affect and shape cell-to-cell variability and heterogeneity in isogenic populations.

For these reasons, we often require the explicitly stochastic formulation of chemical reaction systems, and cannot rely on purely deterministic approaches to model (bio-)molecular reaction systems: often stochasticity is an inalienable aspect of such systems. The most general description of the change in molecular abundances under stochastic dynamics is typically provided by the Chemical Master Equation (CME). But although the CME offers an exact description, it can only be solved for a few, highly idealised cases; therefore, other approaches are needed to simulate and predict the system's behaviour.

One such method is Gillespie's Stochastic Simulation Algorithm (SSA),[2] which offers a direct numerical simulation of the system to obtain a single exact realisation of the stochastic process modelled. But because of its computational cost, it is impossible to model the behaviour of a large population of cells, or of even moderately complex reaction systems. Several variations of the SSA are available that offer computational advantages which make analyses of such systems feasible; all of these systems aim to capture the essential aspects of the stochastic dynamics but in suitable and computationally convenient approximation. Many of these methods divide the system dynamics into a deterministic part and a stochastic part. Hybrid methods,[3,4] for example, divide the species into low and high abundance groups, which are modelled by SSA and ODEs, respectively. Stochastic differential equations (SDE) treat changes in mean concentration deterministically and add a (typically Wiener) noise term to model fluctuations corresponding to each reaction. The linear noise approximation (LNA)[5,6] similarly separates the deterministic trend (drift) and a noise term, providing a quasi-second order approximation of the dynamics of the probability density function. This method offers the exact solution for reaction systems containing only zeroth- and first-order reactions.

However, all these approximation methods rely on the assumption that the abundance of each species is relatively high, in the regime of 100–1000 or more molecules.[7] Low molecular numbers can cause hybrid and SDE methods to take negative values while the LNA fails to correctly follow

[a]Electronic mail: e.lakatos13@imperial.ac.uk
[b]Electronic mail: m.stumpf@imperial.ac.uk

dynamics of highly nonlinear — especially oscillatory or multi-stable — systems. For smaller but complex reaction systems, which are still practically intractable with the SSA, expansion methods have been developed that capture the dynamics throughout the time evolution of moments of the probability distribution. Many approaches include an expansion for specific rate laws or moment orders, for example, the so-called Mass Fluctuation Kinetics (MFK)[8] and the 2MA and 3MA[9] approaches include expressions for the first, second, and third central moments.

Ale *et al.*[10] developed a moment expansion approximation (MEA) method to automatically generate equations of arbitrary order central moments in a computationally efficient manner. This method is suitable to capture the general behaviour of stochastic systems with complex dynamics and 10-100 molecules, which the LNA or conventional MFK approaches typically fail to reproduce adequately. For systems involving only zeroth and first order reactions, the method reproduces the moment equations one would obtain directly from the CME.

Moment closure approximation techniques generally combine an expansion method with a moment closure *Ansatz* to obtain a closed set of differential equations solvable either analytically (this is only possible very rarely) or numerically. The closure of the equations is usually based on the assumption that the variables approximately follow a known distribution for which we already know some relation(s) between the moments, and hence can derive an expression of the higher order moments in terms of lower ones. Based on the properties of normal distribution, Goodman[11] and Whittle[12] pioneered by setting higher order cumulants to zero to close second order equations, which inspired, for example, the work of Matis and Guardiola.[13] Gillespie and colleagues[14,15] proposed a method to obtain moment equations up to any order, listed expressions as the basis of univariate normal, lognormal, and Poisson closure, and demonstrated the algorithm by closing equations of second moments assuming a normal distribution. Krishnarajah *et al.*[16] directly addressed the problem, which arises in the case of highly skewed population distributions when most closure schemes fail as the normal assumption does not hold. They developed second- and third-order beta-binomial, lognormal, and mixture closure techniques for two stochastic epidemic models following a single variable. Later they also described multivariate methods for an extended model.[17] In parallel, several approaches have been developed which do not proceed from a particular distribution. The most straightforward approach is the truncation of the system applied by Lee *et al.*[18] and similar to what we will refer to as zero closure in the manuscript. Keeling[19] derived moment closure in a one-dimensional logistic system for third order moments by formulating the system in terms of multiplicative instead of additive moments. Singh and Hespanha[20–22] derived a systematic construction of moment closure functions based on derivative matching, first in one and then in arbitrary dimensions. In both cases, the derived moment closure functions are consistent with the species being (jointly) lognormally distributed. Hausken and Moxnes[23] set higher mixed moments to zero and assumed a Dirac distribution, while Smadbeck and Kaznessis[24] applied an information theoretic point of view to close the equations based on the assumption that a finite set of moments holds all information about the system.

Although it has been shown that expansions up to three moments tend to provide a more precise approximation than only second order ones,[25] general closure formulas for closing third or higher order equation systems have not been widely used. Furthermore, many of the existing algorithms have been adapted manually to specific systems and appear not to be concerned with (or capable to provide) general applicability without manual intervention. Crucially, there are no automatic approaches[14,26] capable of generating and closing moment equations of arbitrary order that could be applied to different systems, especially those with nonlinear (e.g., rational) rate laws, and different distributions (for closure) without substantial modifications to the method itself.

In this paper, we introduce a new multivariate moment closure approximation technique — consisted of a moment expansion algorithm and generation of multivariate closure expressions to close moment equations — that can be applied to any stochastic kinetic system with no constraints on the dimension, order or (lack of) linearity of the system, and its interactions.[62] Our approximation is based on the work of Ale *et al.*,[10] which is a general framework applying two successive Taylor expansions to express the moment generating function of the system. We extend this method by developing closure schemes following either a multivariate normal, lognormal, or a gamma distribution *Ansatz*. We use properties of these three distributions to derive moment closure formulas to express arbitrary order higher moments in terms of lower order ones, replacing truncations used in the original expansion. Our algorithm takes into account the different levels of the potentially intricate interplay between species captured by higher order mixed moments. Since the definition of a multivariate gamma (MVG) distribution is ambiguous and the existing formulations are not suitable to model arbitrary systems, we also describe a new multivariate gamma distribution generalising existing definitions.

The derived formulas are incorporated into the MEA framework to obtain improved and less costly approximations of nonlinear dynamics. We demonstrate the power of our method on two small oscillatory reaction systems, the p53-Mdm2 negative feedback loop in tumour suppression and the Hes1 system in embryogenesis and on a complex two-compartment model of ERK1/2 phosphorylation dynamics. As the framework, like the original MEA, is completely automated and developed in a user-friendly manner, it is suitable for the analysis of any stochastic models without thorough understanding of the algorithm and also can serve as a starting point for parameter inference[27–30] and model selection,[31,32] distribution reconstruction,[33] sensitivity analysis,[34] experimental design,[35,36] and design of dynamical systems with desired qualitative[37] and quantitative[38,39] behaviour.

II. METHODS

A. Moment expansion

In this section, we provide an overview of the MEA technique and how the computationally complex problem of

following the evolution of the whole stochastic system is translated into solving a set of differential equations tracking the evolution of the moments of the system. For a more detailed derivation, see Ref. 10.

Let us consider a system of N chemical species $(X_1,\dots,X_N)$ and r reactions, where the lth reaction corresponds to the process

$$\underline{s}_{l,1}X_1 + \cdots + \underline{s}_{l,N}X_N \xrightarrow{k_l} \overline{s}_{l,1}X_1 + \cdots + \overline{s}_{l,N}X_N, \quad (1)$$

which occurs at rate k_l and where the number of X_i molecules changes from $\underline{s}_i$ to $\overline{s}_i$. At any given time, the system is described by the probability mass function $P(\mathbf{x};t)$, which specifies the probability of observing the abundances $x_1,\dots,x_N$ of each species. The time evolution of the system is then given by the CME,

$$\frac{dP(\mathbf{x};t)}{dt} = \sum_{l=1}^{r} P(\mathbf{x}-\mathbf{s}_l;t)a_l(\mathbf{x}-\mathbf{s}_l) - P(\mathbf{x};t)a_l(\mathbf{x}), \quad (2)$$

where $\mathbf{s}_l = \overline{\mathbf{s}}_l - \underline{\mathbf{s}}_l$ is the vector of stoichiometric coefficients of the species in reaction l and a_l is the reaction propensity of the aforementioned reaction. The moment expansion method[10] follows $P(\mathbf{x})$ over time by capturing the time evolution of its moments instead, based on the time dependent moment generating function $(m(\boldsymbol{\theta},\mathbf{x}))$ derived from the CME,

$$\frac{dm}{dt} = \sum_{l}\left[(e^{\boldsymbol{\theta}\mathbf{s}_l}-1)\sum_{\mathbf{x}} e^{\boldsymbol{\theta}\mathbf{x}}P(\mathbf{x})a_l(\mathbf{x})\right]. \quad (3)$$

The $\mathbf{n}$th raw moment of the system $\mathbb{E}(\mathbf{X}^{\mathbf{n}}) = \mathbb{E}(X_1^{n_1}\dots X_N^{n_N})$ can be found as the nth order derivative of Eq. (3) with respect to $\boldsymbol{\theta}$. The expressions obtained this way are evaluated by expanding $a_l(\mathbf{x})$ in a Taylor series about the mean giving rise to a differential equation system containing terms of central moments, $\mathbb{E}(\hat{\mathbf{X}}^{\mathbf{n}}) = \mathbb{E}((X_1-\mu_1)^{n_1}\dots(X_N-\mu_N)^{n_N}))$, with μ_i being the mean of the ith species. The relationship between raw and central moments is formulated as

$$\mathbb{E}(\hat{\mathbf{X}}^{\mathbf{n}}) = \sum_{k_1=0}^{n_1}\cdots\sum_{k_N=0}^{n_N}\binom{\mathbf{n}}{\mathbf{k}}(-1)^{(\mathbf{n}-\mathbf{k})}\mu^{(\mathbf{n}-\mathbf{k})}\mathbb{E}(\mathbf{X}^{\mathbf{k}}), \quad (4)$$

where

$$(-1)^{(\mathbf{n}-\mathbf{k})} = (-1)^{(n_1-k_1)}\dots(-1)^{(n_N-k_N)},$$
$$\mu^{(\mathbf{n}-\mathbf{k})} = \mu_1^{(n_1-k_1)}\dots\mu_N^{(n_N-k_N)},$$
$$\binom{\mathbf{n}}{\mathbf{k}} = \binom{n_1}{k_1}\dots\binom{n_N}{k_N}.$$

We obtain a system of ordinary differential equations for the time evolution of central moments by applying Eq. (4) to substitute raw moments with central moments, except for the mean molecular numbers, μ_i.

In systems involving nonlinear connections, each central moment depends on higher order moments, and therefore the set of differential equations generated by moment expansion is, in principle, infinite and unsolvable. To obtain a closed set of equations, we can stop in the evaluation of the Taylor expansion at a given order of m, i.e., set all higher order terms to zero, $\mathbb{E}(\hat{\mathbf{X}}^{\mathbf{n}}) = 0$, if $\sum n_i > m$. In the following, we refer to this procedure for substituting $(m+1)$th (as well as all higher) order terms with zeros as *zero-closure*. An alternative way to close the equation set is to apply moment closure techniques which rely on derived expressions for higher order moments, based on properties of an assumed underlying distribution.

B. Moment closure techniques

Most well-known and widely used distributions are completely described by a small set of parameters, meaning that given the first few moments these parameters can be calculated and used to determine the whole probability distribution (including, obviously, all higher order moments). We use this property to substitute expressions of lower order moments in place of the $(m+i)$th (where $i \geq 1$) moments appearing in the set of ODEs generated by the above expansion method; so in the end we obtain a closed set of differential equations for the moments,

$$\frac{d}{dt}\mathbb{E}(\hat{\mathbf{X}}^{\mathbf{m}}) = f(\boldsymbol{\mu},\dots,\mathbb{E}(\hat{\mathbf{X}}^{\mathbf{m}}),mc^k(\mathbf{X})), \quad (5)$$

where $mc^k(\mathbf{X})$ denotes the moment closure expression based on up to kth order central moments of $\mathbf{X}$. In the following, we derive the expressions for $mc^k(\mathbf{X})$ based on three distributions.

We investigate normal, lognormal, and gamma distributions. The first of these is, of course, the most popular distribution due to its prominent role in the central limit theorem. The other two distributions are considered for their asymmetry and restriction to the positive real numbers, both of which are important features of the systems describing molecular reactions. Normal and lognormal distributions have been used successfully to model distributions in biochemistry,[14,15,20,23] but single cell protein expression levels do follow a gamma distribution[40,41] if considered at the population level. Our closure approaches involve two-parameter distributions, i.e., the first two moments provide a sufficient description of the distribution. As biological systems usually contain several species which have strong influence over each other, we use multivariate definitions of the distributions, in which the different interaction levels between variables are captured in the different mixed order moments. In this case, the two parameters become a vector and a matrix, typically in the form of, respectively, a mean vector and covariance matrix.

For each case, we start the derivation by providing a definition for the corresponding multivariate distribution and its characteristic parameters. Then, we derive formulas for all (both non-mixed and mixed) moments of an arbitrary order m (necessarily, $m \geq 3$) in terms of the aforementioned parameters. Expressions of raw moments are translated into central moment terms using Eq. (4), whereas for the normal distribution, we can obtain the central moments directly. Finally, we give a way to calculate the parameters from means, variances, and covariances of the species, which are considered known, as their time evolution has already been described in the moment expansion step. As a comparison to the multivariate techniques, we also generate univariate closures where only expressions for the marginals are generated, by setting all mixed moments to zero.

1. Multivariate normal closure

A random vector, X, is said to be N-variate normally distributed if every linear combination of its components has a univariate normal distribution. It is described by the density function,

$$f_{\mathbf{X}}(\mathbf{x}) = \frac{1}{\sqrt{(2\pi)^N|\Sigma|}} \exp\left(-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T\Sigma^{-1}(\mathbf{x}-\boldsymbol{\mu})\right), \tag{6}$$

which reduces to the pdf of the univariate normal distribution, if $\Sigma \in \mathbb{R}^{1\times 1}$.

All higher order central moments of the normal distribution can be calculated using Isserlis' theorem,[42,43]

$$\mathbb{E}(\hat{X}_1^{m_1}\dots\hat{X}_N^{m_N}) = 0, \text{ if } \sum_{i=1}^{N} m_i = \text{odd}, \tag{7a}$$

$$\mathbb{E}(\hat{X}_1^{m_1}\dots\hat{X}_N^{m_N}) = \sum_k \prod_{(i,j)\in I_k} \Sigma_{i,j}, \tag{7b}$$

where $\mathbb{E}(\hat{X}_i)$ denotes central moments (i.e., $\hat{X}_i = X_i - \mu_i$) and I_k are the sets formed by partitioning the set $\mathbb{I} = \{\underbrace{1,1,\dots,1}_{m_1},\dots,\underbrace{N,N,\dots,N}_{m_N}\}$ into unordered pairs, with k being the number of such sets. (We illustrate the above formula in Appendix A.)

In this case, the necessary parameters $\boldsymbol{\mu},\Sigma$ can be directly obtained from the first two central moments of the species described by the system,

$$\begin{aligned}\mu_i &= \mathbb{E}(X_i),\\ \Sigma_{ii} &= \mathrm{Var}(X_i),\\ \Sigma_{ij} &= \mathrm{Cov}(X_i,X_j).\end{aligned} \tag{8}$$

2. Multivariate lognormal closure

For the univariate case, a random variable X has a lognormal distribution if its logarithm is normally distributed; thus, an N-variate lognormally distributed random vector ($\mathbf{X}$) is characterised by its normal counterpart ($\mathbf{Y}$),

$$\mathbf{X} = \exp(\mathbf{Y}), \tag{9}$$

and $\mathbf{Y} \sim \mathcal{N}(\boldsymbol{\mu},\Sigma)$.

All higher order (raw) moments can simply be expressed in terms of $\boldsymbol{\mu}$ and Σ using the formula[44]

$$\mathbb{E}(\mathbf{X}^{\mathbf{m}}) = \exp(\mathbf{m}^T\boldsymbol{\mu} + \frac{1}{2}\mathbf{m}^T\Sigma\mathbf{m}), \tag{10}$$

where $\mathbf{m} = [m_1, m_2, \dots, m_N]^T$ and $m_i \in \mathbb{N}$.

Given our observation of up to second order moments of the components, the parameters $\boldsymbol{\mu}$ and Σ of $\mathbf{Y}$ are calculated as

$$\begin{aligned}\Sigma_{ii} &= \ln\left(1 + \frac{\mathrm{Var}(X_i)}{\mathbb{E}(X_i)^2}\right),\\ \mu_i &= \ln(\mathbb{E}(X_i)) - \frac{1}{2}\Sigma_{ii},\\ \Sigma_{ij} &= \ln\left(\frac{\mathrm{Cov}(X_i,X_j)}{\exp(\mu_i+\mu_j+\frac{1}{2}(\Sigma_{ii}+\Sigma_{jj}))} + 1\right).\end{aligned} \tag{11}$$

3. Multivariate gamma closure

The univariate gamma probability density distribution can be parametrized by

$$f_X(x;\alpha,\beta) = \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1}\exp(-x/\beta), \tag{12}$$

where Γ denotes the gamma function, β is a scale parameter, and α is a shape parameter. The (raw) moments of a variable $X \sim \mathrm{Gamma}(\alpha,\beta)$ are determined by the recurrence relation

$$\mathbb{E}(X^m) = (\alpha)_m\beta^m, \tag{13}$$

where $(a)_m = (a+m-1)!/(a-1)! = a(a+1)\dots(a+m-1)$.

There are several ways to define a MVG distribution fulfilling the requirement that individual marginals follow a univariate gamma distribution. Since the sum of two gamma variables with the same scale parameter will again be gamma distributed, it is straightforward to obtain a MVG distribution from linear combination of independent gamma variables. In this paper, we introduce a new MVG definition similar to the ones in Mathai[45] and Furman.[46] Here, in order to allow for well-behaved covariances, we define the MVG in terms of linear combinations of gamma random variables; then, the overlap between the independent gamma components naturally determines the cross terms in the covariance matrix.

Let Y_{kl} ($k = 1\dots N$, $l = k\dots N$) be independent gamma variables with shape and scale parameters α_{kl},β_{kl}. For convenience, we also introduce extra — but not any more independent — gamma variates, Y_{rq} ($r = 2\dots N$, $q = 1\dots(r-1)$), such that they equal Y_{kl} if $k = q$, $l = r$. We will use these additional variables in the following to ease the indexing of Y_{kl} as, for example, Y_{12} and Y_{21} refer to the same variable. Then, $\mathbf{X} = [X_1, X_2 \dots X_N]$,

$$X_i := \sum_{k=1}^{N} \frac{\beta_{ii}}{\beta_{ik}} Y_{ik}, \tag{14}$$

will have an N-dimensional gamma distribution, and $X_i \sim \mathrm{Gamma}(\alpha_i,\beta_i)$, where

$$\begin{aligned}\alpha_i &= \sum_{k=1}^{N}\alpha_{ik},\\ \beta_i &= \beta_{ii}.\end{aligned}$$

Furthermore, variables X_i and X_j will have exactly one component ($Y_{ij} = Y_{ji}$) in common to account for their correlations. Consequently, all non-mixed higher order moments can be calculated using Eq. (13). A closed formula for the mixed moments arises from the generalisation of an alternate expression for the non-mixed moment case

$$\begin{aligned}\mathbb{E}(X_i^m) &= \mathbb{E}\left(\left(\sum_{k=1}^{N}\frac{\beta_{ii}}{\beta_{ik}}Y_{ik}\right)^m\right)\\ &= \mathbb{E}\left(\sum_{\mathbf{k}}\mathbb{C}(m,\mathbf{k})\prod_{r=1}^{N}\left(\frac{\beta_{ii}}{\beta_{ir}}Y_{ir}\right)^{k_r}\right)\\ &= \beta_{ii}^m\left(\sum_{\mathbf{k}}\mathbb{C}(m,\mathbf{k})\prod_r(\alpha_{ir})_{k_r}\right),\end{aligned} \tag{15}$$

where the sum is taken over all combinations of non-negative integers k_r such that they sum up to m, and $\mathbb{C}(m,\mathbf{k})$ is the multinomial coefficient

$$\mathbb{C}(m,\mathbf{k}) = \binom{m}{k_1,\ldots,k_N} = \frac{m!}{k_1!\ldots k_N!}, \tag{16}$$

and where we have made use of the independence of Y_{kl} variables. Similarly, we adapt the vectorised notation of exponents $\mathbf{m}$ used in Eq. (10) and obtain

$$\mathbb{E}(\mathbf{X}^{\mathbf{m}}) = \boldsymbol{\beta}^{\mathbf{m}}\left(\sum_{\mathbf{k}}\prod_i \mathbb{C}(m_i,\mathbf{k})(\boldsymbol{\alpha}_i)_{\mathbf{k}}\right), \tag{17}$$

where

$$\boldsymbol{\beta}^{\mathbf{m}} = \beta_{11}^{m_1}\ldots\beta_{NN}^{m_N},$$
$$(\boldsymbol{\alpha}_i)_{\mathbf{k}} = (\alpha_{i1})_{k_1}\ldots(\alpha_{iN})_{k_N},$$

and the combined product and sum correspond to taking the product of the expanded $X_1^{m_1}\ldots X_N^{m_N}$ terms, with $(\boldsymbol{\alpha}_i)_{\mathbf{k}}$ calculated after the $\alpha_{kl} = \alpha_{lk}$ symmetry is taken into account. We provide an example of the computation of mixed moments in Appendix B. The power of our MVG definition lies in the fact that the covariance of any two variables is directly represented by a single one of the auxiliary parameters, as one can see from the following computation:

$$\begin{aligned}\mathbb{E}(X_iX_j) &= \beta_{ii}\beta_{jj}\left(\sum_{(k,l)\neq(i,j)}\alpha_{ik}\alpha_{jl} + (\alpha_{ij})_2\right),\\ \mathrm{Cov}(X_i,X_j) &= \mathbb{E}(X_iX_j) - \beta_{ii}\beta_{jj}\left(\sum_{k,l}\alpha_{ik}\alpha_{jl}\right)\\ &= \alpha_{ij}\beta_{ii}\beta_{jj},\end{aligned} \tag{18}$$

where the indices k and l go through all values between 1 and N, except when stated otherwise. One can see that in order to calculate the final formulas, we need all α_{kl} values, but only the "diagonal" β_{kk} elements of the scale parameters. To emphasise the similarity with the previous closure techniques, we summarise the descriptive parameters in a vector of scale parameters, $\boldsymbol{\beta}$, and a matrix of shape parameters, $\mathbf{A}$: $\mathbf{A}_{kl} = \alpha_{kl}$. These can be obtained by writing Eqs. (13) and (18) as

$$\begin{aligned}\beta_i &= \mathrm{Var}(X_i)/\mathbb{E}(X_i),\\ \mathbf{A}_{ij} &= \mathrm{Cov}(X_i,X_j)/(\beta_i\beta_j),\\ \mathbf{A}_{ii} &= \mathbb{E}(X_i)/\beta_i - \sum_{k\neq i}\alpha_{ik}.\end{aligned} \tag{19}$$

III. RESULTS

In the following examples, we first investigate the performance of moment closure techniques on two systems both of which exhibit nonlinear dynamics and show oscillatory behaviour. Sometimes even a deterministic model suffices to capture the mean behaviour, but this depends on the parameter values. There are regimes in the parameter space, where deterministic and low-order stochastic approximation methods qualitatively deviate from the true behaviour. In the first two models, we focus on parameter values taken from this regime so that we can explore the improvement our method can provide in the most difficult cases. In our last example, we consider a system exhibiting less complex — non-oscillatory — dynamics and demonstrate how our closure techniques can be applied on systems of many interacting species.

In the examples presented, we stop the Taylor expansion at order $(m+1)$ and eliminate the $(m+1)$th order terms using moment closure. Throughout this section, we will use the following notation: an mth-order moment closure means the ODE system consists of equations for up to mth order moments and the closure was applied in order to develop substituting terms for the $(m+1)$th order moments. As expected when only considering one higher order in the expansion, normal closure is equivalent to zero closure in all even order closures, as odd-order terms are always substituted by zeroes. Therefore, in most of the analysis, we will focus on moment closure approximations based on moments up to an odd order.

A. P53 system

We first illustrate the properties of our closures applied together with MEA in the context of the oscillatory p53-Mdm2 system.[47] The model consists of three variables x_1, x_2, and x_3 corresponding to the proteins, p53, Mdm2 precursor, and Mdm2, respectively; these molecular species are connected through a feedback loop with nonlinear rate laws. The reaction network is described by the stoichiometry matrix

$$S = \begin{bmatrix}1 & -1 & -1 & 0 & 0 & 0\\ 0 & 0 & 0 & 1 & -1 & 0\\ 0 & 0 & 0 & 0 & 1 & -1\end{bmatrix}, \tag{20}$$

and six reaction propensities

$$\begin{aligned}a_1 = k_1,\ a_2 = k_2x_1,\\ a_3 = k_3\frac{x_1x_3}{x_1+k_7},\ a_4 = k_4x_1,\\ a_5 = k_5x_2,\ a_6 = k_6x_3,\end{aligned} \tag{21}$$

where k_1 is the p53 production rate, k_2 is the Mdm2-independent p53 degradation rate, k_3 the saturated p53 degradation rate, k_7 is the p53 threshold for degradation by Mdm2, k_4 is the p53-dependent Mdm2 production rate, k_5 is the Mdm2 maturation rate, and k_6 is the Mdm2 degradation rate.

Despite the sustained oscillations at the single-cell level, as different cells go out of phase, dampened oscillations can be observed in the mean population dynamics. Ale *et al.*[10] have shown that the deterministic approach and the LNA are not able to capture this behaviour, and at least 6 moments need to be included in the standard MEA (equivalent to zero closure). This suggests that the variance and the skewness are not sufficient for capturing the correct mean, and only expanding the moment equations further allows us to observe the behaviour of the system. Here, we use the parameters and initial values from Ale *et al.*[10] — both already adapted to a stochastic system, i.e., rates are in 1/s and initial values are dimensionless quantities representing the total number of molecules in the system — $\mathbf{k} = [90, 0.002, 1.7, 1.1, 0.93, 0.96, 0.01]$, $\mathbf{x}(0)$

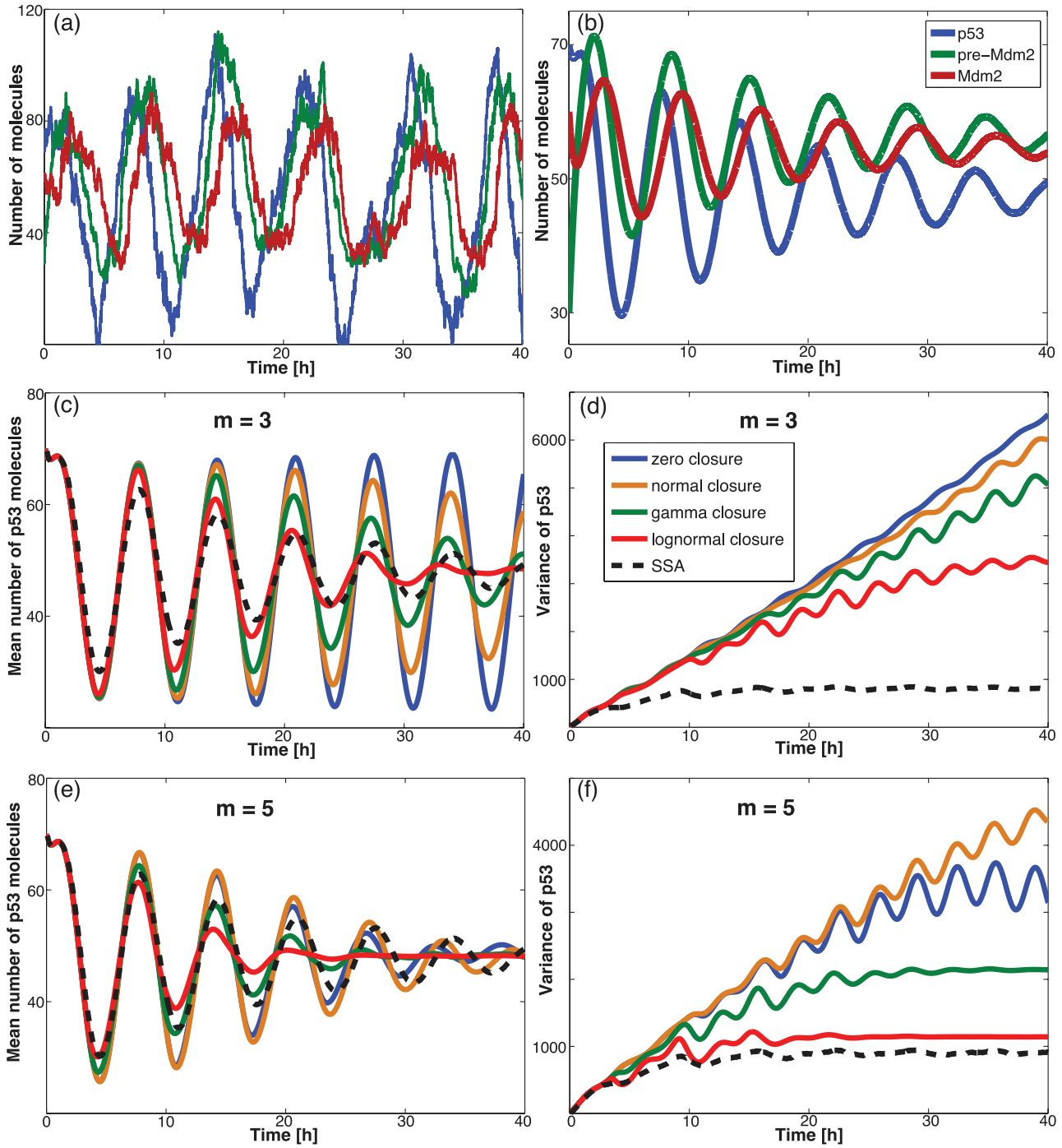

FIG. 1. Study of the p53 system with parameters $\mathbf{k} = [90, 0.002, 1.7, 0.6, 0.96, 0.01]$ and initial conditions $\mathbf{x} = [70, 30, 60]$. (a) A single exact realisation by SSA; colours are as in the legend in part (b). (b) Mean behaviour obtained as the average of 10 000 SSA simulations. (c) and (d) Trajectory of the mean and variance for p53 protein number (x_1) calculated by applying third order zero (blue), normal (orange), gamma (green) and lognormal (red) closure. The dashed black line shows the result of 10 000 SSA runs in correspondence. (e) and (f) Trajectory of the mean and variance of p53 protein number calculated through fifth order moment closures.

= [70, 30, 60] and compare our results to 10 000 stochastic simulations representing an assembly of individual cells (Figures 1(a) and 1(b)).

In Figs. 1(c) and 1(d), we compare third-order moment closure approximations (including the corresponding zero closure) and the true behaviour obtained from averaging 10 000 SSA runs for state x_1. For the first quarter of the trajectory, all closure methods perform similarly, but after that zero closure methods are not able to capture the dampening oscillations while all the other closures show a decrease in amplitude. As expected, normal closure shows the least pronounced improvement, as the real distribution for p53 molecules is highly asymmetric, which suggests that a symmetrical distribution as the Gaussian will not provide a sufficient description. Gamma and lognormal closure both perform well, the latter being closest to the true trajectory up to 25 min, then overestimating the dampening. The quality of the approximation is quantified as the cumulative squared

difference between the respective methods and the SSA trajectory. By this measure, the lognormal closure approximates the true behaviour best, but the error of gamma closure also appears to level off over time. As was shown previously for zero closure, the variance is not correctly estimated for fewer than six moments. Using multivariate lognormal closure, the qualitative behaviour of the variance of x_1 is estimated from three moments just as well as from six moments in the zero closure case. On the other hand, assuming a gamma distribution does only slightly improve the approximation qualitatively, as no levelling-off is observed over the time-course considered here. We also extended the simulated time interval to 160 h and find that the performance of closure methods does not change; distribution-based closures all reach the steady state but show varying dampenings in the trajectory, while zero-closure fails to converge to a steady state and the variance approximated by this method grows exponentially instead of levelling-off.

We also investigate fifth order distribution-based closure methods, which are compared to sixth order zero closure (Figures 1(e) and 1(f)). All approximation methods overestimate the dampening. Lognormal and gamma closure converge to the correct final value, while the trajectory predicted by zero closure for a longer time-period is not constant but an increasing linear function. On the other hand, the variance of p53 is approximated very well, especially by lognormal closure. We also evaluate the difference between multivariate and univariate fifth-order closures and find that the time evolution of variances and covariances computed through univariate techniques diverges more from the multivariate ones than for third-order closures. This suggests that including various amounts of 6th order information in the closure makes a difference in the quality of the approximation, which is in correspondence with the previous findings showing that including 6th order moments significantly improves the estimation of the variance.[10]

We also find that even order moment closures (except for zero closure) result in an ODE system which is not stable enough to be evaluated for the entire time course. We use several ODE solvers to obtain trajectories, all of which fail because the system is too *stiff*. That is, the derivatives for some of the moment equations become too large and hence the time step of the solver needs to be reduced below the minimal allowed step size; however, the time of failure differs amongst the algorithms solvers used. Zero closure produces stable and reasonable trajectories, suggesting that some subsequent moments are so strongly coupled that if we include only an approximation of odd-order terms, we lose relevant information providing the stability of the system. We note that problems with ODE solvers occur frequently with moment expansion approaches, especially as vast parameter spaces are explored as is the case in inference; we suspect that ODEs resulting from MEA (or related approaches) may require specific solvers.

B. HES1 system

Our second example is the feedback loop of the transcription factor Hes1, involved in the regulation of embryonic segmentation, which is also described by a 3-variable, 6-reaction model. The observables are the level of Hes1 mRNA (x_1) and number of Hes1 protein molecules in the cytosol (x_2) and in the nucleus (x_3) — not all of these can, however, be simultaneously measured *in vivo*. The system is formulated in terms of stoichiometries and propensities as[37]

$$S = \begin{bmatrix} -1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 1 & -1 & 0 \\ 0 & 0 & 0 & 0 & 1 & -1 \end{bmatrix}, \tag{22}$$

$$\begin{aligned} a_1 &= k_1 x_1,\ a_2 = \frac{1}{1 + (x_3/k_2)^{k_3}}, \\ a_3 &= k_1 x_2,\ a_4 = k_4 x_1, \\ a_5 &= k_5 x_2,\ a_6 = k_1 x_3, \end{aligned} \tag{23}$$

where k_1 is the Hes1 degradation rate that we assume to be the same for all each species, k_2 is the amount of Hes1 protein in the nucleus at half-maximal transcription of the mRNA, k_3 is the Hill coefficient, k_4 is the Hes1 mRNA translation rate, and k_5 is the rate of nuclear transport of Hes1 protein. The parameter values used (in the above order) are $\mathbf{k} = [0.03, 15, 7, 0.1, 0.008]$, and we start the simulations from $\mathbf{x}(0) = [15, 10, 6]$. As in the case of the p53 system, the results are compared to the average result of 10 000 SSA runs. Similarly to the previous example, sustained oscillations are present in single realisations but appear only transiently in the mean behaviour and the system soon converges to a steady-state value (Figures 2(a) and 2(b)).

The very low copy numbers of some species (on the order of 1-10 molecules) make it harder to follow the evolution of the probability distribution using moment expansion/closure. We again find that including information about odd-order moments without continuing the expansion to the subsequent even order makes the resulting system unstable; hence, odd orders for zero closure and even orders for distribution-based closures fail to produce realistic dynamics, even though the resulting equations are still numerically tractable. On the other hand, Fig. 2(c) shows that similarly to the p53 system, third order moment closure methods are able to capture the characteristic dampening. Although even zero closure converges to a steady state (4 moments are shown in comparison, as the third order closure is unstable), only lognormal closure estimates the steady state value correctly. As for the variances, values are underestimated by the approximation methods, but the shape of the curve is captured in all cases; overall, lognormal and gamma closures perform best.

We also investigate the difference between applying univariate and multivariate closure techniques for this system. In contrast to the p53 system, there is a significant difference between mean values obtained with or without the inclusion of mixed moments, as shown for lognormal closure in Fig. 2(e). In the univariate closure cases, the amplitude of oscillations reduces only slowly over time, while multivariate closures converge much faster to the steady state. In this regime, the dynamics of the Hes1 system are strongly dependent on the interactions between the three species. Interestingly, closure not considering mixed moments performs about as well as

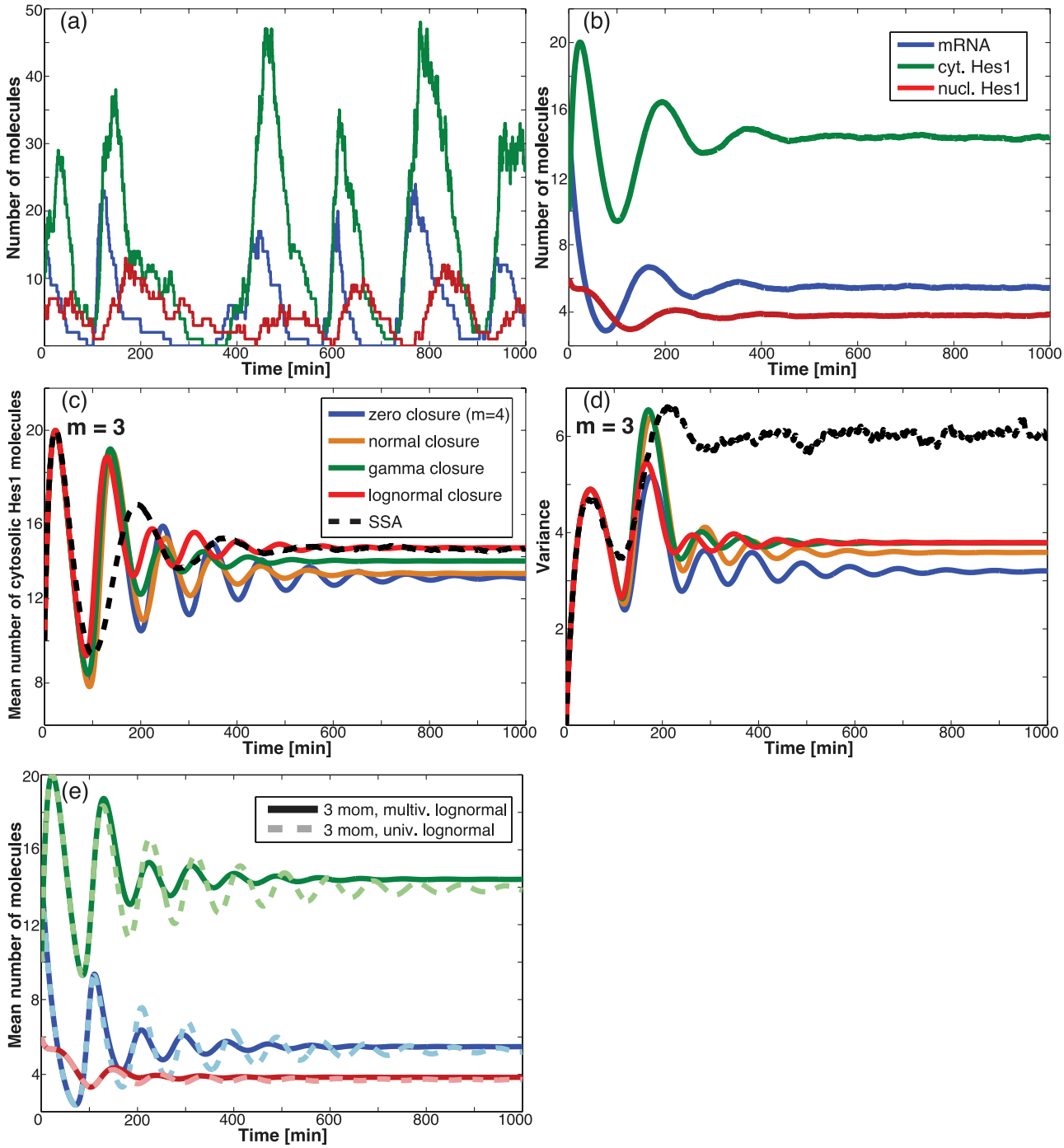

FIG. 2. Study of the Hes1 oscillatory model with parameter $\mathbf{k} = [0.03, 1.5, 7, 0.1, 0.008]$ and initial conditions $\mathbf{x} = [15, 10, 6]$. (a) Single SSA simulation of the system; colours as in part (b). (b) Average of 10 000 SSA realisations. (c) Mean number of Hes1 molecules in cytosol (x_2) obtained from SSA simulations (dashed black line), fourth order zero closure (blue line), and third order normal, gamma, and lognormal moment closures (orange, green, and red lines, respectively). (d) Variance of nuclear Hes1 molecular number calculated as above. (e) Trajectory of the mean for all variables, obtained by univariate (light dashed lines) and multivariate (dark solid lines) third order lognormal closure; colours as in part (b).

fourth order zero closure, although the latter contains more information on fourth order moments.

C. ERK/MEK system

Up to now, we have considered systems with "challenging" dynamics (see Ref. 37). Here, we now demonstrate the advantages of our method on a large complex system, modelling the widely studied mitogen-activated kinase, Erk, and its cognate kinase, Mek. Mitogen-activated protein kinases (MAPK) are involved in regulating cellular fates such as proliferation, differentiation, and apoptosis, and therefore many deterministic and stochastic models are available including various details of regulatory and regulated interactions. We choose a very simplified minimal model by Harrington *et al.*,[48] but even this model contains 7 species and 8 reactions. We reformulate the model in stochastic terms with stoichiometry matrix

TABLE I. Variables and parameter values used for the ERK/MEK system. Variable values represent the number of molecules initially in the system, while parameter values are given in 1/s (molecules/s).

Variable	Value	Parameter	Value
x_1	9240	k_1	$9.524 \cdot 10^{-8}$
x_2	1848	k_2	$8.8 \cdot 10^{-2}$
x_3	0	k_3	$3 \cdot 10^{-1}$
x_4	0	k_4	$9.524 \cdot 10^{-8}$
x_5	7392	k_5	$8.8 \cdot 10^{-2}$
x_6	0	k_6	$2 \cdot 10^{-1}$
x_7	0	k_7	$1.4 \cdot 10^{-2}$
		k_8	$1.4 \cdot 10^{-2}$

$$S = \begin{bmatrix} -1 & 1 & 1 & -1 & 1 & 1 & 0 & 0 \\ -1 & 1 & 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & -1 & -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & -1 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & -1 \end{bmatrix}, \quad (24)$$

and reaction propensities

$$\begin{aligned} a_1 &= k_1 x_1 x_2, \quad a_2 = k_2 x_3, \; a_3 = k_3 x_3, \\ a_4 &= k_4 x_1 x_5, \quad a_5 = k_5 x_6, \; a_6 = k_6 x_6, \\ a_7 &= k_7 x_4, \qquad a_8 = k_8 x_7, \end{aligned} \quad (25)$$

where the variables $(x_i,\ i = 1 \ldots 7)$ denote the absolute number of Mek, Erk1, Mek-Erk1 complex, Erk1p, Erk2, Mek-Erk2 complex, and Erk2p molecules, respectively. We adjusted initial conditions and parameters to fit the above stochastic model assuming an average system size of $9.24 \cdot 10^5$ (based on the 1540 μm^3 average volume of HeLa cells) and also lowered molecular concentrations into a regime where stochastic effects are prevalent; for a list of the values used in the simulations, see Table I. In our analysis, we focus on the levels of Erk1p and Erk2p, i.e., phosphorylated Erk1 and Erk2. In this regime, exact simulation of the stochastic dynamics is typically computationally prohibitive and approximate methods are required if we wish to allow for stochastic dynamics.

We study second order moment closures, compared to an average of 10 000 SSA runs. We find that in this case of relatively simple, at most second order rate laws, the mean behaviour is completely captured by the approximation methods. Hence, in contrast to the previous two examples, higher order moments have negligible influence on the evolution of the mean. As expected, there is no difference between zero and normal closure, as both methods set all third order terms to zero, but also the results from gamma closure are the same. Fig. 3(a) shows the mean ±2 standard deviations obtained by gamma closure (zero and normal closure would result in the same figure) as well as a single SSA realisation for both outputs Erk1p and Erk2p. The average trajectory of the 10 000 SSA realisations completely agrees with the approximated means; hence, for the sake of transparency, they are not shown in Fig. 3(a). The same holds even when we reduce the system size 100 times, having only a couple of molecules from some species in the system. On the other hand, the equation set produced by lognormal closure is numerically unstable and cannot be evaluated by the ODE solvers. This is probably due to the distribution of the molecules having a low skewness compared to its variance; hence, a lognormal assumption would force moments to take unrealistic values and this, together with the fact that most variables, especially x_3 and x_6 change very rapidly in the first few seconds, introduces instability into the system. Similarly, variances produced by closure approximations and calculated from the SSA runs completely match, although the SSA one already shows some uncertainty.

A second order approximation of the system consists of 35 equations, each with between 5 and 20 terms. Third order approximations would produce an equation set of 118 variables with more complex (around 30 terms) right-hand side expressions, that is considerably slower to simulate (not to mention the original costs of determining the equations —

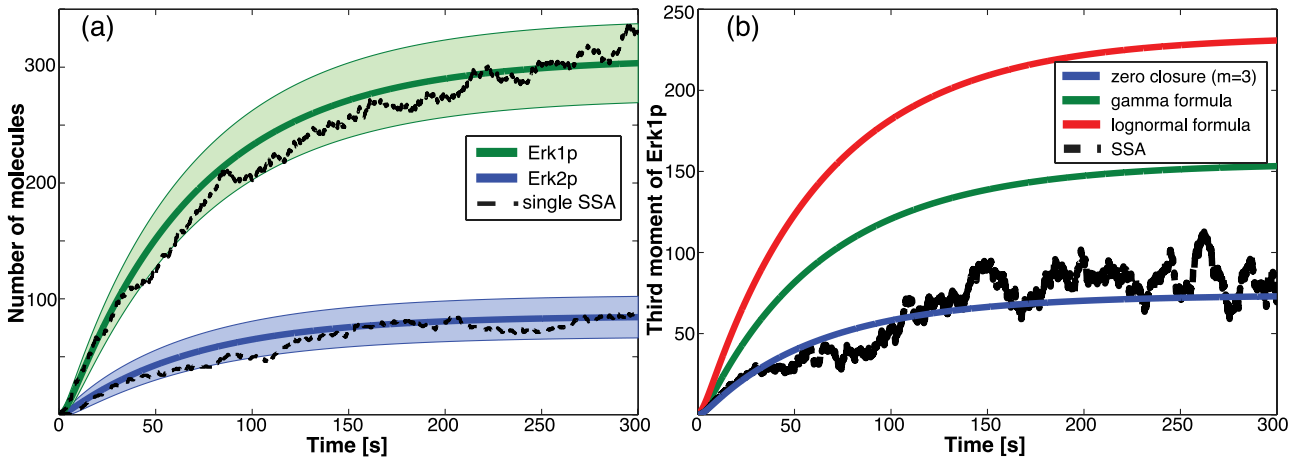

FIG. 3. Study of the simplified Erk/Mek model, parameters, and initial conditions as in Table I. (a) Mean trajectory (solid line) of the two output variables (Erk1p and Erk2p) ±2 standard deviations (shaded areas) calculated by second order gamma closure and a single SSA realisation (black dashed lines). (b) Approximations of the third central moment of Erk1p obtained by full third order zero closure (blue) and applying gamma (green), and lognormal (red) closure formulas based on the full second order moment closure. Normal closure is not shown as it is 0 for the whole time-course.

which is here done automatically using computer algebra); additionally, the risk of introducing numerical errors is also higher. However, deriving third order moments from SSA simulations is not feasible as at least a million simulations would be required to be able to follow the evolution of such moments reliably. In such situations, the formulas used during approximation to close the moment equations can be explicitly applied to estimate moments of higher orders based on available moment values. In Figure 3(b), we plot the third moment of Erk1p calculated from Equations (10) and (17) to obtain the necessary second order moments for the formulas. Since there are very strong fluctuations present in the SSA estimation, we also compare the results to the trajectory obtained by the third order zero closure approximation. Lognormal closure overestimates the actual moment by a factor of three, which supports our previous observation that the instability of lognormal closure arises because this closure *Ansatz* assumes a more highly skewed distribution given the mean and variance/covariance values. Gamma closure also overestimates the moment, but qualitatively follows the change in skewness, hence provides more information on the actual time evolution than normal closure, which is equivalent to assuming constant zero third order moments. Some results suggest that if moment closures can be evaluated by their performance — or stability — on estimating the mean, an improved prediction can be obtained by taking a consensus of the "accepted" third order predictors, e.g., in Fig. 3(a), we see that both gamma and normal closure perfectly match the mean, and the average of trajectories derived from gamma and normal closure formulas would give a more precise approximation of the actual third order moment.

IV. CONCLUSION

In this paper, we have introduced a general moment closure approximation technique based on n-dimensional multivariate forms of normal, lognormal, and gamma distributions. We have derived expressions that automatically determine higher order moments in terms of means, variances, and covariances, which can be used in any application concerning summary statistics of dependent random variables. Here, we have combined the formulas with a general moment expansion approach to follow the time evolution of complex nonlinear systems and shown that moment closure methods are able to approximate the mean behaviour and higher moments in the regime of 5-100 molecules, where many other methods tend to fail.[5]

With recent developments in single cell recording techniques, our approach offers many benefits and allows us to make better use of the information in such experimental data.[49,50] Since our closure-based approximation method transforms the stochastic problem into a set of ODEs, where moments of the probability distribution are treated as new variables, the approach can be directly used in many applications designed for deterministic systems. For example, standard inference schemes[28] can be used to estimate parameters of stochastic systems via conditioning not only on the mean but also variance, skewness, etc. We have shown that the use of moment closure formulas instead of truncation considerably reduces the number of moments required to approximate the correct mean behaviour, while going to the same order improves the estimation of higher moments. The reduced system size can be beneficial in most inference tasks, as these algorithms involve evaluating the system for thousands of parameter sets. A related further application of our framework is in the field of experimental and control design.[35,51,52] This field offers numerous mathematical and engineering tools developed for differential equation systems, which applied to a stochastic model followed with our algorithm could lead to automatic, *in silico* experimental design for biochemical systems, for example, determining the optimal input in order to drive the system into a state with reduced variability. Furthermore, combined with moment-based distribution reconstruction techniques,[53,54] our method can provide a way to access the whole distribution of the system without analytically solving the CME or costly SSA simulations. Since parametric sensitivities are easily obtained from our analysis (see, e.g., Refs. 10 and 34), our framework can also be used to study parameter dependence of the system's evolutions, for example, through deriving Waddington landscapes[55] from the probability distribution.[56] However, note that sometimes high order moments are needed for the unique reconstruction of the distribution, and the overall cost of our algorithm and the reconstruction technique might exceed that of the generation of SSA trajectories.

Our closure methods rely on the assumption that the observed species follow a specific joint distribution; therefore, choosing the best closure technique can be problematic and one has to be aware of the limitations arising from the chosen method. For some systems, prior knowledge might help in determining the best distribution to model the species,[41] but usually we are not so fortunate and can only base our decision on very general properties. For example, normal closure can lead to errors or failure in systems where the skewness of the probability distribution over states is significantly different from zero. Both the p53 and Hes1 systems exhibit such dynamics and normal closure performs much worse than the two asymmetric distributions. In light of our analyses, it appears that lognormal closure should be more successful in cases with highly skewed distributions. In cases where the lognormal *Ansatz* fails, gamma closure might still be applied, but due to positiveness of the univariate gamma distribution higher order mixed moments should all be positive, and hence negatively correlated variables cannot be modelled with this *Ansatz*. However, our implementation of a new MVG distribution allows the system to assign negative values to these moments based on the corresponding covariances. In addition, the current algorithm can be refined by including more moments in the calculation of closure expressions. In cases where the actual distribution cannot be described by any of the distributions considered here, an improved approximation might be obtained by deriving the parameters from different pairs of lower order moments — e.g., by expressing $\mathbf{m}$ and $\mathbf{\Sigma}$ of a lognormal distribution in terms of mean and skewness as well as mean and variance — and using a weighted average of the results. However, this method would either require tighter restrictions on the closure order, m, or considerable computational time.

In case we have no prior information about the distribution of species, a possible way of finding the optimal closure is to calculate all three second order approximations and compare their stability,[57] as incorrect assumptions can lead to unstable systems, such as in the case of the Erk/Mek network. Moreover, if experimental data or data from a couple of SSA simulations are available, the reliability of each closure can be estimated, similar to the estimation of necessary order moments in Ale *et al.*[10] If the system size allows comparing full third order moments (obtained from third order closure) to the ones calculated by the respective assumptions, this computation can also help in deciding which distribution fits the data best. After the distribution assumptions are evaluated with one of the above approaches, the optimal or a mixture of reasonable closures can be used to model the distribution of the system. For example, closure formulas can be used as an estimator of higher order moments even when further expanding the approximation is not feasible or efficient. However, although the three included cases can fit a very broad range of possible molecular distributions, note that there are still some examples which cannot be well approximated, for example, bistable systems and other multimodal distributions. Investigating further distributions, such as Poisson or (negative) binomial distribution, can also improve the approximation as they have been shown to have relevance in some biochemical models, but they lack a general multivariate formulation for which closure formulas could be derived automatically for an arbitrary system.

For some such systems, general criteria have recently been developed[57,58] that allow us to determine whether moment closure approximations can be usefully applied. Whenever nonlinear dynamics and noise interact, it poses a challenge for computationally convenient approximations;[59] approaches such as the one introduced here can easily improve the analysis of such systems compared, for example, to the LNA, but many biologically important processes will defy these attempts, and for these there may be no alternative to costly SSA simulations. Statistical inference is an area in which approximations could have a potentially important role, since even approximations to the true pdf could greatly help when trying to define (approximate) likelihood functions.[6,10,60,61]

ACKNOWLEDGMENTS

E.L. is funded through a Schrödinger Fellowship and A.A. through a MRC Special Training Fellowship in Biomedical Informatics. P.K. and M.P.H.S. gratefully acknowledge funding through HFSP Grant No. RPG0061/2011; M.P.H.S. is a Royal Society Wolfson Research Merit Award holder.

APPENDIX A: EXAMPLE OF ISSERLIS' THEOREM

To illustrate how Isserlis' theorem provides a way to express any even ordered moments of a normally distributed N-dimensional variable, consider the fourth order central moment $\mathbb{E}(\hat{X}_1^2\hat{X}_2\hat{X}_3)$. First, we need every possible partitioning of the set $\{1,1,2,3\}$ into unordered pairs, which in this case will give us three sets composed of two pairs,

$$\{(1,1);(2,3)\};\ \{(1,2);(1,3)\};\ \{(1,3);(1,2)\}.$$

Then, in Eq. (7b), for each set we generate the product $\Sigma_{ij}\cdot\Sigma_{lk}$, where the indices are determined by the pairs in the set, and take the sum of these products to obtain the desired central moment,

$$\mathbb{E}(\hat{X}_1^2\hat{X}_2\hat{X}_3) = \Sigma_{11}\Sigma_{23} + \Sigma_{12}\Sigma_{13} + \Sigma_{13}\Sigma_{12}.$$

APPENDIX B: EXAMPLE OF MIXED MOMENT GAMMA CLOSURE

We show how the steps in the calculation of the compact formula in Eq. (17) follow each other on the raw mixed moment $\mathbb{E}(X_1^2X_2)$ of a system consisted only of these variables ($N = 2$). We first introduce the auxiliary independent gamma variates Y_{11}, Y_{12}, Y_{21}, and Y_{22}, where $Y_{12} = Y_{21}$. As one substitutes X_i with the corresponding variables using Eq. (14), the studied moment takes the form

$$\mathbb{E}(X_1^2X_2) = \mathbb{E}\left(\left(\frac{\beta_{11}}{\beta_{11}}Y_{11} + \frac{\beta_{11}}{\beta_{12}}Y_{12}\right)^2\left(\frac{\beta_{22}}{\beta_{21}}Y_{21} + Y_{22}\right)\right).$$

The first step is to expand the product within the expectation

$$\begin{aligned}\mathbb{E}(X_1^2X_2) = \mathbb{E}\Bigg(&\frac{\beta_{22}}{\beta_{21}}Y_{11}^2Y_{21} + 2\frac{\beta_{11}\beta_{22}}{\beta_{12}\beta_{21}}Y_{11}Y_{12}Y_{21}\\ &+ \frac{\beta_{11}^2\beta_{22}}{\beta_{12}^2\beta_{21}}Y_{12}^2Y_{21} + Y_{11}^2Y_{22}\\ &+ 2\frac{\beta_{11}}{\beta_{12}}Y_{11}Y_{12}Y_{22} + \frac{\beta_{11}^2}{\beta_{12}^2}Y_{12}^2Y_{22}\Bigg),\end{aligned}$$

then apply the equality $Y_{12} = Y_{21}$ (and hence $\beta_{12} = \beta_{21}$). As all Ys are independent, we can take the individual expectation of each term in each product, using Eq. (13), so that the expected value of Y_{12}^3 is $(\alpha_{12})_3\beta_{12}^3$, and so on. For example, the third term of the above expression becomes

$$\mathbb{E}\left(\frac{\beta_{11}^2\beta_{22}}{\beta_{12}^2\beta_{21}}Y_{12}^2Y_{21}\right) = \frac{\beta_{11}^2\beta_{22}}{\beta_{12}^3}\cdot(\alpha_{12})_3\beta_{12}^3 = (\alpha_{12})_3\beta_{11}^2\beta_{22}.$$

As can be seen on the above example, the powers individual Ys are on and the coefficients of the terms in the expanded sum determine the α terms and the final form of the raw moment will be a sum of products of αs, summarised in the combined sum and product of Eq. (17),

$$\begin{aligned}\mathbb{E}(X_1^2X_2) = \Big((&\alpha_{11})_2\alpha_{12} + \alpha_{11}(\alpha_{12})_2 + 2(\alpha_{12})_3\\ &+ (\alpha_{11})_2\alpha_{22} + 2\alpha_{11}\alpha_{12}\alpha_{22} + (\alpha_{12})_2\alpha_{22}\Big)\beta_{11}^2\beta_{22}.\end{aligned}$$

APPENDIX C: EXAMPLES OF MOMENT CLOSURE FORMULAS

We illustrate the closure formulas derived in Section II on an example taken from the third order closure schemes, i.e., when fourth order moments are approximated based on the assumed distribution. Let us consider a system with at least two variables and investigate how $\mathbb{E}(\hat{X}_1^2\hat{X}_2^2)$ can be expressed in terms of lower order moments. This expression does not

depend on any other variables; hence, regardless of the total number of species in the system, the same formula can be substituted into Eq. (5) and applied in all considered systems.

Truncation (zero-closure) means setting all higher order moments to zero, while in all univariate closures mixed moments become zero similarly; therefore, in both cases,

$$\mathbb{E}(\hat{X}_1^2\hat{X}_2^2) = 0.$$

Assuming a normal distribution and applying the expressions derived in Eqs. (7) and (8), we obtain

$$\mathbb{E}(\hat{X}_1^2\hat{X}_2^2) = 2\sigma_{1,2} + \sigma_1\sigma_2,$$

where

$$\begin{aligned}\sigma_{1,2} &= \mathrm{Cov}(X_1, X_2),\\ \sigma_1 &= \mathrm{Var}(X_1),\\ \sigma_2 &= \mathrm{Var}(X_2).\end{aligned}$$

If, on the other hand, we assume a lognormal distribution and apply our formulas accordingly, the above moment becomes

$$\begin{aligned}\mathbb{E}(\hat{X}_1^2\hat{X}_2^2) = {}& \mu_1^2\mu_2^2(\sigma_1 + \sigma_2 - 1) + 4\mu_1\mu_2(\sigma_{1,2} + \mu_1\mu_2)\\ & - 2\left(\frac{\sigma_1}{\mu_1^2} + \frac{\sigma_2}{\mu_2^2} + 2\right)(\sigma_{1,2} + \mu_1\mu_2)^2\\ & + \frac{(\sigma_{1,2} + \mu_1\mu_2)^4(\mu_1^2 + \sigma_1)(\mu_2^2 + \sigma_2)}{\mu_1^4\mu_2^4},\end{aligned}$$

where $\sigma_1, \sigma_2, \sigma_{1,2}$ are as before and μ_1, μ_2 denote the mean of X_1 and X_2.

Finally, using the above notation for means, variances, and covariance, in gamma closure, the expected value of $\hat{X}_1^2\hat{X}_2^2$ is substituted with

$$\mathbb{E}(\hat{X}_1^2\hat{X}_2^2) = 2\sigma_{1,2}^2 + \sigma_1\sigma_2 + \frac{6\sigma_1\sigma_2\sigma_{1,2}}{\mu_1\mu_2}.$$

[1]D. J. Wilkinson, "Stochastic modelling for quantitative description of heterogeneous biological systems," Nat. Rev. Genet. **10**(2), 122–133 (2009).
[2]D. T. Gillespie, "Exact stochastic simulation of coupled chemical reactions," J. Phys. Chem. **81**(25), 2340–2361 (1977).
[3]T. R. Kiehl, R. M. Mattheyses, and M. K. Simmons, "Hybrid simulation of cellular behavior," Bioinformatics **20**(3), 316–322 (2004).
[4]J. Puchalka and A. M. Kierzek, "Bridging the gap between stochastic and deterministic regimes in the kinetic simulations of the biochemical reaction networks," Biophys. J. **86**(3), 1357–1372 (2004).
[5]M. Komorowski, J. Mikisz, and M. P. H. Stumpf, "Decomposing noise in biochemical signaling systems highlights the role of protein degradation," Biophys. J. **104**(8), 1783–1793 (2013).
[6]M. Komorowski, B. Finkenstadt, C. Harper, and D. Rand, "Bayesian inference of biochemical kinetic parameters using the linear noise approximation," BMC Bioinf. **10**(1), 343 (2009).
[7]E. W. J. Wallace, D. T. Gillespie, K. R. Sanft, and L. R. Petzold, "Linear noise approximation is valid over limited times for any chemical system that is sufficiently large," IET Syst. Biol. **6**(4), 102–115 (2012).
[8]C. A. Gómez-Uribe and G. C. Verghese, "Mass fluctuation kinetics: Capturing stochastic effects in systems of chemical reactions through coupled mean–variance computations," J. Chem. Phys. **126**(2), 024109 (2007).
[9]M. Ullah and O. Wolkenhauer, "Investigating the two-moment characterisation of subcellular biochemical networks," J. Theor. Biol. **260**(3), 340–352 (2009).
[10]A. Ale, P. Kirk, and M. P. H. Stumpf, "A general moment expansion method for stochastic kinetic models," J. Chem. Phys. **138**(17), 174101 (2013).
[11]L. A. Goodman, "Population growth of the sexes," in *Mathematical Demography*, Biomathematics (Springer, Berlin, Heidelberg, 1977), Vol. 6, pp. 469–480.
[12]P. Whittle, "On the use of the normal approximation in the treatment of stochastic processes," J. R. Stat. Soc., Ser. B **19**(2), 268–281 (1957).
[13]T. I. Matis and I. G. Guardiola, "Achieving moment closure through cumulant neglect," Math. J. **12**(1), 2:1–2:18 (2010).
[14]C. S. Gillespie, "Moment-closure approximations for mass-action models," IET Syst. Biol. **3**(1), 52–58 (2009).
[15]P. Milner, C. S. Gillespie, and D. J. Wilkinson, "Moment closure approximations for stochastic kinetic models with rational rate laws," Math. Biosci. **231**(2), 99–104 (2011).
[16]I. Krishnarajah, A. Cook, G. Marion, and G. Gibson, "Novel moment closure approximations in stochastic epidemics," Bull. Math. Biol. **67**(4), 855–873 (2005).
[17]I. Krishnarajah, G. Marion, and G. Gibson, "Novel bivariate moment-closure approximations," Math. Biosci. **208**(2), 621–643 (2007).
[18]C. H. Lee, K.-H. Kim, and P. Kim, "A moment closure method for stochastic reaction networks," J. Chem. Phys. **130**(13), 134107 (2009).
[19]M. J. Keeling, "Metapopulation moments: Coupling, stochasticity and persistence," J. Anim. Ecol. **69**(5), 725–736 (2000).
[20]A. Singh and J. P. Hespanha, "Lognormal moment closures for biochemical reactions," in *Proceedings of the 45th IEEE Conference on Decision and Control* (IEEE, 2006), pp. 2063–2068.
[21]A. Singh and J. P. Hespanha, "A derivative matching approach to moment closure for the stochastic logistic model," Bull. Math. Biol. **69**(6), 1909–1925 (2007).
[22]A. Singh and J. P. Hespanha, "Approximate moment dynamics for chemically reacting systems," IEEE Trans. Autom. Control **56**(2), 414–418 (2011).
[23]K. Hausken and J. F. Moxnes, "A closure approximation technique for epidemic models," Math. Comput. Modell. Dyn. Syst. **16**, 555–574 (2010).
[24]P. Smadbeck and Y. N. Kaznessis, "A closure scheme for chemical master equations," Proc. Natl. Acad. Sci. U. S. A. **110**(35), 14261–14265 (2013).
[25]R. Grima, "A study of the accuracy of moment-closure approximations for stochastic chemical kinetics," J. Chem. Phys. **136**(15), 154105 (2012).
[26]J. P. Hespanha, `StochDynTools` — a MATLAB toolbox to compute moment dynamics for stochastic networks of bio-chemical reactions. Available at http://www.ece.ucsb.edu/~hespanha, 2007.
[27]P. Milner, C. S. Gillespie, and D. J. Wilkinson, "Moment closure based parameter inference of stochastic kinetic models," Stat. Comput. **23**(2), 287–295 (2013).
[28]T. Toni, D. Welch, N. Strelkowa, A. Ipsen, and M. P. H. Stumpf, "Approximate Bayesian computation scheme for parameter inference and model selection in dynamical systems," J. R. Soc., Interface **6**(31), 187–202 (2009).
[29]I. G. Johnston, "Efficient parametric inference for stochastic biological systems with measured variability," Stat. Appl. Genet. Mol. Biol. **13**(3), 379–390 (2014).
[30]J. Owen, D. J. Wilkinson, and C. S. Gillespie, "Likelihood free inference for Markov processes: A comparison," Stat. Appl. Genet. Mol. Biol. **14**(2), 189–209 (2015).
[31]P. Kirk, T. W. Thorne, and M. P. H. Stumpf, "Model selection in systems and synthetic biology," Curr. Opin. Biotechnol. **24**(4), 767–774 (2013).
[32]P. Fearnhead, D. Prangle, M. P. Cox, P. J. Biggs, and N. P. French, "Semi-automatic selection of summary statistics for ABC model choice," Stat. Appl. Genet. Mol. Biol. **13**(1), 67–82 (2014).
[33]L. R. Mead and N. Papanicolaou, "Maximum-entropy in the problem of moments," J. Math. Phys. **25**, 2404–2417 (1984).
[34]M. Komorowski, M. J. Costa, D. A. Rand, and M. P. H. Stumpf, "Sensitivity, robustness, and identifiability in stochastic chemical kinetics models," Proc. Natl. Acad. Sci. U. S. A. **108**(21), 8645–8650 (2011).
[35]J. Liepe, S. Filippi, M. Komorowski, and M. P. H. Stumpf, "Maximizing the information content of experiments in systems biology," PLoS Comput. Biol. **9**(1), e1002888 (2013).
[36]D. Silk, P. Kirk, C. P. Barnes, T. Toni, and M. P. H. Stumpf, "Model selection in systems biology depends on experimental design," PLoS Comput. Biol. **10**(6), e1003650 (2014).
[37]D. Silk, P. D. W. Kirk, C. P. Barnes, T. Toni, A. Rose, S. Moon, M. J. Dallman, and M. P. H. Stumpf, "Designing attractive models via automated identification of chaotic and oscillatory dynamical regimes," Nat. Commun. **2**, 489 (2011).
[38]C. P. Barnes, D. Silk, X. Sheng, and M. P. H. Stumpf, "Bayesian design of synthetic biological systems," Proc. Natl. Acad. Sci. U. S. A. **108**(37), 15190–15195 (2011).
[39]P. Smadbeck and Y. N. Kaznessis, "Chemical master equation closure for computer-aided synthetic biology," Methods Mol. Biol. (Clifton, N.J.) **1244**, 179–191 (2015).

[40]Y. Taniguchi, P. J. Choi, G.-W. Li, H. Chen, M. Babu, J. Hearn, A. Emili, and X. S. Xie, "Quantifying e. coli proteome and transcriptome with single-molecule sensitivity in single cells," Science **329**(5991), 533–538 (2010).
[41]N. Friedman, L. Cai, and X. S. Xie, "Linking stochastic dynamics to population distribution: An analytical framework of gene expression," Phys. Rev. Lett. **97**(16), 168302 (2006).
[42]L. Isserlis, "On a formula for the product-moment coefficient of any order of a normal frequency distribution in any number of variables," Biometrika **12**(1-2), 134–139 (1918).
[43]A. Stefanek, M. C. Guenther, and J. T. Bradley, "Normal and inhomogeneous moment closures for stochastic process algebras," in 10th Workshop on Process Algebra and Stochastically Timed Activities 2011 (PASTA'11), September 2011.
[44]E. L. Crow and K. Shimizu, *Lognormal Distributions: Theory and Applications*, Statistics, Textbooks and Monographs (M. Dekker, New York, 1988), Vol. 88.
[45]A. M. Mathai and P. G. Moschopoulos, "On a multivariate-gamma," J. Multivar. Anal. **39**, 135–153 (1991).
[46]E. Furman, "On a multivariate gamma distribution," Stat. Probab. Lett. **78**, 2353–2360 (2008).
[47]N. Geva-Zatorsky, N. Rosenfeld, S. Itzkovitz, R. Milo, A. Sigal, E. Dekel, T. Yarnitzky, Y. Liron, P. Polak, G. Lahav, and U. Alon, "Oscillations and variability in the p53 system," Mol. Syst. Biol. **2**, 2006.0033 (2006).
[48]H. A. Harrington, M. Komorowski, M. Beguerisse-Díaz, G. M. Ratto, and M. P. H. Stumpf, "Mathematical modeling reveals the functional implications of the different nuclear shuttling rates of erk1 and erk2," Phys. Biol. **9**(3), 036001 (2012).
[49]P. Kügler, "Moment fitting for parameter inference in repeatedly and partially observed stochastic biological models," PLoS One **7**(8), e43001 (2012).
[50]B. Munsky, B. Trinh, and M. Khammash, "Listening to the noise: Random fluctuations reveal gene network parameters," Mol. Syst. Biol. **5**(1), 318 (2009).
[51]E. G. Gilbert, "The decoupling of multivariable systems by state feedback," SIAM J. Control **7**(1), 50–63 (1969).
[52]J. F. Apgar, J. E. Toettcher, D. Endy, F. M. White, and B. Tidor, "Stimulus design for model selection and validation in cell signaling," PLoS Comput. Biol. **4**(2), e30 (2008).
[53]V. John, I. Angelov, A. A. Oencuel, and D. Thevenin, "Techniques for the reconstruction of a distribution from a finite number of its moments," Chem. Eng. Sci. **62**, 2890–2904 (2007).
[54]J. Hasenauer, V. Wolf, A. Kazeroonian, and F. J. Theis, "Method of conditional moments (mcm) for the chemical master equation : A unified framework for the method of moments and hybrid stochastic-deterministic models," J. Math. Biol. **69**(3), 687–735 (2013).
[55]C. H. Waddington, *The Strategy of the Genes: A Discussion of Some Aspects of Theoretical Biology* (Allen & Unwin, London, 1957).
[56]C. Li and J. Wang, "Quantifying Waddington landscapes and paths of non-adiabatic cell fate decisions for differentiation, reprogramming and trans-differentiation," J. R. Soc., Interface **10**(89), 20130787 (2013).
[57]P. Smadbeck and Y. N. Kaznessis, "On a theory of stability for nonlinear stochastic chemical reaction networks," J. Chem. Phys. **142**(18), 184101 (2015).
[58]D. Schnoerr, G. Sanguinetti, and R. Grima, "Validity conditions for moment closure approximations in stochastic chemical kinetics," J. Chem. Phys. **141**(8), 084103 (2014).
[59]K. Erguler and M. P. H. Stumpf, "Statistical interpretation of the interplay between noise and chaos in the stochastic logistic map," Math. Biosci. **216**, 90–99 (2008).
[60]A. Golightly and D. J. Wilkinson, "Bayesian inference for stochastic kinetic models using a diffusion approximation," Biometrics **61**(3), 781–788 (2005).
[61]A. Golightly and D. J. Wilkinson, "Bayesian parameter inference for stochastic biochemical network models using particle Markov chain Monte Carlo," Interface Focus **1**(6), 807–820 (2011).
[62]A Python package that implements the method described here is available from our web page http://www.theosysbio.bio.ic.ac.uk.