Parameter estimation and bifurcation analysis of stochastic models of gene regulatory networks: tensor-structured methods

Shuohao Liao*, Tomáš Vejchodský†, Radek Erban*

Abstract. Stochastic modelling provides an indispensable tool for understanding how random events at the molecular level influence cellular functions. In practice, the common challenge is to calibrate a large number of model parameters against the experimental data. A related problem is to efficiently study how the behaviour of a stochastic model depends on its parameters, i.e. whether a change in model parameters can lead to a significant qualitative change in model behaviour (bifurcation). In this paper, tensor-structured parametric analysis (TPA) is presented. It is based on recently proposed low-parametric tensor-structured representations of classical matrices and vectors. This approach enables simultaneous computation of the model properties for all parameter values within a parameter space. This methodology is exemplified to study the parameter estimation, robustness, sensitivity and bifurcation structure in stochastic models of biochemical networks. The TPA has been implemented in Matlab and the codes are available at http://www.stobifan.org.

1 Introduction

Many cellular processes are influenced by stochastic fluctuations at the molecular level which can be modelled by stochastic simulation algorithms for chemical reaction networks [47,61]. Stochastic models can capture both the intrinsic noise (caused by relatively low numbers of biomolecules inside cells) and the extrinsic noise coming from random fluctuations of the external environment [15, 51]. The external environment is noisy, nonspecific, and it can potentially act globally on many parameters. However, its influence can differ from cell to cell [56]. In this paper, we present a tensor-structured parametric analysis (TPA) which can be used to understand how the complex structure of gene regulatory networks exploits both intrinsic and extrinsic sources of noise. We illustrate major application areas of the TPA by studying several biological models with increasing complexity. We leave the discussion of technical computational details of the underlying methods to Appendix A.

We start with the problem of parameter estimation [40] in Section 2. Some biophysical parameters can be directly measured, whereas some others might have no direct biochemical interpretations and need to be inferred from the abundance of measurements over time. Optimal parameters are usually sought as minimisers of a distance between model outputs and experimental measurements. Non-statistical methods define the distance in terms of an objective functional that measures goodness how well the predicted measurement fits to the actual measurement [30]. Alternatively, a statistical (or Bayesian) approach is based on the likelihood concept to measure the distances. This might be beneficial as it estimates both parameter values and associated level of uncertainties [22]. However, under the stochastic context, both the non-statistical and statistical approaches yield distance measures that are merely analytically tractable, and have to be approximated by Monte Carlo simulations [8]. Consequently, tuning parameters in high-dimension parameter space tends to be more computationally expensive. Another potential challenge is whether at all unique parameter values

--

*Mathematical Institute, University of Oxford, Radcliffe Observatory Quarter, Woodstock Road, Oxford OX2 6GG, United Kingdom; emails: liao@maths.ox.ac.uk, erban@maths.ox.ac.uk
† Institute of Mathematics, Czech Academy of Sciences, Zitna 25, 115 67 Praha 1, Czech Republic; e-mail: vejchod@math.cas.cz
can be determined from available data. This is known as the problem of identifiability. Inappropriate choice of the distance measure may yield ranges of parameter values with equally good fit, i.e. the parameters being not identifiable [25]. It is therefore desirable to develop computational techniques that enable efficient explorations of the high-dimensional parameter space.

The second application area of the TPA is presented in Section 3 where we investigate sensitivity and bifurcation analysis of stochastic chemical reaction networks. Sensitivity quantifies the dependence of certain quantity of interest on continuous changes in model parameters, whereas bifurcation is defined as a qualitative (topological) transformation in the behaviour of the quantity during continuous change of model parameters. The bifurcation structure of stochastic models has been studied for relatively low-dimensional chemical networks [16, 29]. However, stochastic bifurcation structure has remained unclear for many complex gene regulatory networks, because existing approaches have prohibitive computational cost for high-dimensional reaction systems. There has been an increased interest in computational tools for sensitivity analysis in recent years. These tools can be divided into two categories: gradient-based and gradient-free methods. The gradient-based methods, assess sensitivity by (partial) derivatives of a target function [3, 32, 39, 58]. These methods usually compute finite differences using Monte Carlo simulations or by solving the chemical master equation. The gradient-free methods rely on information theory and measure the sensitivity using Fisher information matrix [34, 48].

The TPA can also be used for robustness analysis as it is shown in Section 4. A system is robust if it maintains its function for a range of magnitudes of both intrinsic and extrinsic noise. Stochastic models are ideal tools for investigating the interplay between the two major sources of noise, because they allows us to directly modulate model parameters according to the extrinsic fluctuations. This has been done for example in [56], where the standard stochastic simulation algorithm has been extended to include time-varying parameters. Unfortunately, extrinsic fluctuations are nonspecific. Thus, to rigorously examine the robustness, many hypotheses have to be tested. Monte Carlo simulations designed for this purpose can be computationally intensive, due to high dimensionality of the problem. Considering different types of extrinsic noise with varying mean, variance or other moments, can yield problems of even higher dimension than for the sensitivity analysis, where the only limitation is the dimensionality of the parameter space.

The high dimensionality is the major impediment for the efficient use of existing computational tools for parameter estimation and bifurcation analysis. If one uses classical matrix-based computational methods, then memory requirements and the computational complexity grow exponentially with the number of dimensions which is known as the “curse of dimensionality” [45]. In this paper, we show how the existing parametric analysis equipped with tensor data structures can be used to solve both low- and high-dimensional problems. We refer to this approach (TPA) as tensor-structured, because it makes use of tensors, multidimensional arrays of real numbers. For tensors, we can define standard algebraic operations similar to standard matrix operations such that the resulting tensor calculus enables efficient computation. The key feature of the TPA is the low parametric representation of tensors. These representations are known as tensor formats. Over the last decade, a number of tensor formats have been proposed. For example the canonical polyadic decomposition [10], the Tucker decomposition [64], the tensor train decomposition [44], and the hierarchical Tucker decomposition [23]. For further details we refer to review article [33]. Tensor-structured algorithms were previously applied to the chemical master equation [12, 31]. Regulation networks with up to six chemical species, including λ-phage [57] and enzymatic futile cycle [53], were studied using the chemical master equation to emphasize the sub-linear growth of complexity with respect to the dimension in tensor-structured computations. Tensors have also been used in biological applications for inferring the connectivity structure of biological networks from large data sets [2, 42]. The numerical frame-
work of the TPA is discussed in more detail in Appendix A. The central concept, low parametric representation of tensors based on the separation of dimensions [7], is the most difficult part to achieve numerically, because the numerical properties such as the conditioning, stability, and accuracy of tensor formats worsen with the dimension. In this paper we concentrate on the solution of the chemical Fokker-Planck equation [20], as a continuous approximation to a Markov jump process that can be exactly described by the chemical master equation [24].

We demonstrate the capabilities of the TPA framework by investigating four model systems: bistable switch in 5-dimensional Schögl model [55], oscillations in 7-dimesional cell cycle model [65], neurons excitability in 6-dimensional FitzHugh-Nagumo system [63], and isomeric transformation in 20-dimensional reaction chain [21]. All presented examples can be formulated as general well-mixed chemically-reacting systems of $N$ distinct chemical species in a reactor of constant volume size $V$ which react through $M$ chemical reactions

$$
\nu_{j,1}X_1 + \cdots + \nu_{j,N}X_N \xrightarrow{k_j} \nu_{j,1}^+X_1 + \cdots + \nu_{j,N}^+X_N, \quad j = 1, 2, \ldots, M, \quad (1)
$$

where $X_1, \ldots, X_N$ denote the chemical species, and $\nu_{j,i}^+$ and $\nu_{j,i}^-$ are the stoichiometric coefficients. The rate of each chemical reaction is characterised by its kinetic rate constant $k_j$. The time evolution of system (1) can be computed using the Gillespie stochastic simulation algorithm [18] or its equivalent formulations [9, 17]. The state of the system is given by the vector $\vec{x} = (x_1, \ldots, x_N)^T$, where the nonnegative integer $x_i$ corresponds to the number of molecules of chemical species $X_i$, $i = 1, 2, \ldots, N$.

We study the stationary distribution $p(\vec{x} | \vec{k})$ of the probability that in a long time the system is in the state $\vec{x}$ given the kinetic rates values $\vec{k} = (k_1, k_2, \ldots, k_M)^T$. A crucial step is to compute $p(\vec{x} | \vec{k})$ simultaneously for different values of kinetic rates $\vec{k}$. This is achieved by directly solving the parametric stationary Fokker-Planck equation (13) in a separated tensor representation. In what follows, we first illustrate significant memory and computational savings of the tensor-based approach in comparison with the standard matrix-vector based approach. The following sections then illustrates how to use the TPA for parameter estimation, robustness analysis, and sensitivity and bifurcation analysis.

1.1 Computational performance

The TPA maintains affordable computational and memory requirements for all four problems we consider. The model parameters are taken from the literature (further information about the reactions, parameter values and discretizations of the model systems is given in Appendix B). Table 1 shows the memory requirements and computational times to compute the stationary distribution $p(\vec{x} | \vec{k})$ using the separated tensor-representation of data. The sum of the number of chemical species $N$ and the number of varied kinetic rate parameters $M_0$ gives the total dimension of the problem. Computational times are of orders of hours up to several days. In Table 1 we show CPU times for assembling the parametric Fokker-Planck operator in seconds (Ta) and the total computational time of solving the stationary problem in minutes (Ts) for each system.

Memory requirements to store the parametric stationary distribution in the separated tensor-structured data format, Mem(TPA), are of orders $10^5$–$10^7$ bytes. Using the classical matrix-based approach, we would need at least $10^6$ times more memory just to store the stationary distribution. Table 1 shows the estimated minimal memory requirements of methods which use the standard vector data formats. They include the finite state projection method, Mem(FSP), for the parametric chemical master equation, and the finite difference method, Mem(FDM), for the parametric Fokker-Planck equation. These memory requirements have been estimated as products of numbers of discrete
Table 1: Comparison of matrix-based and tensor-based parametric analysis.

| Reaction system     | Dimensionality | Matrix-based methods | Tensor-based method |
|---------------------|----------------|----------------------|---------------------|
|                     | $N$ $M_0$ $N + M_0$ | Mem(FSP) Mem(FDM) | Mem(TPA) $Ta$ $Ts$ |
| Schlögl 1           | 1 4 5           | $2.68 \times 10^{13}$ $2.74 \times 10^{11}$ | $2.07 \times 10^{5}$ $0.5$ $30$ |
| Cell cycle          | 6 1 7           | $6.68 \times 10^{17}$ $7.04 \times 10^{13}$ | $1.00 \times 10^{7}$ $3.3$ $6433$ |
| FitzHugh-Nagumo 2   | 2 4 6           | $6.38 \times 10^{14}$ $1.75 \times 10^{13}$ | $4.02 \times 10^{5}$ $1.7$ $37$ |
| Chain reaction 20   | 20 0 20         | $1.20 \times 10^{44}$ $1.53 \times 10^{54}$ | $7.28 \times 10^{5}$ $1.3$ $283$ |

states times the product of numbers of parameter values. They equal to numbers of entries of the corresponding tensors (17). In Table 1 we see that the memory requirements of standard matrix-vector methods vary in ranges $10^{13}–10^{44}$ and $10^{11}–10^{54}$, respectively. These values are far beyond the limits of the available hardware.

## 2 Parameter estimation

Parameter estimation of stochastic chemical reaction networks is usually addressed by determining kinetic rate values that minimize some measure of the error between some measurable quantities and the corresponding model predictions. Ideally, if experimental sample distribution $\hat{p}(\vec{x})$ is available, a natural measurement of the error would be a distance function between this sample distribution and the model outcome, $p(\vec{x} | \vec{k})$, i.e.,

$$J(\vec{k}) = \int_{\Omega^x} \left( \hat{p}(\vec{x}) - p(\vec{x} | \vec{k}) \right)^2 \, d\vec{x},$$

(2)

where $\Omega^x$ stands for the state space. In practice, we rarely infer stochastic networks directly via minimising (2), because compressing real biological time series data into probability distribution $\hat{p}(\vec{x})$ is computationally intensive. Although multi-way data compression methods might be useful for this purpose, they are normally used for no more than three dimensions [35]. It is therefore advantageous to make use of other statistical information that can be easily computed from the experimental time series.

A straightforward option is to use the method of moments [26], where optimal parameters are identified by equating the moments of $p(\vec{x} | \vec{k})$ to the corresponding empirical moments. The distance function (2) is then formulated as [36]

$$J(\vec{k}) = \sum_{i \in \mathcal{L}} \beta_i \left( \frac{\hat{\mu}_i - \mu_i(\vec{k})}{\mu_i} \right)^2,$$

(3)

where $\hat{\mu}_i$ are the $i$-th order empirical moments, $\mu_i(\vec{k})$ are the $i$-th order moments of $p(\vec{x} | \vec{k})$ and $\mathcal{L}$ is a finite index set. The weights, $\beta_i$, can be chosen by a modeller to attribute different relative importances to moments. Under the TPA framework, the tensor structure enables to evaluate the distance function $J(\vec{k})$ in (3) simultaneously for all possible combinations of parameter values within the parameter space, but only with linear complexity with respect to the number of parameters (see Appendix A). Thus estimating parameters using the method of moments is particularly suitable for the TPA.
2.1 Identifiability

We begin with the identifiability analysis of the Schlögl chemical system [55] which consists of one chemical species and four mass-action chemical reactions:

$$3X \xrightleftharpoons{\kappa_1}{\kappa_2} 2X, \quad \emptyset \xrightleftharpoons{\kappa_3}{\kappa_4} X.$$  \hfill (4)

Given parameter values in Table 3, the Gillespie stochastic simulation algorithm is used to compute time series data (up to time $10^7$) which we use as pseudoexperimental data for testing the TPA. A short segment of this simulation is plotted in Figure 1(a).

If we consider steady-state data, then the Schlögl model is only partially identifiable. This means that it is theoretically not possible to find all the true parameter values by minimising the distance function (3) from any steady-state data. The reason is that multiplication of all parameters, $k_j$ for $j = 1, \ldots, 4$, by a constant yields only a scaling of the temporal dynamics whereas the stationary distribution stays unchanged, see Eq. (15). Therefore we will only be identifying the values of three parameters $k_1/k_4$, $k_2/k_4$ and $k_3/k_4$.

We first divide the model parameters into two pairs, $k_1$-$k_3$ and $k_2$-$k_4$, and evaluate the distance function $J(\vec{k})$ in (3) for one pair at a time with the other pair being fixed at true values given in Table 3. In both cases, we consider the distance function with the first three moments, i.e.

$$J(\vec{k}) = \sum_{i=1}^{3} \beta_i \left( \frac{\hat{\mu}_i - \mu_i(\vec{k})}{\hat{\mu}_i} \right)^2.$$  \hfill (5)

The weights $\beta_i$ and the empirical values of moments $\hat{\mu}_i$ are provided in Appendix B, Table 4. The weights are chosen in such a way that the contributions of the different orders of moments in (5) are of similar magnitude within the parameter space. The distance function (5) possesses a well distinguishable global minimum at the true values and, thus, the model is identifiable in both cases. This is illustrated in Figure 1, where the colour scale indicates the values of the distance function (5).

Both pairs of kinetic parameters can be inferred from the pseudo-experimental data, but not in the deterministic scenario. As one might expect in the deterministic rate equations, the mean concentration is the only quantity being captured, and the distance function (3) only measures the discrepancy between the sample and model mean. As a result, the corresponding distance function $J(\vec{k})$ (indicated by blue contour lines in Figure 1, panels (b) and (c)) attains its minimal values on a curve in the 2D parameter space. This can be explained by the collinearity between the model parameters (see equation (15)). For collinear relations, we need at least $M$ quantities to estimate the value of $M$ parameters. Parameters of stochastic models are better identifiable than the deterministic rate equations, because they can make use of the second and higher moments of the stationary distribution.

We also demonstrate the differences between the deterministic and stochastic models in their response to parameter perturbations. We evaluate the distance function $J(\vec{k})$ in (5) for $\beta_2 = \beta_3 = 0$, i.e. the stochastic model only uses the value of mean in the TPA. This distance function is visualized in Figure 1, panels (b) and (c), using green contour lines. It again represents a non-identifiable situation, where the minimum of the distance functional is attained on a straight line. The difference between the contour lines of the deterministic (blue) and stochastic (green) model indicates the fundamental difference in the geometry of the problem of parameter estimation for the two types of models. Such a difference explains the fact that deterministic sensible parameter values are likely
to lead a poor behaviour of the stochastic model, both in terms of average behaviour and the noise level [67]. This agrees with [34], where the discrepancies between the stochastic and deterministic models are computed using the linear noise approximation. The advantage of the TPA is that it estimates stationary distributions for all parameter values in one computation. This enables a direct comparison of statistical moments in (3).

2.2 Admissible parameter values

When the model is identifiable, it seems reasonable to expect that the unique minimizer of $J(\vec{k})$ is a ‘good’ estimate of the true values. Practically, measurement noise might shift the minimizer away from the true values. Therefore, we search for a set of admissible parameter values $\vec{k}$, satisfying $J(\vec{k}) < J_{TOL}$ for a prescribed tolerance level $J_{TOL}$. Having the values of $J(\vec{k})$ for all $\vec{k}$ within a (subset of) parameter space represented in the tensor structured data, we can search for the admissible parameter values by any general optimization algorithms. Suitable strategy seems to be a greedy search [38] combined with the numerical continuation [1].

We now identify the admissible parameter values for the Schlögl model with a good fit to the pseudoexperimental data, i.e., those yielding the distance function to be below $J_{TOL} = 0.25\%$. The results are visualised in Figure 2, panels (a)–(d), where each of the panels corresponds to the case when we fix one parameter at its true value and estimate the other three. The considered ranges of parameters are ±3% variations around the true values (Table 5). The presented experiments exhibit substantial differences in the number of admissible parameter values under the same tolerance level. Namely, the number of admissible parameter values in Figure 2(a) is approximately 5.3 times smaller than in Figure 2(b). Similarly, it is 2.2 times and 7.7 times less comparing to panels (c) and (d), respectively. This indicates that the model system (4) is more sensitive in response to variations in $k_1$. Parameter $k_4$ in Figure 2(b) features “sloppiness” in parameter sensitivity [25], since tuning $k_4$ within most part of its inspected range would yield quantitatively similar steady state distributions (moments would match within the error tolerance), provided $k_2$ is fixed at its true value. Sloppy parameters are commonly encountered in inferring biochemical models [11]. They are characterised
Figure 2: (a–d) Circular representation [66] of estimated parameter combinations. Each spoke represents the corresponding parameter range listed in Table 5. The true parameter values are specified by the dashed line. Each triangle (or polygon in general) of a fixed colour corresponds to one admissible parameter set with tolerance \( J_{\text{TOL}} = 0.25\% \). Each panel shows the situation with one parameter fixed at its true value. (e) Set of parameters yielding splitting probability that the number of molecules \( X \) is below 230 to be 47.61\% with tolerance \( J_{\text{TOL}} = 5\% \). The value of \( k_4 \) is fixed at its true value.

by the fact that the vicinity of the minimum of the distance function \( J(\vec{k}) \) along the coordinate \( k_4 \) is mostly flat, and one requires evaluating \( J(\vec{k}) \) with very accuracy to achieve an optimum. We have shown that under the tensor-structured framework the distance function can be accurately and simultaneously evaluated for ranges of parameter values. Therefore, the number of admissible parameter values can be reduced by decreasing the tolerance \( J_{\text{TOL}} \).

2.3 Parameter estimation using splitting probability

Measuring statistical moments, especially higher order moments, in a reliable, unbiased, and efficient way is still a challenging task for biological experiments [4]. In this section we illustrate that the input of the TPA does not have to be given in terms of moments. We use the time series data (Figure 1(a)) to estimate the probability that the number of molecules of \( X \) is below 230. This measurement (estimated value 47.61 \%) does not contain sufficient amount of information to infer the values of system parameters. However, the TPA can use this information (see Appendix A.4) to estimate admissible combinations of model parameters. They lie on a surface shown in Figure 2(e). The value of \( k_4 \) is fixed at its true value in this example.
Figure 3: (a) Schematic description of the cyclin-cdc2 interactions. (b) Sensitivity indicators $S_6(\Theta)$ calculated by (6) for 64 equidistant nodes within the range $[0.25, 0.4]$ of parameter $k_6$ (see Table 7). Three observables are considered: stationary distribution (blue), the probability that there is more than 400 molecules of active MPF (red) and the average number of active MPF (orange). The deterministic and stochastic bifurcation points are indicated by dot-dashed line ($k_6 = 0.2694$) and dashed line ($k_6 = 0.3032$), respectively. Stationary distributions of cdc2-cyclin-p (M) and p-cdc2-cyclin-p (pM) are plotted in (c) at the deterministic bifurcation point, and in (d) at the stochastic bifurcation point.

3 Sensitivity and bifurcation analysis

We now exemplify the tensor-structured sensitivity and bifurcation analysis using a model of fission yeast cell cycle control developed by Tyson [65]. The model describes cyclin-cdc2 interactions as it is schematically shown in Figure 3(a)). Free cyclin molecules combine rapidly with phosphorylated cdc2, to form the dimer MPF (cdc2-cyclin-p), which is immediately inactivated by phosphorylation process. The inactive MPF (p-cdc2-cyclin-p) can be converted to active MPF by autocatalytic dephosphorylation. The active MPF in excess breaks down into cdc2 molecules and phosphorylated cyclin. The phosphorated cyclin is later subject to proteolysis, and cdc2 will get phosphorylated to repeat the cycle. The list of reactions and the corresponding rate values are given in Table 6. The reaction rate $k_6$ corresponds to the breakdown of the active MPF complex. A deterministic model has suggested that, when $k_6$ is at its low values, the system displays a stable steady state, referring to the metaphase arrest of unfertilized eggs [28], while when $k_6$ increases, the system is driven into rapid cell cycling exhibiting autonomous oscillations [65]. We will study how the sensitivity and bifurcation structures of the stochastic cell cycle model depend on parameter $k_6$ close to its deterministic bifurcation point.
3.1 Sensitivity analysis

The sensitivity analysis is conducted to quantify the influence of a single parameter on the system behaviour. An observable is sensitive to the value of the parameter, if its small change yields a significant change of the observable. Sensitivity analysis is related to the robustness analysis (see Section 4), where the simultaneous change of all parameters is considered. The sensitivity indicator for an observable quantity $\Theta$ with respect to the parameter $k_j$ is defined as [54]

$$S_j(\Theta) = \frac{d\Theta(k_j)}{dk_j} \frac{k_j}{\Theta(k_j)},$$

which is often approximated by

$$S_j(\Theta) \approx \frac{\Theta(k_j + \Delta k_j) - \Theta(k_j)}{\Delta k_j} \frac{k_j}{\Theta(k_j)}.$$  \hspace{1cm} (6)

where $\Delta k_j$ is a small change in the parameter value. The TPA enables to compute the finite difference (6) with much less computational costs, compared to Monte Carlo simulations, because the steady state probability is simultaneously computed for a range of parameter values (see Appendix A.4).

In this section, we evaluate the sensitivity indicators with respect to the parameter $k_6$ of the cell cycle model for the following three observables: mean concentration, “oscillation amplitude” and steady state distribution of active MPF. The observable $\Theta$ is for the case of “oscillation amplitude” defined as the probability that the molecular population of the active MPF exceeds 400. The steady state distribution is not a scalar quantity and, thus, formula (6) does not directly apply. Instead, we compare the change of the steady state distribution in the 6-dimensional state space, and measure it in $L^2$-norm:

$$S_j(\Theta) \approx \frac{\parallel \Theta(k_j + \Delta k_j) - \Theta(k_j) \parallel_{L^2}}{\Delta k_j} \frac{k_j}{\parallel \Theta(k_j) \parallel_{L^2}}.$$ \hspace{1cm} (7)

In Figure 3(b), we show that, within the considered range of values of $k_6$, the sensitivity in the probability distribution (blue curve) dominates in magnitude over the sensitivity indicators for the oscillation amplitude (red curve) and the mean MPF (orange). This agrees with the intuition, because the probability distribution contains a global information about the system and any sensitivity of the system is reflected by a change of the probability distribution, and consequently recorded by the $L^2$ norm indicator (7). The derived quantities, like the oscillation amplitude and the mean value, are only based on a limited information. Thus, it is possible that a change of certain aspects of the system does not influence them. As a result we may observe small sensitivity indicators.

3.2 Bifurcation analysis

According to [65], the deterministic cell cycle model has a bifurcation point $k_6 = 0.2694$, where a limit cycle is born from a steady state. Stochastically, we define bifurcations as qualitative (topological) transformations of the stationary distribution. For the cell cycle model, the stochastic bifurcation occurs at $k_6 = 0.3032$, from which point the steady state probability density transits from a unimodal shape (single maximum, see Figure 3(c) or Figure 4(a)) to a distribution with a doughnut-shaped region of high probability (see Figure 3(d) or Figure 4(d)). In particular, the stochastic bifurcation appears for higher values of $k_6$ than the deterministic bifurcation. We can infer the difference between the deterministic and stochastic bifurcation from the sensitivity analysis. If the value of $k_6$ approaches the stochastic bifurcation point from the left (see Figure 3(b)), the sensitivity indicator for the stationary distribution increases. Once the system passes through the stochastic bifurcation point,
Figure 4: (a–d) Marginal stationary distributions of the phosphorylated cyclin (YF), the inactive MPF (pM) and the active MPF (M), see Figure 3(a). (e–h) Marginal stationary distributions of the cyclin (Y), the phosphorylated cdc2 (CP) and the inactive MPF (pM). Each column corresponds to the same value of the bifurcation parameter $k_6$, which from the left to the right are 0.25, 0.3, 0.35 and 0.4, respectively. All figures show $\log_2 p$ for better visualisation.

the sensitivity decreases. This behaviour indicates that the steady state distribution undergoes more significant changes when the parameter value is close to the stochastic bifurcation point.

Using the computed tensor-structured parametric probability distribution, we visualise the stochastic bifurcation structure of the cell cycle model in Figure 4. As the bifurcation parameter $k_6$ increases, the expected oscillation tube is formed and amplified in the marginalised YF-pM-M state space (Figure 4, panels (a)–(d)). In panels (e)–(h) of Figure 4, the marginal distribution in the Y-CP-pM subspace is plotted. We observe that this observable changes from unimodal (Figure 4(e)) to bimodal distribution (Figure 4(f)). Some cell cycle models in the literature have been proposed to show the oscillatory mechanism [65], while others have two stable steady states [50, 69]. We see that, by simply considering different observables of one stochastic cell cycle model (Figure 3(a)), the stochastic model can appear to have both oscillations and bimodality.

4 Analysis of robustness

Gene regulatory networks are subject to the extrinsic noise which is manifested by fluctuations of parameter values [49]. This extrinsic noise originates from interactions of the modelled system with other stochastic processes in the cell or its surrounding environment. We can naturally include extrinsic fluctuations under the tensor-structured framework. For a general chemically-reacting system (1), we consider the copy numbers $X_1, \ldots, X_N$ as intrinsic variables and reaction rates $k_1, \ldots, k_M$ as extrinsic variables. Total stochasticity is quantified by the stationary distribution of the intrinsic variables, $p(\vec{x})$. We assume that the invariant probability density of extrinsic variables, $q(\vec{k})$, does not depend on the values of intrinsic variables $\vec{x}$. Then the law of total probability implies that the
stationary probability distribution of intrinsic variables is given by

\[ p(\bar{x}) = \int_{\Omega^k} p(\bar{x} | \bar{k}) q(\bar{k}) \, d\bar{k}, \]  

where \( \Omega^k \) is the parameter space and \( p(\bar{x} | \bar{k}) \) represents the invariant density of intrinsic variables conditioned on constant values of kinetic parameters, see the definition below Eq. (1). If distributions \( q(\bar{k}) \) of extrinsic variables can be determined from high quality experimental data then the stationary density can be computed directly by (9). If not, the TPA framework enables to test the behaviour of genetic regulatory networks for different hypothesis about the distribution of the extrinsic variables. The advantage of the TPA is that it efficiently computes the high-dimensional integrals in (9), see Appendix A.4.
4.1 Extrinsic noise in FitzHugh-Nagumo model

We consider the effect of extrinsic fluctuations on an activator-inhibitor oscillator with simple negative feedback loop: the FitzHugh-Nagumo neuron model which is presented in Figure 5(a). Self-autocatalytic positive feedback loop activates the $X_1$ molecules, which are further triggered by the external signal. The species $X_2$ is enhanced by the feedforward connection and it acts as an inhibitor that turns off the signalling.

In our computational examples, we assume that $q(\vec{k}) = q_1(k_1) q_2(k_1) \ldots q_M(k_M)$, i.e. the invariant distributions of rate constants $k_1, k_2, \ldots, k_M$ are independent. Then (8) reads as follows

$$p(\vec{x}) = \int_{\Omega^k} p(\vec{x} | \vec{k}) q_1(k_1) \cdots q_M(k_M) d\vec{k}. \quad (9)$$

Extrinsic variability in the FitzHugh-Nagumo system is studied in four prototypical cases of $q_i$, $i = 1, 2, \ldots, M$: (i) Dirac delta, (ii) normal, (iii) uniform, and (iv) bimodal distributions, as shown Figure 5(b). Since these distributions have zero mean, the extrinsic noise considered is not biased. We can then use this information about extrinsic noise to simulate the stationary probability distribution of intrinsic variables by formula (9).

When the extrinsic noise is omitted, the inhibited and excited states are linked by a volcano-shaped oscillatory probability distribution (Figure 5(c)). At the inhibited state, $X_1$ molecules first get activated from the positive feedback loop, and then excite $X_2$ molecules by feedforward control (Figure 5(a)). The delay between the excitability of the two molecular species gives rise to the convex path switching from inhibited state to excited state (Figure 5(c)). The delay in the response of $X_2$ molecules is also applied to explain the convexity of the way back. However, if the Gaussian noise is introduced to the extrinsic variables, then the convexity becomes less visible in the resulting marginal distribution (Figure 5(d)). This suggests that, once $X_1$ molecules get excited or inhibited, $X_2$ molecules require less time to respond.

Genetic networks with stronger negative feedback regulation gain higher potential to reduce the stochasticity. This argument has been both theoretically analysed [60,62], and experimentally tested for a plasmid-borne system [6]. We have shown that the extrinsic noise reduces the delay caused by the feedback loop (Figure 5(d)). If we further increase the variability of the extrinsic noise, then the delay caused by the feedback loop is further reduced. (Figure 5(e)). In the case of the bimodal extrinsic fluctuation, the path linking the inhibited and excited states even shrinks into an almost straight line (Figure 5(f)). This means that, for the same level of the inhibitor $X_2$, the number of the activator $X_1$ is lower, i.e., the increasing variability of extrinsic noise results in a stronger feedback regulation in the long time scale. Consequently, the stochasticity of the constitutive system with enhanced negative feedback gets attenuated, as the majority of steady state density accumulates around the two states (Figure 5(f)). Therefore, the presented robustness analysis shows that the performance of stochastic gene regulatory networks with negative feedback regulation can substantially benefit from the extrinsic noise.

5 A chemical reaction system in 20 dimensions

To demonstrate the potential of the TPA to explore high dimensional gene regulatory networks, we consider a reaction chain of 20 molecular species interacting through 40 reversible isomerization reactions:

$$\emptyset \overset{k_0}{\rightarrow} X_1 \overset{k_1}{\leftrightarrow} X_2 \overset{k_2}{\leftrightarrow} X_3 \overset{k_3}{\leftrightarrow} \cdots \overset{k_{18}}{\leftrightarrow} X_{19} \overset{k_{19}}{\leftrightarrow} X_{20} \overset{k_{20}}{\rightarrow} \emptyset, \quad (10)$$
where \( k_0 = 12 \), \( k_i = 0.2 \) for \( i = 1, \ldots, 20 \), and \( k_{-j} = 0.1 \) for \( j = 1, \ldots, 19 \). In order to solve it, we implement a multilevel approach, where the steady state probability density is first approximated on a coarse grid, and then it is interpolated to a finer grid, where it serves as the initial guess for the iterative solver (see Appendix A.3). The results are plotted in Figure 6. Panel (c) shows the convergence of the total error which is given as \( \| p_{\text{exact}} - p_h \|_{L^2} \), where \( p_h \) denotes the approximation and \( p_{\text{exact}} \) is the exact solution which can be analytically computed for the model (10) as the multidimensional Poisson distribution [29]. This example illustrates the advantage to solve the chemical Fokker-Planck rather than the chemical master equation. Even for the newly proposed tensor-based numerical methods [43, 45], it is generally difficult to solve the tensor-structured master equation due to the non-symmetry and bad conditioning of the operator. Although it is possible to improve the conditioning by shortening the time-step [31], the Fokker-Planck approximation has the flexibility in choosing the grid size. This is a distinctive advantage from the computational point of view, because it enables to control the accuracy and use of acceleration strategies, such as the presented multilevel approach.

### 6 Discussion

The aim of this paper is to introduce an innovative computational framework of the tensor-structured parametric analysis (TPA) for stochastic models of chemical systems. The framework utilises a low-parametric tensor-structured data format to model, store and calculate the results throughout separate computations in individual dimensions. This enables efficient exploration of features in high dimensional spaces. So far, parametric studies of high-dimensional problems have mainly been performed by Monte Carlo methods. The TPA is an alternative to these stochastic methods enabling to acquire information not accessible by other methods.

We have shown that the TPA can be used for parameter estimation, robustness, sensitivity and bifurcation analysis. In parameter estimation, we show that statistical properties (like moments and splitting probability) can be easily derived from the tensor-structured solution. Such flexibility increases the identifiability of model parameters than simply comparing the mean values. At the same time, the TPA speeds up the optimisation process of the parameter estimation as it computes
steady state solution simultaneously for ranges of parameter values, and we do not need to repeatedly simulate the solution for each optimization step.

In robustness analysis, the TPA provides a general strategy to combine the intrinsic and extrinsic sources of noise by simultaneously varying several or all model parameters according to arbitrary distributions. In the sensitivity and bifurcation analysis, we have shown that the stochastic bifurcation induces qualitative changes in both the steady state probability density, as well as in the model sensitivities. We also visualised the stochastic bifurcation structure in the marginalised state space, where the stochastic system exhibits two types of behaviour, oscillations and bistability, at the same time. Thus, the intrinsic noise itself is capable to induce bifurcation structure which might not exist in the deterministic description.

The TPA can be easily adapted to the analysis and design of experiments effectively probing the behaviour of a population of cells. A frequent scenario is that certain reaction rates are strongly correlated to different stages of the cell cycle [52]. Thus, a collection of genetically identical cells, or an organism, might behave differently than any prediction using constant values of reaction rates. A more reasonable approach is to incorporate these uncertainties of reaction rates into the model analysis, as we did in Figure 5. Discrepancies between models with and without extrinsic noise (like among Figure 5, panels (c)–(f)) may also appear between the single-cell behaviour and the collective behaviour of cells.

The TPA can be also used for model selection. The structure of biochemical networks is usually not completely known and models have to be hypothesized based on a limited amount of experimentally measured data. The unknown values of the rate constants in the hypothesized models can be considered as random variables with estimated distributions [36]. The tensor-structured framework then enables to evaluate the marginal distribution (9) for all hypothesized models and to identify the one with the best fit to the experimentally measured data. The distinctive advantage of the TPA lies in its flexibility to incorporate both the intrinsic and extrinsic random nature of biochemical networks.

The performance of the TPA essentially relies on the choice of the particular tensor format, as well as on the efficiency of the numerical solution of tensor-structured linear algebraic systems. Development of suitable tensor formats has been an active field of research during the last decade. In line with the recently proposed methods for the chemical master equation [12, 31], we utilize the tensor train format. While some aspects of the tensor-based computation are heuristic, the approach achieves dramatic storage savings over the traditional matrix-based methods (Table 1). The TPA is efficient because of three strategies: (i) the Fokker-Planck approximation that provides the flexibility in discretising the state space as opposed to the exact description by the chemical master equation; (ii) a new adaptive shifted inverse power method that overcomes the major drawback of the classic procedure in choosing an appropriate shift value (see Appendix A.3); and (iii) the alternating minimal energy method with guaranteed local convergence [13]. To our knowledge, no attempt has been made so far to find the stationary distribution of biochemical networks of comparable complexity to the cell cycle model (Figure 3(a)). We have also achieved to solve the 20 dimensional reaction chain (10), which seems to be difficult for other tensor-structured algorithms. In Figure 6(c) we touched the important issue of the accuracy of the separable tensor approximations. We will address this issue in a future publication.

Acknowledgement. The research leading to these results has received funding from the European Research Council under the European Community’s Seventh Framework Programme (FP7/2007-2013) / ERC grant agreement No. 239870 and from the People Programme (Marie Curie Actions) of the European Union’s Seventh Framework Programme (FP7/2007-2013) under REA grant agreement no. 328008. Radek Erban would also like to thank the Royal
Society for a University Research Fellowship; Brasenose College, University of Oxford, for a Nicholas Kurti Junior Fellowship; and the Leverhulme Trust for a Philip Leverhulme Prize. Tomáš Vejchodský also acknowledges the support of RVO: 67985840.

References

[1] E. Allgower and K. Georg. *Numerical continuation methods*, volume 13. Springer-Verlag Berlin, 1990.

[2] O. Alter and G. Golub. Reconstructing the pathways of a cellular system from genome-scale signals by using matrix and tensor computations. *PNAS*, 102(49):17559–17564, 2005.

[3] D. Anderson. An efficient finite difference method for parameter sensitivities of continuous time Markov chains. *SIAM Journal on Numerical Analysis*, 50(5):2237–2258, 2012.

[4] S. Bajikar, C. Fuchs, A. Roller, F. Theis, and K. Janes. Parameterizing cell-to-cell regulatory heterogeneities via stochastic transcriptional profiles. *PNAS*, 201311647, 2014.

[5] N. Barkai and S. Leibler. Biological rhythms: Circadian clocks limited by noise. *Nature*, 403(6767):267–268, 2000.

[6] A. Becskei and L. Serrano. Engineering stability in gene networks by autoregulation. *Nature*, 405(6786):590–593, 2000.

[7] G. Beylkin and M. Mohlenkamp. Numerical operator calculus in higher dimensions. *PNAS*, 99(16):10246–10251, 2002.

[8] R. Boys, D. Wilkinson, and T. Kirkwood. Bayesian inference for a discretely observed stochastic kinetic model. *Statistics and Computing*, 18(2):125–135, 2008.

[9] Y. Cao, D. Gillespie, and L. Petzold. The slow-scale stochastic simulation algorithm. *Journal of Chemical Physics*, 122(1):014116, 2004.

[10] J. Carroll and J. Chang. Analysis of individual differences in multidimensional scaling via an n-way generalization of “Eckart–Young” decomposition. *Psychometrika*, 35(3):283–319, 1970.

[11] B. Daniels, Y. Chen, J. Sethna, R. Gutenkunst, and C. Myers. Sloppiness, robustness, and evolvability in systems biology. *Current Opinion in Biotechnology*, 19(4):389–395, 2008.

[12] S. Dolgov, B. Khoromskij, et al. Tensor-product approach to global time-space-parametric discretization of chemical master equation. 2012.

[13] S. Dolgov and D. Savostyanov. Alternating minimal energy methods for linear systems in higher dimensions. Part I: SPD systems. arXiv:1301.6068 [math.NA], 2013.

[14] D. Donoho et al. High-dimensional data analysis: The curses and blessings of dimensionality. *AMS Math Challenges Lecture*, pages 1–32, 2000.

[15] M. Elowitz, A. Levine, E. Siggia, and P. Swain. Stochastic gene expression in a single cell. *Science*, 297(5584):1183–1186, 2002.

[16] R. Erban, S.J. Chapman, I. Kevrekidis, and T. Vejchodský. Analysis of a stochastic chemical system close to a SNIPER bifurcation of its mean-field model. *SIAM Journal on Applied Mathematics*, 70(3):984–1016, 2009.

[17] M. Gibson and J. Bruck. Efficient exact stochastic simulation of chemical systems with many species and many channels. *Journal of Physical Chemistry A*, 104(9):1876–1889, 2000.

[18] D. Gillespie. Exact stochastic simulation of coupled chemical reactions. *Journal of Physical Chemistry*, 81(25):2340–2361, 1977.

[19] D. Gillespie. A rigorous derivation of the chemical master equation. *Physica A: Statistical Mechanics and its Applications*, 188(1):404–425, 1992.

[20] D. Gillespie. The multivariate Langevin and Fokker–Planck equations. *American Journal of Physics*, 64(10):1246–1257, 1996.

[21] D. Gillespie. The chemical Langevin and Fokker-Planck equations for the reversible isomerization reaction. *Journal of Physical Chemistry A*, 106(20):5063–5071, 2002.

[22] A. Golightly and D. Wilkinson. Bayesian sequential inference for stochastic kinetic biochemical network models. *Journal of Computational Biology*, 13(3):838–851, 2006.
[23] L. Grasedyck. Hierarchical singular value decomposition of tensors. *SIAM Journal on Matrix Analysis and Applications*, 31(4):2029–2054, 2010.
[24] R. Grima, P. Thomas, and A. Straube. How accurate are the nonlinear chemical Fokker-Planck and chemical Langevin equations? *Journal of Chemical Physics*, 135(8):084103, 2011.
[25] R. Gutenkunst, J. Waterfall, F. Casey, K. Brown, Christopher R Myers, and James P Sethna. Universally sloppy parameter sensitivities in systems biology models. *PLoS Computational Biology*, 3(10):e189, 2007.
[26] L. Hansen. Large sample properties of generalized method of moments estimators. *Econometrica: Journal of the Econometric Society*, pages 1029–1054, 1982.
[27] A. Hilfinger and J. Paulsson. Separating intrinsic from extrinsic fluctuations in dynamic biological systems. *PNAS*, 108(29):12167–12172, 2011.
[28] T. Hunt. Under arrest in the cell cycle. *Nature*, 342(6249):483–484, 1989.
[29] T. Jahnke and W. Huisinga. Solving the chemical master equation for monomolecular reaction systems analytically. *Journal of Mathematical Biology*, 54(1):1–26, 2007.
[30] K. Jaqaman and G. Danuser. Linking data to models: data regression. *Nature Reviews Molecular Cell Biology*, 7(11):813–819, 2006.
[31] V. Kazeev, M. Khammash, M. Nip, and C. Schwab. Direct solution of the chemical master equation using quantized tensor trains. *PLoS Computational Biology*, 10(3):e1003359, 2014.
[32] D. Kim, B. Debusschere, and H. Najm. Spectral methods for parametric sensitivity in stochastic dynamical systems. *Biophysical Journal*, 92(2):379–393, 2007.
[33] T. Kolda and B. Bader. Tensor decompositions and applications. *SIAM Review*, 51(3):455–500, 2009.
[34] M. Komorowski, M. Costa, D. Rand, and M. Stumpf. Sensitivity, robustness, and identifiability in stochastic chemical kinetics models. *PNAS*, 108(21):8645–8650, 2011.
[35] J. Kruskal. Three-way arrays: rank and uniqueness of trilinear decompositions, with application to arithmetic complexity and statistics. *Linear Algebra and Its Applications*, 18(2):95–138, 1977.
[36] G. Lillici and M. Khammash. Parameter estimation and model selection in computational biology. *PLoS Computational Biology*, 6(3):e1000696, 2010.
[37] R. Lynch, J. Rice, and D. Thomas. Direct solution of partial difference equations by tensor product methods. *Numerische Mathematik*, 6(1):185–199, 1964.
[38] P. Marttinen, J. Corander, P. Törönen, and L. Holm. Bayesian search of functionally divergent protein subgroups and their function specific residues. *Bioinformatics*, 22(20):2466–2474, 2006.
[39] J. McGill, B. Ogunnaike, and D. Vlachos. Efficient gradient estimation using finite differencing and likelihood ratios for kinetic Monte Carlo simulations. *Journal of Computational Physics*, 231(21):7170–7186, 2012.
[40] B. Munsky, B. Trinh, and M. Khammash. Listening to the noise: random fluctuations reveal gene network parameters. *Molecular Systems Biology*, 5(1), 2009.
[41] F. Nobile, R. Tempone, and C. Webster. A sparse grid stochastic collocation method for partial differential equations with random input data. *SIAM Journal on Numerical Analysis*, 46(5):2309–2345, 2008.
[42] L. Omberg, G. Golub, and O. Alter. A tensor higher-order singular value decomposition for integrative analysis of DNA microarray data from different studies. *PNAS*, 104(47):18371–18376, 2007.
[43] I. Oseledets. Approximation of $2^d \times 2^d$ matrices using tensor decomposition. *SIAM Journal on Matrix Analysis and Applications*, 31(4):2130–2145, 2010.
[44] I. Oseledets. Tensor-train decomposition. *SIAM Journal on Scientific Computing*, 33(5):2295–2317, 2011.
[45] I. Oseledets and E. Tyrtyshnikov. Breaking the curse of dimensionality, or how to use SVD in many dimensions. *SIAM Journal on Scientific Computing*, 31(5):3744–3759, 2009.
[46] J. Owedyk. On the master equation with time-dependent transition probabilities. *Physics Letters A*, 109(4):152–154, 1985.
[47] E. Ozbudak, M. Thattai, I. Kurtser, A. Grossman, and A. van Oudenaarden. Regulation of noise in the expression of a single gene. *Nature Genetics*, 31(1):69–73, 2002.
[48] Y. Pantazis, M. Katsoulakis, and D. Vlachos. Parametric sensitivity analysis for biochemical reaction networks based on pathway information theory. *BMC Bioinformatics*, 14(1):311, 2013.
[49] J. Paulsson. Summing up the noise in gene networks. *Nature*, 427(6973):415–418, 2004.
A Methods

All models studied by the TPA are given in terms of well-mixed chemical systems where the system state changes according to the chemical reactions (1). Probability that a reaction occurs is determined by the propensity function

$$\alpha_j(\vec{x},k_j) = k_j \tilde{\alpha}_j(\vec{x}), \quad j = 1, 2, \ldots, M,$$

with the non-parametric part $\tilde{\alpha}_j(\vec{x})$ given by

$$\tilde{\alpha}_j(\vec{x}) = \exp \left[ \left( 1 - \sum_{i=1}^{N} \nu_{j,i}^- \right) \log V \right] \prod_{i=1}^{N} \nu_{j,i}^+ \left( \frac{x_i}{\nu_{j,i}^+} \right),$$

where $\nu_{j,i}^+$ and $\nu_{j,i}^-$ are parameters related to the stoichiometry of the reaction.
where parenthesis stand for the binomial coefficient and $V$ is the volume of the reactor.

The stationary distribution $p(\vec{x} | \vec{k})$, where $\vec{k} = (k_1, \ldots, k_M)^T$ is the vector of kinetic rates can be computed as the exact solution of the chemical master equation [19]. However, for the computational reasons, we will approximate it by the solution of the stationary chemical Fokker-Planck equation [20]. It can be simply expressed as

$$0 = \mathcal{A}(\vec{x}, \vec{k}) p(\vec{x} | \vec{k}),$$

where

$$\mathcal{A}(\vec{x}, \vec{k}) p(\vec{x} | \vec{k}) = -\sum_{i=1}^{N} \frac{\partial}{\partial x_i} \left( \sum_{j=1}^{M} \nu_{j,i} \alpha_{j}(\vec{x}, k_j) p(\vec{x} | \vec{k}) \right) + \frac{1}{2} \sum_{i,i'=1}^{N} \frac{\partial^2}{\partial x_i \partial x_{i'}} \left( \sum_{j=1}^{M} \nu_{j,i} \nu_{j,i'} \alpha_{j}(\vec{x}, k_j) p(\vec{x} | \vec{k}) \right),$$

is the parametric Fokker-Planck operator.

We will use the tensor structures to compute $p(\vec{x} | \vec{k})$ simultaneously for ranges of reaction rate values. The first step to achieve this is to split the model parameters from the state variables in a multiplicative relationship between the kinetic rate parameters. In the context of parameter estimation this means that a suitable approximations having this suitable product form.

Under these assumptions, the splitting of the parametric Fokker-Planck operator (15) implies a perfect collinear relationship between the kinetic rate parameters. In the context of parameter estimation this means that a single constraint (like mean and variance) restricts the original $M$-dimensional parameter space to an $(M-1)$-dimensional subspace of parameter values that comply with the constraint. Therefore, if a direct comparison between model and sample is not possible, a necessary condition to statistically infer $M$ parameters of a stochastic system is to define at least $M$ constraints.

### A.1 Tensorization

We will consider the state variable $\vec{x}$ in a bounded domain $\Omega^x \subset (0, \infty)^N$. Similarly, the kinetic rates $\vec{k}$ are considered in $\Omega^k \subset [0, \infty)^M$. In order to utilize the tensor structures, we have to assume that $\Omega^x = I_1 \times \cdots \times I_N$ and $\Omega^k = J_1 \times \cdots \times J_M$, where $I_d = (a_d^1, b_d^1)$, $d = 1, 2, \ldots, N$ are open intervals and $J_d = [a_d^k, b_d^k]$ are closed intervals.

For simplicity, we consider homogeneous Dirichlet boundary conditions on the boundary of $\Omega^x$. This, however, means that (13) has only the trivial solution $p(\vec{x} | \vec{k}) = 0$. Therefore, we approximate the stationary distribution by the (normalized) eigenfunction of $\mathcal{A}(\vec{x}, \vec{k})$ corresponding to the eigenvalue closest to zero. If $\Omega^x$ is sufficiently large and if the probability of having molecular numbers close to zero is negligible then this eigenvalue is almost zero and the Dirichlet boundary conditions do not cause any substantial error.
The chemical Fokker-Planck operator (16) is discretized in $\Omega^x$ by the finite difference method. However, finite element method can be used as well. In any case, we consider tensor grids [41] in both $\Omega^x$ and $\Omega^k$. The tensor grid in $\Omega^x$ has nodes $(x_{i_1,i_2,\ldots,x_{N,i_N}})$, $i_d = 1,2,\ldots,n_d$, $d = 1,2,\ldots,N$. There are $n_d$ points $x_d,i_d = a_d^x + i_d h_d^x$, $i_d = 1,2,\ldots,n_d$, in every $\mathcal{I}_d$ with the grid size $h_d^x = (b_d^x - a_d^x)/(n_d + 1)$, $d = 1,2,\ldots,N$. Similarly, we define tensor grid $(k_{1,j_1,\ldots,k_{M,j_M}})$ in $\Omega^k$, where $k_{\ell,j_\ell} = a_{\ell}^k + (j_\ell - 1)h_{\ell}^k$, $j_\ell = 1,2,\ldots,m_\ell$, form a uniform partition of $\mathcal{I}_\ell$ with the grid size $h_{\ell}^k = (b_{\ell}^k - a_{\ell}^k)/(m_\ell - 1)$, $\ell = 1,2,\ldots,M$. Note that the boundary points $a_d^x$ and $b_d^x$, $d = 1,2,\ldots,N$, are not present in the tensor grid due to the vanishing Dirichlet boundary conditions.

The values of the stationary distribution $p(\bar{x}|\bar{k})$ at the nodal points are organized as an $(N + M)$-dimensional tensor $\mathbf{p} \in \mathbb{R}^{n_1 \times \cdots \times n_N \times m_1 \times \cdots \times m_M}$ with entries

$$p_{i_1,\ldots,i_N,j_1,\ldots,j_M} = p(x_{1,i_1},\ldots,x_{N,i_N}|k_{1,j_1},\ldots,k_{M,j_M}). \quad (17)$$

In the traditional matrix-vector approach, we would organize the entries of $\mathbf{p}$ into a long vector. However, the tensor structure is more natural, because it corresponds to the original physical position of the nodes within the state and parameter space [14]. Finally, let us note that if $n = n_1 = \cdots = n_N$ and $m = m_1 = \cdots = m_M$ then there is $n^N m^M$ entries in the tensor $\mathbf{p}$. Thus, the number of memory places to store the tensor $\mathbf{p}$ grows exponentially with $N$ and $M$. In the next subsection we, present the main idea of the separated representation of tensors that allows to solve this problem.

A.2 Separation of dimensions

The main idea of the separated (or low-parametric) representation is to approximate a tensor $\mathbf{p}$ by a sum of rank-one tensors:

$$\mathbf{p} \approx \sum_{r=1}^{R} \phi_1^{[r]} \otimes \cdots \otimes \phi_N^{[r]} \otimes \psi_1^{[r]} \otimes \cdots \otimes \psi_M^{[r]}, \quad (18)$$

where $\phi_d^{[r]} \in \mathbb{R}^{n_d}$, $d = 1,2,\ldots,N$ and $\psi_\ell^{[r]} \in \mathbb{R}^{m_\ell}$, $\ell = 1,2,\ldots,M$ are factor vectors, $R$ is known as the separation rank, and symbol $\otimes$ denotes the tensor product of vectors [37]. Let us recall that the tensor product $\vec{v}_1 \otimes \cdots \otimes \vec{v}_N$ of vectors $\vec{v}_d \in \mathbb{R}^{n_d}$, $d = 1,2,\ldots,N$, is defined as a tensor $\mathbf{v} \in \mathbb{R}^{n_1 \times \cdots \times n_N}$ with entries $v_{i_1,\ldots,i_N} = v_{1,i_1} \cdots v_{N,i_N}$.

Representation (18) has the potential to solve high-dimensional problems. Indeed, if we consider for simplicity $n = n_1 = \cdots = n_N$ and $m = m_1 = \cdots = m_M$ then the representation (18) requires to store $nmNMR$ numbers only. For moderate values of $R$ this is substantially less than the number of entries of $\mathbf{p}$. Moreover, low-parametric representations such as (18) enable to perform algebraic operations in an efficient way (see A.4).

The accuracy of the separated representation (18) depends on the choice of the factor vectors and on the size of the tensor rank $R$. Clearly, the higher rank enables higher accuracy, but requires higher computational and storage costs. In practical computations, the rank $R$ is dynamically controlled using algorithms for tensor truncation, see Section A.3.

Let us note that the representation (18) is known as the canonical polyadic decomposition [35]. However, due to reasons connected with the stability of the tensor truncation algorithms, it is not suitable for actual computation and more stable tensor formats have to be employed [44]. We have introduced the canonical polyadic decomposition (18) due to its simplicity to illustrate the main idea of the separate representation of tensors.

For certain simple problems, like birth-death process, the separable representation of the stationary distribution is trivial and can be derived explicitly. However, in general, we have to compute the stationary distribution in the form (18). To achieve this, we need to express the discretized Fokker-Planck operator in a separable form as well. Based on the structure of $\mathcal{A}(\bar{x}|\bar{k})$ in (15), the discretization of the parametric Fokker-Planck operator divides into two step: decomposing the non-parametric part and decomposing the parametric part.
A.2.1 Decomposition of the non-parametric part

We use the finite differences to discretize the derivatives in the non-parametric operators \( \mathbf{A}^{[j]}(\mathbf{x}) \) in (16), see e.g. [59]. However, the separated tensor representation never requires high-dimensional difference stencils. Instead, just one-dimensional differences are needed. Further, since the standard finite difference discretizations of differential operators yield matrices, we organize their entries naturally into tensors. In this situation we speak about tensor matrices and denote them in capital bold font. The idea is exactly the same as in (17), where we organized a long vector into a tensor.

Thus, the finite difference matrix approximating the non-parametric operator \( \mathbf{A}^{[j]}(\mathbf{x}) \) in (16) can be expressed as the following tensor matrix:

\[
\mathbf{A}^{[j]} = -\sum_{i=1}^{N} \nu_{j,i} \mathbf{G}^{[i,j]} + \frac{1}{2} \sum_{i,i^\prime=1}^{N} \nu_{j,i} \nu_{j,i^\prime} \mathbf{F}^{[i,i^\prime,j]}, \quad j = 1,2,\ldots,M,
\]

where tensor matrices \( \mathbf{G}^{[i,j]} \) and \( \mathbf{F}^{[i,i^\prime,j]} \) refer to tensor-structured discretizations of the summands in the first and second sums in (15), respectively, and are determined by

\[
\mathbf{G}^{[i,j]} = \nu_{j,i} \mathbf{H}_{1}^{[j]} \otimes \cdots \otimes D_{i} \mathbf{H}_{1}^{[j]} \otimes \cdots \otimes \mathbf{H}_{N}^{[j]},
\]

\[
\mathbf{F}^{[i,i^\prime,j]} = \nu_{j,i^\prime} \mathbf{H}_{1}^{[j]} \otimes \cdots \otimes D_{i} \mathbf{H}_{1}^{[j]} \otimes D_{i^\prime} \mathbf{H}_{1}^{[j]} \otimes \cdots \otimes \mathbf{H}_{N}^{[j]}.
\]

where the volume scaling coefficients is \( \nu_{j,i} = \exp \left[ (1 - \sum_{i=1}^{N} \nu_{j,i}^{-1}) \log V \right] \). Here, \( \mathbf{H}_{1}^{[j]} \in \mathbb{R}^{n_{1} \times n_{1}} \) and \( D_{i} \in \mathbb{R}^{n_{i} \times n_{i}} \) for \( i = 1,\ldots,N \) and \( j = 1,\ldots,M \) are matrices and, thus, the tensor product \( \otimes \) works in the same way as the Kronecker product. Matrix \( D_{i} \) is the central difference matrix with entries \(-1/(2h_{x}^{2})\) and \(1/(2h_{x}^{2})\) distributed along its super- and sub-diagonal, respectively. Matrix \( \mathbf{H}_{1}^{[j]} \) is diagonal with diagonal entries \( \mathbf{H}_{1}^{[j]}(\ell,\ell) = (\nu_{j,i}^{-1})(\nu_{j,i}^{-1}) \) for \( \ell = 1,\ldots,n_{i} \). We observe that tensor matrices \( \mathbf{G}^{[i,j]} \) and \( \mathbf{F}^{[i,i^\prime,j]} \) are expressed in a separated representation similar to (18) with the separation rank \( R = 1 \). Consequently, the non-parametric operator \( \mathbf{A}^{[j]} \) in (19) admits separable representation of rank \( R = N^{2} + N \). Thus, any further algebraic operation on \( \mathbf{A}^{[j]} \) would contribute to the overall complexity growing quadratically in terms of number of chemical species.

A.2.2 Decomposition of the parametric part

Having the low-parametric discrete tensor-structured representations (19) of the non-parametric operators \( \mathbf{A}^{[j]} \), we can easily derive a discrete tensor-structured representation of the parametric Fokker-Planck operator (15):

\[
\mathbf{A} = \mathbf{A}^{[1]} \otimes \mathbf{K}_{1} \otimes \mathbf{I}_{2} \otimes \cdots \otimes \mathbf{I}_{M} + \mathbf{A}^{[2]} \otimes \mathbf{I}_{1} \otimes \mathbf{K}_{2} \otimes \cdots \otimes \mathbf{I}_{M} + \cdots + \mathbf{A}^{[M]} \otimes \mathbf{I}_{1} \otimes \mathbf{I}_{2} \otimes \cdots \otimes \mathbf{K}_{M},
\]

where \( \mathbf{K}_{j} \in \mathbb{R}^{m_{j} \times m_{j}} \) denotes a diagonal matrix whose diagonal entries correspond to the grid nodes of the \( j \)-th parameter, i.e., \( \mathbf{K}_{j}(\ell,\ell) = k_{j,\ell} \) for \( \ell = 1,\ldots,m_{j} \) and \( j = 1,\ldots,M \).

Thus, in (20) we succeeded to express the discretized parametric Fokker-Planck operator in a low-parametric tensor representation with separation rank \( M(N^{2} + N) \). This rank grows linearly with the number of chemical reactions \( M \) and quadratically with the number of chemical species \( N \). Then, the parametric steady state distribution of the form (18) is solved as the eigenvector of \( \mathbf{A} \) corresponding to the eigenvalue closest to zero (see Section A.3).

A.3 Solving the stationary CFPE in tensor format

Let \( \mathbf{A} \) be the tensor-structured parametric Fokker-Planck operator assembled in Equation (20), and our goal is to approximate the stationary distribution by the eigenvector corresponding to the eigenvalue closest
to zero:
\[ p = \arg \min_{\hat{p}} \{|\lambda| : A\hat{p} = \lambda\hat{p}\}. \quad (21) \]

A standard method to find the required eigenpair of \( A \) is the shifted inverse power method, and here we modify the original algorithm for better implementations in tensor-structured computations.

**Adaptive inverse power algorithm.** The main building block is the fact that, beginning with an initial guess \( p_0 \) and a shift value \( \sigma_k \), the shifted inverse scheme,
\[ (A - \sigma I)p_{k+1} = \frac{p_k}{\|p_k\|}, \quad k = 0, 1, \ldots \quad (22) \]
would converge to the eigenvector corresponding to the eigenvalue closest to the chosen shift \( \sigma \), provided that the eigenvalue is distinct. Since the Fokker-Planck operator has one trivial eigenvalue with all other spectrum distributed in the negative half plane, we adaptively choose a non-negative \( \sigma \) dynamically based on the performance of the tensor linear solver.

We apply the alternating minimum energy method (AMEN) [13] to solve the linear system (22). Given an initial \( N\)-dimensional tensor \( p \), rather than solve the system directly, the AMEN method minimise the residual in single dimension at a time with other dimensions fixed, and alternates the dimension from 1 to \( N \). The entire sweep repeats until some convergence criterion is satisfied. Typically, smaller shift \( \sigma \) makes the whole inverse power method converge faster to the steady state solution, however, within each inverse iteration (22), the AMEN may require many sweeps to achieve a reasonable tolerance. Thus, our strategy is to double the shift value \( \sigma \) when the solver reach certain upper threshold, and half \( \sigma \) to seek for better convergence for the whole procedure when the AMEN converges with only a few sweeps.

Another extension arises from a feature of tensor-structured data format. The tensor separation rank \( R \) can increase rapidly over successive algebraic operations, making the representation untenable. To avoid uncontrollable growth of the separation rank throughout the computation, we need to reduce it by adaptively changing the involving factor vectors while maintaining the required accuracy. This procedure is usually called tensor truncation:
\[ \hat{p} = \Gamma(p) \quad (23) \]
where operator \( \Gamma \) is the truncation operator, and \( \text{rank}(\hat{p}) \ll \text{rank}(p) \). Although finding the optimal tensor separation rank is still an open question in ongoing research, the tensor train format, together with its SVD-based tensor truncation algorithm, to be a stable and useful prototype for our implementations, and we refer the readers to [44] for further details.

Consequently, our adaptive shifted inverse power method is summarised as follows:

**Step 0.** Initiate: the initial guess, \( p_0 \), the shift value \( \sigma_0 \), the stopping criteria \( \epsilon \), maximum number of AMEN sweeps in each inverse iteration \( N_{\text{max}} \), and threshold to increase and decrease the shift value, \( N_{\text{in}} \) and \( N_{\text{de}} \).

**Step 1.** Solve the \( k \)-th tensor-structured inverse iteration (22) up to \( N_{\text{max}} \) sweeps.

**Step 2.** Check the number of sweep \( N_{\text{comp}} \) for the AMEN solver to converge:
2a. If \( N_{\text{comp}} > N_{\text{in}} \), let \( \sigma = 2\sigma \) and jump back to Step 1.
2b. If \( N_{\text{de}} < N_{\text{comp}} \leq N_{\text{in}} \), go to Step 3.
2c. If \( N_{\text{comp}} \leq N_{\text{de}} \), \( \sigma = \sigma/2 \) and go to Step 3.

**Step 3.** Truncate the tensor separation rank as in (23)

**Step 4.** Check the stopping criteria:
4a. If \( \|p_{k+1} - p_k\| > \epsilon \), let \( p_{k+1} = p_k \) and \( k = k + 1 \), and jump to Step 1.
4b. If \( \|p_{k+1} - p_k\| \leq \epsilon \), return \( p_{k+1} \) and exist.
Multi-level acceleration  When the dimensionality of the problem is large or the genetic network involves many non-linearities, the adaptive scheme discussed above may converge slowly, because on a fixed grid size, the AMEN requires very large shift value $\sigma$ to solve (22). Thus, we now consider a multilevel scheme to accelerate the solution process. The concept is straightforward. The system (21) is first solved on a coarse grid with grid size $2h$, and interpolate the approximated stationary solution to a fine grid with grid size $h$ as an initial guess. The method continues to solve the system on a finer grid until some convergence criteria are achieved.

A key ingredients in the multilevel approach is the interpolation, or prolongation, matrix that transfer the solution on a coarse grid to a fine grid. Owing to the structure of the separable tensor format, the prolongation operator possesses a nice rank-one tensor structure. Let $N$-dimensional tensor $p \in \mathbb{R}^{n_1 \times \cdots \times n_N}$ be the function values on an $N$-dimensional tensor grid with $n_k$, $k = 1, \ldots, N$, grid points along each direction, the prolongation operator $P[k]$ to the $k$-th dimension is defined as

$$P[k] = I \otimes \cdots \otimes P_{n_k}^{2n_k} \otimes \cdots \otimes I,$$

where $P_{n_k}^{2n_k} \in \mathbb{R}^{2n_k \times n_k}$ is the one-dimensional interpolation matrix defined by

$$P_{n_k}^{2n_k} = \frac{1}{2} \begin{pmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & \ddots \end{pmatrix}.$$

If tensor $p$ has the rank-$R$ separated representation as (7), the complexity to interpolate a single dimension is of order $O(n)$, and for a full interpolation over $N$-dimensional tensor grid, the total operations are bounded by $O(nN)$. We summarise the multi-level accelerated adaptive inverse power method as below:

**Step 0.** Initial grid size on the coarsest grid $h_1$, and corresponding initial guess $p_0^{(1)}$, maximum number of grid levels $L_{\text{max}}$, error tolerance $\epsilon^{(l)}$, and let $l = 1$.

**Step 1.** Solve the system (21) on the $l$-th level with initial guess $p_0^{(l)}$, using the adaptive inverse power method. Return the solution $p_k^{(l)}$ that satisfies the error tolerance $\|p_k^{(l)} - p_{k-1}^{(l)}\| \leq \epsilon^{(l)}$.

**Step 2.** Check the level index:

2a. If $l < L_{\text{max}}$, interpolate the solution $p_k^{(l)}$ to a finer grid by successively apply the prolongation operator $P_{[k]}$ in (24) to each dimension. Let $p_0^{(l+1)} = p_k^{(l)}$, $\epsilon^{(l+1)} = \epsilon^{(l)}/2$, $l = l + 1$, and go to Step 1.

2b. If $l = L_{\text{max}}$, return the solution $p_k^{(l)}$ and exist.

Multilevel approach is used in Section 5 to analyse the 20-dimensional chemical system (10). CPU times for each grid size are shown in Table 2.

| Level | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|-------|---|---|---|---|---|---|---|
| No. of nodes | 8 | 16 | 32 | 64 | 128 | 256 | 512 |
| Grid size $h$ | 20 | 10 | 5 | 2.5 | 1.25 | 0.625 | 0.3125 |
| CPU time ($\times 10^3$ sec) | 2.82 | 7.02 | 2.84 | 2.3 | 2.25 | 0.08 | 0.10 |
Implementation. Our implementations are coded in MATLAB Release 2013b, and are part of the Stochastic Bifurcation Analyzer toolbox publicly available at http://www.stobifan.org. The source code relies on the Tensor Train Toolbox [44]. Simulations are performed on a 64-bit Linux desktop equipped with Quad-Core AMD Opteron(tm) Processor 8356 × 16 and 63 GB RAM.

A.4 Elementary tensor operations

The previous section describes how the tensor-structured parametric solution \( p \) of the form (18) is sought implicitly, and this section, we discuss computational details to post-process the tensor solution for parametric analysis. In fact, all of our proposed analysis is based on high-dimensional integration, which is operated through the \( k \)-mode product described below.

Tensor multiplication: the \( k \)-mode product [33]. Let \( p \in \mathbb{R}^{n_1 \times \cdots \times n_N} \) be an \( N \)-dimensional tensor, the \( k \)-mode product of \( p \) with a vector \( \vec{q} \in \mathbb{R}^{k} \) is denoted by \( p \times_k \vec{q} \) and is of size \( n_1 \times n_{k-1} \times 1 \times n_{k+1} \times \cdots n_N \). Elementwise, we have

\[
(p \times_k \vec{q})_{i_1,\ldots,i_{k-1},1,i_{k+1},\ldots,i_N} = \sum_{j_k=1}^{n_k} p_{i_1,\ldots,i_{k-1},j_k,i_{k+1},\ldots,i_N} \vec{q}_{j_k},
\]

Further, if \( p \) can be written as a rank-\( R \) tensor, i.e.,

\[
p = \sum_{r=1}^{R} \vec{\phi}_1^{[r]} \otimes \cdots \otimes \vec{\phi}_N^{[r]},
\]

then the \( k \)-mode product can be evaluated through \( R \) one-dimensional inner products:

\[
p \times_k \vec{q} = \sum_{r=1}^{R} \vec{\phi}_1^{[r]} \otimes \cdots \otimes \vec{\phi}_{k-1}^{[r]} \otimes \langle \vec{\phi}_k^{[r]} \cdot \vec{q} \rangle \otimes \vec{\phi}_{k+1}^{[r]} \otimes \cdots \otimes \vec{\phi}_N^{[r]},
\]

which requires \( n_k R \) arithmetic operations.

The \((l_1,\ldots,l_N)\)-th order moment computation. In the parametric analysis, evaluating the cost function \( J(\vec{k}) \) in (3) requires extracting the higher-order moment values for different parameter combinations:

\[
\mu_{l_1,\ldots,l_N} = \int_{\Omega} x_1^{l_1} x_2^{l_2} \cdots x_N^{l_N} p(\vec{x} | \vec{k}) d\vec{x}.
\]

For tensor-structured parametric solution in (18), this can be done simultaneously for all parameter sets through successive applying the mode product introduced in (26) as

\[
\mu_{l_1,\ldots,l_N} = h_1^{\vec{x}} h_2^{\vec{x}} \cdots h_N^{\vec{x}} \left( p \times_1 \vec{x}_1^{l_1} \times_2 \vec{x}_2^{l_2} \times_3 \cdots \times_M \vec{x}_M^{l_M} \right),
\]

where \( \vec{x}_d = (x_{d,1}, \ldots, x_{d,n_d})^T \) denotes a vector whose entries equal to the grid points along the \( d \)-th dimension in the state space and \( h_d^{\vec{x}} \) is the grid size. The computational complexity of (30) is \( \mathcal{O}(nNR) \).
Combining intrinsic and extrinsic fluctuations. In (9), we have shown that the invariant probability distribution of intrinsic variables in presence of extrinsic noise can be approximated by the weighted integral of the invariant probability distributions without extrinsic noise, under the assumption that there exists a time scale separation between the intrinsic and extrinsic fluctuations. Provided the experimental measured distribution of parameters are \( q_j(k_j) \) for \( j = 1, \ldots, M \), the invariant probability of intrinsic variables is given in (9). For tensor-structured solution in (18), this can be numerically achieved by

\[
\mathbf{p}_x = h_1^k h_2^k \ldots h_M^k \left( \mathbf{p} \times_{N+1} \mathbf{q}_1 \times_{N+2} \mathbf{q}_2 \times_{N_3} \cdots \times_{N+M} \mathbf{q}_M \right),
\]

where the entries of vectors \( \mathbf{q}_j \) for \( j = 1, \ldots, M \) represent the values of \( q_j(k_j) \) at the discrete node points \( k_{j,1}, k_{j,2}, \ldots, k_{j,m} \) and \( h_1^k, \ldots, h_M^k \) are grid sizes in the parameter space. If \( m_1 = \cdots = m_M = m \) then we can easily express the complexity of evaluating the \( M \)-dimensional integral (9) in this way as \( \mathcal{O}(mMR) \), which scales linearly with the separation rank \( R \), number of parameters \( M \), and the number of grid nodes \( m \) along each dimension in the parameter space.

Computing transition probability and oscillation amplitude. In the parameter estimation (Figure 2) and sensitivity analysis (Figure 3), we illustrate the results based on transition probability and oscillation amplitude that are extracted from tensor-structured parametric solution. In fact, the computational idea is similar as above. For instance, we wish to estimate probability that, in steady state distribution, the \( l \)-th chemical species stays below certain threshold \( \bar{x}_l \). To achieve this, one can first integrate out all the other dimensions in the state space, and integrate the \( l \)-th dimension up to \( \bar{x}_l \), i.e.,

\[
p(x_l \leq \bar{x}_l | \mathbf{k}) = \int_{a_1^x}^{b_1^x} \cdots \int_{a_{l-1}^x}^{b_{l-1}^x} \int_{a_l^x}^{\bar{x}_l} \cdots \int_{a_N^x}^{b_N^x} p(\mathbf{x} | \mathbf{k}) \, d\mathbf{x}
\]

In tensor structure, we use \( N \) mode products to compute \( p(x_l \leq \bar{x}_l | \mathbf{k}) \) simultaneously for all parameter combinations by

\[
\mathbf{p}_\bar{x}_l = h_1^x h_2^x \cdots h_N^x \left( \mathbf{p} \times_1 \mathbf{1} \times_2 \mathbf{1} \times_3 \cdots \times_{l-1} \mathbf{1} \times_l \mathbf{1} \times_{l+1} \mathbf{1} \times_{l+2} \cdots \times_N \mathbf{1} \right),
\]

where \( \mathbf{1} \) denotes a vector of all ones. Entries of \( \mathbf{1}_{\bar{x}_l} \) equal to 1 if the corresponding grid point is smaller or equal to \( \bar{x}_l \), while its other entries are zero.

### B Tables: Model specifications

| Index | Reaction | Kinetic rate\(^a\) | True value |
|-------|----------|------------------|------------|
| 1     | 3\(X \rightarrow 2X\) | \(k_1/V^2\) | \(k_1 = 2.5 \times 10^{-4}\) |
| 2     | 2\(X \rightarrow 3X\) | \(k_2/V\) | \(k_2 = 0.18\) |
| 3     | \(\emptyset \rightarrow X\) | \(k_3 \times V\) | \(k_3 = 2250\) |
| 4     | \(X \rightarrow \emptyset\) | \(k_4\) | \(k_4 = 37.5\) |

\(^a\)The reacting volume is set to \( V = 1 \) unit.
Table 4: Moments estimated from stochastic simulation of the Schlögl model.

| Moment order | Value       | Weight |
|--------------|-------------|--------|
| 1            | $\hat{\mu}_1 = 261.32$ | $\beta_1 = 1$ |
| 2            | $\hat{\mu}_2 = 2.03 \times 10^4$ | $\beta_2 = 100$ |
| 3            | $\hat{\mu}_3 = -2.04 \times 10^5$ | $\beta_3 = 0.001$ |

Table 5: Properties of molecular and rate variables in the Schlögl model.

| Type | Notation | Range           | No. of nodes |
|------|----------|-----------------|--------------|
| Species | X        | [0, 1000]       | 1024         |
| Rate   | $k_1$    | $[2.43 \times 10^{-4}, 2.58 \times 10^{-4}]$ | 128          |
| Rate   | $k_2$    | [0.17, 0.19]    | 128          |
| Rate   | $k_3$    | [2134, 2266]    | 128          |
| Rate   | $k_4$    | [36.08, 36.08]  | 128          |

Table 6: Overview of kinetic reactions of the cell cycle model.

| Index | Reaction | Kinetic rate | Parameter(s) |
|-------|----------|--------------|--------------|
| 1     | $\emptyset \rightarrow Y$ | $k_1 \times V$ | $k_1 = 0.015$ |
| 2     | $Y \rightarrow \emptyset$ | $k_2$ | $k_2 = 0$ |
| 3     | $CP + Y \rightarrow pM$ | $k_3/V$ | $k_3 = 200$ |
| 4     | $pM \rightarrow M$ | $k'_4 + k_4 (M/V)^2$ | $k_4 = 180, k'_4 = 0.018$ |
| 5     | $M \rightarrow pM$ | $k_5 \times tP$ | $k_5 = 0, tP = 0.001$ |
| 6     | $M \rightarrow C_2 + YP$ | $k_6$ | N/A |
| 7     | $YP \rightarrow \emptyset$ | $k_7$ | $k_7 = 0.6$ |
| 8     | $C_2 \rightarrow CP$ | $k_8 \times tP$ | $k_8 = 1000$ |
| 9     | $CP \rightarrow C_2$ | $k_9$ | $k_9 = 1000$ |

*aThe volume corresponds to a single cell and is set to $V = 5000$ units.*

Table 7: Properties of molecular and rate variables in the cell cycle model.

| Type | Name          | Notation | Range           | No. of nodes |
|------|---------------|----------|-----------------|--------------|
| Species | cdc2       | C2       | [2230, 4990]    | N/A*         |
| Species | cdc2-P      | CP       | [10, 70]        | 256          |
| Species | p-cyclin-cdc2-p | pM      | [0, 1500]       | 256          |
| Species | p-cyclin-cdc2 | M        | [0, 1200]       | 256          |
| Species | cyclin      | Y        | [20, 70]        | 256          |
| Species | p-cyclin    | YP       | [0, 700]        | 256          |
| Rate   | degradation rate of active MPF | $k_6$ | [0.25, 0.4] | 64 |

*aDiscretisation of cdc2 is not applicable here, since this variable is eliminated by the conservation law of cdc2 assumed by the original author.*
Table 8: Properties of molecular and rate variables in the FitzHugh-Nagumo model.

| Type   | Notation | Range       | No. of nodes |
|--------|----------|-------------|--------------|
| Species| $X_1$    | $[0,1800]$  | 256          |
| Species| $X_2$    | $[0,700]$   | 256          |
| Rate   | $k_1$    | $[0.17,0.23]$ | 128         |
| Rate   | $k_2$    | $[0.952,0.1288]$ | 128     |
| Rate   | $k_3$    | $[2.125,2.875]$ | 128        |
| Rate   | $k_4$    | $[0.0892,0.1207]$ | 128      |

Table 9: Overview of kinetic reactions of the FitzHugh-Nagumo model.

| Index | Reaction       | Kinetic rate$^a$ | Mean value |
|-------|----------------|------------------|------------|
| 1     | $X_1 \rightarrow 2X_1$ | $(X_1 - k_1 \times V)(V - X_1)$ | $k_1 = 0.2$ |
| 2     | $X_1 \rightarrow \emptyset$ | $X_2$ | N/A         |
| 3     | $\emptyset \rightarrow X_1$ | $k_2 \times V$ | $k_2 = 0.112$ |
| 4     | $X_2 \rightarrow \emptyset$ | $k_3 \times k_4$ | $k_3 = 2.5$ |
| 5     | $X_1 \rightarrow X_1 + X_2$ | $k_4$ | $k_4 = 0.105$ |

$^a$The system volume is $V = 2000$ units.