



OPEN ACCESS

EDITED BY

Claudio Angione,
Teesside University, United Kingdom

REVIEWED BY

Marissa Renardy,
Applied BioMath, United States
Daniela Besozzi,
University of Milano-Bicocca, Italy

*CORRESPONDENCE

Christian Fleck,
✉ christian.fleck@fdm.uni-freiburg.de

RECEIVED 18 April 2024

ACCEPTED 27 August 2024

PUBLISHED 03 October 2024

CITATION

Deneer A, Molenaar J and Fleck C (2024)
Spectral expansion methods for prediction
uncertainty quantification in systems biology.
Front. Syst. Biol. 4:1419809.
doi: 10.3389/fsysb.2024.1419809

COPYRIGHT

© 2024 Deneer, Molenaar and Fleck. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Spectral expansion methods for prediction uncertainty quantification in systems biology

Anna Deneer¹, Jaap Molenaar¹ and Christian Fleck^{2*}

¹Mathematical and Statistical Methods Group, Wageningen University and Research, Wageningen, Netherlands, ²Freiburg Center for Data Analysis and Modeling, University of Freiburg, Freiburg, Germany

Uncertainty is ubiquitous in biological systems. For example, since gene expression is intrinsically governed by noise, nature shows a fascinating degree of variability. If we want to use a model to predict the behaviour of such an intrinsically stochastic system, we have to cope with the fact that the model parameters are never exactly known, but vary according to some distribution. A key question is then to determine how the uncertainties in the parameters affect the model outcome. Knowing the latter uncertainties is crucial when a model is used for, e.g., experimental design, optimisation, or decision-making. To establish how parameter and model prediction uncertainties are related, Monte Carlo approaches could be used. Then, the model is evaluated for a huge number of parameters sets, drawn from the multivariate parameter distribution. However, when model solutions are computationally expensive this approach is intractable. To overcome this problem, so-called spectral expansion (SE) methods have been developed to quantify prediction uncertainty within a probabilistic framework. Such SE methods have a basis in, e.g., computational mathematics, engineering, physics, and fluid dynamics, and, to a lesser extent, systems biology. The computational costs of SE schemes mainly stem from the calculation of the expansion coefficients. Furthermore, SE effectively leads to a surrogate model which captures the dependence of the model on the uncertainty parameters, but is much simpler to execute compared to the original model. In this paper, we present an innovative scheme for the calculation of the expansion coefficients. It guarantees that the model has to be evaluated only a restricted number of times. Especially for models of high complexity this may be a huge computational advantage. By applying the scheme to a variety of examples we show its power, especially in challenging situations where solutions slowly converge due to high computational costs, bifurcations, and discontinuities.

KEYWORDS

systems biology, computational systems biology, mathematical modelling, spectral expansion, surrogate models

1 Introduction

Every mathematical model in systems biology is subject to uncertainty and incomplete knowledge (Geris and Gomez-Cabrero, 2016; Gutenkunst et al., 2007; Mitra and Hlavacek, 2019; van Mourik et al., 2014). This can be in the form of unknown model structure, unknown model parameters and imperfect experimental data. Characterizing and quantifying these sources is crucial, as the uncertainty can translate into inaccuracies in the model predictions. Information about the quality of model predictions is vital when

applied as support for decision-making or optimization routines such as experimental design and parameter estimation (Raue et al., 2013). The aim of uncertainty quantification (UQ) is to determine the likeliness of certain outcomes, given that some aspects of the system under study are not (exactly) known.

Generally, uncertainty is distinguished into two classes (Bruno, 2007; Ghanem et al., 2017; Le Maître and Knio, 2010). The first class is so-called aleatoric uncertainty. Aleatoric uncertainty stems from the intrinsic variability found in the system under consideration, for this reason it is also referred to as statistical uncertainty. For example, in the case of parameter estimation, this uncertainty is related to the fact that parameters may essentially vary over the system components (e.g., cells) (Elowitz et al., 2002), so that for the system as a whole only a distribution of parameter values can be estimated, and not one precise value per parameter. Gene expression noise is often a source of varying conditions (e.g., initial conditions or protein production rates) in cells resulting in variability of process parameters such as the protein production rate (Thattai and van Oudenaarden, 2001; Elowitz et al., 2002; Blake et al., 2003; Lev, 2014; Schultheiß Araújo et al., 2017).

In contrast, the second class of uncertainty, termed epistemic (or systemic) uncertainty, is caused by a lack of information (Bruno, 2007; Ghanem et al., 2017; Le Maître and Knio, 2010). For example, in parameter estimation, uncertainty in the estimates may be caused by imperfect data sets that contain noisy, incoherent, or even missing data points (Sullivan, 2015). In such cases, the uncertainty can in principle be reduced by performing extra experiments (Banga and Balsa-Canto, 2008; Kreutz and Timmer, 2009; Barz et al., 2010), but practical restrictions often prohibit this.

In biological systems both types of uncertainty are typically present (Geris and Gomez-Cabrero, 2016). In terms of modelling, both are usually dealt with by employing a probabilistic framework (Kirk et al., 2016), in which model parameters are represented according to a probability density function (PDF) (van Mourik et al., 2014; Bruno, 2007). The choice of the type of PDF and the corresponding distribution of parameters is usually based on previous knowledge. For example, the case of a completely unknown parameter is described by a uniform distribution on a broad (positive) interval. In other cases a parameter could be known to follow a normal or lognormal distribution with known mean and variance, established in previously performed experiments (Tsigkinopoulou et al., 2018).

Among the field of UQ, Monte Carlo (MC) methods are most commonly used (Barbu and Zhu, 2020; James, 1980). In an MC approach the parameter PDFs are sampled and model responses for each sample recorded, thus providing a distribution of model outcomes and an indication of the uncertainty therein (e.g., by analyzing the distribution moments). These methods are simple in their implementation and are widely applicable. However, for models that have a large number of parameters or are computationally expensive, these MC procedures are often not feasible (Bruno, 2007; Le Maître and Knio, 2010).

As an alternative to MC, meta-modelling techniques, also referred to as surrogate modelling, are frequently adopted to deal with models that would otherwise be intractable. Support vector machines (Noble, 2004), artificial neural networks (Tadeusiewicz, 2015) and Bayesian networks (Wilkinson, 2007) are examples meta-modelling techniques used in systems biology. In this work, we focus

on spectral expansion (SE) methods, an approach that is widely used in engineering systems (Ghanem et al., 2017; Le Maître and Knio, 2010) and to a lesser extent in biological systems (Martin-Casas and Mesbah, 2016; Paulson et al., 2019; Streif et al., 2014; Renardy et al., 2018) for UQ purposes. In such an approach the model response is represented in the form of a series expansion. The advantage of such a representation is that an approximation of the model response is obtained for *all* values of the parameters at once. This allows immediate evaluation of statistics of the model outcome, either analytically or through sampling of the stochastic parameters, which can be done significantly faster than through MC methods for models that are problematic and computationally expensive (Xiu and George, 2002).

These advantages come at the cost of the need to calculate expansion coefficients. For this, two classes of spectral methods are in use. In the first class the governing equations of the model are reformulated such that each variable is represented by a spectral expansion. This results in a system of differential equations for the expansion coefficients and is known as intrusive spectral projection (Bruno, 2007). In the second class, based on so-called non-intrusive spectral projection, the expansion coefficients are determined without changing the original model equations (Eldred et al., 2008). This is the line followed in this paper. The advantage of a non-intrusive approach is that it requires only straightforward deterministic model evaluations and does not involve any reformulation of the model. It is particularly attractive in case of models where intrusive methods would become too laborious.

In the past, SE methods have been shown to converge very slowly or not at all for models involving non-smooth functions (Joel Chorin, 1974; Le Maître et al., 2004). This is indeed a critical challenge for biological models, which often show complex, non-linear behaviour such as bifurcations and spatial discontinuities. In this work, we provide a scheme for non-intrusive spectral projection that may overcome these problems. It is easy to implement and we show its power through applying it to a number of various biological models. The scheme straightforwardly allows the use of different basis functions for the expansion, such as Haar wavelets in case of models with bifurcations. We also present a simple segmentation method resulting in a piecewise continuous approximation of the model at hand which helps to deal with complex models avoiding the necessity of high order expansions. The examples treated in this paper have been chosen such that each one shows how a specific problem can be overcome. The MATLAB code used for each example is publicly available on GitHub (see Section 8)

2 Methods

2.1 Spectral expansion

Let us consider a model Ω that depends on a vector of stochastic input parameters θ . The model response Y can be any chosen quantity, e.g., the concentration of one of the model components or a function thereof. Here, the uncertainty parameters θ_l , $l = 1, \dots, U$, are assumed to be uncorrelated and thus independently distributed, each with PDF $P_l(\theta_l)$. So,

$$Y = \Omega(\theta), \quad \theta \sim P(\theta), \quad (1)$$

where $P(\boldsymbol{\theta})$ is the joint probability density function (PDF) for all U uncertainty parameters: $P(\boldsymbol{\theta}) = \prod_{l=1}^U P_l(\theta_l)$. The case of correlated parameters is discussed in [Section 2.2.3](#). For reasons of clarity, we restrict the explanation in this section to the univariate case, i.e., Y and θ , as shown in [Equation 1](#), are scalars. In [Section 2.2](#) we show how to deal with more than one uncertainty parameter.

The underlying model could be of any type, e.g., an ODE, a PDE, an algebraic, or a statistical model. This implies that Y may also depend on time and space. The challenge is to analyse the behaviour of Y as a function of θ . In cases where the numerical evaluation of the underlying model takes a considerable amount of computational time, this tends to obstruct any form of comprehensive analysis. In this paper we present a method that aims at making this tractable. The idea is to replace the original model by a meta-model, which is achieved by representing the output Y in terms of an expansion. This meta-model can be constructed such that it represents the underlying model to a high degree of accuracy, but having the advantage that it is much faster to evaluate than the original model.

This meta-model can be used to determine the distribution of the model response Y or reconstruct the function accurately at given points in the parameter space. In the spectral expansion (SE) approach the model response is represented by

$$Y(\theta) = \sum_{n=0}^{\infty} c_n \phi_n(\theta), \quad (2)$$

where c_n are the expansion coefficients (which can be time and/or space dependent) and ϕ_n are functions that are orthonormal with respect to the distribution of the uncertainty parameters as weight functions for an inner product, as dealt with in [Supplementary Material S1](#). For example, suitable basis functions for uniformly distributed parameters are Legendre polynomials, while for normally distributed parameters Hermite polynomials qualify. For any practical purpose the expansion given in [Equation 2](#) needs to be truncated to a certain degree:

$$Y^s(\theta) = \sum_{n=0}^N c_n \phi_n(\theta), \quad (3)$$

where N is the truncation degree. For the meta-model Y^s N expansion coefficients have to be calculated. The advantage of [Equation 3](#) is that the statistics of the model response Y can be evaluated very fast, either analytically or through sampling of the parameters θ . The main computational cost of the SE approach stems from the computation of the coefficients c_n . Below, we provide an easy-to-implement scheme for the calculation of these coefficients.

The most commonly used method for determining the coefficients in the SE approach is through Gaussian quadrature schemes ([Bruno, 2007](#)). Here, we propose an alternative scheme. It is applicable to any set of orthonormal functions, allowing the flexibility to tackle different modelling challenges. A key feature in the scheme is the introduction of the symmetric matrix:

$$\hat{B}_{n,m} = \int \phi_n(\theta) \theta \phi_m(\theta) P(\theta) d\theta. \quad (4)$$

Since this matrix is symmetric, its eigenvalues $\lambda^{(l)}$, $l = 1, 2, \dots$, are real and its eigenvectors $u^{(l)}$ orthonormal. In the [Supplementary Material S1](#) we show that these eigenvalues and eigenvectors can be

used to derive an explicit expression for the coefficients c_n . After substitution of this expression, [Equation 3](#) then reads as

$$Y^s(\theta) = \sum_{l=1}^N Y(\lambda^{(l)}) u_1^{(l)} \psi_l^s(\theta), \quad (5)$$

where

$$\psi_l^s(\theta) \equiv \sum_{n=0}^N u_{n+1}^{(l)} \phi_n(\theta). \quad (6)$$

The striking point here is that this expansion requires evaluation of the model only N times, namely, for each of the eigenvalues $\lambda^{(l)}$, $l = 1, \dots, N$. Note that all terms in the expansion that do not depend on the uncertainty parameter θ can be calculated in advance, so once and for all. This saves computation time for any future application. For models that take a long time to evaluate the use of [Equation 5](#) is a very fast alternative, compared to e.g., a Monte-Carlo approach. What's more, statistical moments like mean and variance follow directly from the expansion coefficients ([Bruno, 2007](#)), thus requiring no further calculation. Note that the expansion in [Equation 5](#) is only exact in the limit $N \rightarrow \infty$. Taking a finite value for N introduces inaccuracy. Therefore, N must be chosen with care and it is often not obvious beforehand which value of N will give reliable results. In the examples below we have chosen N by visual comparison with e.g., an analytical solution or with MC solutions. In cases when a high level of accuracy is needed a more robust comparison or measure of convergence will be required, see for example, ([Vincent et al., 2013](#)) where a leave-one-out technique is used as a stopping criterion. In [Section 2.2.4](#), we propose an easy-to-use scheme to deal with cases where a high degree of expansion might result in infeasible computation times.

2.2 Practical aspects

Here, we treat some specific aspects of the method presented above.

2.2.1 PDFs and basis functions

We already mentioned Legendre and Hermite polynomials as possible basis functions for SE. Legendre polynomials are defined over the interval $[-1, 1]$ and are orthogonal with respect to the uniform distribution $\mathcal{U}(-1, 1)$. The Hermite polynomials are defined over the whole real axis \mathbb{R} and are orthogonal with respect to the Gaussian distribution $\mathcal{N}(0, 1)$, with mean 0 and standard deviation 1. Both polynomials can be normalized with appropriate prefactors to ensure orthonormality. These two families of polynomials are most often used to represent biological parameters. The uniform distribution is typically applied in uninformed cases and the lognormal distribution in cases where there is prior information available on a parameter. In practice, sometimes other classical orthogonal polynomial families are appropriate, e.g., Laguerre polynomials that are related to Gamma distributions. The method is also known as Polynomial Chaos Expansion in the case of polynomial basis functions, first introduced by Norbert Wiener in 1938 ([Wiener, 1938](#)). However, note that non-polynomial functions may also be applied in SE methods, such as spherical harmonics and wavelets. The

approach presented in this paper can be used for any set of orthonormal functions.

As mentioned above, Legendre are related to the standard uniform distribution $\mathcal{U}(-1, 1)$ and Hermite polynomials to the standard normal distribution $\mathcal{N}(0, 1)$. In practice, however, the biological parameters often do not fit these precise restrictions. In some cases this can be repaired by applying an isoprobabilistic transformation, as we will show underneath. For example, to obtain a normally distributed random variable k with mean μ and variance σ , so $k \sim \mathcal{N}(\mu, \sigma)$, starting from a $\theta \sim \mathcal{N}(0, 1)$, we need the transform

$$k = \mu + \sigma\theta, \quad \theta \sim \mathcal{N}(0, 1). \tag{7}$$

To obtain a uniformly distributed random variable k on the interval $[a, b]$, so $k \sim \mathcal{U}(a, b)$, from $\theta \sim \mathcal{U}(-1, 1)$, we need the transform

$$k = (b + a)/2 + (b - a)\theta/2, \quad \theta \sim \mathcal{U}(-1, 1). \tag{8}$$

A lognormally distributed random variable $k \sim \text{Lognormal}(\mu, \sigma)$ is obtained from $\theta \sim \mathcal{N}(0, 1)$ via the transformation

$$k = \mu \exp\left[\alpha\theta - \frac{\alpha^2}{2}\right], \quad \theta \sim \mathcal{N}(0, 1), \tag{9}$$

Equations 7–9 show the most common transformations. Other transformations are, of course, possible as well. where $\alpha = \sqrt{\ln(1 + \frac{\sigma^2}{\mu^2})}$.

2.2.2 Multiple uncertainty parameters

Typically, biological models involve more than one random parameter, which means that the SE basis $\{\phi_n(\theta), n \in \mathbb{N}^M\}$ is multivariate. Extending Equation 5 to the M -dimensional case is straightforward:

$$Y^s(\theta_1, \dots, \theta_M) = \sum_{l_1=1}^N \dots \sum_{l_M=1}^N Y(\lambda^{(l_1)}, \dots, \lambda^{(l_M)}) u_1^{(l_1)} \dots u_M^{(l_M)} \psi_{l_1}^s(\theta_1) \dots \psi_{l_M}^s(\theta_M). \tag{10}$$

As other similar approaches, SE suffers from the curse of dimensionality (Ghanem et al., 2017; Xiu, 2007). Note from Equation 10 that the number of times the model has to be evaluated scales as N^M , where N is the expansion order and M the number of uncertainty parameters.

2.2.3 Correlated uncertainty parameters

To handle correlated parameters we propose transforming the basis functions. Here, we discuss the case of two correlated parameters; in Supplementary Material S2 we show the general case. Let $P(\theta_1, \theta_2)$ be the common distribution for the uncertainty parameters θ_1 and θ_2 , and let $P(\theta_1)$ and $P(\theta_2)$ be the corresponding marginalised distributions. Assuming that $P(\theta_1, \theta_2) \neq 0, \forall \theta_1, \theta_2$, we may write the new basis functions as (for details see Supplementary Material S2):

$$\psi_{l_1, l_2}^s(\theta_1, \theta_2) = \sqrt{\frac{P(\theta_1)P(\theta_2)}{P(\theta_1, \theta_2)}} \psi_{l_1}^s(\theta_1) \psi_{l_2}^s(\theta_2), \tag{11}$$

where the functions ψ_n^s are orthonormal with respect to $P(\theta)$ and are given by Equation 6. The expansion of a function depending on correlated parameters can be written as:

$$Y^s(\theta_1, \theta_2) = \sqrt{\frac{P(\theta_1)P(\theta_2)}{P(\theta_1, \theta_2)}} \sum_{l_1, l_2} Y(\lambda^{(l_1)}, \lambda^{(l_2)}) \omega_{l_1, l_2} \psi_{l_1}^s(\theta_1) \psi_{l_2}^s(\theta_2), \tag{12}$$

with

$$\omega_{l_1, l_2} = \int \sqrt{P(\theta_1)P(\theta_2)P(\theta_1, \theta_2)} \psi_{l_1}^s(\theta_1) \psi_{l_2}^s(\theta_2) d\theta_1 d\theta_2. \tag{13}$$

2.2.4 Segmentation

In cases where either an expansion to high order is needed to obtain the requested approximation accuracy or the model is computationally expensive, one needs to switch to optimised or more effective methods (Le Maître and Knio, 2010; Sullivan, 2015). Models that show complex response surfaces (e.g., bifurcations) will require high order expansions to capture the inherent complexity. To overcome these problems, we present here a simple and easy to implement scheme that segments the parameter intervals into subintervals, yielding a piecewise continuous approximation of the original function. Within each of these segments we then perform a separate expansion. In this approach we have to deal with a trade-off: the number of expansions is increased, but per expansion we have a (much) lower order of expansion. Below, we argue why the second positive aspect greatly counterbalances the first negative aspect.

To determine the segments we define a scaling function g_m with $M \in \mathbb{N}$ and $L \in \mathbb{R}$ by:

$$g_m: [-L, L] \rightarrow \mathcal{I}_m = \left[\frac{(2m-1)L}{2M+1}, \frac{(2m+1)L}{2M+1} \right], m \in [-M, M] \subset \mathbb{Z},$$

$$\theta \mapsto g_m(\theta) = \frac{2mL}{2M+1} + \frac{\theta}{2M+1}. \tag{14}$$

This scaling function g_m divides the interval $[-L, L]$ into $2M + 1$ segments \mathcal{I}_m of equal length. Of course, L must be larger than or equal to any value of θ . For example, consider the case $L = 1$. Then, the whole interval is $[-1, 1]$. For a segmentation granularity of $M = 1$, this interval is divided into three subintervals: $\mathcal{I}_{-1} = [-1, -0.33]$, $\mathcal{I}_0 = [-0.33, 0.33]$ and $\mathcal{I}_{+1} = [0.33, 1]$. The expansion of Y on any subinterval \mathcal{I}_m is given by:

$$Y(g_m(\theta)) = \sum_l Y(g_m(\lambda^{(l)})) u_1^{(l)} \psi_l(\theta). \tag{15}$$

Upon a variable transformation $y = g_m(\theta)$, Equation 15 becomes:

$$Y(y) = \sum_l Y(g_m(\lambda^{(l)})) u_1^{(l)} \psi_l(g_m^{-1}(y)): y \in \mathcal{I}_m. \tag{16}$$

After segmentation, the expansion on the interval $[-L, L]$ as a whole is a superposition of the expansions on the subintervals:

$$Y(y) = \sum_m \sum_l Y(g_m(\lambda^{(l)})) u_1^{(l)} \chi_m(y) \psi_l(g_m^{-1}(y)): y \in [-L, L], \tag{17}$$

where $\chi_m(y)$ is an indicator function for selecting the correct segment:

$$\chi_m(y) = \begin{cases} 1: & y \in \mathcal{I}_m, \\ 0: & y \notin \mathcal{I}_m. \end{cases} \quad (18)$$

We can also define an index function to select $m^* \in [-M, M]$ for which $\chi_m(y) = 1$:

$$m^* = z(y) = \left\lfloor \frac{(2M+1)y}{2L} + \frac{1}{2} \right\rfloor \text{ for } |y| \leq L. \quad (19)$$

Using both index functions we can finally write the segmented reconstruction as:

$$Y(y) = \sum_I Y(g_z(y)(\lambda^{(I)})) u_1^{(I)} \psi_1(g_z^{-1}(y)(y)): y \in [-L, L]. \quad (20)$$

The number of model solutions required now scales as $(2M+1)^K N^K$, where K is the number of uncertainty parameters, M the segmentation granularity, and N the expansion order. The reduced accuracy by using a lower order expansion is compensated by evaluating the model more often, as a result of zooming in. Expanding up to the N -th order for p uncertainty dimensions requires solving the system N^p times. Reconstruction requires the summation of N^{2p} terms. Therefore, it is advantageous to keep N as low as possible. Normally, the reconstruction error is large for low N , but this is mitigated by segmentation. Segmentation requires to evaluate the system $(2M+1)^p N^p$ times, but due to segmentation N can be taken much smaller.

To illustrate this with an example, we take a system with 2 species of interest and 5 uncertainty parameters θ_i . The expansion order is taken as $N = 8$. This implies summing over $2 \times 8^{10} = 2,147,483,648$ terms per time point and per parameter set $(\theta_1, \dots, \theta_5)$. In the case of segmented expansion, we can choose a lower N , for example, $N = 3$ with a segmentation granularity of $M = 1$. The number of terms to be summed over is $2 \times 3^{10} = 118,098$. This is dramatically more efficient and stems from the fact that one only has to determine the segment in which the parameter set (θ_1, θ_5) falls and choose the corresponding expansion coefficients. Example II in Section 3.2 shows that segmentation may indeed be very beneficial.

2.2.5 Haar wavelet expansion

Traditional SE methods are known to have difficulties with capturing discontinuous behaviour (Joel Chorin, 1974; Le Maître et al., 2004). Spectral convergence is only observed when solutions are sufficiently regular and continuous. Just like Fourier expansions, SE suffers from Gibbs phenomena at discontinuities, resulting in slow convergence (Le Maître and Knio, 2010). Haar wavelets have been suggested to overcome these difficulties (Le Maître et al., 2004; Sullivan, 2015). In contrast to global basis functions like the aforementioned polynomial systems, wavelet representations lead to localized decompositions, resulting in increased robustness at the cost of a slower convergence rate (Sullivan, 2015; Le Maître and Knio, 2010). Here, we discuss that Haar wavelets can be easily incorporated in the framework presented above and in Example IV in Section 3.4 we show how they can be applied in practice.

As mother wavelet we take

$$\psi^W(y) = \begin{cases} 1 & \text{for } 0 \leq y < \frac{1}{2} \\ -1 & \text{for } \frac{1}{2} \leq y < 1 \\ 0 & \text{otherwise} \end{cases} \quad (21)$$

By introducing a scaling factor j and a sliding factor k , we may construct the wavelet family

$$\psi_{j,k}^W(y) = 2^{j/2} \psi^W(2^j y - k), \quad j = 0, 1, \dots; \quad k = 0, \dots, 2^j - 1. \quad (22)$$

Given the uncertainty parameter θ with its cumulative distribution function $F(\theta)$, we define the basis functions as

$$W_{j,k}(\theta) \sim \psi_{j,k}^W(F(\theta)). \quad (23)$$

By concatenating the indices j and k into one index $i \equiv 2^j + k$, we may expand the meta-model Y similarly as we did in Equation 2:

$$Y(\theta) = \sum_{n=0}^{\infty} c_n W_n(\theta). \quad (24)$$

2.2.6 Sensitivity analysis

In sensitivity analysis one quantifies the effects of changes in the parameters on the variability of the model response. Here, we show how our SE approach allows for sensitivity analysis in an elegant way. In the case of local sensitivity analysis, small parameter variations around a certain point in parameter space are used to determine the effect on the model output (Brian, 2013). This sensitivity is estimated via calculation of the partial derivatives of model output with respect to parameters, evaluated in that point (Ingalls, 2008). Alternatively, global sensitivity approaches do not specify a specific point in parameter space (Saltelli, 2008). For example, Sobol indices are a popular sensitivity measure as they provide a measure of global sensitivity and accurate information for most models (Ilya, 2001). Sobol indices are based on the decomposition of the variance of the output Y as a function of the contribution of the parameters (and possibly their combination), also called the ANalysis of VAriance, or ANOVA (Ilya, 2001). Thanks to the orthonormality of basis functions in SE, Sobol indices can be determined analytically from the coefficients of the SE (Sudret, 2008; Blatman and Sudret, 2010a). So, once these coefficients are known, one gets the Sobol indices nearly for free. Given the SE of output Y , the total variance of the model output is given by

$$\hat{D} = \sum_{i \in I - \{0\}} c_i^2, \quad (25)$$

where I is the multi-index set of all variables and c_i the expansion coefficients. The 0th coefficient is not included as this is a constant. The partial variance is given by

$$D_{\theta_i} = \sum_{i \in I_{\theta_i}} c_i^2, \quad (26)$$

where I_{θ_i} is the multi-index set of parameter θ_i , i.e., where the i^{th} term in the multi-index is larger than 0. The Sobol indices are then given by

$$S_{\theta_i} = \frac{D_{\theta_i}}{D}, \quad (27)$$

In this way the relative contribution of parameter θ_i to the variance of the output is easily calculated.

2.3 Monte Carlo sampling

In the examples in [Section 3.5](#) we compare SE with Monte Carlo sampling. For efficiency reasons we apply Quasi Monte Carlo methods using Sobol sequences, since these show a faster rate of convergence than standard sequences of pseudorandom numbers (Sobol, 1990). In order to test for convergence, we use the so-called blocking method. In this method the error is estimated in a straightforward manner. The quantity of interest (i.e., the model response Y) is divided into several groups (or blocks). Then, for each block, we determine the moment of interest (e.g., mean or variance). The spread (variance) of values between blocks gives an estimate of the error. Finally, to test for convergence, we choose a threshold on the between-block variance. Note that the convergence rate of MC methods do not depend on the dimensionality of the parameter space that is being sampled, unlike SE. MC merely scales by \sqrt{N} , where N is the number of samples (Christian, 2004).

2.4 Summary of implementation

In this section we provide an overview of the steps needed to arrive at a meta-model using SE:

1. Determine which of the model parameters is stochastic in nature and decide upon an appropriate PDF for such parameters.
2. Choose a truncation degree N .
3. Based on the PDF in the previous steps, calculate the appropriate basis functions $\phi_n(\theta)$, $n = 0, 1, 2 \dots, N$.
4. Determine the $N \times N$ matrix \hat{B} as defined in [Equation 4](#). For example, for Legendre polynomials \hat{B} reads as

$$\hat{B}_{n,m} = \frac{n}{\sqrt{2n+1}\sqrt{2m+1}}\delta_{nm+1} + \frac{m}{\sqrt{2n+1}\sqrt{2m+1}}\delta_{nm-1}, \quad (28)$$

where $n, m = 0, 1, 2 \dots, N$.

For Hermitian polynomials \hat{B} reads as:

$$\hat{B}_{n,m} = \sqrt{n}\delta_{nm+1} + \sqrt{m}\delta_{nm-1}. \quad (29)$$

5. Calculate the eigenvalues $\lambda^{(l)}$, $l = 1, 2, \dots, N$ and orthonormal eigenvectors $u^{(l)}$.
6. Calculate $\psi_i^s(\theta) = \sum_{n=0}^N u_{n+1}^{(i)} \phi_n(\theta)$.
7. Calculate $Y(\lambda^{(l)})$, $l = 1, 2, \dots, N$ by evaluating the model N times.
8. Arrive at the metamodel $Y^s(\theta) = \sum_{l=1}^N Y(\lambda^{(l)}) u_1^{(l)} \psi_l^s(\theta)$.
9. Eventually, apply post-processing through, e.g., sensitivity analysis.

3 Results

To test the performance of the present SE approach in biological simulations, we have chosen six typical examples. Through these

examples, we show how to deal with several challenges usually encountered in systems biology.

The first example has only one uncertainty parameter. Its simplicity allows comparison between the results of our approach with an exact solution.

The second example concerns a biochemical reaction network and is higher dimensional, i.e., it contains more than one uncertainty parameters. We use it to highlight the advantages of segmentation.

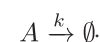
The third example is the glycolytic oscillator, which shows bifurcations, i.e., different dynamic behaviour for different parameter sets (Strogatz, 1994). We use it to demonstrate the power of global sensitivity analysis, which in the SE framework can be achieved without significant additional computational costs once the SE coefficients have been calculated. In addition, this example allows us to show the use of mixed expansions, since the parameter PDFs follow different distributions. This leads to a combination of different families of basis functions, thus highlighting the flexibility of the SE approach when applied to varying input uncertainties.

The fourth and fifth examples have a spatial dimension. First, we consider the Schnakenberg model which is a well-known model of pattern formation and comes with challenges such as shifts from non-patterning to patterning regions (D Murray, 2007). In this example we demonstrate the advantage of using Haar wavelets over polynomial basis functions for systems with bifurcations. Second, we study a model describing pattern formation in plants, more specifically patterning of the hairs found on top of leaves, so-called trichomes (Bouyer et al., 2008). In this model we show how to adapt the approach such that the computational costs are reduced as much as possible by carefully choosing the quantity of interest, without changing the standard set of steps.

The final example is a model of plasmid transfection, where we consider correlated parameter distributions. This model predicts the distribution of two different plasmid constructs among a population of dividing cells. Upon division the plasmids are distributed among the daughter cells according to a bivariate Poisson distribution.

3.1 Example I. Exponential decay: comparing performance of SE to MC and an analytical solution

For this example case, we consider the simple reaction consisting of one decaying species:



Its dynamics is described by $A(t) = A_0 \exp^{-kt}$, where A_0 is the initial concentration of A at $t = 0$ and k the rate of decay. We test the SE method against: 1) the exact, analytical solution and 2) the classical Monte Carlo approach.

We assume that k is distributed according to a lognormal distribution with known mean and variance, i.e., $k \sim \text{Lognormal}(\mu, \sigma)$, and we choose $\mu = 0.5, \sigma = 0.2$. The PDF for k is shown in [Figure 1](#) and the derivation for the exact PDF for A is given in the [Supplementary Material S2](#).

Next, we determine how the uncertainty in k propagates through the model and affects concentration $A(t)$. To that end, we expand

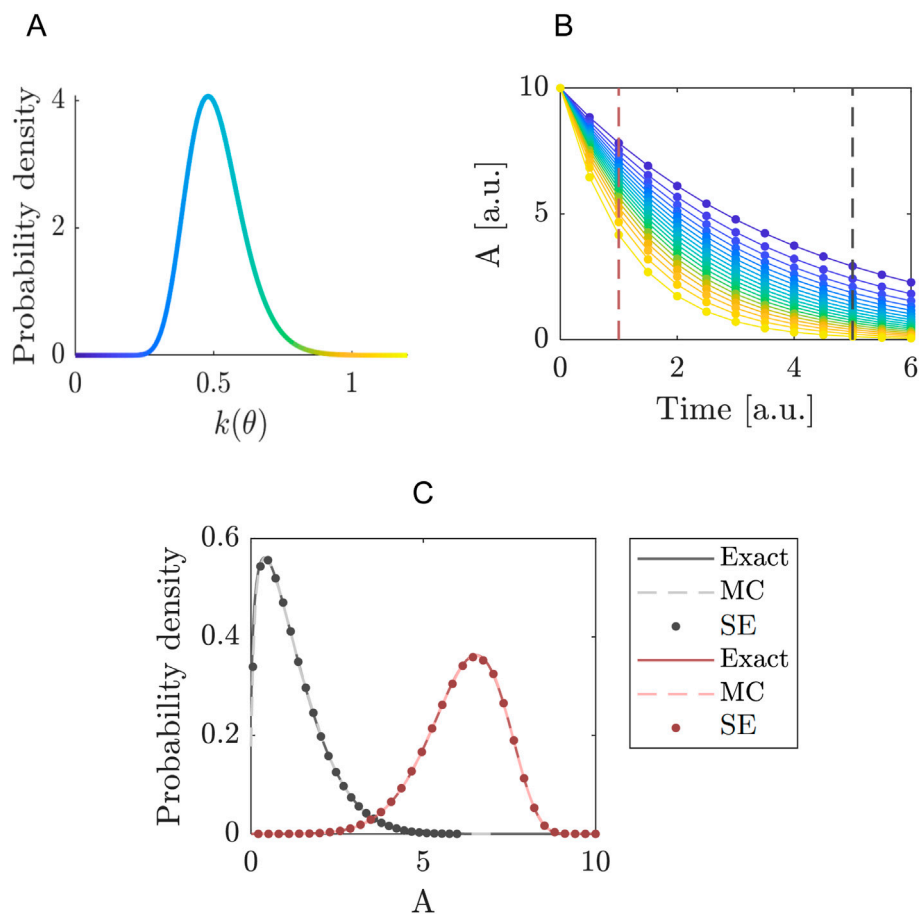


FIGURE 1

Quantifying the uncertainty propagated by the decay rate in the exponential decay model. **(A)** Probability density function of the decay rate $k(\theta)$ with $\mu = 0.5$ and $\sigma = 0.2$. The colour gradient corresponds to the value of $k(\theta)$. **(B)** The concentration of A up to $t = 6$ seconds. The solid lines indicate the analytical solution and the dots indicate the reconstruction using SE with Hermite basis functions and an expansion order $N = 5$. The colour for each of the solutions correspond to the colour of the line in **(A)**, which indicates the value of $k(\theta)$ used for each of the depicted solutions. **(C)** The dashed lines in **(B)** indicate a vertical cross-section along the model response space at $t = 1$ for the red line, and at $t = 5$ for the grey line, determined through three methods. First, the exact dynamics of the model (solid lines), second through MC sampling using the exact model dynamics (dashed lines) and finally, through MC sampling of the reconstructed function as obtained through SE (dots).

the function $A(t) = A_0 \exp^{-kt}$ in terms of Hermitian polynomials, following the steps given in the summary and using the matrix derived for Hermitian polynomials given in Equation 29. Using Equation 9, k is transformed into a standard normal variable θ . To arrive at the meta-model $Y^s(\theta)$ we truncate the expansion to order N , as shown in Equation 3. Choosing N is not straightforward and will involve some experimentation. In Figure 1 we compare results for $N = 5$ to the analytical solution for different k values. This shows that for this expansion order the reconstructed function accurately matches the analytical solution. Note that this is only based on visual inspection.

In Figure 1, we focus on the distributions of $A(t = 1)$ and $A(t = 5)$. This is achieved by sampling the SE using a large sample set χ of reduced (i.e., standard normally distributed) variables θ , $\chi_{sim} = \{\theta_j, j = 1, \dots, n_{sim}\}$. The truncated series is then evaluated onto this sample: $Y_{sim}^s = \{\eta_j = \sum_{n=0}^N c_n \phi_n(\theta_j), j = 1, \dots, n_{sim}\}$. These PDFs are obtained by kernel smoothing (Bowman and Azzalini, 1997) using a sample set with $n_{sim} = 10^6$, drawn from the standard normal distribution with $\mu = 0$ and $\sigma = 1$. The kernel density estimator is given by

$$\hat{f}_Y(y) = \frac{1}{n_{sim}h} \sum_{j=1}^{n_{sim}} K\left(\frac{y - \eta_j}{h}\right), \quad (30)$$

with kernel function $K(t) = \frac{1}{\sqrt{2\pi}} \exp^{-t^2/2}$ and bandwidth h , which is determined by Silverman's rule of thumb (Silverman, 1986). Figure 1 shows that both MC and SE perform well (using kernel smoothing with the kernel defined in Equation 30) in reproducing.

3.2 Example II. Biochemical reaction network: segmentation to deal with higher dimensions

In this example we present a simple model with multiple uncertainty parameters. It allows us to illustrate the computational advantage of segmentation as explained in Section 2.2.4. The model describes the dynamics of two proteins x_1 and x_2 which bind together to form a dimer x_3 . We consider the following reactions:

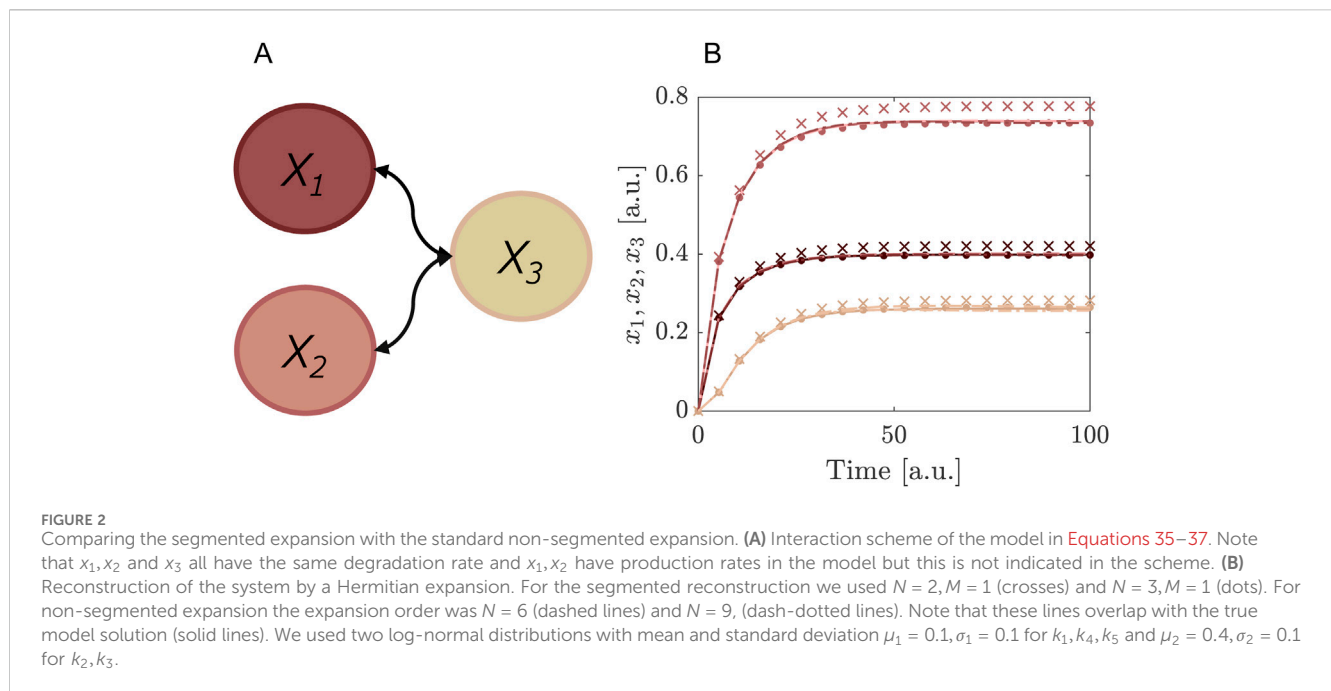


TABLE 1 Benchmarking of segmented and non-segmented expansion. An overview of the number of model evaluations N_λ , the number of summation terms N_Σ and the time in seconds spent on summation (t_Σ), for different orders N of expansion, both segmented $M = 1$ and non-segmented $M = 0$. The last column highlights the speed-up factor when segmentation is used (keeping N_λ constant), i.e., $N = 2, M = 1$ is roughly 1,000 times faster than $N = 6$.

N	M	N_λ	N_Σ	t_Σ [s]	Speed-up factor
6	0	7,776	1.81E + 08	31.96	
9	0	59,049	1.05E + 10	5,402	
2	1	7,776	3.07E + 03	0.03	1,065
3	1	59,049	1.77E + 05	0.17	31,776



In this network, the proteins x_1 and x_2 are produced at rates k_1 and k_4 . Proteins x_1 and x_2 reversibly bind to form species x_3 , with binding rate k_2 and unbinding rate k_3 . All three proteins are degraded at the rate k_5 . These interactions are visualized in a

reaction scheme in Figure 2. The ODEs for the system given by Equations 31–34 are:

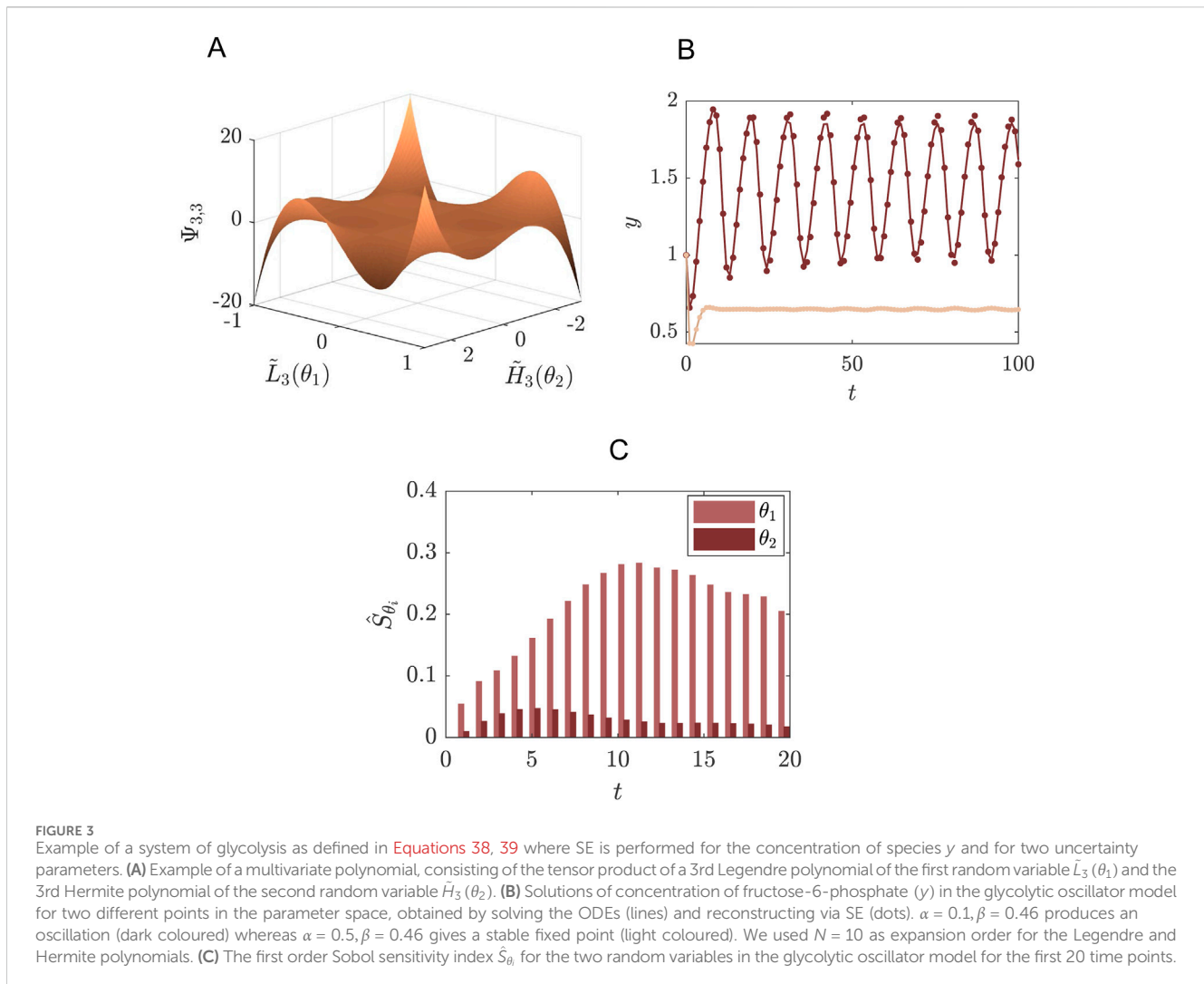
$$\dot{x}_1 = k_1 - k_2 x_1 x_2 + k_3 x_3 - k_5 x_1 \tag{35}$$

$$\dot{x}_2 = k_4 - k_2 x_1 x_2 + k_3 x_3 - k_5 x_2 \tag{36}$$

$$\dot{x}_3 = k_2 x_1 x_2 - k_3 x_3 - k_5 x_3. \tag{37}$$

We use for the parameters $k_1 - k_5$ log-normal distributions and expand the functions $x_1 - x_3$ in terms of Hermitian polynomials. In Figure 2 we compare the results of the segmented expansion, see Equations 14–20, with the non-segmented expansion and the exact results. For this comparison we chose the degree of expansion and segmentation granularity such that the same number of model evaluations were required. We found that the subsequent summation to reconstruct the solutions for the differential equations improved by factors of 1,000–30,000 when using the segmented expansion. See Table 1. As mentioned in the Methods section, this improvements stems from the large reduction in the number of terms to be summed over in the segmented case compared to the non-segmented expansion.

Dividing the parameter intervals into smaller sub-intervals is a relatively straightforward and simple way to circumvent huge computation times. Other, more intricate methods have been developed to tackle models with an even larger amount of parameters (Blatman and Sudret, 2010b; Xiu, 2007; Nobile et al., 2008). For example, using an adaptive algorithm that is based on classical statistical learning tools can result in a “sparse” SE, that consists of only the significant coefficients in the expansion, thereby reducing the computational cost. This method has been tested on models of stochastic finite element analysis with up to 21 parameters (Blatman and Sudret, 2010b).



3.3 Example III. Glycolytic oscillator: different uncertainty PDFs and global sensitivity

Living cells obtain energy by breaking down sugar in the biochemical process called glycolysis. In yeast cells, this glycolysis was observed to behave in an oscillatory fashion, where the concentration of various intermediates were increasing and decreasing within a period of several minutes (Hess and Boiteux, 1968). This glycolytic oscillator can be modelled as a two-component system with a negative feedback (Strogatz, 1994):

$$\dot{x} = -x + \alpha y + x^2 y \quad (38)$$

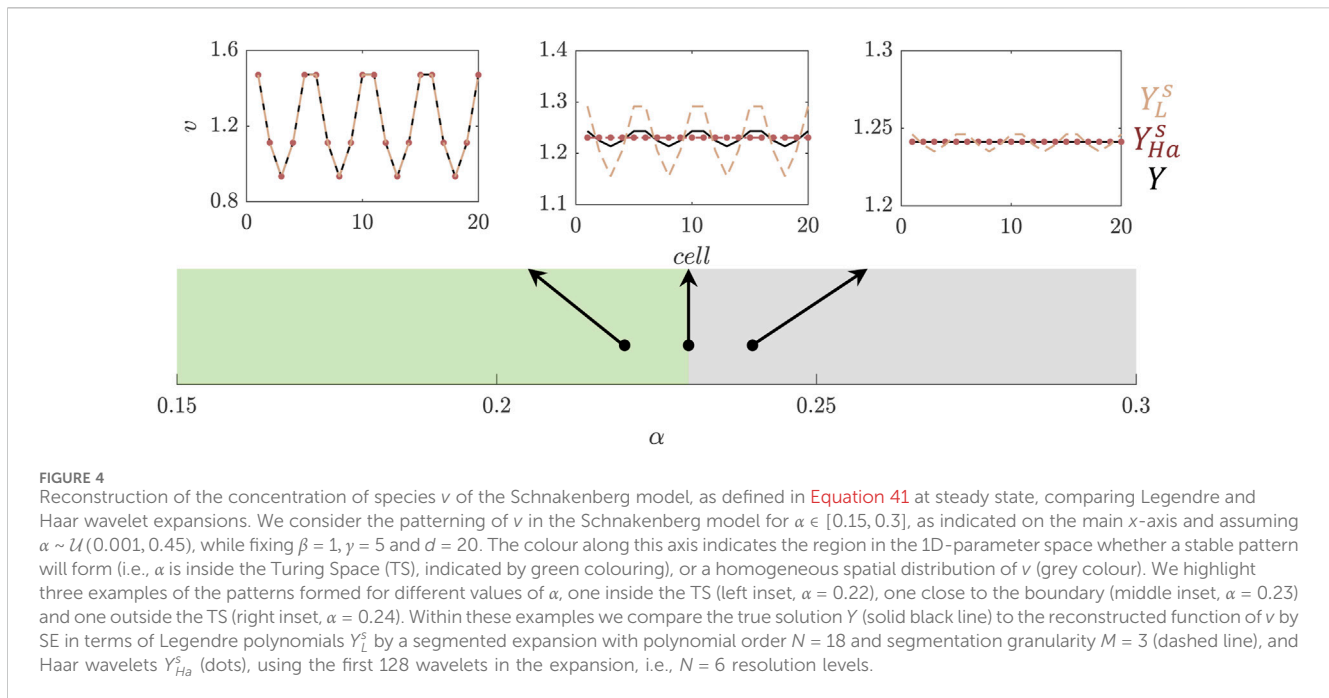
$$\dot{y} = \beta - \alpha y - x^2 y. \quad (39)$$

where x and y are the concentrations of ADP (adenosine diphosphate) and F6P (fructose-6-phosphate) and α, β are kinetic parameters. Depending on the values of α and β the system will be in a stable limit cycle or a stable fixed point (Strogatz, 1994). In this example, we assume α to be uniformly distributed on the interval $[0.1, 0.5]$ and β to follow a lognormal distribution with $\mu = 0.3$ and $\sigma = 0.1$. Because the two uncertainty parameters come from different distributions, we use as basis functions multivariate

polynomials $\Psi_{N,M}$ which are tensor products of the univariate polynomials. In this case, Legendre polynomials are used to expand $\alpha(\theta_1)$ and Hermite polynomials for $\beta(\theta_2)$, where $\theta_1 \sim \mathcal{U}(-1, 1)$ and $\theta_2 \sim \mathcal{N}(0, 1)$. Note that the eigenvalues and eigenvectors used in the expansion are readily available from the matrices given in Equations 28, 29, corresponding to Legendre and Hermite polynomials, respectively. This results in a mixed polynomial for the overall expansion. As illustration, we show in Figure 3 the shape of the product of a 3rd order Legendre polynomial \tilde{L}_3 and a 3rd order Hermite polynomial \tilde{H}_3 , giving $\Psi_{3,3} = \tilde{L}_3 \cdot \tilde{H}_3$.

The distributions of the uncertainty parameters were chosen such that they include the bifurcation point from stable limit cycle to the stable fixed point (Figure 3). For the purpose of this example we are interested in the concentration of y only and therefore reconstruct this model response using SE. A good approximation is obtained with a truncation degree of the SE of $N = 10$. This value is relatively high, due to the bifurcation in the system. However, this case shows that convergence can be reached using SE despite such challenges. It is important to note that the computational costs are still quite low.

In post-processing we may use the SE coefficients to determine the first order Sobol indices for the parameters α and β at each time



point, providing a representation of the global sensitivity based on variance decomposition. The Sobol indices are readily available from the SE coefficients, as shown in Section 2.2.6. They have the advantage of being global measures of sensitivity. In Figure 3 we show the first order Sobol indices given in Equation 25–27 for the first and second random variable. They indicate the contribution to the total output variance of either θ_1 or θ_2 individually. Higher-order terms would give an indication of interaction effects between θ_1 and θ_2 , which are also readily available from the SE coefficients but are not considered here for brevity.

3.4 Example IV. Schnakenberg model: dealing with spatial discontinuities

In this example we introduce a spatial component. We consider the Schnakenberg model, which is one of the simplest, but yet realistic two-species system that can produce spatially oscillating solutions and therefore has become a prototype for reaction diffusion systems. The Schnakenberg model consists of the following (dimensionless) equations for the concentrations u and v of the two species (D Murray, 2007)

$$\dot{u} = \nabla^2 u + \gamma(\alpha - u + u^2 v), \quad (40)$$

$$\dot{v} = d\nabla^2 v + \gamma(\beta - u^2 v). \quad (41)$$

α, β are reaction rates, γ a scale parameter and d the ratio of diffusion constants between the species u and v . The species u is auto-catalytically produced by the $u^2 v$ term in Equation 40, whereby species v is consumed. There are certain combinations of the parameters α, β, γ and d for which the system will exhibit a stable pattern (Liu et al., 2013); this region of parameter space is called the Turing space (TS). For illustrational purposes we limit the number of uncertainty parameters to one: the parameter α , fixing the other parameters at $\beta = 1, \gamma = 5$ and

$d = 20$. We assume α to be distributed as $\alpha \sim \mathcal{U}(0.001, 0.45)$ and determine the TS for a range of α (Figure 4) using linear stability analysis [for details see (D Murray, 2007; Liu et al., 2013)]. To that end, the model is simulated on a 1D grid of 20 cells. We focus on the concentration of species v at steady state and compare expansions in terms of Legendre polynomials and of Haar wavelets.

SE using polynomial basis functions is known for being inaccurate in regions that contain discontinuities (Le Maître and Knio, 2010; Le Maître et al., 2004; Joel Chorin, 1974). In this example, the lack of convergence in SE can be seen along the boundary of the patterning space (TS) in Figure 4, where the expansion by Legendre polynomials is indicated by the dashed lines. For the reconstruction of concentration v in terms of Legendre polynomials, we used a segmented expansion with $N = 18$ and a segmentation granularity of $M = 3$, leading to a total of 126 model evaluations used in the expansion. To show that Haar wavelets perform much better in such a region, we also consider an expansion in terms of Haar wavelets (Equations 21–24). As resolution level we take $N = 6$, which means a total of $N_w = 128$ wavelets are used in the expansion. In Figure 4 the performances of Legendre polynomials and Haar wavelets are compared in the vicinity of $\alpha = 0.23$ (middle inset, Figure 4), showing that the Haar-wavelets provide an improvement in accuracy at the bifurcation point, while using the same number of model evaluations (i.e., the same amount of information and computational cost) for the expansion.

3.5 Example V. Trichome patterning: dealing with spatial discontinuities

As an extra example of pattern formation we consider a model that describes trichomes. Trichomes are hairs found on the epidermal layer of leaves. In *Arabidopsis Thaliana* these trichomes form a regular pattern, where each trichome is

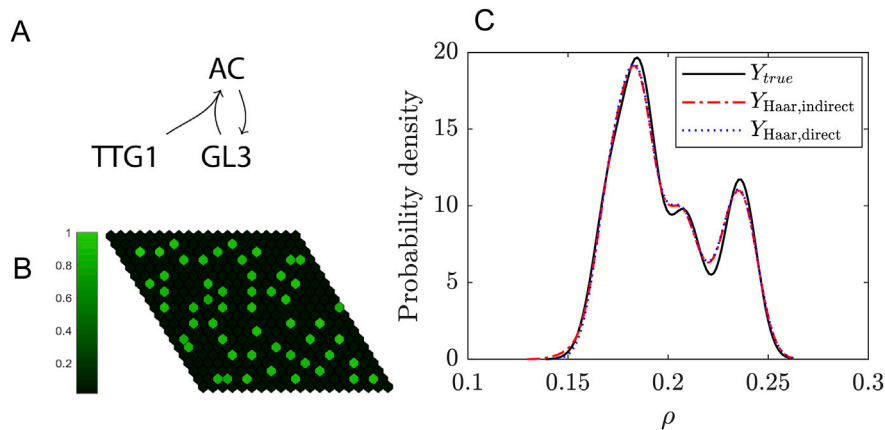


FIGURE 5

Uncertainty quantification for the trichome system. (A) Schematic of the model. (B) Example of a simulation of the model on a 20-by-20 grid. The colorbar indicates the levels of the activating complex (AC) in the cells relative to the maximum. (C) Probability density function of the trichome density in the Turing Space using either the indirect (dashed red line) or direct expansion (dotted blue line) method and for comparison the solution of the real model (solid black line). A resolution level of 3 (i.e., a total of 16 wavelets) has been used for the expansion.

separated by around three to four epidermal cells (Hülkamp, 2004). The model studied here consists of three proteins and their interactions which can explain features of trichome patterning (Bouyer et al., 2008; Pesch and Hülkamp, 2004). Protein TRANSPARENT TESTA GLABRA1 (TTG1) binds to the transcription factor GLABRA3 (GL3) which together form a trichome-promoting complex, called the activating complex (AC) (Bouyer et al., 2008). Experimental data suggests that TTG1 is depleted from cells neighbouring a trichome (Bouyer et al., 2008). For this reason the interaction between TTG1 and GL3 is modelled in a substrate-depletion form (Figure 5), where TTG1 acts as a substrate for the formation of AC (Bouyer et al., 2008). After non-dimensionalisation this model consists of four parameters, none of which have been experimentally determined, highlighting the substantial amount of uncertainty within this model (Pesch and Hülkamp, 2009; Scheres, 2000). Here, we examine the propagation of uncertainty in the parameters to the predicted pattern.

The trichome patterning is described by the following set of coupled ODEs (Bouyer et al., 2008):

$$\dot{\text{TTG1}}_j = \alpha - \lambda \text{TTG1}_j - \text{TTG1}_j \text{GL3}_j + \delta \hat{L} \text{TTG1}_j \quad (42)$$

$$\dot{\text{GL3}}_j = \beta \text{AC}_j^2 - \text{GL3}_j - \text{TTG1}_j \text{GL3}_j \quad (43)$$

$$\dot{\text{AC}}_j = \text{TTG1}_j \text{GL3}_j - \text{AC}_j, \quad (44)$$

where α , λ , δ and β are parameters in the model and \hat{L} describes the coupling between the cells. The subscript j indicates the j^{th} cell. We solve Equations 42–44 for 400 cells, grouped on a hexagonal grid of 20 by 20 cells.

In this example we focus on the parameter α , the basal production for TTG1. We assume this parameter to be uniformly distributed on the interval [0.4, 0.9]. We are interested in the number of trichomes that are predicted by the model, therefore we consider the trichome density ρ (total number of trichomes divided by the total number of cells in the simulated tissue) as the model response of interest. The number of trichomes is determined by simulating the system until steady state is reached and counting

the number of cells for which the concentration of AC exceeds a threshold. The amount of AC is considered to be an indicator for trichome cell fate in plants. However, the biological threshold for this is unknown. We set this threshold to the half-maximum of AC in the system. This leads to the following description of trichome density:

$$\rho = \frac{|\mathcal{T}|}{N}, \quad (45)$$

where \mathcal{T} is the set of cells which exceed the AC threshold, \mathcal{J} is the set of all cells on the grid, N is the total number of cells and $|\mathcal{T}|$ is the cardinality of \mathcal{T} .

Our present goal is to study the uncertainty in ρ as a result of the uncertainty in α . To this end, we employ two different approaches. For both approaches we first transform α to a standard uniform variable by $\alpha = T^{-1}(\theta)$, using the transform function for a uniform variable given in Equation 8. The first approach, referred to as the indirect approach, is the same as used in Example IV in Section 3.4. To reconstruct the concentration at steady state for all cells, we expand the concentration of AC in terms of Haar wavelets. From the result we may determine ρ . In this process we discriminate between cases where there is a pattern and where there is no pattern. Through linear stability analysis we determine beforehand whether a pattern will form or not, i.e., whether the chosen parameter set is in the Turing Space (TS) (D Murray, 2007). For a certain realisation θ we can determine ρ (Equation 45) by:

$$\rho(\theta) = \begin{cases} \frac{|\mathcal{T}^s|}{N} & \text{if } \theta \in \text{TS} \\ 0 & \text{if } \theta \notin \text{TS} \end{cases}, \quad (46)$$

where \mathcal{T}^s is the set of trichomes as determined from the reconstructed AC concentration profile.

Our second, direct approach is to directly reconstruct ρ as

$$\rho^s(\theta) = \sum_{l=1}^N Y(\lambda^{(l)}) u_1^{(l)} \psi_l^s(\theta). \quad (47)$$

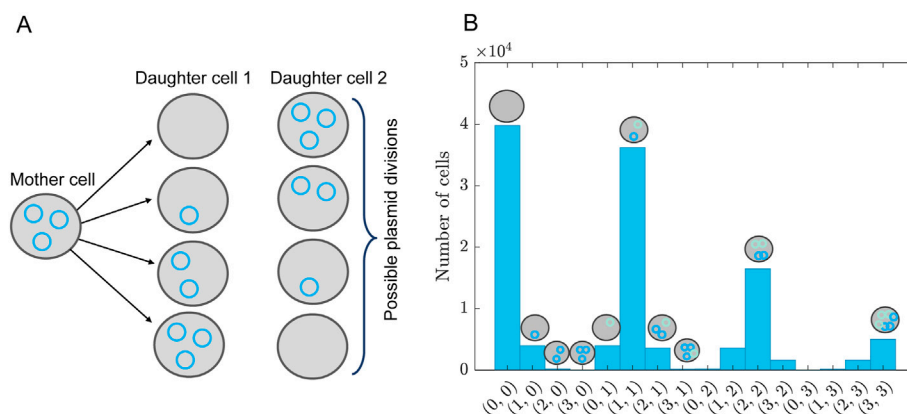


FIGURE 6 Plasmid distribution model. **(A)** Schematic overview of plasmid distribution upon cell division, showing all possible combinations of how to divide three plasmids among two daughter cells. This is for a single type of plasmid. The probability of finding these combinations is given by Equation 56. **(B)** Distribution of plasmid combinations for two types of plasmids, assuming a bivariate Poisson distribution given by Equation 59 with $\lambda = 1$ and $\xi = 0.9$ and a population of 100.000 cells. The x-axis indicates the number of type n and m plasmids as a tuple (n, m) , visualized in a few example cells above the bars, where the light-blue circles indicate plasmids of type n and the dark blue plasmids of type m . The y-axis shows the number of cells that have a certain amount of n, m -type plasmids.

Similarly, as we did for $\rho(\theta)$ we can define $Y(\lambda^{(l)})$ as

$$Y(\lambda^{(l)}) = \begin{cases} \frac{|T|}{N} & \text{if } \lambda^{(l)} \in TS \\ 0 & \text{if } \lambda^{(l)} \notin TS \end{cases} \quad (48)$$

In other words, we only solve the system and determine the trichome density if the parameter set falls within the Turing space. This lends robustness to the SE for the non-smooth parts of the function $Y(\theta)$ and at the same time limits the amount of simulations to be performed, as the non-patterning parameter combinations need not be solved for.

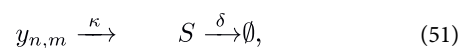
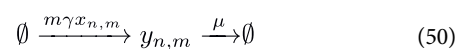
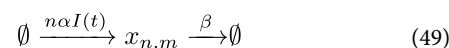
One of the nice features of this Example is that it illustrates there are multiple ways in which the uncertainty in the output can be captured: first, an indirect method (Equation 46) where the model output consists of concentration profiles from which the pattern features have to be extracted in post-processing, and second, by taking the density as model output (Equations 47, 48). In Figure 5 results for the indirect and direct approaches are compared. We conclude that both have similar levels of accuracy and in both cases the expansions converge to the real solution at resolution level $N = 6$, which means a summation of 128 wavelets. The PDF in Figure 5 is constructed using 10^3 samples which costs 4.7 s for the wavelet reconstruction as opposed to 80.9 s for solving the full model a 1,000 times in an MC approach.

3.6 Example VI. Plasmid transfection: dealing with correlated parameters

It can happen that in parameter space a structure occurs, i.e., that the multivariate joined probability cannot be written as a product of univariate distributions. In this section we exemplify how to handle such a case. As an example we choose the transfection of mammalian cells (e.g., Human embryonic kidney cells, HEK293) with two plasmids: plasmid pl_1 with the construct for induction

(e.g., through chemical or optogenetics (Yoshida and Sato, 2009)), and plasmid pl_2 with the reporter construct. In short, one of the plasmids is required to activate the cell (pl_1), the other is used for read-out (pl_2). In the part below we describe a model that predicts the distribution of both types of plasmids among a population of cells, assuming that the plasmids will have to be distributed among daughter cells upon division (Figure 6A). We assume that this distribution of plasmids is accurately described by a bivariate Poisson distribution (example given in Figure 6B), as we will argue below.

Denoting by n the number of pl_1 plasmids and by m the number of pl_2 plasmids in a specific cell, the corresponding reaction scheme reads:



where $x_{n,m}$ and $y_{n,m}$ are the concentrations of molecule x and y , resp., in a cell with plasmid composition nm . S is the concentration of the reporter molecule in the bulk (i.e., the read-out produced by the reporter construct of plasmid pl_2). α and γ are the corresponding production rates per plasmid and β, μ are the degradation rates. $I(t)$ is the external induction signal (i.e., the signal required to activate the plasmid pl_1). In what follows we will set $I(t) = 1$ for $t \geq 0$.

For sake of clarity we keep the system simple and ignore all complicating effects, such as gene expression noise, maturation of the reporter construct (e.g., GFP). However, for transient transfections of the mammalian cells, we need to take into

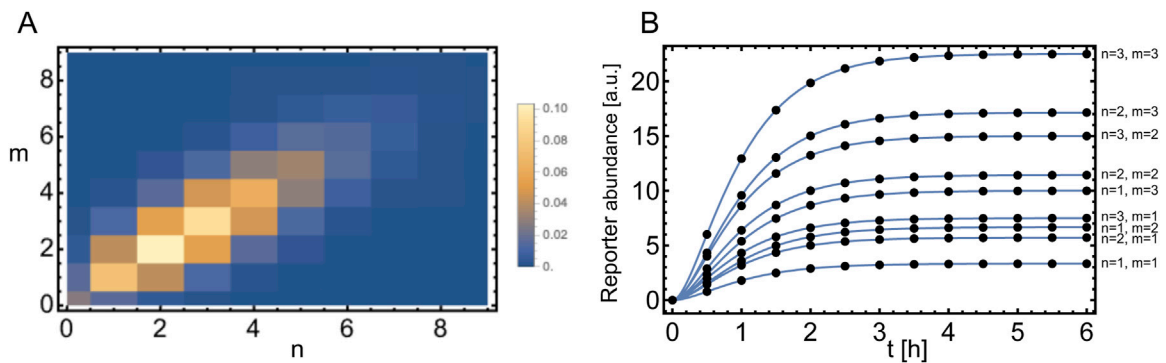


FIGURE 7 Expansion with respect to a bivariate Poisson distribution. **(A)** Bivariate Poisson distribution given by Equation 59 with $\lambda = 3$ and $\xi = 2.4$. **(B)** Comparison of the exact solution for $y_{n,m}(t)$ given by Equation 53 (solid lines) to the spectral expansion with $N = 6$ (dots). Parameters: $\alpha = 2 (Vh)^{-1}$, $\beta = 2 h^{-1}$, $\gamma = 6 h^{-1}$, $\eta = 0.5 V^{-1}$, $\mu = 1.5 h^{-1}$, $\kappa = 0 h^{-1}$.

account the distribution of plasmids among the cells, which is not constant due to cell division. The set of Ordinary Differential Equations describing the reactions given by Equations 49–51 read:

$$\dot{x}_{n,m} = \alpha xI(t) - \beta x_{n,m} \tag{52}$$

$$\dot{y}_{n,m} = \frac{m\gamma x_{n,m}}{1 + \eta x_{n,m}} - (\mu + \kappa)y_{n,m} \tag{53}$$

$$\dot{Q}_{n,m} = \omega \left(1 - \frac{\|Q\|_1}{N_{\max}} \right) \sum_{p,q} L_{n,m}^{p,q} Q_{p,q} \tag{54}$$

$$\dot{S} = \nu \kappa \sum_{n,m} Q_{n,m} y_{n,m} - \delta S. \tag{55}$$

The parameter η captures the saturation of the expression capacity of the cells. We set $\eta = 0.2 V^{-1}$ throughout the calculations. V is the unit of the volume of the cells, e.g., $V = \mu m^3$. $Q_{n,m}$ is the number of cells that have n, m plasmids which are growing by ω . To follow the partition of plasmids upon division, the tensor L is introduced and described in Equations 57, 58 below. $\|Q\|_1 = \sum_{n,m} Q_{n,m} = N_{\text{tot}}$ which is the total number of cells at time t and N_{\max} is the maximum number of cells due to environmental constraints. The growth rate is set to $\omega = 0.034 h^{-1}$ based on a doubling time of 20 h as being measured for HEK293T cells (Tan et al., 2021; Zhang et al., 2024). S is the reporter [e.g., SEAP (Berger et al., 1988)] accumulated in the bulk due to secretion by the cells with rate κ . For non-secreted reporters (e.g., YFP) for single cell measurements, κ is set to zero. ν is a volume correction factor, set to $\nu = 10^{-8}$.

We assume that the rate of dilution of the plasmids corresponds to the growth rate of the cells, i.e., plasmids are lost upon division. In reality, the plasmids are also degraded, but we assume the time scale of this degradation is much longer than the experimental duration and can therefore be ignored. To model the partition of plasmids upon cell division we assume that any partition of the number of plasmids is possible, see, for example, Figure 6A. The probability of finding a certain partition is given by the binomial distribution

$$P_n^k = \theta_n^k \binom{k}{n} (1-q)^{k-n} q^n, \tag{56}$$

where n is the number of plasmids in the mother cell, k the number of plasmids in one of the daughter cells and q the probability, which for equal cell division is $q = 0.5$. θ_n^k is the discrete unit step function

with $\theta_n^k = 1$ for $k \geq n$ and $\theta_n^k = 0$ else. We can write the rate of change in the pools of cells as

$$\dot{Q}_{n,m} = 2 \sum_{k,p} P_n^k P_m^p Q_{k,p} - Q_{n,m} = \sum_{k,p} (2P_n^k P_m^p - \delta_n^k \delta_m^p) Q_{k,p}. \tag{57}$$

$\delta_n^k = 1$ for $k = n$ and $\delta_n^k = 0$ else. From this, we can define the tensor used in Equation 54:

$$\dot{Q}_{n,m} = \Omega \left(2 \sum_{k,p} P_n^k P_m^p Q_{k,p} - Q_{n,m} \right) = \Omega \sum_{k,p} (2P_n^k P_m^p - \delta_n^k \delta_m^p) Q_{k,p}. \tag{58}$$

We treat the plasmid uptake as a Poisson process, i.e., the number of plasmids inside a cell is Poisson distributed. Assuming that the mean number of plasmid taken up is the same for pl_1 and pl_2 and a correlation exists between the uptake of two plasmids, the distribution is given by (Berkhout and Plug, 2004):

$$P(n, m) = \frac{(\lambda - \xi)^m}{(\lambda - \xi)^n} \frac{(-1)^n \xi^n}{n!m!} U \left(-n, 1 - n + m, -\frac{(\lambda - \xi)^2}{\xi} \right) e^{-2\lambda + \xi}, \tag{59}$$

where U is Kummer’s confluent hypergeometric function (Olver, 2010), λ is the mean number of plasmids taken up, and $0 \leq \xi \leq \lambda$ is the correlation parameter. Note that for $\xi \rightarrow 0$ we find:

$$\lim_{\xi \rightarrow 0} P(n, m) = P(n)P(m), \quad \text{with } P(x) = \frac{\lambda^x}{x!} e^{-\lambda}. \tag{60}$$

The correlation between n and m reads:

$$\text{corr}(n, m) = \frac{\text{cov}(n, m)}{\sqrt{\text{var}(n)\text{var}(m)}} = \frac{\xi}{\lambda}. \tag{61}$$

In Figure 7A we show $P(n, m)$ for $\lambda = 3$ and $\xi = 2.4$. The asymmetry due to the correlation $\text{corr}(n, m) = 0.8$ can be clearly seen. The orthonormal basis function $\phi_p(n)$ with respect to the Poisson distribution read (Ogura, 1972):

$$\phi_p(n) = {}_2F_0 \left(-p, -n; ; -\frac{1}{\lambda} \right) \sqrt{\frac{\lambda^p}{p!}}. \tag{62}$$

${}_2F_0$ is a generalised hypergeometric function (Olver, 2010). The matrix \hat{B} defined in Equation 4 can be calculated analytically and is given by:

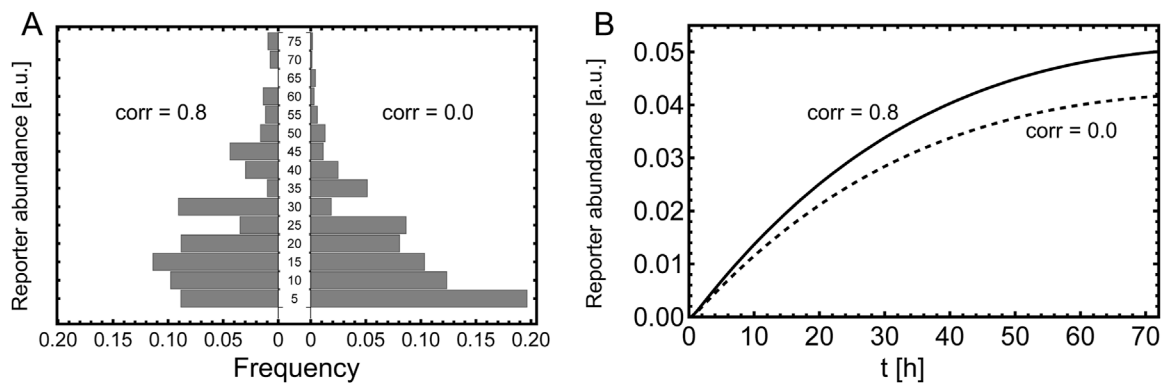


FIGURE 8

Measurement of the reporter. Comparison of correlated and un-correlated plasmid uptake. (A) Histogram of the single cell readout, measured 6 h after induction. The parameters are the same as stated in Figure 7 and the histogram is calculated according Equation 64 with $g = 5$ (B) Reporter abundance in the bulk measured for 3 days. The parameters are the same as in A, besides $\kappa = 3 h^{-1}$, $\delta = 0.01 h^{-1}$ and $\nu = 10^{-8}$. We used for the calculations the expansion of $y_{n,m}$ with $N = 6$.

$$\hat{B}_{n,m} = (n + \lambda)\delta_{nm} - \sqrt{n\lambda}\delta_{n-1,m} - \sqrt{m\lambda}\delta_{n,m-1}. \quad (63)$$

Note the dependence on the parameter λ , in contrast to the matrices given in Equations 28, 29.

We first consider a system in which the mammalian cells do not secrete the reporter molecule, e.g., YFP. In this case we set $\kappa = 0$, and consequently $S(t) = 0, \forall t$. In Figure 7B we show $y_{n,m}(t)$ for different plasmid compositions and compare the expansion of $y_{n,m}(t)$, using Equations 11–13 and Equations 59–63, of the order $N = 6$ (dots) to the exact results (solid lines). A question of interest is what is the measured distribution of fluorescence intensities of the mammalian cells. The distribution of reporter molecule abundance is given by:

$$P(z, t) = \sum_{n,m}^{\infty} \delta_{z, \chi_{n,m}(t)} \frac{Q_{n,m}(t)}{\|Q\|_1}, \quad (64)$$

where $\chi_{n,m}(t)$ is a binning function: $\chi_{n,m}(t) = g \lfloor y_{n,m}(t)/g \rfloor$ (other binning methods are also possible, of course). Note that for sake of simplicity we ignore the transient phase after cell division in which the dynamic of $y_{n,m}$ adopts to the new, reduced plasmid composition. This is also a reasonable simplification considering that the steady state of $y_{n,m}$ is roughly reached after 4 h, as can be seen in Figure 7B, in contrast to a cell doubling time of 20 h. In Figure 8A we show the distribution $P(z, t = 50)$ for $g = 5$ for $\xi = 2.4$ (left) and $\xi = 0$ (right); we used the expansion with $N = 6$ for the calculations in Equation 64.

Next, we consider a segregated reporter, e.g., SEAP. In this case the reporter abundance in the bulk is measured instead of single cell measurements. The temporal evolution of the bulk reporter is governed by Equations 52–55. In Figure 8B one can see that the dynamics of the reporter is clearly different due to correlation of the plasmid uptake.

4 Discussion and conclusion

In this paper we have presented an efficient and widely-applicable version of spectral expansion to quantify the effect of parameter uncertainty on model outcomes. The present scheme is based on non-intrusive spectral projection and makes use of expansions in

terms of some orthonormal set of basis functions, e.g., polynomial functions or wavelets. The orthonormal properties of those basis functions are utilized to develop a novel scheme to determine the expansion coefficients. It has the attractive property that it is computationally very fast. The scheme is similar to the Golub-Welsch algorithm known from Gaussian quadrature (Golub and Welsch, 1969). In the latter procedure the roots of the polynomials used play an essential role (Bruno, 2007). In our new scheme this role is played by the eigenvalues of a matrix, which can be calculated very fast. Our method does not require modification of the model equations. This is in general the main advantage of non-intrusive methods: there is no need to recast the model into a probabilistic framework. Instead, the random behaviour of parameters is accounted for through a set of deterministic simulations of the process for a restricted number of parameter values. These values are chosen such that they reflect the uncertainties in the parameters. To test the performance of our method we applied it to a number of different models (Example I) a model of exponential decay (Example II) a biochemical reaction network (Example III) the glycolytic oscillator (Example IV) the Schnakenberg model, and (Example V) a trichome model. Examples IV and V deal with spatial pattern formation, and, finally, Example VI illustrates SE for correlated parameters. For each test case, the results of the SE are compared to, if available, analytical solutions and/or Monte Carlo simulations. In these comparisons we mostly focus on the accuracy of SE. Although the computational advantage is an important reason for using SE techniques, we do not focus on that aspect since it has already been extensively explored elsewhere; see, for example, (Zein et al., 2013; Bruno, 2007; Fajraoui et al., 2017).

The accuracy of the reconstruction by SE depends on the choice of expansion order and the appropriate choice of basis functions. While the latter choice is determined by the PDFs of the input parameters, the choice of expansion order has to be chosen by the user. For example, in Examples I and III we chose $N = 5$ and $N = 10$, respectively. These choices were based on careful observation of the convergence properties of the method.

In some cases the expansion order has to be chosen prohibitively large. For such a situation we propose an extended approach that segments the parameter interval into subintervals, essentially

zooming in on these sub-intervals such that a lower expansion order can be used in each sub-interval. In Example II we have shown that this segmentation approach can greatly reduce the computational costs, thus providing a way to circumvent the curse of dimensionality. Such adaptations are nearly always required in high-dimensional cases and segmentation is a relatively simple and straight-forward method to tackle dimensionality problems yielding a piecewise continuous approximation to the original function. It is an alternative for so-called sparse SE methods, that utilise only a small subset of the basis functions in order to limit the amount of model evaluations (Blatman and Sudret, 2010b; Xiu, 2007; Nobile et al., 2008).

Convergence of the SE may be poor in regions of the parameter space around a bifurcation (Ghanem et al., 2017), due to the use of smooth basis functions to represent non-smooth model behaviour. This effect is often illustrated by the Gibbs phenomenon in Fourier expansions, where the spectral basis consisting of smooth sine and cosine functions is not suitable, giving rise to slow and even lack of correct convergence. Since smooth functions like the Hermite and Legendre polynomials will fail to describe steep or discontinuous solutions, we explored the use of Haar wavelets. Wavelets naturally allow localised decompositions and this leads to more robust behaviour (Le Maître et al., 2004). We show in Example IV (Schnakenberg model) the advantages of using Haar wavelets over polynomials by focusing on the region in parameter space where the system jumps from spatially heterogeneous to spatially homogeneous dynamics. Although around this bifurcation point an expansion in terms of Haar wavelets also turned out to show slow convergence, the accuracy of the expansion is much better than when Legendre polynomials are used. In the vicinity of bifurcations Haar wavelets thus provide a useful tool for biological systems which feature discontinuities. In Example V (trichome pattern formation) we have highlighted the flexibility of the method: some quantities, e.g., the scalar quantity of trichome density, can either be directly expanded or indirectly. By making use of that adaptability the number of model evaluations can be reduced while the level of accuracy is maintained. Finally, in Example VI we illustrate how to handle correlated parameters by means of correlated plasmid uptake by mammalian cells. We show how single-cell or bulk readout can be calculated using SE.

Overall, the approach presented here consists of a number of easy-to-implement steps and is applicable to a great variety of systems that would be computationally costly when analysed in the context of uncertainty quantification in the usual way. We therefore believe that this approach could provide a valuable asset for the toolkit of computational systems biology.

References

- Banga, J. R., and Balsa-Canto, E. (2008). Parameter estimation and optimal experimental design. *Essays Biochem.* 45, 195–209. doi:10.1042/BSE0450195
- Barbu, A., and Zhu, S.-C. (2020). *Monte Carlo methods*. Singapore: Springer Singapore. doi:10.1007/978-981-13-2971-5
- Barz, T., Arellano-García, H., and Wozny, G. (2010). Handling uncertainty in model-based optimal experimental design. *Industrial and Eng. Chem. Res.* 49 (12), 5702–5713. doi:10.1021/ie901611b
- Berger, J., Hauber, J., Hauber, R., Geiger, R., and Cullen, B. R. (1988). Secreted placental alkaline phosphatase: a powerful new quantitative indicator of gene expression in eukaryotic cells. *Gene* 66 (1), 1–10. doi:10.1016/0378-1119(88)90219-3
- Berkhout, P., and Plug, E. (2004). A bivariate Poisson count data model using conditional probabilities. *Stat. Neerl.* 58 (3), 349–364. ISSN 1467-9574. doi:10.1111/j.1467-9574.2004.00126.x
- Blake, W. J., Kaern, M., Cantor, C. R., and Collins, J. J. (2003). Noise in eukaryotic gene expression. *Nature* 422 (6932), 633–637. doi:10.1038/nature01546
- Blatman, G., and Sudret, B. (2010a). Efficient computation of global sensitivity indices using sparse polynomial chaos expansions. *Reliab. Eng. and Syst. Saf.* 95 (11), 1216–1229. doi:10.1016/j.ress.2010.06.015
- Blatman, G., and Sudret, B. (2010b). An adaptive algorithm to build up sparse polynomial chaos expansions for stochastic finite element analysis. *Probabilistic Eng. Mech.* 25 (2), 183–197. doi:10.1016/j.probengmech.2009.10.003

Data availability statement

The code is publicly available at: <https://github.com/AnnaDeneer/SpectralExpansion>.

Author contributions

AD: Formal Analysis, Investigation, Methodology, Writing—original draft, Writing—review and editing. JM: Conceptualization, Methodology, Writing—original draft, Writing—review and editing. CF: Conceptualization, Funding acquisition, Project administration, Resources, Supervision, Writing—original draft, Writing—review and editing, Methodology.

Funding

The author(s) declare that financial support was received for the research, authorship, and/or publication of this article. CF received funding from FET-Open research and innovation actions grant under the European Union's Horizon 2020 (CyGenTiG; grant agreement 801041).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The author(s) declared that they were an editorial board member of Frontiers, at the time of submission. This had no impact on the peer review process and the final decision.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fsysb.2024.1419809/full#supplementary-material>

- Bouyer, D., Geier, F., Kragler, F., Schnittger, A., Pesch, M., Wester, K., et al. (2008). Two-dimensional patterning by a trapping/depletion mechanism: the role of TTG1 and GL3 in *Arabidopsis* trichome formation. *PLoS Biol.* 6 (6), e141. ZSCC: NoCitationData [s0]. doi:10.1371/journal.pbio.0060141
- Bowman, A. W., and Azzalini, A. (1997). *Applied smoothing techniques for data analysis: the kernel approach with S-Plus illustrations*, Vol. 18. Oxford: OUP.
- Brian, P. I. (2013). *Mathematical Modeling in systems biology. An introduction*. MIT Press. 0-262-01888-8.
- Bruno, S. (2007). Uncertainty propagation and sensitivity analysis in mechanical models—contributions to structural reliability and stochastic spectral methods. *Habilit. Dir. Des. Rech. Univ. Blaise Pascal, Clermont-Ferrand, Fr.* 147.
- Christian, P. (2004). “Robert and George Casella,” in *Monte Carlo statistical methods* (New York, New York, NY: Springer Texts in Statistics. Springer). 978-1-4419-1939-7 978-1-4757-4145-2. doi:10.1007/978-1-4757-4145-2
- D Murray, J. (2007). *Mathematical biology: I. An introduction*, Vol. 17. Springer Science and Business Media.
- Eldred, M., Webster, C., and Constantine, P. (2008). “Evaluation of non-intrusive approaches for wiener-askay generalized polynomial chaos,” in *49th AIAA/ASME/ASCE/AHS/ASC structures, structural dynamics, and materials conference, 16th AIAA/ASME/AHS adaptive structures conference, 10th AIAA non-deterministic approaches conference, 9th AIAA gossamer spacecraft forum, 4th AIAA multidisciplinary design optimization specialists conference*, 1892.
- Elowitz, M. B., Levine, A. J., Siggia, E. D., and Swain, P. S. (2002). Stochastic gene expression in a single cell. *Sci. (New York, NY)* 297 (5584), 1183–1186. doi:10.1126/science.1070919
- Fajraoui, N., Marelli, S., and Sudret, B. (2017). On optimal experimental designs for sparse polynomial chaos expansions. *arXiv Prepr. arXiv:1703.05312*.
- Geris, L., and Gomez-Cabrero, D. (2016). *Uncertainty in biology*. Springer.
- Ghanem, R., Higdon, D., and Owahdi, H. (2017). *Handbook of uncertainty quantification*, Vol. 6. Springer.
- Golub, G. H., and Welsch, J. H. (1969). Calculation of gauss quadrature rules. *Math. Comput.* 23 (106), 221–230. doi:10.2307/2004418
- Gutenkunst, R. N., Waterfall, J. J., Casey, F. P., Brown, K. S., Myers, C. R., and Sethna, J. P. (2007). Universally sloppy parameter sensitivities in systems biology models. *PLoS Comput. Biol.* 3 (10), 1871–1878. doi:10.1371/journal.pcbi.0030189
- Hess, B., and Boiteux, A. (1968). Mechanism of glycolytic oscillation in yeast, i. aerobic and anaerobic growth conditions for obtaining glycolytic oscillation. *Biol. Chem.* 349 (2), 1567–1574. doi:10.1515/bchm2.1968.349.2.1567
- Hülkamp, M. (2004). Plant trichomes: a model for cell differentiation. *Nat. Rev. Mol. Cell Biol.* 5 (6), 471–480. doi:10.1038/nrm1404
- Ilya, M. S. (2001). Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates. *Math. Comput. Simul.* 55 (1-3), 271–280. doi:10.1016/S0378-4754(00)00270-6
- Ingalls, B. (2008). Sensitivity analysis: from model parameters to system behaviour. *Essays Biochem.* 45, 177–193. ISSN 0071-1365. doi:10.1042/bse0450177
- James, F. (1980). Monte Carlo theory and practice. *Rep. Prog. Phys.* 43 (9), 1145–1189. doi:10.1088/0034-4885/43/9/002
- Joel Chorin, A. (1974). Gaussian fields and random flow. *J. Fluid Mech.* 63 (1), 21–32. doi:10.1017/s0022112074000991
- Kirk, P., Silk, D., and Stumpf, M. P. H. (2016). “Reverse engineering under uncertainty,” in *Uncertainty in biology* (Springer), 15–32.
- Kreutz, C., and Timmer, J. (2009). Systems biology: experimental design. *FEBS J.* 276 (4), 923–942. doi:10.1111/j.1742-4658.2008.06843.x
- Le Maître, O., and Knio, O. M. (2010). *Spectral methods for uncertainty quantification: with applications to computational fluid dynamics*. Springer Science and Business Media.
- Le Maître, O. P., Knio, O. M., Najm, H. N., and Ghanem, R. G. (2004). Uncertainty propagation using wiener-haar expansions. *J. Comput. Phys.* 197 (1), 28–57. doi:10.1016/j.jcp.2003.11.033
- Lev, S. (2014). Noise in biology. *Rep. Prog. Phys.* 77 (2), 026601–026629. doi:10.1088/0034-4885/77/2/026601
- Liu, P., Shi, J., Wang, Y., and Feng, X. (2013). Bifurcation analysis of reaction-diffusion schnakenberg model. *J. Math. Chem.* 51 (8), 2001–2019. doi:10.1007/s10910-013-0196-x
- Martin-Casas, M., and Mesbah, A. (2016). Discrimination between competing model structures of biological systems in the presence of population heterogeneity. *IEEE Life Sci. Lett.* 2 (3), 23–26. doi:10.1109/lsls.2016.2644645
- Mitra, E. D., and Hlavacek, W. S. (2019). Parameter estimation and uncertainty quantification for systems biology models. *Curr. Opin. Syst. Biol.* 18, 9–18. doi:10.1016/j.coisb.2019.10.006
- Nobile, F., Tempone, R., and Webster, C. G. (2008). A sparse grid stochastic collocation method for partial differential equations with random input data. *SIAM J. Numer. Analysis* 46 (5), 2309–2345. doi:10.1137/060663660
- Noble, W. S. (2004). Support vector machine applications in computational biology. *Kernel methods Comput. Biol.* 71 (92), 71–92. doi:10.7551/mitpress/4057.003.0005
- Ogura, H. (1972). Orthogonal functionals of the Poisson process. *IEEE Trans. Inf. Theory* 18 (4), 473–481. ISSN 0018-9448. doi:10.1109/TIT.1972.1054856
- Olver, F. W. J. (2010). in *NIST handbook of mathematical functions hardback and CD-ROM* (Cambridge University Press). Available at: [https://books.google.de/books?id=3115Ph1Q38C.National institute of standards, and technology \(U.S.\)](https://books.google.de/books?id=3115Ph1Q38C.National%20institute%20of%20standards%20and%20technology%20(U.S.))
- Paulson, J. A., Martin-Casas, M., and Ali, M. (2019). Fast uncertainty quantification for dynamic flux balance analysis using non-smooth polynomial chaos expansions. *PLoS Comput. Biol.* 15 (8), e1007308. doi:10.1371/journal.pcbi.1007308
- Pesch, M., and Hülkamp, M. (2004). Creating a two-dimensional pattern *de novo* during *Arabidopsis* trichome and root hair initiation. *Curr. Opin. Genet. and Dev.* 14 (4), 422–427. doi:10.1016/j.gde.2004.06.007
- Pesch, M., and Hülkamp, M. (2009). One, two, three models for trichome patterning in *Arabidopsis*? *Curr. Opin. plant Biol.* 12 (5), 587–592. doi:10.1016/j.pbi.2009.07.015
- Raue, A., Schilling, M., Bachmann, J., Matteson, A., Schelke, M., Kaschek, D., et al. (2013). Lessons learned from quantitative dynamical modeling in systems biology. *PLoS one* 8 (9), e74335. doi:10.1371/journal.pone.0074335
- Renardy, M., Yi, T.-Mu, Xiu, D., and Chou, C.-S. (2018). Parameter uncertainty quantification using surrogate models applied to a spatial model of yeast mating polarization. *PLoS Comput. Biol.* 14 (5), e1006181. doi:10.1371/journal.pcbi.1006181
- Saltelli, A. (2008). *Global sensitivity analysis: the primer*. Chichester, England ; Hoboken, NJ: John Wiley. 978-0-470-05997-5. OCLC: ocn180852094.
- Scheres, B. (2000). Non-linear signaling for pattern formation? *Curr. Opin. plant Biol.* 3 (5), 412–417. doi:10.1016/S1369-5266(00)00105-9
- Schultheiß Araújo, I., Pietsch, J. M., Keizer, E. M., Grees, B., Balkunde, R., Fleck, C., et al. (2017). Stochastic gene expression in *Arabidopsis thaliana*. *Nat. Commun.* 8 (1), 2132–2139. doi:10.1038/s41467-017-02285-7
- Silverman, B. W. (1986). *Density estimation for statistics and data analysis*, Vol. 26. New York, United States: CRC Press.
- Soboń, I. M. (1990). Quasi-Monte Carlo methods. *Prog. Nucl. Energy* 24 (1-3), 55–61. doi:10.1016/0149-1970(90)90022-w
- Streif, S., Petzke, F., Ali, M., Findeisen, R., and Braatz, R. D. (2014). Optimal experimental design for probabilistic model discrimination using polynomial chaos. *IFAC Proc. Vol.* 47 (3), 4103–4109. doi:10.3182/20140824-6-za-1003.01562
- Strogatz, S. H. (1994). “Nonlinear dynamics and chaos: with applications to physics, biology, chemistry, and engineering,” in *Studies in nonlinearity* (Addison-Wesley Pub). 978-0-201-54344-5.
- Sudret, B. (2008). Global sensitivity analysis using polynomial chaos expansions. *Reliab. Eng. and Syst. Saf.* 93 (7), 964–979. doi:10.1016/j.res.2007.04.002
- Sullivan, T. J. (2015). *Introduction to uncertainty quantification*. Springer International Publishing. ISBN ISBN 978-3-319-23395-6.
- Tadeusiewicz, R. (2015). Neural networks as a tool for modeling of biological systems. *Bio-Algorithms Med-Systems* 11 (3), 135–144. doi:10.1515/bams-2015-0021
- Tan, E., Chin, C. S. H., Lim, Z. F. S., and Ng, S. K. (2021). HEK293 cell line as a platform to produce recombinant proteins and viral vectors. *Front. Bioeng. Biotechnol.* 9, 796991. ISSN 2296-4185. doi:10.3389/fbioe.2021.796991
- Thattai, M., and van Oudenaarden, A. (2001). Intrinsic noise in gene regulatory networks. *PNAS* 98 (15), 8614–8619. doi:10.1073/pnas.151588598
- Tsigkinopoulou, A., Hawari, A., Uttley, M., and Breiting, R. (2018). Defining informative priors for ensemble modeling in systems biology. *Nat. Protoc.* 13 (11), 2643–2663. doi:10.1038/s41596-018-0056-z
- van Mourik, S., Ter Braak, C., Stigter, H., and Molenaar, J. (2014). Prediction uncertainty assessment of a systems biology model requires a sample of the full probability distribution of its parameters. *PeerJ* 2, e433. doi:10.7717/peerj.433
- Vincent, D., Sudret, B., and Deheeger, F. (2013). Metamodel-based importance sampling for structural reliability analysis. *Probabilistic Eng. Mech.* 33, 47–57. doi:10.1016/j.probenmech.2013.02.002
- Wiener, N. (1938). The homogeneous chaos. *Am. J. Math.* 60 (4), 897–936. doi:10.2307/2371268
- Wilkinson, D. J. (2007). Bayesian methods in bioinformatics and computational systems biology. *Briefings Bioinforma.* 8 (2), 109–116. doi:10.1093/bib/bbm007
- Xiu, D. (2007). Efficient collocation approach for parametric uncertainty analysis. *Commun. Comput. Phys.* 2 (2), 293–309.
- Xiu, D., and George, Em K. (2002). The wiener-askay polynomial chaos for stochastic differential equations. *SIAM J. Sci. Comput.* 24 (2), 619–644. doi:10.1137/s1064827501387826
- Yoshida, N., and Sato, M. (2009). Plasmid uptake by bacteria: a comparison of methods and efficiencies. *Appl. Microbiol. Biotechnol.* 83 (5), 791–798. ISSN 0175-7598. doi:10.1007/s00253-009-2042-4
- Zein, S., Colson, B., and Glineur, F. (2013). An efficient sampling method for regression-based polynomial chaos expansion. *Commun. Comput. Phys.* 13 (4), 1173–1188. doi:10.4208/cicp.020911.200412a
- Zhang, L., Gao, J., Zhang, Xi, Wang, X., Wang, T., and Zhang, J. (2024). Current strategies for the development of high-yield HEK293 cell lines. *Biochem. Eng. J.* 205, 109279. ISSN 1369703X. doi:10.1016/j.bej.2024.109279