Preconditioned Bayesian Regression for Stochastic Chemical Kinetics

Alen Alexanderian,[1] Francesco Rizzi,[2] Muruhan Rathinam,[3] Olivier P. Le Maître,[4] Omar M. Knio[2]

[1]Institute for Computational Engineering and Sciences
The University of Texas at Austin
Austin, TX 78712

[2]Department of Mechanical Engineering and Materials Science
Duke University
Durham, NC 27708

[3]Department of Mathematics and Statistics
University of Maryland, Baltimore County
Baltimore, MD 21250

[4]LIMSI-CNRS
BP 133, Bt 508
F-91403 Orsay Cedex, France

Corresponding Author: Omar M. Knio
Department of Mechanical Engineering and Materials Science
Duke University
Durham, NC 27708

Phone: +1 919 660 5344
Fax: +1 919 660 8963
Email: omar.knio@duke.edu

Submitted to: *Journal of Scientific Computing*
June 2012

Revised May 2013

Abstract

We develop a preconditioned Bayesian regression method that enables sparse polynomial chaos representations of noisy outputs for stochastic chemical systems with uncertain reaction rates. The approach is based on the definition of an appropriate multiscale transformation of the state variables coupled with a Bayesian regression formalism. This enables efficient and robust recovery of both the transient dynamics and the corresponding noise levels. Implementation of the present approach is illustrated through applications to a stochastic Michaelis-Menten dynamics and a higher dimensional example involving a genetic positive feedback loop. In all cases, a stochastic simulation algorithm (SSA) is used to compute the system dynamics. Numerical experiments show that Bayesian preconditioning algorithms can simultaneously accommodate large noise levels and large variability with uncertain parameters, and that robust estimates can be obtained with a small number of SSA realizations.

1 Introduction

In this paper, we develop Bayesian preconditioning approaches for computing spectral representations of the of the dependence of the mean and variance of the trajectory of a stochastic chemical system on uncertain reaction rates. For a fixed set of reaction rates, simulation of such systems is performed using Gillespie's Stochastic Simulation Algorithm (SSA). Straightforward Monte Carlo approaches, which involve generating a sufficiently large sample of the random reaction rate vectors and then running a large number of SSA simulations to estimate the first moments for each sample vector are exceedingly demanding, especially for high-dimensional stiff systems. Furthermore, assessment of the impact of individual uncertain reaction rates on the solution variability (global sensitivity) is not trivial within such simulation approaches. Such considerations motivate development of suitable and efficient computational approaches.

An increasingly popular class of methods for quantifying uncertainty in dynamical systems utilizes Polynomial Chaos (PC) expansions of the model variables [2, 11, 17, 27–29, 31, 32, 36, 37, 43, 47, 51]. These methods provide a representation of the model variables in terms of a (truncated) spectral expansion in an appropriate orthogonal polynomial basis. Once available, the spectral expansion can be readily used to efficiently compute the statistical properties of the model variables, and to characterize the sensitivity of specific observables with respect to specific components of the random reaction rate vector.

In general, there are two major approaches for computing the spectral coefficients in a PC expansion, namely intrusive or Galerkin methods and non-intrusive or sampling based techniques. The intrusive method involves reformulating the original set of dynamical system equations through its Galerkin or pseudospectral projection onto the approximation space spanned by the PC basis [17, 27]. One then has to solve a large coupled system of equations for the time evolution of the PC coefficients. However, for the class of systems considered in this work, which are intrinsically stochastic and can only be simulated when the reaction rates are prescribed, we have no explicit governing equations available for the quantities of interest (QoIs), namely the solution moments, to perform the Galerkin projection. On the other hand, the non-intrusive methods share the common characteristic of not requiring the reformulation of the original dynamical system, and thus enable the use of already existing solvers or simulators. For example, in the so-called non-intrusive spectral projection (NISP) method, the PC coefficients of the QoIs are computed via numerical integration, where only computations of the QoIs are needed over the nodes of an appropriate quadrature formula (cf. Section 2.2). PC methods in general, and NISP in particular, face the so-called curse of dimensionality. In the context of NISP, the curse of dimensionality refers to the rapid increase in the number of quadrature points needed for accurate computation of the PC coefficients. In many cases, this can be mitigated through sparse tensorization and sparse grid quadratures, combined with adaptive techniques [13, 14, 25, 26, 30].

Another well-known difficulty with time-dependent systems concerns the ubiquitous observation that the spectrum of a straightforward PC representation can be broad with significant evolutions in time; this makes it difficult to approximate quantities of interest via fixed (low) order PC expansions [28]. For example, in the context of a chemical system,

some realizations of the system trajectory may reach equilibrium quickly, whereas others may exhibit a transient behavior over much larger timescales. Such decorrelation of system trajectories leads to energetic higher order modes in a PC representation; see e.g. [36] for specific illustrations. In [2], a preconditioning strategy for NISP was developed to address the issue of broad spectra in the context of stiff chemical systems having no source of stochasticity other than the reaction rates, i.e., chemical systems described by their limiting reaction rate equations (RREs). Preconditioning involves using appropriate multiscale linear or affine transformation of the state variables, so as to control the variance (with respect to the uncertain model parameters) of the scaled variable, and to maintain a tight PC spectrum over time. It was shown in [2] that such an approach enables a low-order PC basis that remains suitable over time. As stated previously, we focus in this work on the problem of quantifying uncertainties in the reaction rates of stochastic chemical systems simulated using SSAs, i.e., without passing to a continuous limit. Such models lead to several challenges, arising due to the compound impact of high fluctuation levels (intrinsic stochasticity due to finite system size), and of the steep dependence of the system response on uncertain rates. In particular, a straightforward NISP approach can be problematic especially when sparse SSA sampling must be utilized to estimate the QoI at the quadrature nodes; in these situations, accuracy may severely deteriorate due to large sampling errors [44].

We aim at addressing these hurdles through Bayesian preconditioning and regression algorithms to simultaneously accommodate large fluctuations and steep dependence of QoIs on uncertain inputs. Specifically, we seek to exploit the capability afforded by Bayesian regression [38, 39, 41, 44, 45], to accommodate noisy signals, namely by defining a suitable transform (depending on the uncertain inputs only) that is later applied to the original noisy signals (SSA realizations). This defines a collection of noisy scaled variables that are also captured using a Bayesian regression formalism. Questions that we seek to address concern (1) whether this methodology is successful in leading to sparse low order representations of the noisy signals, (2) whether one can thus represent the response of both the mean and variance of noisy SSA realizations, and (3) whether reasonable estimates of the dependence of mean signals and fluctuation levels on the random inputs can be obtained based on moderate number (viz. tens as opposed to several thousands) of realizations of the stochastic chemical system. These questions are motivated, in particular, by our desire to explore the development of efficient surrogates that can be used to couple stochastic simulation algorithms with fluctuating uncertain mesoscale models. In these situations, the stochastic surrogates of microscale regions must not only provide a suitable representation of the mean dynamics and their dependence on uncertain inputs, but also of the corresponding fluctuation levels.

The structure of the paper is as follows. In Section 2, we fix the notation and briefly discuss the background concepts used throughout the paper. In Section 3, we briefly describe the idea of Bayesian regression for computing a spectral approximation of expected trajectory of a chemical system. A simple example, involving a birth-death process is also provided for further motivation. Section 4 contains the main contributions of this paper, where we begin by a brief outline of the preconditioning strategy of [2] in Section 4.1 and proceed by describing the method of preconditioned Bayesian regression in Section 4.2. We present numerical results for two model problems: a stochastic Michaelis-Menten system

with three uncertain parameters (Section 5), and a stochastic genetic positive feedback loop system, having nine uncertain reaction rates (Section 6). Concluding remarks are provided in Section 7.

2 Background

2.1 Polynomial chaos

In this paper, we will encounter stochastic dynamical systems with finitely many random inputs, parameterized by a finite collection of real-valued independent random variables $\vartheta^1, \vartheta^2, \ldots, \vartheta^N$. Denoting the range of each ϑ^i by Θ^i, and letting $\Theta = \Theta^1 \times \Theta^2 \times \cdots \times \Theta^N \subseteq \mathbb{R}^N$, we will be working in the probability space $(\Theta, \mathcal{B}(\Theta), F_{\vartheta})$, where $\mathcal{B}(\Theta)$ denotes the Borel σ-algebra on Θ and F_{ϑ} is the joint distribution function of the random N-vector $\vartheta = (\vartheta^1, \ldots, \vartheta^N)^T$. We denote the expectation of a random variable $g : \Theta \to \mathbb{R}$ by $\langle g \rangle = \int_{\Theta} g(s)\, dF_{\vartheta}(s)$. The space $L^2(\Theta, dF_{\vartheta})$ of square integrable random variables is endowed with the inner product $(\cdot, \cdot) : L^2(\Theta, dF_{\vartheta}) \times L^2(\Theta, dF_{\vartheta}) \to \mathbb{R}$, given by $(f, g) = \int_{\Theta} f(s)g(s)\, dF_{\vartheta}(s) = \langle fg \rangle$. In what follows, with a slight abuse of notation, we continue denoting points in $\Theta \subseteq \mathbb{R}^N$ by ϑ.

In this work, we will focus on the case $\vartheta^i \overset{iid}{\sim} \mathcal{U}(-1, 1)$, where $\mathcal{U}(-1, 1)$ denotes the uniform distribution on interval $[-1, 1]$, and rely on representing observables $g \in L^2(\Theta) \equiv L^2(\Theta, dF_{\vartheta})$, $dF_{\vartheta} \equiv (1/2)^N\, d\vartheta$, in terms of expansions of the form:

$$g = \sum_{k=0}^{\infty} g_k \Psi_k, \tag{1}$$

where $\{\Psi_k\}_0^{\infty}$ is a complete orthogonal set consisting of N-variate Legendre polynomials [1]. The expansion (1) is known as the Wiener-Legendre polynomial chaos expansion [7, 24, 27, 52, 54] of g. In practical computations, we will be approximating $g(\vartheta)$ with a truncated PC expansion,

$$g(\vartheta) \doteq \sum_{k=0}^{P} g_k \Psi_k(\vartheta), \tag{2}$$

where P is finite and depends on the truncation strategy adopted. We consider truncations based on the total degree of the polynomials in the series; that is, if we let p to denote the largest polynomial degree in the expansion, then P depends on the stochastic dimension N and expansion order p according to $1 + P = \dfrac{(N+p)!}{N!p!}$.

2.2 Non Intrusive Spectral Projection (NISP)

The truncated PC expansion in (2) can be defined as the $L^2(\Theta)$-projection of g onto the span of the PC basis $\{\Psi_0, \Psi_1, \ldots, \Psi_P\}$. Exploiting the orthogonality of the basis, the coefficients g_k are therefore given by:

$$g_k = \frac{\langle g\Psi_k \rangle}{\langle \Psi_k^2 \rangle}, \quad k = 0, \ldots, P. \tag{3}$$

Since the moments $\langle \Psi_k^2 \rangle$ can be computed analytically, the determination of coefficients g_k amounts to the evaluation of the moments $\langle g\Psi_k \rangle$. The NISP method involves computation

of these moments through quadrature,

$$\langle g\Psi_k\rangle \equiv \int_\Theta g(s)\Psi_k(s)\,dF_\vartheta(s) \doteq \sum_{j=1}^{N_q} w_j g(\vartheta_j)\Psi_k(\vartheta_j), \tag{4}$$

where $\vartheta_j \in \Theta$ and w_j are the nodes and weights of an appropriate quadrature formula. Note that the same set of nodes is used to compute all the coefficients g_k, so the complexity of NISP scales with N_q, the number of nodes where one has to compute g. To reduce the cost of NISP, we shall rely on sparse quadrature rules [16,33,34] based on Smolyak's formula [48].

Given a quadrature formula, to evaluate (4) we need to compute $g(\vartheta_q)$ for all $\vartheta_q \in \mathcal{S}$, where

$$\mathcal{S} = \{\vartheta_j\}_{j=1}^{N_q} \subset \Theta,$$

is the corresponding set of quadrature nodes. Let $\mathsf{K} \in \mathbb{R}^{(P+1)\times N_q}$ be the NISP matrix defined by,

$$K_{ij} = \frac{w_j\Psi_i(\vartheta_j)}{\langle\Psi_i^2\rangle}, \quad i = 0,\ldots,P, j = 1,\ldots,N_q,$$

and denote by $\mathbf{d}$ the vector with coordinates $d_j = g(\vartheta_j)$. Then the vector $\mathbf{g} = (g_0,\ldots,g_P)^T$ of the spectral coefficients is given by

$$\mathbf{g} = \mathsf{K}\mathbf{d}.$$

2.3 Stochastic simulation algorithm

In this section, we provide a brief outline of SSA [10,18–22,35]. Consider a chemical system with n species $S_1,\ldots,S_n$, and let $X_i(t)$ be the number of molecules of S_i at time t. The vector $\mathbf{X}(t) = (X_1(t),\ldots,X_n(t))^T$ is the state vector of the system. Assuming M reactions, we denote by a_k and $\boldsymbol{\nu}_k$ the propensity function and the stoichiometric vector corresponding to k^{th} reaction respectively. Recall that the stoichiometric vector (or state-change vector), $\boldsymbol{\nu}_k \in \mathbb{R}^n$ determines the change in state vector if k^{th} reaction takes place; that is, the k^{th} reaction has the effect of changing the state vector $\mathbf{X}(t)$ to $\mathbf{X}(t) + \boldsymbol{\nu}_k$. The propensity function $a_k = a_k\big(\mathbf{X}(t)\big)$ on the other hand has the interpretation that the probability of reaction k taking place in the infinitesimal time interval $[t, t+dt)$ is $a_k\big(\mathbf{X}(t)\big)dt$.

The exact simulation of the stochastic chemical system is done using SSA outlined in Algorithm 1. Note that the system trajectory is a piecewise constant function (over time) and therefore, can be interpolated exactly on any given time mesh. This will later prove convenient for the purpose of estimating the state vector at a fixed time mesh.

Algorithm 1 Gillespie SSA

Specify initial state $\mathbf{X}(0)$
Specify final time T_f
Specify stoichiometric vectors $\boldsymbol{\nu}_k$, $k = 1, \ldots, M$
Define propensity functions a_k, $k = 1, \ldots, M$
Let $t = 0$
while $t < T_f$ **do**
  Evaluate $a_k\big(\mathbf{X}(t)\big)$, $k = 1, \ldots, M$
  Set $a_0(t) = \displaystyle\sum_{k=1}^{M} a_k\big(\mathbf{X}(t)\big)$
  Draw iid $\mathcal{U}(0,1)$ random variables r_1 and r_2
  Set j to be the smallest integer such that $r_1 < \dfrac{1}{a_0(t)} \displaystyle\sum_{k=1}^{j} a_k\big(\mathbf{X}(t)\big)$ {pick the reaction to fire}
  Set $\tau = \dfrac{1}{a_0(t)} \ln\left(\dfrac{1}{r_2}\right)$ {the reaction time}
  Set $\mathbf{X}(t+\tau) = \mathbf{X}(t) + \boldsymbol{\nu}_j$ {update the state vector}
  Set $t = t + \tau$ {update the time step}
end while

3 Bayesian regression for chemical kinetics

In this section, we provide a brief overview of Bayesian regression and its application to spectral representation of the trajectory of a stochastic chemical system. We begin by setting up the basic formulation of a Bayesian regression problem in Section 3.1. Then, we illustrate the application of Bayesian regression for inferring a PC model for the expected trajectory of a stochastic chemical system in Section 3.2. Finally, in Section 3.3, we discuss a more detailed formulation of the problem for chemical systems, which will be used later for Bayesian regression of the preconditioned variables of a stochastic chemical system.

3.1 Bayesian regression

Consider a function $g \in L^2(\Theta)$ as in Section 2.2. In the context of stochastic chemical systems, for a given ϑ, $g(\vartheta)$ will be a quantity of interest that can only be estimated with limited accuracy, through finite SSA sampling of the model. In this context, a more robust way of estimating the PC expansion for g is through Bayesian regression [6, 8, 15]. The Bayesian approach [38,39,41,44] views the coefficients g_k as random variables, which provides a natural framework for accommodating intrinsic fluctuations or noise in the estimates of $g(\vartheta)$. Assuming some prior knowledge on the set of PC coefficients $\mathbf{g} = (g_0, g_1, \ldots, g_P)^T$ in (3), collecting a set of model observations, $\mathbf{d} \in \mathbb{R}^m$, with $d_i = g(\vartheta_i)$ for $\vartheta_i \in \Theta$, $i = 1, \ldots, m$, and adopting an appropriate noise model (and thus an appropriate likelihood) allows for invoking Bayes' theorem,

$$\Pi(\mathbf{g}|\mathbf{d}) \propto \mathcal{L}(\mathbf{d}|\mathbf{g})\pi(\mathbf{g})$$

where $\Pi(\mathbf{g}|\mathbf{d})$ is the posterior distribution, $\mathcal{L}(\mathbf{d}|\mathbf{g})$ is the likelihood and $\pi(\mathbf{g})$ is the prior. Bayes' theorem thus represents a process of refining the prior knowledge, $\pi(\mathbf{g})$, on the coefficients based on (new simulation) data, $\mathbf{d}$.

The likelihood Let $\mathbf{d} = (d_1, \ldots, d_m)^T$ be the vector of observed data, i.e., noisy estimates of g, at sampling points $\vartheta_1, \ldots, \vartheta_m \in \mathcal{S} \subset \Theta$. The likelihood $\mathcal{L}(\mathbf{d}|\mathbf{g})$ represents the probability of observing the data $\mathbf{d}$, given the parameter values (coefficients) $\mathbf{g}$. We assume an additive error model of the form

$$d_i = g(\vartheta_i) + \varepsilon_i, \quad i = 1, \ldots, m,$$

where ε_i are independent Gaussian random variables $\mathcal{N}\big(0, \sigma^2(\vartheta_i)\big)$, accounting for the "noise" in the estimate of g, and σ^2 is a hyper-parameter. In the present case, since g denotes the mean of a stochastic signal, and ε_i is the sampling error, the choice of a Gaussian model is well justified based on the central limit theorem. The independence assumption on ε_i is justified as we will rely on independent model realizations when estimating $g(\vartheta_i)$. While in general σ^2 depends on $\vartheta \in \mathcal{S}$, in cases where the noise amplitude is small and does not change significantly throughout the parameter space, one may use a simplified model where ε_i are iid $\mathcal{N}(0, \sigma^2)$ with a single hyper-parameter σ^2. To keep the presentation simple, we

consider the case where ε_i are iid $\mathcal{N}(0,\sigma^2)$ in this section. The more general case, where the dependence of σ^2 on ϑ is accounted for, will be addressed later in Section 4.

With the present assumptions, we have that $d_i \sim \mathcal{N}(g(\vartheta_i),\sigma^2)$, $i=1,\ldots,m$, and thus, the likelihood is given by:

$$\mathcal{L}(\mathbf{d}|\mathbf{g}) = \prod_{i=1}^{m} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\Big\{ -\frac{\big(d_i - g(\vartheta_i)\big)^2}{2\sigma^2} \Big\}. \tag{5}$$

The choice of prior and the application of Bayes theorem Let π_k denote the prior on g_k, $k=0,\ldots,P$. In what follows, we assume a non-informative uniform prior on the coefficients g_k. Specifically, we assume $g_k \sim \mathcal{U}(-I,I)$ with a sufficiently large I,

$$\pi_k(x) = \begin{cases} \dfrac{1}{2I}, & x \in [-I,I], \\ 0, & \text{otherwise.} \end{cases}$$

As for the hyper-parameter σ^2, we assume a Jeffreys prior in the form [15]

$$\pi_{\text{Jeff}}(x) \propto \begin{cases} 1/x, & \text{if } x > 0, \\ 0, & \text{otherwise.} \end{cases}$$

Using the above assumptions, Bayes' theorem gives

$$\Pi(\mathbf{g},\sigma^2|\mathbf{d}) \propto \mathcal{L}(\mathbf{d}|\mathbf{g}) \times \prod_{k=0}^{P} \pi_k(g_k) \times \pi_{\text{Jeff}}(\sigma^2),$$

where $\Pi(\mathbf{g},\sigma^2|\mathbf{d})$ is the joint posterior and $\mathcal{L}(\mathbf{d}|\mathbf{g})$ is specified in (5). The problem then reduces to sampling the target posterior above. To this end, we use a Markov Chain Monte Carlo (MCMC) method based on adaptive Metropolis [3, 4, 23, 40].

Remark 3.1 In the present application, the prior on the spectral coefficients is a non-informative uniform distribution. However, there are situations where one may want to use a different prior on the coefficients. We will revisit this issue later when discussing inference of degenerate spectra.

Remark 3.2 The choice of the sample set (or the design set) $\mathcal{S} \subset \Theta$ is important. If the process of collecting observations is cheap, one can choose a sufficiently large Monte Carlo sample. Other types of sampling such as Latin Hypercube sampling, Quasi Monte Carlo sequences, or Sobol sequences are also possible. However, in applications we consider, we generally let $\mathcal{S}$ be a set of sparse quadrature nodes one would use in a standard NISP procedure. The exact choice of the sparse quadrature nodes depends on the choice of the basic one-dimensional quadrature used to build the multi-dimensional sparse quadrature formula. In the present work, we use a one-dimensional Gauss-Kronrod-Patterson quadrature to build multi-dimensional sparse grids. Table 1 lists the number of sampling points needed as the dimension and level of quadrature increases.

	$l=1$	$l=2$	$l=3$	$l=4$	$l=5$
$d=2$	5	9	17	33	65
$d=3$	7	19	39	87	135
$d=4$	9	33	81	193	385
$d=5$	11	51	151	391	903
$d=6$	13	73	257	737	1889
$d=7$	15	99	407	1303	3655
$d=8$	17	129	609	2177	6657
$d=9$	19	163	871	3463	11527

Table 1: The number of sampling points needed for a Smolyak sparse grid based on the Gauss-Kronrod-Patterson rule.

3.2 Application to chemical kinetics

In this section, we illustrate the application of Bayesian regression in obtaining a PC representation of the mean trajectory of a chemical system simulated via SSA. Here we specifically want to find the expected number of a specific molecule, S, in the system at a given time.

Let $X(t)$ be the number of S molecules at a time t, and note that $X = X(t,\omega)$ is a random variable, where $\omega \in \Omega$ represents the intrinsic randomness in the system. Additionally if the reaction rates are uncertain, we would have, $X = X(t,\omega,\vartheta)$, where $\vartheta \in \Theta$ parameterizes the reaction rates. To be more precise, $X(t)$ is defined on a product probability space, $(\Omega,\mathcal{F},\nu)\otimes(\Theta,\mathcal{B}(\Theta),F_\vartheta)$, where $(\Omega,\mathcal{F},\nu)$ is the probability space on which an individual stochastic system trajectory is defined. Let us consider the expected number of S molecules at time t, given $\vartheta\in\Theta$:

$$\mu(t,\vartheta) = \mathsf{E}_\omega\left(X(t,\cdot,\vartheta)\right) \equiv \int_\Omega X(t,\omega,\vartheta)\, d\nu(\omega). \tag{6}$$

We seek to approximate $\mu(t,\vartheta)$ with a truncated PC expansion:

$$\mu(t,\vartheta) \doteq \sum_{k=0}^{P} \mu_k(t)\Psi_k(\vartheta), \quad \vartheta\in\Theta. \tag{7}$$

Of course, the average in (6) cannot be computed analytically in general; thus, we estimate $\mu(t,\vartheta)$ using Monte Carlo sampling:

$$\mu(t,\vartheta) \approx \hat{\mu}(t,\vartheta) = \frac{1}{N_g}\sum_{\ell=1}^{N_g} X^{(\ell)}(t,\vartheta) \tag{8}$$

where $\left\{X^{(\ell)}(t,\vartheta)\right\}_1^{N_g}$ denote SSA realizations of $X(t,\cdot,\vartheta)$, and N_g is the sample size (i.e. the number of SSA replicas). If a sufficiently large N_g is used, the estimator $\hat{\mu}(t,\vartheta)$ for $\mu(t,\vartheta)$ given above can be used to compute its PC representation using NISP. In particular, we can choose an appropriate sample $\mathcal{S} = \{\vartheta_1,\ldots,\vartheta_{N_q}\}\subseteq\Theta$, and use Algorithm 2 below to get the spectral representation in (7).

Algorithm 2 NISP algorithm to compute the spectral representation $\mu(t,\vartheta) \doteq \sum_{k=0}^{P} \mu^k(t)\Psi_k(\vartheta)$

for $j = 1, 2, \ldots, N_q$ **do**
  for $\ell = 1, 2, \ldots, N_g$ **do**
    $X^{(\ell)}(t,\vartheta_j) = \mathsf{SSA}(t,\vartheta_j)$
  end for
  $\hat{\mu}(t,\vartheta_j) = \frac{1}{N_g}\sum_{\ell=1}^{N_g} X^{(\ell)}(t,\vartheta_j)$
end for
Form an appropriate NISP matrix K
Define data vector $\mathbf{d}(t) = \left(\hat{\mu}(t,\vartheta_1),\ldots,\hat{\mu}(t,\vartheta_{N_q})\right)^T$
Compute $(\mu_0(t),\ldots,\mu_P(t))^T = \mathsf{K}\mathbf{d}(t)$

One of the main challenges with Algorithm 2 is the need for a large N_g to get a sufficiently accurate estimate of the SSA sample averages. Compounding the difficulty is the increase in complexity of NISP as the number, N, of random parameters and also the order of the expansion are increased. As previously mentioned, one way to mitigate the complexity of NISP is to use a sparse quadrature. We would also like to compute a reasonably accurate spectral representation, with a low number of SSA runs per each $\vartheta_j \in \mathcal{S}$. This, however, leads to "noisy" sample averages $\hat{\mu}(t,\vartheta_j)$, which are known to cause numerical instabilities with sparse quadrature methods [44]. Therefore, the NISP approach outlined in Algorithm 2 is limited to systems with a few random parameters and low intrinsic noise.

Alternatively, we wish to explore whether the Bayesian regression strategy outlined in the previous section can be employed to reduce the computational cost, by allowing a low number of SSA realizations (viz. $N_g \sim 30$ instead of $N_g \sim 10{,}000$ in (8)). In particular, we can compute the spectral coefficients $\{\mu_k(t)\}_{k=0}^P$ in (7) as follows: First, we collect observations,

$$d_j(t) = \hat{\mu}(t,\vartheta_j), \quad \vartheta_j \in \mathcal{S}.$$

Next, as described in Section 3.1, we assume an additive error model of the form,

$$\hat{\mu}(t,\vartheta_j) = \mu(t,\vartheta_j) + \varepsilon_j(t),$$

and exploit the Bayesian formalism to infer the distribution of the spectral coefficients of $\mu(t,\cdot)$. We start by assuming a (generally non-informative uniform) prior on the coefficients $\mu_k(t)$, and use the likelihood,

$$\mathcal{L}\left(\{\hat{\mu}(t,\vartheta_j)\}_{j=1}^{N_q} \,\middle|\, \{\mu_k(t)\}_{k=0}^P\right) = \prod_{j=1}^{N_q} \frac{1}{\sqrt{2\pi\sigma^2(t)}} \exp\left\{-\frac{\left(\hat{\mu}(t,\vartheta_j) - \sum_k \mu_k(t)\Psi_k(\vartheta_j)\right)^2}{2\sigma^2(t)}\right\}. \tag{9}$$

This leads to the following Bayesian update of the coefficients $\{\mu_0(t),\ldots,\mu_P(t)\}$,

$$\Pi\left(\{\mu_k(t)\}_{k=0}^P, \sigma^2 \middle| \{\hat{\mu}(t,\vartheta_j)\}_{j=1}^{N_q}\right) \propto \mathcal{L}\left(\{\hat{\mu}(t,\vartheta_j)\}_{j=1}^{N_q} \middle| \{\mu_k(t)\}_{k=0}^P\right) \times \pi\left(\{\mu_k(t)\}_{k=0}^P, \sigma^2(t)\right). \tag{10}$$

Here $\pi(\{\mu_k(t)\}_{k=0}^P, \sigma^2(t))$ is the presumed prior on the spectral coefficients and the hyper-parameter σ^2, modeled as

$$\pi\left(\{\mu_k(t)\}_{k=0}^P, \sigma^2(t)\right) = \Big(\prod_{k=0}^{P} \pi_k(\mu_k(t))\Big) \times \pi_{\text{Jeff}}\left(\sigma^2(t)\right),$$

where $\pi_k(\mu_k(t))$ is the prior on the k^{th} spectral coefficient, and $\pi_{\text{Jeff}}(\sigma^2)$ is Jeffreys prior on the noise variance $\sigma^2(t)$ in the estimate of the mean.

In our calculations, we rely on an adaptive MCMC sampler to sample the posterior distribution, and use either the maximum a posteriori probability (MAP) estimate or the mean posterior as the point estimate for the spectral coefficients. We close this subsection by providing a simple example to illustrate the idea of Bayesian regression for chemical kinetics.

An illustrative example Here we consider following birth-death system:

$$\emptyset \xrightarrow{c_1} S$$
$$S \xrightarrow{c_2} \emptyset$$

We denote by $X(t)$ the number of S molecules at time t. Figure 1 shows a number of realizations of the system trajectory with fixed reaction rates $(c_1, c_2) = (10,\, 0.5)$.

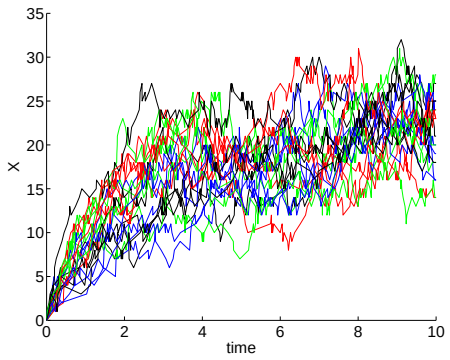


Figure 1: Realizations of the system trajectory for the birth-death system.

As mentioned before, our goal is to incorporate parametric uncertainty in the system, and we are interested in approximating

$$\mu(t, \vartheta) = \int_\Omega X(t, \omega, \vartheta)\, d\nu(\omega).$$

For this simple example, an analytical solution is available for $\mu(t, \vartheta)$ and is given by

$$\mu(t, \vartheta) = X_0 \exp\big\{-c_2(\vartheta)t\big\} + \frac{c_1(\vartheta)}{c_2(\vartheta)}\Big(1 - \exp\big\{-c_2(\vartheta)t\big\}\Big).$$

In the computations, we used $X_0 = 0$, and let

$$c_1 \sim \mathcal{U}(8, 12), \qquad c_2 \sim \mathcal{U}(0.2, 0.8).$$

These ranges were chosen so that a sufficiently wide variation in the system is observed.

Having the analytical expression for $\mu(t, \vartheta)$ allows for a basic test of the proposed Bayesian regression approach. To illustrate this, we fix our attention on the solution at $t = 10$, and use a cubic PC expansion of the analytical solution for purposes of comparison (note that the same analysis can be readily applied to explore the system at different times). We then collect observations by selecting a sampling of the parameter space, and computing $\hat{\mu}(t, \vartheta)$ with a small N_g, specifically $N_g = 20$, for each $\vartheta \in \mathcal{S}$. The sampling points correspond to the quadrature nodes of a fully tensorized 2D Gauss quadrature with five nodes in each direction. Table 2 shows the true coefficients (computed by expanding the analytical solution in a PC basis at $t = 10$) and the inferred coefficients.

Mode #	True Model	Mean posterior
0	22.2905	22.1427
1	4.4581	4.6335
2	-13.7026	-13.6830
3	0	-0.2902
4	-2.7405	-2.8497
5	5.1459	5.1576
6	0	0.1714
7	0	0.7118
8	1.0292	0.7867
9	-1.5698	-1.6620

Table 2: Comparison of the spectral coefficients of the true model at $t = 10$ with $(\mu_0(t = 10), \ldots, \mu_9(t = 10))$, estimated using (10).

Figures 2(a) and 2(b) respectively show the inferred response surface (blue) and the analytic one (red), and the relative error between the surfaces. Also, using the MCMC sample of the posterior, we computed the correlations between the inferred spectral coefficients (not shown) and observed little correlations between them. The results indicate that in the present case, Bayesian regression leads to accurate estimates, even though a small SSA sample size N_g is used. In Section 4, we investigate the extension of the present Bayesian approach to the preconditioning methods originally proposed in [2], to accommodate more complex and stiff stochastic chemical systems.

3.3 Toward a more detailed formulation

In this section, we consider a more detailed formulation for the Bayesian regression for stochastic chemical systems. Recall that in the previous section, we form a data vector $\mathbf{d}$ whose entries are the observed sample averages $\hat{\mu}(t, \vartheta)$ computed with a coarse SSA sampling

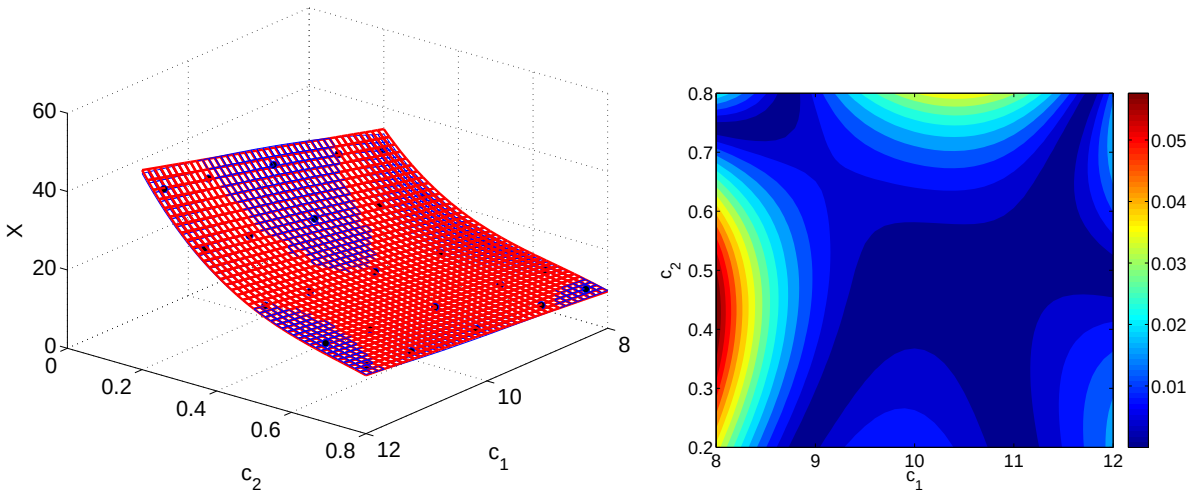


Figure 2: Bayesian regression for Birth-Death process: (a) true (red) and inferred (blue) surfaces of $\mu(t,\cdot)$ at $t = 10$, where the mean is plotted for $(c_1, c_2) \in [8, 12] \times [0.2, 0.8]$ (the space of random parameters), and the blue circles indicate the observed data; (b) relative difference between true and inferred surface for $(c_1, c_2) \in [8, 12] \times [0.2, 0.8]$.

of the system for each $\vartheta \in \mathcal{S}$. While this approach is straightforward, by averaging over the SSA samples, it discards valuable data that reflects the fluctuation levels. (The latter can be essential, for instance, if one seeks to couple SSAs with mesoscale models, such as fluctuating hydrodynamics.) One may therefore be motivated to retain all the raw data from the SSA sampling of the system in the inference problem. In the context of a stochastic chemical system, for a fixed time t, we have the raw observed data $X^{(\ell)}(t, \vartheta_j)$, $j = 1, \ldots, N_q$, $\ell = 1, \ldots, N_g$. Hence, we may use the data matrix, $\mathsf{D} = \mathsf{D}(t)$ defined by:

$$D_j^\ell = X^{(\ell)}(t, \vartheta_j), \quad j = 1, \ldots, N_q, \quad \ell = 1, \ldots, N_g.$$

Then, instead of the approach in the previous section, we may assume,

$$D_j^\ell = \mu(t, \vartheta_j) + \varepsilon_j^\ell, \tag{11}$$

where the discrepancies ε_j^ℓ are independent $\mathcal{N}\Big(0, \Sigma^2(t, \vartheta_j)\Big)$; that is, $D_j^\ell \sim \mathcal{N}\big(\mu(t, \vartheta_j), \Sigma^2(t, \vartheta_j)\big)$. This setup has the advantage of retaining all the SSA replicas of the system in the inference process; moreover, this formulation allows for estimation of the noise amplitude in X, as the $\Sigma^2(t, \vartheta_j)$ in this case approximates the variance of $X(t, \cdot, \vartheta_j)$.

The noise model used in the formulation in the previous section is motivated by the Central Limit Theorem. On the other hand, the Gaussian noise model in (11) is mainly motivated by the central limit results for Markov processes, as made rigorous in [12]. (See in particular, Theorems 2.1 and 2.3 in Chapter 11 of [12].) While the central limit results in [12] apply asymptotically as the system size becomes very large, in many chemical systems, even with a moderate number of molecules present (in order of say 10 molecules), a Gaussian noise model is observed to be a reasonable approximation. As an example, we consider the distribution of X in the birth-death system at $t = 10$ in Figure 3.

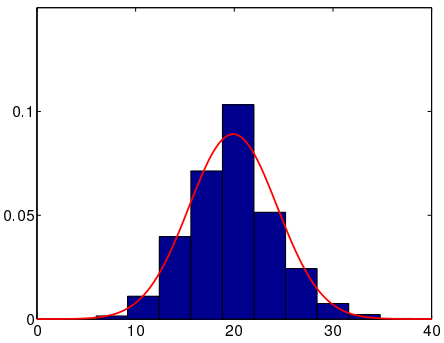


Figure 3: Distribution of X at $t = 10$ (the histogram) and a normal fit (the solid line) computed using sample mean and standard deviation based on 10,000 SSA samples.

We observed similar results for the other model problems discussed in the present work, and shall show some sample results when reporting numerical results. However, we would like to point out that such approximations do not hold in general when the system size is small; notable exceptions include systems that show bistability such as the Schlögl system [46]. The methods discussed in this paper are thus not directly applicable to such systems. In the present work, we do not use the noise model (11) for the state variable X directly. Instead, we utilize this approach to infer the mean and variance of a corresponding preconditioned variable (cf. Section 4.2), where a noise model of form (11) was found to be appropriate. In Section 5, we illustrate the validity of a Gaussian noise model used by showing the distribution of both the state variable at a given time and its preconditioned counterpart.

Remark 3.3 An alternative (not explored in the present work) to the simultaneous inference of the mean and variance through the noise model in (11) is to infer a response surface for the expected value of X using the method described in Section 3.1, and to consider the variance of X as a separate inference variable. The SSA realizations of X can be used to obtain observed values for the variance of X. Then using an appropriate noise model, one can infer a response surface for the variance using the Bayesian regression framework formulated in Section 3.1. A potential problem concerning the implementation of this alternative concerns the anticipated need of a large number of SSA samples, required to sufficiently reduce the noise in observed values of variance, so that the inference would be successful in computing a reasonably accurate response surface. Note that in the preconditioned Bayesian regression method developed in the present work, instead of working with the state variable X we perform a Bayesian regression for a transformed version of X. Thus, treating the variance as an additional observable may require computation of a separate preconditioner to alleviate the corresponding parametric stiffness. This points out another advantage of our formulation in Section 3.3, which requires computing a single preconditioner.

Remark 3.4 For the cases where a Gaussian noise model is not applicable, an immediate remedy is to replace the noise model with a more appropriate distribution, leading to a different likelihood. While such extensions are not pursued in the present work, we mention a useful special case. Specifically, when a log-normal distribution is an appropriate model for X, we may consider the variable $Z(t, \vartheta) = \log X(t, \vartheta)$ and assume a Gaussian noise model

for Z. More precisely, we define the data matrix D through,

$$D_j^\ell = Z^{(\ell)}(t, \vartheta_j), \quad j = 1, \ldots, N_q, \quad \ell = 1, \ldots, N_g,$$

and use the model:

$$D_j^\ell \sim \mathcal{N}\Big(\mu_Z(t, \vartheta_j), \sigma_Z^2(t, \vartheta_j)\Big).$$

After inferring response surfaces for mean and variance of Z through $\mu_Z(t, \vartheta_j)$ and $\sigma_Z^2(t, \vartheta_j)$, we may then obtain response surfaces for mean and variance of X by using the well known formulas,

$$\mu_X(t, \vartheta) = \exp\big\{\mu_Z(t, \vartheta) + \frac{1}{2}\sigma_Z^2(t, \vartheta)\big\},$$

and

$$\sigma_X^2(t, \vartheta) = \Big(\exp\big\{\sigma_Z^2(t, \vartheta)\big\} - 1\Big) \exp\big\{2\mu_Z(t, \vartheta) + \sigma_Z^2(t, \vartheta)\big\}.$$

Note that the idea of inferring a transformed quantity, in place of the original X, is precisely the idea underlying the preconditioned approach discussed in the next section, and the examples shown later indeed exhibit near Gaussian distributions compared to the distribution of the original (un-preconditioned) X.

4 Stochastic preconditioning

In this section, we extend the preconditioning method developed in [2] to stochastic chemical systems with uncertain reaction rates. As mentioned earlier, the method in [2] addresses the issue of widening of spectra of the state variables over time in the context of stiff chemical systems having no source of stochasticity other than the reaction rates. As such, the method could be applied to the RREs of a stochastic chemical system. Here we consider the problem of quantifying uncertainties in the reaction rates for stochastic chemical systems without passing to the corresponding RREs. To this end, we combine the preconditioning approach from [2] with Bayesian regression to accommodate large fluctuations and steep dependence of state variables on uncertain inputs. Our method results in response surfaces for the mean $\mu(t,\vartheta)$ and variance $\Sigma^2(t,\vartheta)$ of the state variable, involving low order representations of the moments of the preconditioned variables and the preconditioner. We begin by recalling in Section 4.1 the basic idea of the method developed in [2] . Then, we develop the method of preconditioned Bayesian regression in Section 4.2, where all the steps are discussed in detail.

4.1 The case of no intrinsic noise

Let us consider the dynamical system,

$$\dot{X} = F(X;q), \qquad X(0) = X_0.$$

The state variable $X(t)$ is in $\mathbb{R}^n$ and $q \in \mathbb{R}^N$ denotes a vector of model parameters. Here we consider the case of uncertain parameters, with $q = q(\vartheta)$, $\vartheta \in \Theta \subseteq \mathbb{R}^N$. Then, assuming no further source of uncertainty, a generic component of the state vector X, denoted by x, is defined on $[0, T_{fin}] \times \Theta$, and $x = x(t,\vartheta)$, can be approximated via a truncated PC expansion,

$$x(t,\vartheta) \doteq \sum_{k=0}^{P} x_k(t)\Psi_k(\vartheta).$$

A major difficulty in computing such a spectral approximation concerns the widening of spectra over time, which makes approximation of x with low order PC expansions difficult (see e.g. [36]). In [28] a time transformation approach was proposed to allow long time integration of the system resulting from a Galerkin (intrusive) reformulation of the problem. In [2], a preconditioning strategy for NISP was formulated, where instead of a direct projection of each state variable component x into a PC basis, one works with a transformed variable y in a scaled time τ,

$$y\Big(\tau(t,\vartheta),\vartheta\Big) = \Phi\Big[x(t,\vartheta)\Big],$$

where $\tau(t,\vartheta) = \dfrac{t}{\hat{t}(\vartheta)}$. The choice of transformation Φ and the time scaling factor $\hat{t}$ depends on the behavior of the realizations of x; see [2] for specific examples. In this paper, we consider linear scalings of the form:

$$y\Big(\frac{t}{\hat{t}(\vartheta)},\vartheta\Big) = \frac{1}{\hat{c}(\vartheta)}x(t,\vartheta), \tag{12}$$

where the choice of the time and amplitude scaling factors $\hat{t}$ and $\hat{c}$ will be discussed in the context of the applications presented. The basic idea behind such a transformation is to exploit the fact that in many chemical systems, the family of curves, $\{x(\cdot,\vartheta)\}_{\vartheta\in\Theta}$ are in a sense similar, and thus one can define an appropriate transformation that maps $\{x(\cdot,\vartheta)\}_{\vartheta\in\Theta}$ to a family of curves $\{y(\tau,\vartheta)\}_{\vartheta\in\Theta}$ that cluster together. For example, $x(t,\vartheta)$ may exhibit a a fast initial increase in building up to a distinct maximum followed by a slow monotonic decay; the time at which the maximum concentration is reached as well as the magnitude of the maximum concentration could vary drastically as ϑ ranges over Θ, which is the main source of parametric stiffness in such problems. In the context of the example just described, one can define the time scaling factor $\hat{t}(\vartheta)$ as the time taken for $x(t,\vartheta)$ to return to half of the maximum concentration, and define the amplitude scaling factor as $\hat{c}(\vartheta) = \max_t x(t,\vartheta)$. Then, one can synchronize $\{x(\cdot,\vartheta)\}_{\vartheta\in\Theta}$, by defining a transformed variable through (12). For further examples and detailed discussions, see [2].

It is helpful to keep in mind that preconditioning is done in the random parameter space Θ, and its purpose is to enable a low order PC representation of QoIs over time. However, as further discussed below, another beneficial feature is that in presence of intrinsic stochasticity the preconditioned variables have reduced noise amplitudes. In the next section, we extend the preconditioning formulation from [2] to the systems with intrinsic noise, with particular attention to stochastic chemical kinetics. In particular, the preconditioning is applied to stochastic system realizations (i.e. the observed data) and then Bayesian regression is used to infer the spectral coefficients of the scaled variable from the preconditioned data. We will also infer the noise amplitude of the scaled variable and use it to recover the variance of the state variable.

Remark 4.1 We would like to point out that the preconditioning strategy described here is not directly applicable to the cases where x does not depend continuously on the parameter ϑ. In such cases, a possible extension of the present preconditioning strategy could follow a domain decomposition approach, where the sample space Θ is decomposed into sub-domains where x depends continuously on ϑ; then, a preconditioner can be computed for each sub-domain of Θ.

4.2 Preconditioned Bayesian regression for stochastic systems

As mentioned before, for stochastic chemical systems with parametric uncertainty, a straightforward NISP can lead to poor approximations. Implementation of stochastic preconditioning would also be problematic, due to the stochastic nature of the signals, which makes it difficult to define suitable transforms. To address these challenges, we develop a Bayesian preconditioning approach, and highlight its application to compute suitable response surfaces for the mean, $\mu(t,\vartheta)$, and variance, $\Sigma^2(t,\vartheta)$, of the state variables of the stochastic chemical system. In what follows, we will be working with state variables and their corresponding scaled (preconditioned) counterparts. For convenience, we will denote by X a generic state variable (the number of molecules of a certain system species) and by Y its preconditioned counterpart, and note that $X = X(t,\omega,\vartheta)$ and $Y = Y(\tau,\omega,\vartheta)$ for $\omega\in\Omega$ and

$\vartheta \in \Theta$. The major steps of the presently developed Bayesian preconditioning approach are as follows: (1) Computation of the preconditioner; (2) discretization of the preconditioned variable; (3) Bayesian regression of the preconditioned variable; (4) recovery of the moments of the state variable. A flowchart depicting the process is given in Figure 4.

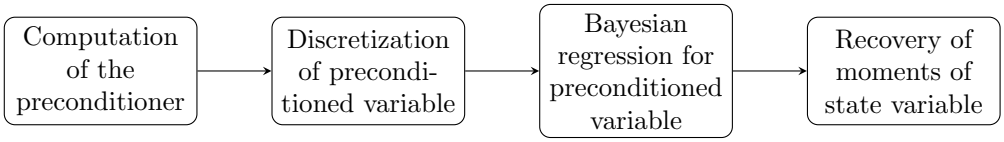


Figure 4: The major steps of preconditioned Bayesian regression.

4.2.1 Computing the preconditioner

The first step is to compute the preconditioner. In particular, let us focus on the case of linear scaling,

$$Y(\frac{t}{\hat{t}(\vartheta)}, \omega, \vartheta) = \frac{1}{\hat{c}(\vartheta)} X(t, \omega, \vartheta), \quad \omega \in \Omega, \vartheta \in \Theta. \tag{13}$$

The scaling factors (i.e. the preconditioner) will be computed to address the ill conditioning (large variance) associated with parametric uncertainty. Specifically, we seek a transformation that is independent of ω, specified in terms of truncated PC expansions of the scaling factors,

$$\hat{c}(\vartheta) \doteq \sum_{k=0}^{P} \hat{c}_k \Psi_k(\vartheta), \qquad \hat{t}(\vartheta) \doteq \sum_{k=0}^{P} \hat{t}_k \Psi_k(\vartheta). \tag{14}$$

We explore the following two strategies to compute the PC representations for $\hat{c}$ and $\hat{t}$:

Option 1: Use a Bayesian regression approach to infer the distribution of the spectral coefficients, and use the MAP estimate (or mean posterior) of the spectral coefficients in (14).

Option 2: Compute the PC representation of the scaling factors $\hat{c}$ and $\hat{t}$, based on the trajectory of the corresponding RREs for the chemical system.

We illustrate both the above strategies in the examples that follow. The second option may be particularly attractive because it is simpler computationally and also provides response surfaces for the scaling factors that are not affected by the intrinsic fluctuations of the SSA.

4.2.2 Discretization of the scaled variable

Let $Y^{(\ell)}(\tau, \vartheta)$ denote a realization of the scaled variable corresponding to parameter vector ϑ; the superscript ℓ enumerates the SSA replicas of the scaled variable for a given ϑ. The Bayesian regression of the scaled variable is done by first generating a collection of observed

scaled variables on a uniform τ (scaled time) mesh. In particular, we let τ_i, $i = 0, \ldots, N_\tau$ be a deterministic mesh of a suitable interval of the scaled time domain, and use Algorithm 3 to compute

$$Y^{(\ell)}(\tau_i, \vartheta_j), \quad i = 0, \ldots, N_\tau,\ j = 1, \ldots, N_q,\ \ell = 1, \ldots, N_g.$$

Here, N_g is the number of SSA replicas, taken to be the same for each ϑ_j, and $\{\vartheta_j\}_1^{N_q}$ is a suitable sample in Θ.

Algorithm 3 Algorithm for discretization of $Y(\tau, \vartheta)$

```
Select a sample set S ⊆ Θ
Let N_q = |S|
Set scaled time mesh τ_i, i = 0, ..., N_τ
for j = 1 to N_q do
  [c̄, t̄] = GetScalings(ϑ_j)                    {compute scaling factors}
  for ℓ = 1 to N_g do
    Compute X^(ℓ)_{i,j} = X^(ℓ)(t̄ · τ_i, ϑ_j),   i = 0, ..., N_τ
    Y^(ℓ)(τ_i, ϑ_j) = (1/c̄) X^(ℓ)_{i,j},   i = 0, ..., N_τ
  end for
end for
```

Remark 4.2 The procedure **GetScalings**(ϑ_j) computes the scaling factors $\hat{c}$ and $\hat{t}$ at ϑ_j. In the present work, this amounts to evaluating the PC expansions for $\hat{c}$ and $\hat{t}$ which are computed using one of the options listed in Section 4.2.1 and are available prior to running Algorithm 3.

Remark 4.3 To compute $X^{(\ell)}_{i,j} = X^{(\ell)}(\bar{t}\cdot\tau_i, \vartheta_j)$, $i = 0, \ldots, N_\tau$, we rely on the ability afforded by SSA to interpolate (exactly) in time the SSA realizations of the system trajectory.

4.2.3 Bayesian regression for the mean and variance of the scaled variable

Consider a fixed τ, let μ_Y and Σ_Y^2 denote the mean and variance of the scaled variable Y, and let $\mathsf{D} = \mathsf{D}(\tau)$ denote the data matrix defined by:

$$D_j^\ell = Y^{(\ell)}(\tau, \vartheta_j), \quad j = 1, \ldots, N_q, \quad \ell = 1, \ldots, N_g,$$

We assume,

$$D_j^\ell = \mu_Y(\tau, \vartheta_j) + \varepsilon_j^\ell,$$

where the discrepancies ε_j^ℓ are independent $\mathcal{N}\Big(0, \Sigma_Y^2(\tau, \vartheta_j)\Big)$; thus,

$$D_j^\ell \sim \mathcal{N}\Big(\mu_Y(\tau, \vartheta_j), \Sigma_Y^2(\tau, \vartheta_j)\Big).$$

Working in a Bayesian regression framework, we seek to infer the PC coefficients for μ_Y,

$$\mu_Y(\tau, \vartheta) \doteq \sum_{k=0}^{P} \mu_Y^k(\tau)\Psi_k(\vartheta),$$

as well as a PC model for the hyper-parameter Σ_Y^2 as follows. We will restrict our attention to a first-order expansion of the standard deviation Σ_Y, namely

$$\Sigma_Y(\tau,\vartheta) \doteq \Sigma_Y^0(\tau) + \Sigma_Y^1(\tau)\vartheta^1 + \Sigma_Y^2(\tau)\vartheta^2 + \ldots + \Sigma_Y^N(\tau)\vartheta^N.$$

The likelihood is given by,

$$\mathcal{L}\Big(\mathsf{D}|\big\{\mu_Y^k(\tau)\big\}_0^P, \big\{\Sigma_Y^k(\tau)\big\}_0^N\Big) = \prod_{j=1}^{N_q}\prod_{\ell=1}^{N_g} \frac{1}{\sqrt{2\pi\Sigma_Y^2(\tau,\vartheta_j)}} \exp\left\{-\frac{\big(D_j^\ell - \mu_Y(\tau,\vartheta_j)\big)^2}{2\Sigma_Y^2(\tau,\vartheta_j)}\right\}.$$

As for the prior, we will use a non-informative uniform prior on the mean mode. Since we expect a degenerate spectrum for Y [2], we will rely on a Gaussian prior for the higher order modes, with a mean around zero and a small standard deviation.

4.2.4 Recovery of the state variables

As done in [2], we can recover the state variable from the preconditioned variable Y at a given time t^* via an inverse transform. In the present application, we seek to recover the mean and variance of the state variable X. The recovery of the first moment is given by,

$$\begin{aligned}
\mathsf{E}_\omega\Big(X(t^*,\vartheta)\Big) &= \Phi^{-1}\Big(\mu_Y\big(\tau^*(\vartheta),\vartheta\big)\Big)\\
&= \hat{c}(\vartheta)\mu_Y\big(\tau^*(\vartheta),\vartheta\big)\\
&\approx \Big(\sum_k \hat{c}_k(\vartheta)\Psi_k(\vartheta)\Big)\Big(\sum_k \mu_Y^k\big(\tau^*(\vartheta)\big)\Psi_k(\vartheta)\Big)\\
&=: \tilde{\mu}(t^*,\vartheta),
\end{aligned}$$

where

$$\tau^*(\vartheta) = \frac{t^*}{\hat{t}(\vartheta)} \approx \frac{t^*}{\sum_k \hat{t}_k\Psi_k(\vartheta)}.$$

Note that since the PC coefficients of μ_Y are computed on a fixed τ grid, this process requires an interpolation of the coefficients in the scaled time domain. The process of computing $\tilde{\mu}_X(t^*,\vartheta^*)$ for a $t^* \in [0, T_{fin}]$ and a $\vartheta^* \in \Theta$ is outlined in Algorithm 4.

The recovery of the variance is also possible through a similar process. To compute an approximate response surface for Σ_X^2 we proceed as follows. For $(t^*,\vartheta) \in [0,T_{fin}]\times\Theta$, we have $\tau^*(\vartheta) = t/\hat{t}(\vartheta)$ and note that

$$\mathsf{Var}_\omega\Big(X(t^*,\vartheta)\Big) = \mathsf{Var}_\omega\Big(\hat{c}(\vartheta)Y(\tau^*,\vartheta)\Big) = \hat{c}(\vartheta)^2\mathsf{Var}_\omega\Big(Y(\tau^*,\vartheta)\Big).$$

Now, using the approximation,

$$\mathsf{Var}_\omega\Big(Y(\tau^*,\vartheta)\Big) \approx \Sigma_Y^2(\tau^*,\vartheta),$$

Algorithm 4 Procedure for computing $\tilde{\mu}_X(t^*, \vartheta^*)$ at $t^* \in [0, T_{fin}]$ and $\vartheta^* \in \Theta$

$[\bar{c}, \bar{t}] = \textbf{GetScalings}(\vartheta^*)$ {compute scaling factors from PC expansions}
$\tau^* = t^*/\bar{t}$ {scaled time}
Find τ_l such that $\tau^* \in [\tau_l, \tau_{l+1}]$
for $k = 0$ to P **do**
$\mu_Y^k(\tau^*) = \mu_Y^k(\tau_l) + \left[\dfrac{\mu_Y^k(\tau_{l+1}) - \mu_Y^k(\tau_l)}{\Delta\tau}\right](\tau^* - \tau_l)$ {Interpolate $\mu_Y^k(\tau^*)$}
end for {loop on PC modes}
Compute $\tilde{\mu}_X(t^*, \vartheta^*) = \bar{c} \cdot \sum_k \mu_Y^k(\tau^*(\vartheta^*))\Psi_k(\vartheta^*)$

obtained from the Bayesian regression approach described above and utilizing the spectral representation of Σ_Y and $\hat{c}$ we obtain

$$\mathsf{Var}_\omega\Big(X(t^*, \vartheta)\Big) \approx \Big(\sum_k \hat{c}_k(\vartheta)\Psi_k(\vartheta)\Big)^2 \Big(\sum_k \Sigma_Y^k(\tau^*)\Psi_k(\vartheta)\Big)^2 =: \tilde{\Sigma}^2(t^*, \vartheta).$$

As before, interpolation procedure can be used to get the coefficients $\Sigma_Y^k(\tau)$ for a given τ in the scaled time mesh.

5 Michaelis-Menten system

In this section, we consider a well known problem from biochemistry, the Michaelis-Menten system:

$$\begin{aligned} S_1 + S_2 &\xrightarrow{c_1} S_3 \\ S_3 &\xrightarrow{c_2} S_1 + S_2 \\ S_3 &\xrightarrow{c_3} S_4 + S_2 \end{aligned}$$

The system describes the creation of the product S_4 through binding of the enzyme S_2 with substrate S_1. The complex, S_3, is an intermediate species that can either dissociate back into the enzyme and substrate (the second reaction) or it can decompose to the product S_4 and enzyme S_2. There is a great amount of literature discussing the Michaelis-Menten dynamics. For example, see [53] where a brief discussion and a set of illustrative numerical experiments are provided. As before, we let $X_i(t)$ denote the number of molecules of species S_i in the system for $i = 1, 2, 3, 4$. A typical behavior of the system, initialized with fixed values for X_1 and X_2 and vanishing X_3 and X_4 is shown in Figure 5. Plotted are selected SSA realizations corresponding to fixed values of the rate parameters, $(c_1, c_2, c_3) = (0.0017, 10^{-3}, 0.125)$.

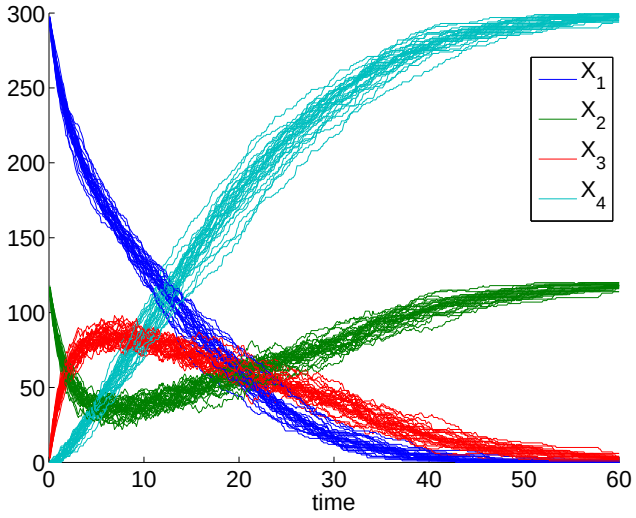


Figure 5: SSA realizations of the Michaelis-Menten system.

In the computations below, we consider independent uncertain rate constants that are distributed as:

$$c_1 \sim \mathcal{U}(0.0015, 0.003), \quad c_2 \sim \mathcal{U}(10^{-4}, 10^{-2}), \quad c_3 \sim \mathcal{U}(0.1, 0.15),$$

and parametrized using canonical random variables $\vartheta^i \overset{iid}{\sim} \mathcal{U}(-1, 1)$, $i = 1, \ldots, 3$. The sample set $\mathcal{S}$ consists of $N_q = 87$ realizations, corresponding to the nodes of a level 4 Smolyak cubature based on the Gauss-Patterson rule. This choice followed a refinement study (not shown) that considered different cubature levels and showed that the selected resolution leads to reasonable estimates of both the mean signals and of the corresponding variability.

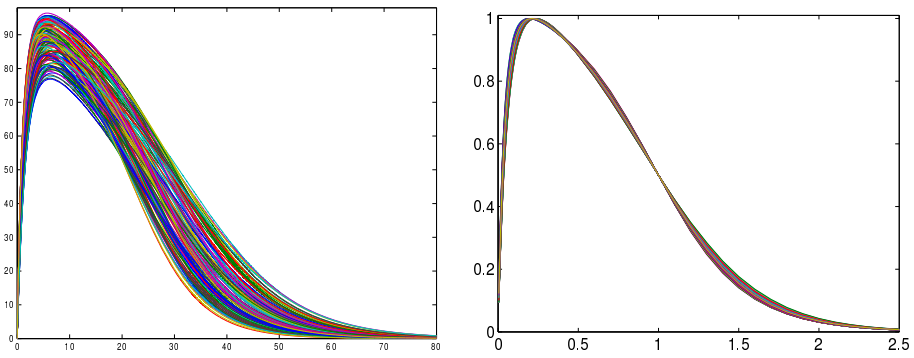


Figure 6: Left: Unscaled variable realizations (over time); Right: scaled variable realizations (over scaled time).

In addition, we shall focus for brevity on the evolution of X_3, and consequently drop the species subscript in the remainder of this section.

5.1 Parametric stiffness and preconditioning

In this section, we illustrate the phenomenon of parametric stiffness, which entails increases in the number of PC terms needed to represent time-dependent trajectories accurately as time progresses. We illustrate this for the Michaelis-Menten system with the ranges of uncertainty specified above. To focus on parametric uncertainty alone, we consider the corresponding system of RREs, describing the system in thermodynamic limit:

$$\begin{cases} \dot{x}_1 = -k_1 x_1 x_2 + k_2 x_3 \\ \dot{x}_2 = -k_1 x_1 x_2 + (k_2 + k_3) x_3 \\ \dot{x}_3 = k_1 x_1 x_2 - (k_2 + k_3) x_3, \\ \dot{x}_4 = k_3 x_3, \end{cases}$$

Focusing on the third species of the system, $x(t, \vartheta) = X_3(t, \vartheta)$, we use the following definition of the preconditioned variable, $y\left(\frac{t}{\hat{t}(\vartheta)}, \vartheta\right) = \frac{1}{\hat{c}(\vartheta)} x(t, \vartheta)$,

$$\hat{c}(\vartheta_i) = \max_t x(t, \vartheta_i), \quad \hat{t}(\vartheta_i) = \text{ time taken to return to } \frac{1}{2}\hat{c}(\vartheta_i).$$

The plots in Figure 6 show realizations of the unscaled variable x (left) and the scaled variable (right). In Figure 7, we plot the PC coefficients of x at selected points in time (top row) and the PC coefficients of the preconditioned variable y at selected (scaled) times (bottom row). Note that in this case, which exhibits moderate parametric stiffness, we see the excitation of the higher modes of the PCE for the unscaled variable. In contrast, the scaled variable maintains a tight spectrum. More severe cases of such a phenomenon are observed for example in [2] in context of a hydrogen oxidation chemical system.

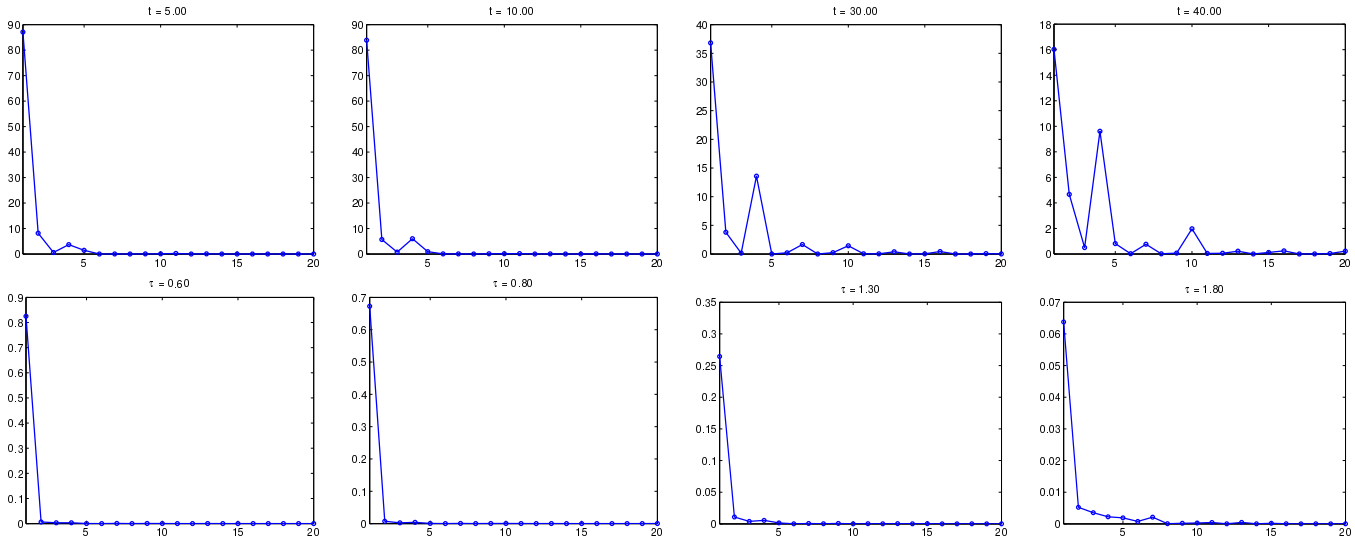


Figure 7: Absolute value of the PC coefficients for the unscaled state variable (top), and for the scaled variable (bottom). A third order expansion is used in the computations.

5.2 Stochastic scaling factors

Figure 8 shows the time evolution of the sample averages,

$$\hat{\mu}(t,\vartheta_i) = \frac{1}{N_g}\sum_{\ell=1}^{N_g} X^{(\ell)}(t,\vartheta_i), \quad \vartheta_i \in \mathcal{S}.$$

For each $\vartheta_i \in \mathcal{S}$, we define

$$\hat{C}(\vartheta_i) = \max_t \hat{\mu}(t,\vartheta_i), \quad \hat{T}(\vartheta_i) = \text{ time to return to } \frac{1}{2}\hat{C}(\vartheta_i),$$

and note that $\hat{C}$ and $\hat{T}$ are also affected by sampling noise, since a finite, and relatively small value of $N_g = 30$ is used.

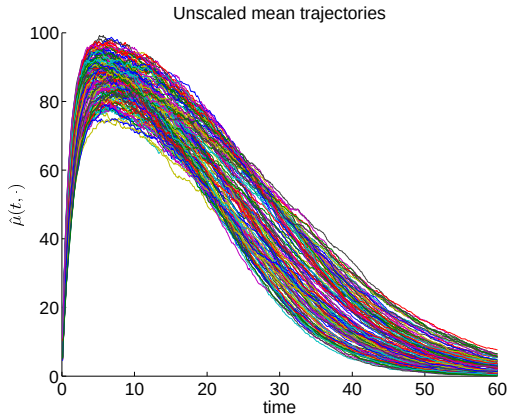


Figure 8: Evolution of the sample averages $\hat{\mu}$, estimated using $N_g = 30$ SSA realizations. Curves are generated for individual values of ϑ.

To infer the PC expansion of $\hat{c}(\vartheta)$, we first define the data vector

$$\mathbf{d} = \left(\hat{C}(\vartheta_1), \hat{C}(\vartheta_2), \ldots, \hat{C}(\vartheta_{N_q})\right)^T,$$

and use the model,

$$\begin{aligned} d_j &= \hat{c}(\vartheta_j) + \varepsilon_j \\ &= \Big[\sum_{k=0}^{P} \hat{c}_k \Psi_k(\vartheta_j)\Big] + \varepsilon_j, \quad j = 1,\ldots,N_q, \end{aligned}$$

where in our computations using ε_j as iid $\mathcal{N}(0,\sigma_{\hat{c}}^2)$ was found to be adequate. The Bayesian regression scheme outlined earlier is then applied to infer the PC coefficients. The computation of spectral representation of $\hat{t}$ is performed in a similar fashion. Figure 9 depicts the resulting distributions and spectra (PC coefficients) of $\hat{c}(\vartheta)$ and $\hat{t}(\vartheta)$ based on their MAP estimates. We see here a rather moderate stretching of time and amplitude scales. We shall see a more severe case in Section 6. Also, note that these scaling factors admit sparse low order PC representations.

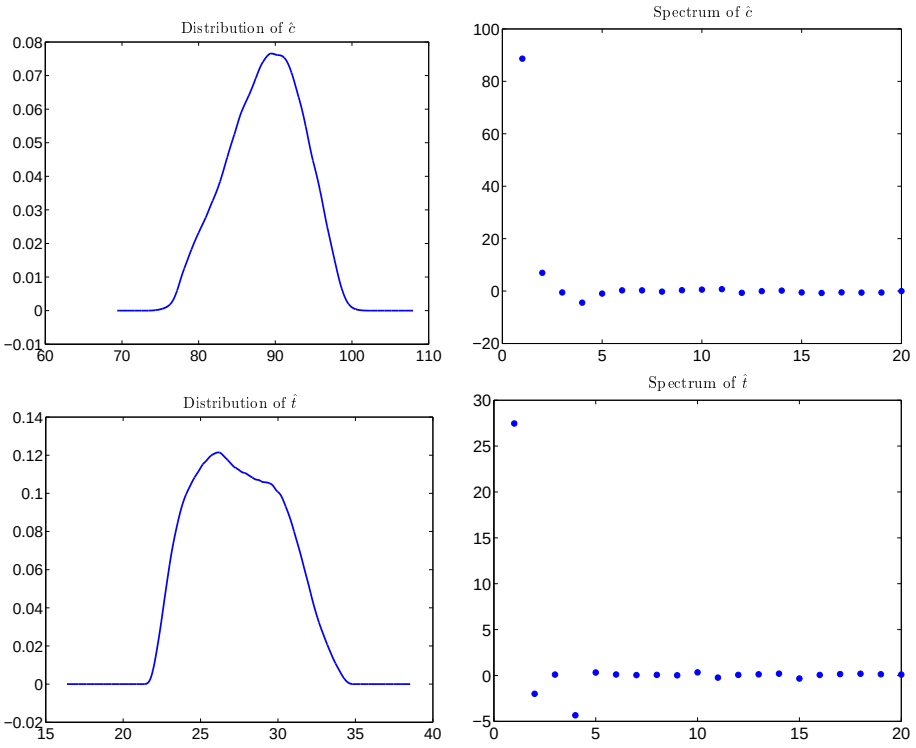


Figure 9: Distribution and spectra for the MAP estimates of the scaling factors. In top row, we show the PDF and the spectral coefficients for $\hat{c}(\vartheta)$, and in the bottom row we provide the corresponding plots for $\hat{t}(\vartheta)$.

5.3 Preconditioned variable

Figure 10 shows the system realizations and their preconditioned counterparts. The latter are obtained using (13), based on the MAP estimate of the spectral coefficients of $\hat{c}$ and $\hat{t}$, and are discretized on a uniform τ grid using Algorithm 3. Shown are the raw N_g SSA replicas for every $\vartheta_i \in \mathcal{S}$ on the top row, whereas the bottom row depicts the corresponding sample averages $\hat{\mu}$ and scaled version $\hat{\mu}_Y$. The figure clearly illustrates the impact of the transformation, namely in reducing noise levels as well as absorbing the variation induced by the parametric uncertainty.

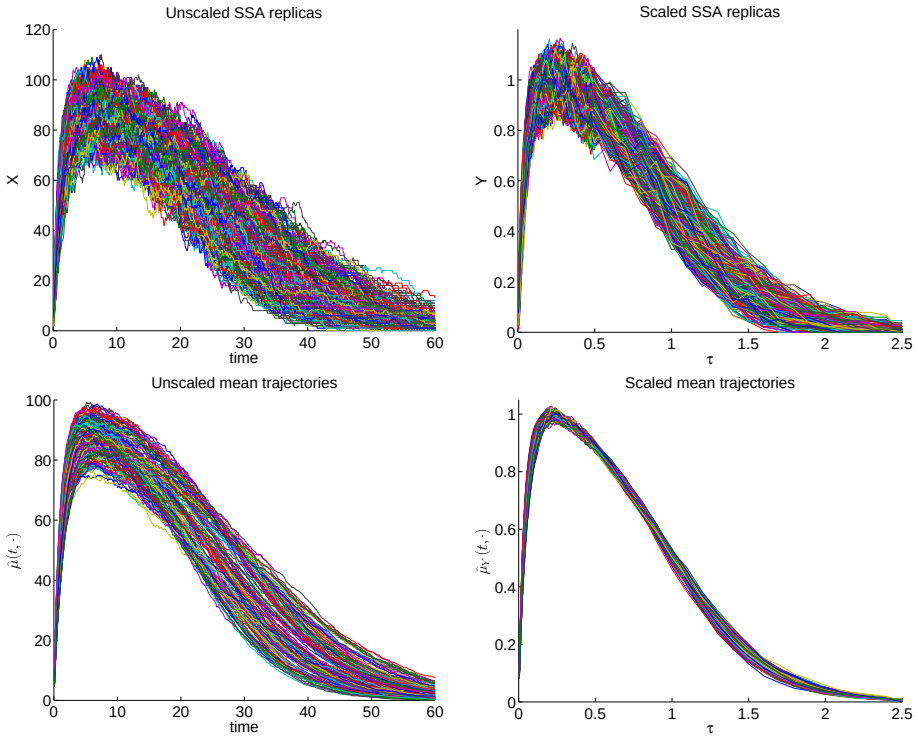


Figure 10: Preconditioning of stochastic dynamics. Top row: SSA replicas of the state variable $X(t, \omega, \vartheta_i)$ and the scaled variable $Y(\tau, \omega, \vartheta_i)$ for each $\vartheta_i \in \mathcal{S}$. Bottom row: mean trajectories, $\hat{\mu}(t, \vartheta_i)$ of the state variable and scaled version $\hat{\mu}_Y(\tau, \vartheta_i)$ for the $\vartheta_i \in \mathcal{S}$.

On the Gaussian approximation of the scaled variables As described in Section 4.2, we perform a Bayesian regression of the scaled variable where a Gaussian noise model is employed. To further motivate this choice, we illustrate the distribution of both the state variable and the corresponding scaled variable for a fixed ϑ and a fixed time. In particular, in Figure 11, we show the distribution of both the state variable $X(t_0, \cdot, \overline{\vartheta})$ at $t_0 = 10$ and the corresponding preconditioned variable, $Y(\overline{\tau}, \cdot, \overline{\vartheta})$, at nominal parameter values $\overline{\vartheta} = (0, 0, 0)$ (note also that $\overline{\tau} = t_0/\hat{t}(\overline{\vartheta})$). We note from the Figure that both the unscaled and scaled variables are well approximated by a Gaussian. Such observations further motivate the use of Gaussian noise model in the Bayesian regression as described in the previous section.

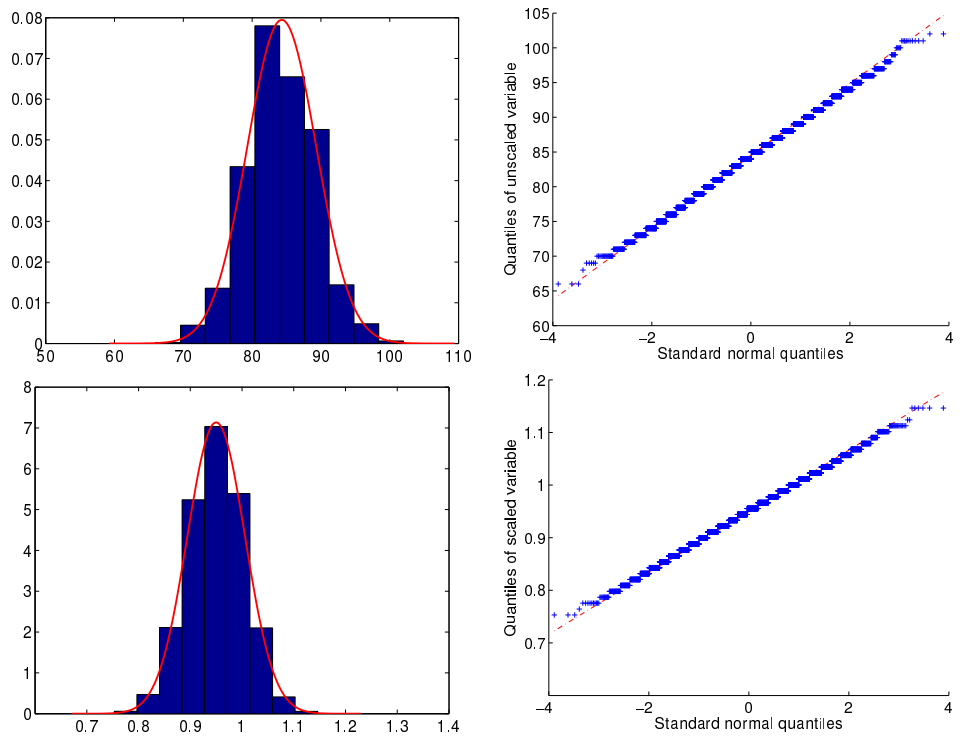


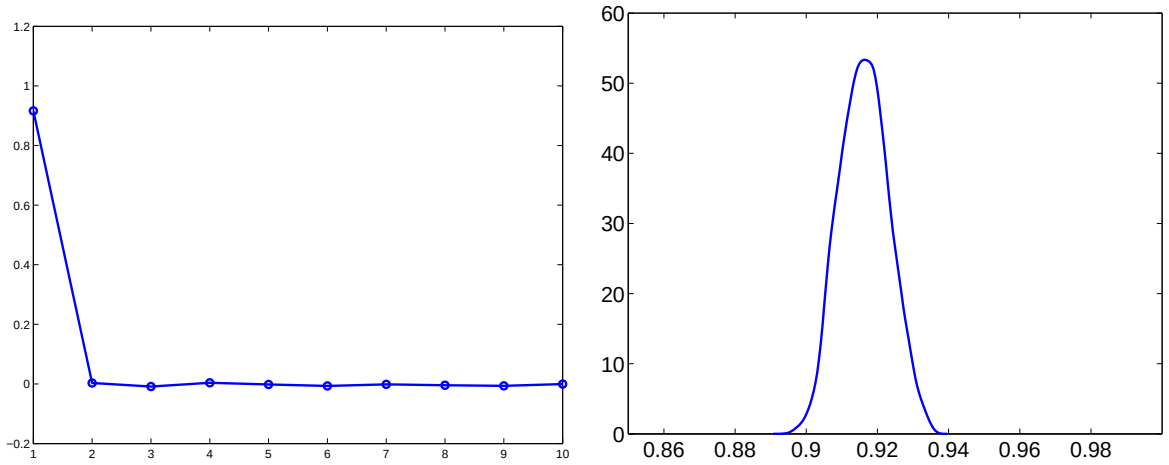
Figure 11: Top row compares the distribution of the state variable to that of Gaussian, whereas the bottom row depicts the same analysis for the scaled variable. In each case we show both a qq-plot, and also the comparison of the distribution of the variable in question with the corresponding normal fit. Here the normal fit is a Gaussian with mean and variance given by the sample mean and variance.

Bayesian regression of the scaled variable We use Bayesian regression (as described in Section 4.2.3) to infer a second order PC expansion for $\mu_Y(\tau, \vartheta)$. In Figures 12(a) and 12(b), we show the mean posterior of PC coefficients for $\mu_Y(\tau, \cdot)$, and the corresponding probability density function (PDF) for the PC approximation of $\mu_Y(\tau, \vartheta)$ at $\tau = 0.45$ respectively. These plots illustrate the goals of preconditioning which aims for a sparse low order PC representation for the preconditioned variable. The estimation of $\Sigma_Y^2(\tau, \vartheta)$ was also done as described in Section 4.2.3, that is, a linear expansion was inferred for $\Sigma_Y(\tau, \vartheta)$.

5.4 Recovery of the mean and variance for the state variable

Here we illustrate recovery of the first moments of the state variable by computing $\tilde{\mu}(t, \vartheta)$ and $\tilde{\Sigma}^2(t, \vartheta)$ at $t^* = 10$, using the procedure described in Section 4.2.4. Also, to verify the predictions we consider a Monte Carlo sample $\mathcal{S}_{mc} \subset \Theta$, and compute MC estimates for $\mu(t, \vartheta)$ and $\Sigma^2(t, \vartheta)$ for all $\vartheta \in \mathcal{S}_{mc}$, using a fine ω-sampling, based on 10^4 SSA replicas. This allows for measuring the accuracy of the response surfaces $\tilde{\mu}(t, \vartheta)$ and $\tilde{\Sigma}^2(t, \vartheta)$. In the computations that follows, we let $\mathcal{S}_{mc}$ be a set of 1000 points (distributed uniformly) in Θ.

In Figure 13, we show the distributions of $\tilde{\mu}(t^*, \vartheta)$ and $\tilde{\Sigma}^2(t^*, \vartheta)$ as compared to the



(a) (b)

Figure 12: (a) Inferred (mean posterior) spectral coefficients of $\mu_Y(\tau, \cdot)$, (b) the PDF for $\mu_Y(\tau, \vartheta)$. The results are generated for $\tau = 0.45$.

distributions of the Monte Carlo estimates. Note that the label "inferred" refers to the PDF constructed by sampling the response surfaces for the mean and variance, whereas the label "observed" refers to the corresponding Monte Carlo estimates (computed for all $\vartheta \in \mathcal{S}_{mc}$).

To get a sense of the mean square error, we computed an estimate of the relative $L^2(\Theta)$ error, at $t^* = 10$, as follows:

$$\frac{\|\mu(t^*,\cdot) - \tilde{\mu}(t^*,\cdot)\|_{L^2(\Theta)}}{\|\mu(t^*,\cdot)\|_{L^2(\Theta)}} \approx \frac{\left(\frac{1}{N}\sum_{\vartheta\in\mathcal{S}_{mc}} \left(\mu(t^*,\vartheta) - \tilde{\mu}(t^*,\vartheta)\right)^2\right)^{1/2}}{\left(\frac{1}{N}\sum_{\vartheta\in\mathcal{S}_{mc}} \mu(t^*,\vartheta)^2\right)^{1/2}} \approx 9.4 \times 10^{-3},$$

and

$$\frac{\left\|\Sigma^2(t^*,\cdot) - \tilde{\Sigma}^2(t^*,\cdot)\right\|_{L^2(\Theta)}}{\|\Sigma^2(t^*,\cdot)\|_{L^2(\Theta)}} \approx \frac{\left(\frac{1}{N}\sum_{\vartheta\in\mathcal{S}_{mc}} \left(\Sigma^2(t^*,\vartheta) - \tilde{\Sigma}^2(t^*,\vartheta)\right)^2\right)^{1/2}}{\left(\frac{1}{N}\sum_{\vartheta\in\mathcal{S}_{mc}} \Sigma^2(t^*,\vartheta)^2\right)^{1/2}} \approx 6.8 \times 10^{-2}.$$

The low error levels in the recovered first moments (less than 1% for the mean and around 6.8% for variance), as well as the close comparison of the distributions of $\tilde{\mu}$ and $\tilde{\Sigma}^2$ with that of their pure Monte Carlo estimates shows the utility of the preconditioned Bayesian regression framework; we note that the present method allows for recovery the moments of the state variable, with only 30 SSA replicas, in agreement with MC estimates based on 10,000 SSA replicas for each point $\vartheta \in \Theta$.

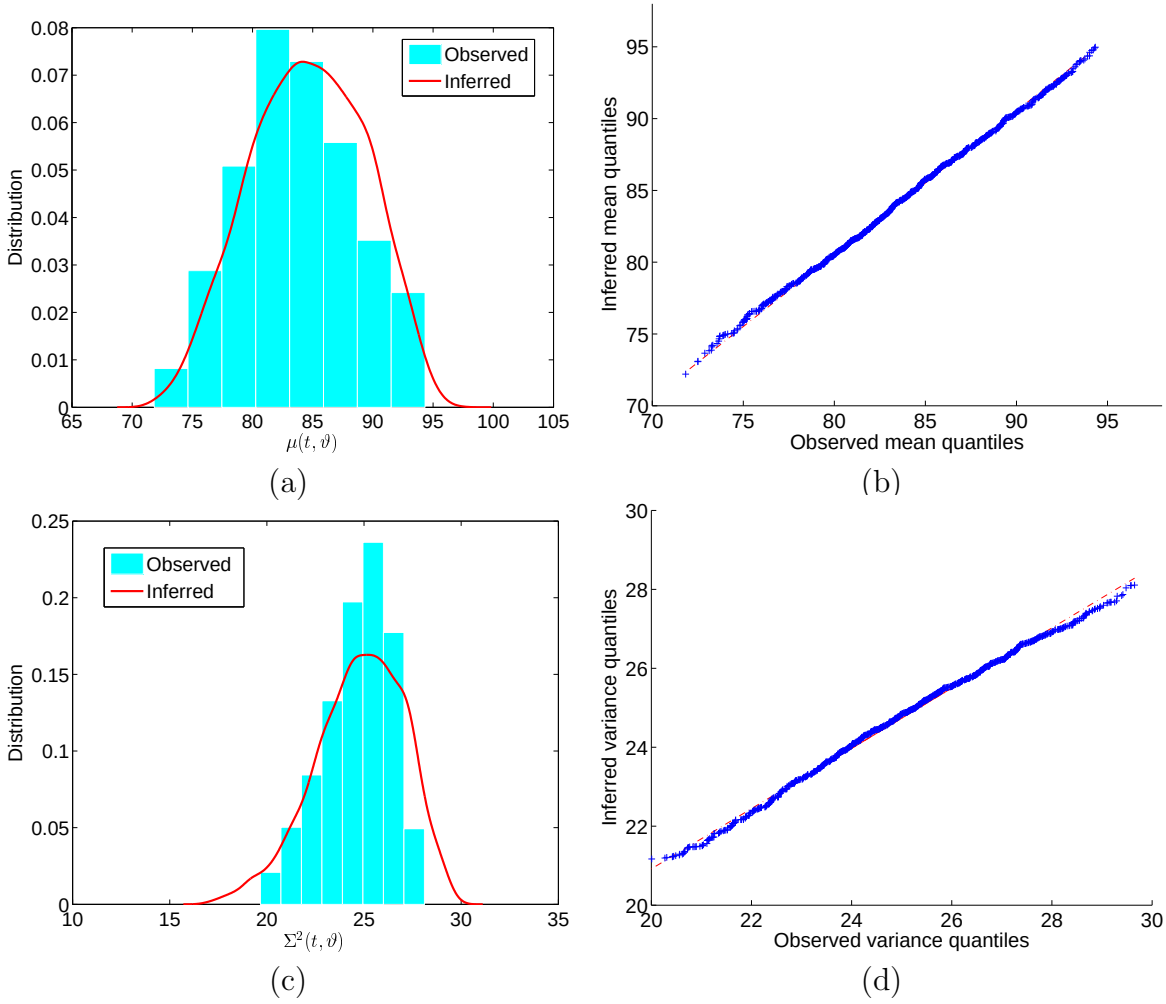


Figure 13: Recovery of mean and variance of the state variable at $t = 10$: (a) Comparing the PDF for $\tilde{\mu}(t, \vartheta)$ (computing via inference) versus the PDF of the observed model mean at the sampling points, (b) qq-plot comparing the distribution of the inferred and observed mean. (c) and (d) compare the recovered $\tilde{\Sigma}^2(t, \vartheta)$ versus the observed model variance at sampling points.

6 Genetic loop

In this section, we consider a more complex biochemical system modeling a genetic positive feedback loop [5]:

$$\begin{aligned} x + x &\underset{c_2}{\overset{c_1}{\rightleftharpoons}} y \\ y + d_0 &\underset{c_4}{\overset{c_3}{\rightleftharpoons}} d_r \\ d_0 &\xrightarrow{c_5} d_0 + m \\ d_r &\xrightarrow{c_6} d_r + m \\ m &\xrightarrow{c_7} m + x \\ x &\xrightarrow{c_8} \emptyset \\ m &\xrightarrow{c_9} \emptyset \end{aligned} \tag{15}$$

Here x is a protein monomer and y is a protein dimer. The quantity d_0 denotes the promoter sites that are free of the dimer, d_r denotes the promoter sites that are bound to the protein, and m is the mRNA. The first two reactions are reversible, and respectively describe the dimerization of the protein monomer, and the binding and unbinding processes of the protein dimer to the promoter site. The remaining reactions are irreversible, and describe the processes of transcription, translation, and degradation. Note that as mentioned in [5], the first two reactions have much faster time scales than the remaining ones. In what follows, the initial conditions are fixed:

$$x(0) = 10, \quad y(0) = 0, \quad d_0(0) = 20, \quad d_r(0) = 0, \quad m(0) = 0.$$

As for parametric uncertainty, all nine reaction rate parameters are assumed uncertain, and uniformly distributed around the "nominal" (mean) rate vector:

$$c = (50, 1000, 50, 1000, 1, 10, 3, 1, 6)^T,$$

with the width of each interval equal to 20% of the corresponding nominal value. In choosing the nominal reaction rates above, we followed [5]. Also, for the computations in this section, we let the sample set $\mathcal{S} \subseteq \mathbb{R}^9$ to consists of $N_q = 163$ realizations, corresponding to the nodes of a level 2 Smolyak quadrature based on the Gauss-Patterson rule (in nine dimensions). For brevity, we focus on a single quantity of interest, namely the state variable $y(t)$.

6.1 RRE-based scaling parameters

As mentioned before, one possible way of choosing the scaling factors used to precondition the stochastic dynamics is to compute them based on the corresponding RREs. The motivation behind this strategy is two-fold. In the first place, simulation of the RREs is cheaper than that of the stochastic system. Secondly, since the preconditioner is designed to scale and approximately collapse the large variation caused by the parametric uncertainty, the

RREs provide a robust means to estimate the scaling factors that is not affected by statistical convergence of the SSA. Note that results obtained using Bayesian regression (not shown) were found to be consistent with the results presented below. Thus, though it is not essential, the RRE approach provides a convenient and effective means to estimate the scaling parameters of the preconditioner.

For the present genetic loop mechanism, the reaction rates equations describing the limiting behavior of the system are given by:

$$\begin{cases} \dot{x} = 2k_2 y - 2k_1 x^2 + k_7 m - k_8 x, \\ \dot{y} = k_1 x^2 - k_2 y + k_4 d_r - k_3 y d_0, \\ \dot{d_0} = k_4 d_r - k_3 y d_0, \\ \dot{d_r} = k_3 y d_0 - k_4 d_r, \\ \dot{m} = k_5 d_0 + k_6 d_r - k_9 m, \end{cases}$$

where we have denoted the species concentrations in the RREs using same symbols as in the SSA, but have used $k_1, \ldots, k_9$ to denote the reaction rates in the ODE system. Assuming a unit volume, we have $k_1 = c_1/2$, and $k_i = c_i$, $i = 2, \ldots, 9$.

Figure 14(a) depicts the evolution of $y(t)$ for realizations of the random parameters. Based on these realizations, the scaling factors are defined by:

$$\hat{c}(\vartheta) = \lim_{t\to\infty} y(t,\vartheta), \qquad \hat{t}(\vartheta) = \text{ time taken } y(t,\vartheta) \text{ to reach } \frac{1}{2}\hat{c}(\vartheta),$$

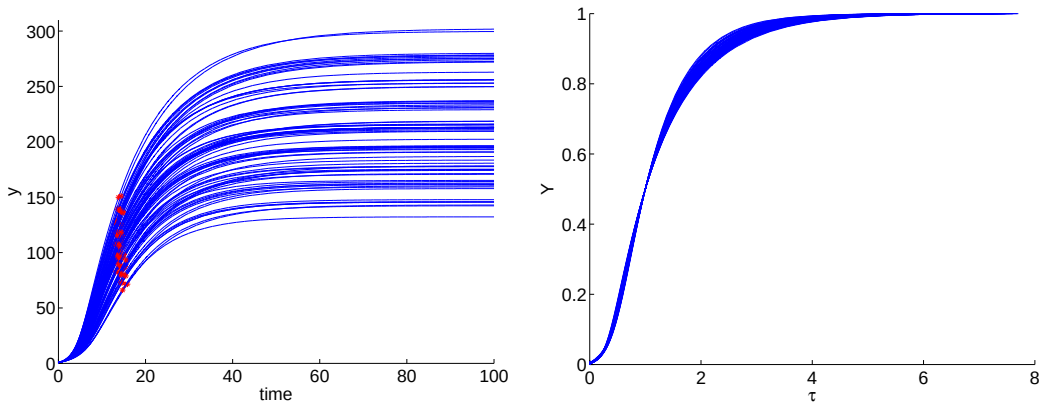
and the scaled variable is obtained according to:

$$Y(t/\hat{t}(\vartheta),\vartheta) = \frac{1}{\hat{c}(\vartheta)} y(t,\vartheta).$$

As shown in Figure 14(b), the preconditioner is quite effective at collapsing the large variation in the response of y with ϑ.

The result of projection of $\hat{c}$ and $\hat{t}$ in a PC basis (third order expansion) is depicted in Figure 15. We observe a severe stretching of amplitude scaling factors $\hat{c}$, whereas the time scaling factors $\hat{t}$ seem to be moderately stretched. These results should also be compared with the behavior of the system realizations in Figure 14(a). However, as seen in Figure 15, a sparse low order representation can be used to accurately estimate both $\hat{c}$ and $\hat{t}$. In fact, noting the shape of the distributions for both $\hat{c}$ and $\hat{t}$ suggests that expanding the logarithm of these quantities in a PC basis (as done in [2]) would allow even a linear expansion, attainable at fairly small computational effort; that is, one may compute the expansion coefficients for $\log \hat{c} \doteq \alpha_0 + \alpha_1\vartheta^1 + \cdots + \alpha_9\vartheta^9$ and then use $\hat{c}(\vartheta) \doteq \exp\left(\alpha_0 + \alpha_1\vartheta^1 + \cdots + \alpha_9\vartheta^9\right)$.

For the RREs, the discretization of the scaled variable and the recovery of the state variables can be done as described in detail in [2]. In addition to efficient estimates of the statistical properties of the state variable, the availability of suitable PC representations also affords a global sensitivity analysis, namely by computing Sobol indices through sampling of the recovered response surfaces. We illustrate this by estimating the total sensitivity indices



(a) (b)

Figure 14: (a) The time evolution of realizations of $y(t,\cdot)$; here the red marker shows the position of $\hat{t}(\vartheta)$ for each realization. (b) The scale variable Y versus the stretched time scale τ for the same realizations.

S_i^{tot} (for the variable y of the RREs), for $i = 1,\dots,9$. Recall that S_i^{tot} quantifies the total contribution of ϑ_i to the variance in y. For further details on Sobol sensitivity indices see for example [9, 42, 49, 50]. We report in Figure 16, the time evolution of S_i^{tot}, $i = 1,\dots,9$, for the variable y over time. Note that the reaction rates $c_1,\dots c_4$ corresponding to the forward and reverse rates of the dimerization and binding/unbinding reactions in (15), are the major contributors to the variance in the initial stages of the reactions; however, as time evolves, these "fast" reactions cease to play such a dominant role, and the other reaction rates become more influential to the solution variability. This is consistent with the analysis of the time-scales of the deterministic RREs [5].

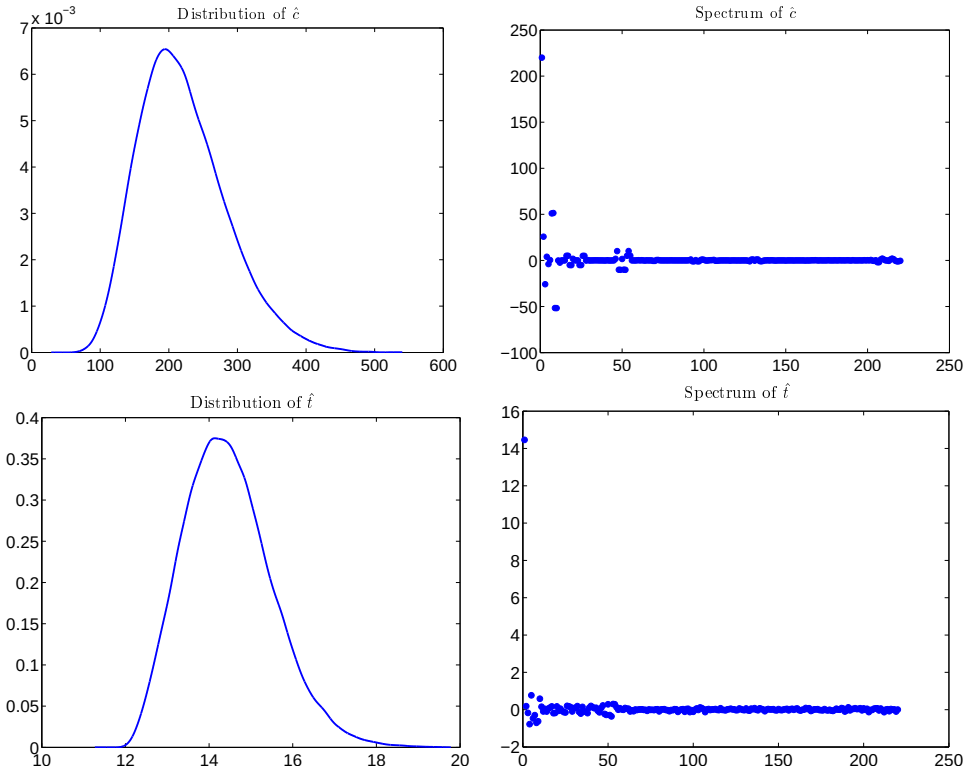


Figure 15: Distribution (left) and spectra (right) for the scaling factors. In top row, we show the PDF and the spectral coefficients for $\hat{c}$, and in the bottom row we provide the corresponding plots for $\hat{t}$.

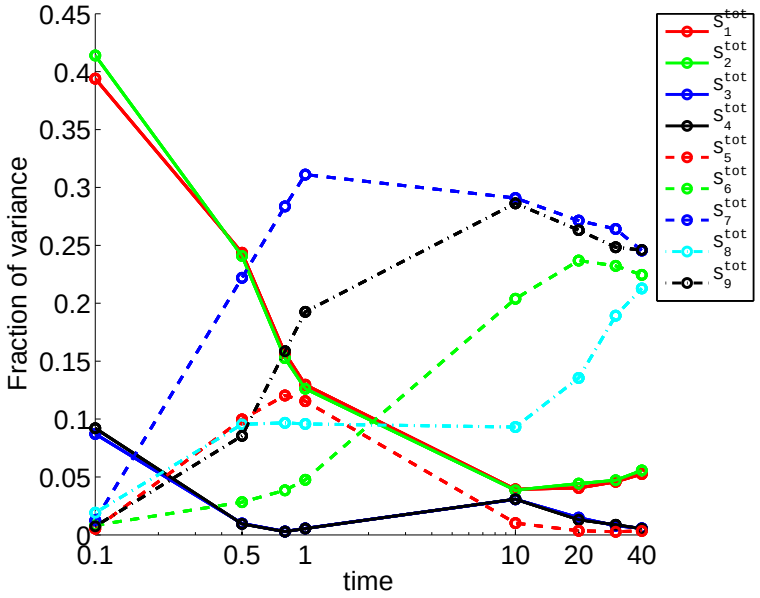


Figure 16: The total sensitivity indices over time. We used a logarithmic scale for the time axis to show the evolution of the indices more clearly.

6.2 Recovery of the mean and variance for the state variable

With the RRE based scaling factors, we use Algorithm 3 to discretize the SSA replicas of the scaled variable on a uniform scaled time mesh, and perform a Bayesian regression to infer the PC representations of the moments of the scaled variable. In Figure 17, we show the state variable, X, and its preconditioned counterpart, Y, in the top row, and show the averaging of X and Y over the SSA samples in the bottom row.

We then recover the mean and variance of the state variable as described in Section 4.2.4. As in the case of Michaelis-Menten system, owing to preconditioning, the scaled variables admit sparse low order PC representation. In this present example, we used a linear PC expansion for the mean and variance of the preconditioned variable.

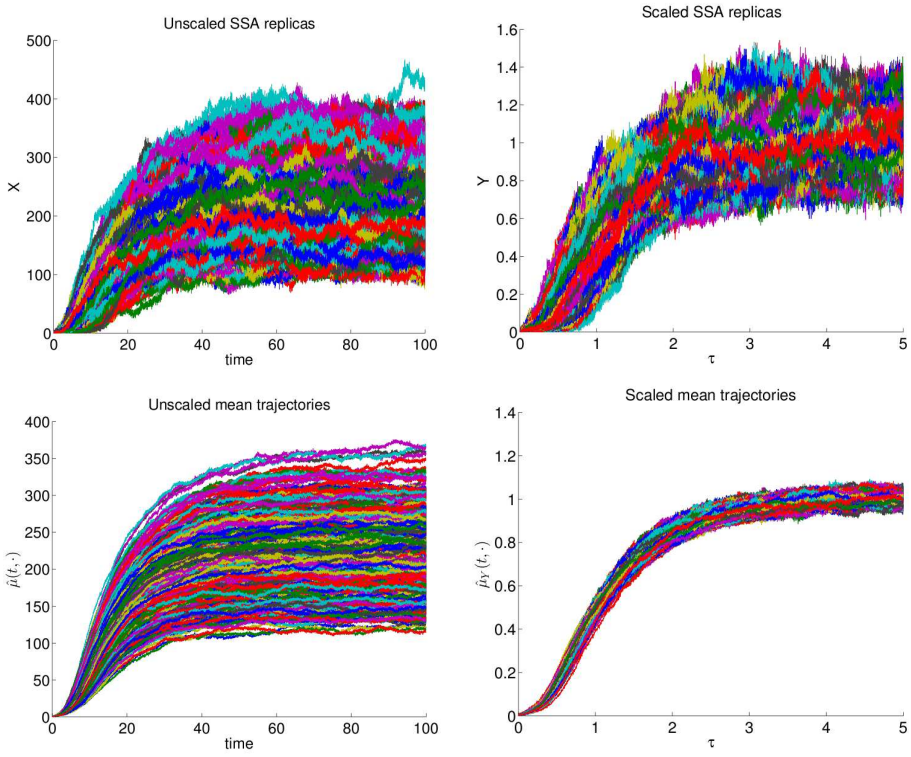


Figure 17: Preconditioning of stochastic dynamics: In top row, we plot the evolution of individual SSA replicas for state variable X and the scaled variable Y. In the bottom row we consider (observed) mean trajectories, $\hat{\mu}$, for state variable and $\hat{\mu}_Y$ for the scaled variable (in scaled time).

The genetic loop system is a stiff system with very small timescales, resulting in very small time steps in the SSA algorithm. This in turn makes the simulation of the system computationally expensive. We applied our algorithm with only 50 SSA replicas per each $\vartheta \in \mathcal{S}$. To illustrate the application of our algorithm, we provide results of recovering mean and variance at $t = 30$ for the genetic loop example. We also adopt a similar strategy used for the Michaelis-Menten example to validate the results. In particular, we chose a Monte

Carlo sample $\mathcal{S}_{mc} \subset \Theta$ consisting of 500 points and computed a Monte Carlo estimate for the mean, $\mu(t, \vartheta)$ and variance $\Sigma^2(t, \vartheta)$, by computing 1000 SSA replicas of the system for each $\vartheta \in \mathcal{S}_{mc}$.

Figure 18 shows results of recovering the mean, at $t = 30$. To further appreciate the wide range of values taken by $\mu(t, \vartheta)$ as ϑ varies over Θ we also show scatter plots which plot the recovered mean against the corresponding Monte Carlo estimates. We also compute estimates of relative $L^2(\Theta)$ errors as before. For the approximation to the mean, the estimated error was around 2%, indicating an accurate recovery. As for the approximation

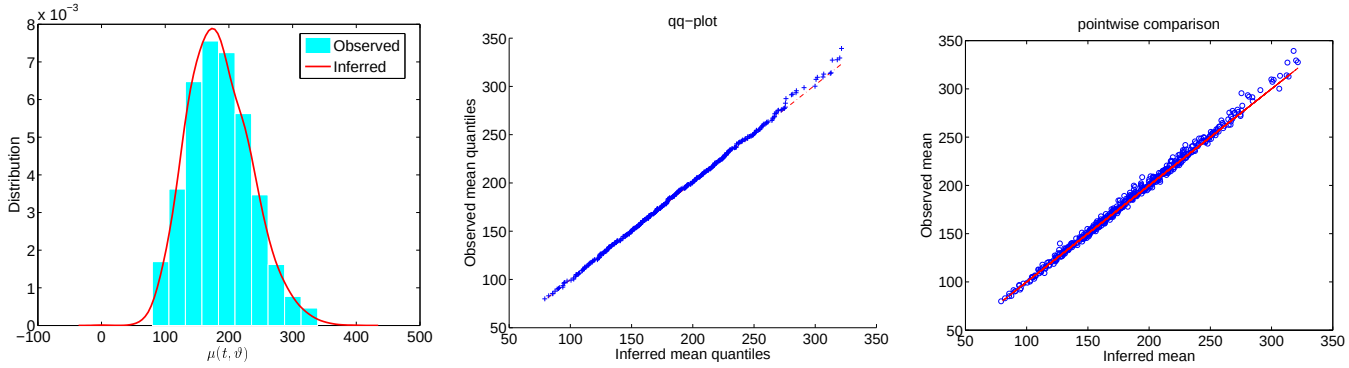


Figure 18: Recovery of mean at $t = 30$: We compare the recovered mean response surface $\tilde{\mu}(t, \vartheta)$ and the observed mean (computed with 1000 SSA realizations per $\vartheta \in \mathcal{S}_{mc}$). The left plot compares the PDF constructed by sampling $\tilde{\mu}(t, \vartheta)$ versus the observed distribution obtained from the corresponding Monte Carlo estimates, the middle plot compares the respective distributions through a qq-plot, and the right plot provides a pointwise comparison.

to the variance, we first point out that the present system exhibits considerably larger noise levels as compared to the example considered in the previous section; this is depicted in Figure 19 where we plot Monte Carlo estimates of the variance $\Sigma^2(t, \vartheta)$ over time, for each $\vartheta \in \mathcal{S}_{mc}$.

The results of noise recovery (at $t = 30$) are given in Figure 20. The estimated relative root-mean-square error is around 10%, and the corresponding estimate for the standard deviation (in physical units) is around 5%. These are reasonable estimates, given that for the present sample size the MC estimates exhibit comparable fluctuations. From the plots comparing the distribution of the recovered response surface against the observed variance, we note that the approximations seem to remain accurate except at the extreme tails of the distribution.

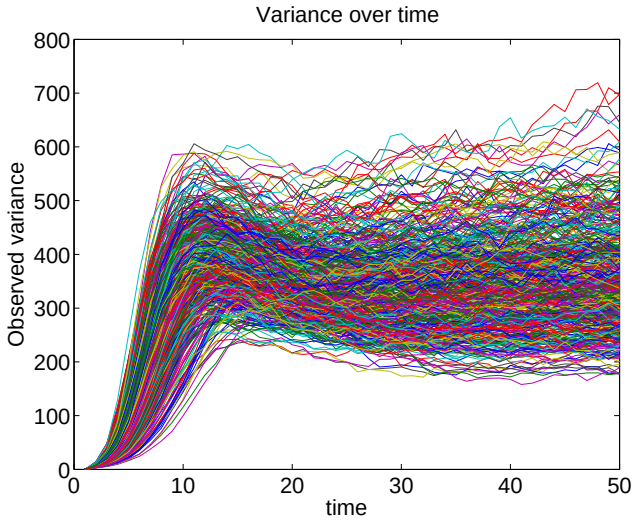


Figure 19: Observed variance over time. Each curve corresponds to a point $\vartheta \in \mathcal{S}_{mc}$ and shows the time evolution of the Monte Carlo estimate to $\Sigma^2(t, \vartheta)$.

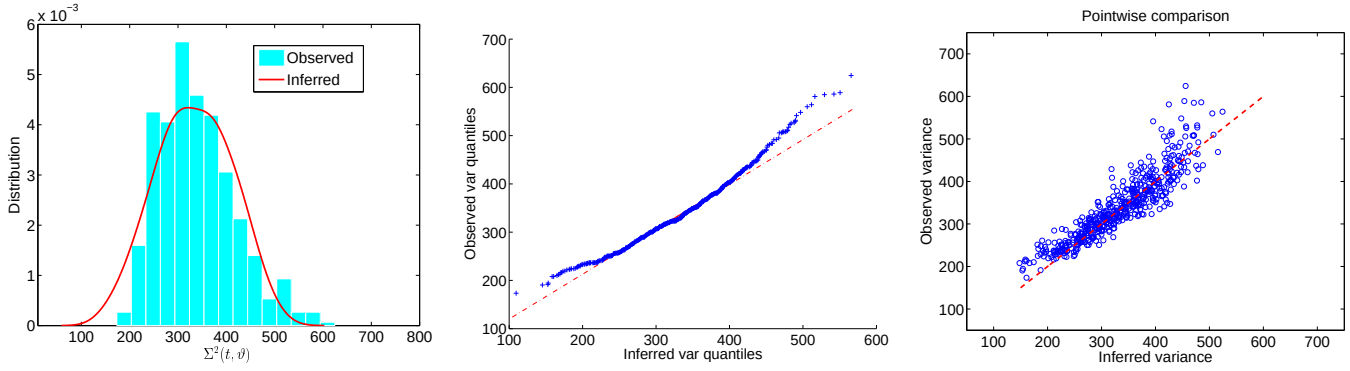


Figure 20: Recovery of variance at $t = 30$: We compare the recovered variance response surface $\tilde{\Sigma}^2(t, \vartheta)$ and the observed variance, computed with 1000 SSA realizations per $\vartheta \in \mathcal{S}_{mc}$. The left plot compares the PDF constructed by sampling $\tilde{\Sigma}^2(t, \vartheta)$ versus the observed distribution obtained from the corresponding Monte Carlo estimates, the middle plot compares the respective distributions through a qq-plot, and the right plot provides a pointwise comparison.

7 Conclusions

A preconditioned Bayesian regression method was developed for computing the mean and variance of the state variables of stochastic chemical systems with uncertain reaction rates. The method extends the stochastic preconditioning approach introduced in [2] for deterministic systems. It utilizes appropriate transformations (preconditioning) of the state variables and infers a PC model for the transformed (preconditioned) variables via Bayesian regression. Two possibilities were considered for computing the preconditioner, which is characterized by time and amplitude scaling factors. The first approach uses Bayesian regression, whereas PC expansions of the scaling factors are determined in the second using deterministic realizations of the RREs for the chemical system. One then proceeds to discretize the resulting preconditioned variables on a scaled time mesh, and Bayesian regression is used to infer PC response surfaces for their moments. The process results in response surfaces for the expected trajectory and variance of the stochastic system, involving combinations of deterministic PC models for the preconditioners and stochastic PC models of the preconditioned variables.

The approach is illustrated for the case of a Michaelis-Menten dynamics and a genetic positive feedback loop. Both of the examples show large noise amplitudes, the latter also exhibiting steep response to uncertainty in the reaction rates. Computational experiments show that Bayesian preconditioning algorithms can simultaneously accommodate large variation with uncertain inputs and high fluctuation levels, and that robust estimates can be obtained with a moderate number of SSA samples. This includes the dependence of both the mean and variance of the state variables on the uncertain inputs. This constitutes a significant advantage, since performing a large number of SSA replicas in high dimensions can be prohibitively expensive.

Finally, we briefly outline potential extensions that are motivated by the present experiences. We first note that in the implementations outlined in this paper, we have relied on sparse grids for sampling the random parameter space. One immediate avenue of exploration is to employ adaptive sparse grids and adaptive PC models. Such strategies, which can significantly reduce the computational cost of a straightforward NISP approach, appear to be promising in the present context also. Another interesting possibility arises from the fact that, owing to preconditioning, one expects the preconditioned variables to admit sparse low order expansions. Thus, the use of sparsity priors to guide the construction of a sparse PC basis may provide an attractive approach especially in high dimensional settings. An additional avenue concerns the representation of the model noise. The linear model used for the model noise, though adequate in the examples considered in this paper can be replaced with more general models. Such constructions are the subject of ongoing research.

Acknowledgements

This work was supported by the US Department of Energy (DOE) under Award Numbers DE-SC0001980 and DE-SC0008789. The work of OLM is partially supported by the French Agence Nationale pour la Recherche (Project ANR-2010-Blan-0904) and the GNR MoMaS funded by Andra, Brgm, Cea, Edf, and Irsn. Finally, would like to thank the reviewer for helpful comments on improving this manuscript.

References

[1] M. Abramowitz and I.A. Stegun. Handbook of mathematical functions. Dover, New York, 9 edition, 1972.

[2] A. Alexanderian, O.P. Le Maître, H.N. Najm, M. Iskandarani, and O.M. Knio. Multiscale stochastic preconditioners in non-intrusive spectral projection. Journal of Scientific Computing, 50:306–340, 2012.

[3] C. Andrieu and E. Moulines. On the ergodicity properties of some adaptive MCMC algorithms. Ann. Appl. Probab., 16:1462–1505, 2006.

[4] Y. F. Atchadé and J. S. Rosenthal. On adaptive Markov chain Monte Carlo algorithms. Bernoulli, 11:815–828, 2005.

[5] M.R. Bennett, D. Volfson, L. Tsimring, and J. Hasty. Transient dynamics of genetic regulatory networks. Biophysical Journal, 92(10):3501–3512, 2007.

[6] G.E.P. Box and G.C. Tiao. Bayesian inference in statistical analysis. Addison-Wesley Publishing Co., Reading, Mass.-London-Don Mills, Ont., 1973. Addison-Wesley Series in Behavioral Science: Quantitative Methods.

[7] R. H. Cameron and W. T. Martin. The orthogonal development of non-linear functionals in series of fourier-hermite functionals. Ann. Math., 48:385–392, 1947.

[8] B.P. Carlin and T.A. Louis. Bayesian methods for data analysis. Texts in Statistical Science Series. CRC Press, Boca Raton, FL, third edition, 2009.

[9] T. Crestaux, O.P. Le Maitre, and J.-M. Martinez. Polynomial chaos expansion for sensitivity analysis. Reliability Engineering & System Safety, 94(7):1161 – 1172, 2009. Special Issue on Sensitivity Analysis.

[10] W. E, D. Liu, and E. Vanden-Eijnden. Nested stochastic simulation algorithms for chemical kinetic systems with multiple time scales. Journal of Computational Physics, 221(1):158 – 180, 2007.

[11] O.G. Ernst, A. Mugler, H.-J. Starkloff, and E. Ullmann. On the convergence of generalized polynomial chaos expansions. ESAIM: Mathematical Modelling and Numerical Analysis, 46:317–339, 2012.

[12] S.N. Ethier and T.G. Kurtz. Markov processes. Characterization and convergence. Wiley Series in Probability and Mathematical Statistics, 1986.

[13] R. Tempone F. Nobile and C.G. Webster. An anisotropic sparse grid stochastic collocation method for partial differential equations with random input data. SIAM J. Numer. Anal., 46(5):2411–2442, 2008.

[14] B. Ganapathysubramanian and N. Zabaras. Sparse grid collacation schemes for stochastic natural convection problems. J. Comput. Phys., 225:652–685, 2007.

[15] A. Gelman, J.B. Carlin, H.S. Stern, and D.B. Rubin. Bayesian data analysis. Texts in Statistical Science Series. Chapman & Hall/CRC, Boca Raton, FL, second edition, 2004.

[16] T. Gerstner and M. Griebel. Numerical integration using sparse grids. Numer. Algorithms, 18:209–232, 1998.

[17] R.G. Ghanem and P.D. Spanos. Stochastic Finite Elements: A Spectral Approach. Dover, 2002. 2nd edition.

[18] D.T. Gillespie. A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. Journal of Computational Physics, 22(4):403 – 434, 1976.

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

[20] D.T. Gillespie. A rigorous derivation of the chemical master equation. Physica A: Statistical Mechanics and its Applications, 188(13):404 – 425, 1992.

[21] D.T. Gillespie. Approximate accelerated stochastic simulation of chemically reacting systems. The Journal of Chemical Physics, 115(4):1716–1733, 2001.

[22] D.T. Gillespie and L.R. Petzold. Improved leap-size selection for accelerated stochastic simulation. The Journal of Chemical Physics, 119(16):8229–8234, 2003.

[23] H. Haario, E. Saksman, and J. Tamminen. An adaptive metropolis algorithm. Bernoulli, 7:223–242, 2001.

[24] S. Janson. Gaussian Hilbert Spaces. Cambridge University Press, 1997.

[25] A. Keese. Numerical Solution of Systems with Stochastic Uncertainties : A General Purpose Framework for Stochastic Finite Elements. PhD thesis, Tech. Univ. Braunschweigh, 2004.

[26] A. Keese and H.G. Matthies. Numerical methods and Smolyak quadrature for nonlinear stochastic partial differential equations. Technical report, Institute of Scientific Computing TU Braunschweig Brunswick, 2003.

[27] O.P. Le Maître and O.M. Knio. Spectral Methods for Uncertainty Quantification With Applications to Computational Fluid Dynamics. Scientific Computation. Springer, 2010.

[28] O.P. Le Maître, L. Mathelin, O.M. Knio, and M.Y. Hussaini. Asynchronous time integration for polynomial chaos expansion of uncertain periodic dynamics. Discrete and Continuous Dynamical Systems, 28(1):199–226, 2010.

[29] O.P. Le Maître, H.N. Najm, P.P. Pebay, R.G. Ghanem, and O. M. Knio. Multiresolution-analysis scheme for uncertainty quantification in chemical systems. SIAM Journal on Scientific Computing, 29(2):864–889, 2007.

[30] X. Ma and N. Zabaras. An adaptative hierarchical sparse grid collocation algorithm for the solution of stochastic differential equations. J. Comp. Phys., 2009. in press.

[31] H.N. Najm. Uncertainty quantification and polynomial chaos techniques in computational fluid dynamics. Ann. Rev. Fluid Mech., 41:35–52, 2009.

[32] H.N. Najm, B. Debusschere, Y. Marzouk, S. Widmer, and O.P. Le Maître. Uncertainty quantification in chemical systems. Int. J. Num. Eng., 80(6):789–814, 2009.

[33] K. Petras. On the smolyak cubature error for analytic functions. Advances in Computational Mathematics, 12:71–93, 2000.

[34] K. Petras. Fast calculation in the smolyak algorithm. Num. Algo., 26:93–109, 2001.

[35] M. Rathinam, L.R. Petzold, Y. Cao, and D.T. Gillespie. Stiffness in stochastic chemically reacting systems: The implicit tau-leaping method. The Journal of Chemical Physics, 119(24):12784–12794, 2003.

[36] M.T. Reagan, H.N. Najm, B.J. Debusschere, O.P. Le Maître, O.M. Knio, and R.G. Ghanem. Spectral stochastic uncertainty quantification in chemical systems. Combustion Theory and Modelling, 8:607–632, 2004.

[37] M.T. Reagan, H.N. Najm, R.G. Ghanem, and O.M. Knio. Uncertainty quantification in reacting flow simulations through non-intrusive spectral projection. Combustion and Flame, 132:545–555, 2003.

[38] F. Rizzi, H.N. Najm, B.J. Debusschere, K. Sargsyan, M. Salloum, H. Adalsteinsson, and O.M. Knio. Uncertainty quantification in MD simulations. Part I: Forward propagation. Multiscale Model. Simul., submitted, 2011.

[39] F. Rizzi, H.N. Najm, B.J. Debusschere, K. Sargsyan, M. Salloum, H. Adalsteinsson, and O.M. Knio. Uncertainty quantification in MD simulations. Part II: Bayesian inference of force-field parameters. Multiscale Model. Simul., submitted, 2011.

[40] G.O. Roberts and J.S. Rosenthal. Examples of adaptive MCMC. J. Comput. and Graph. Stat., 18:349–367, 2009.

[41] M. Salloum, K. Sargsyan, R. Jones, B. Debusschere, H.N. Najm, and H. Adalsteinsson. A stochastic multiscale coupling scheme to account for sampling noise in atomistic-to-continuum simulations. Multiscale Model. Simul., 10:550–584, 2012.

[42] A. Saltelli. Sensitivity analysis for importance assessment. Risk Analysis, 22(3):579–590, 2002.

[43] K. Sargsyan, B. Debusschere, H.N. Najm, and O.P. Le Maître. Spectral representation and reduced order modeling of the dynamics of stochastic reaction networks via adaptive data partitioning. SIAM J. Sci. Comput., 31:4395–4421, 2010.

[44] K. Sargsyan, B. Debusschere, H.N. Najm, and Y. Marzouk. Bayesian inference of spectral expansions for predictability assessment in stochastic reaction networks. Journal of Computational and Theoretical Nanoscience, 6(10):2283–2297, 2009.

[45] K. Sargsyan, C. Safta, B. Debusschere, and H. Najm. Multiparameter spectral representation of noise-induced competence in bacillus subtilis.

IEEE/ACM Transactions on Computational Biology and Bioinformatics, 2012. http://doi.ieeecomputersociety.org/10.1109/TCBB.2012.107.

[46] F. Schlögl. On thermodynamics near a steady state. Zeitschrift für Physik A Hadrons and Nuclei, 248(5):446–458, 1971.

[47] D.A. Sheen and H. Wang. Combustion kinetic modeling using multispecies time histories in shock-tube oxidation of heptane. Combustion and Flame, 158:645–656, 2011.

[48] S.A. Smolyak. Quadrature and interpolation formulas for tensor products of certain classes of functions. Dokl. Akad. Nauk SSSR, 4:240–243, 1963.

[49] I.M Sobol. Global sensitivity indices for nonlinear mathematical models and their monte carlo estimates. Mathematics and Computers in Simulation, 55(1-3):271 – 280, 2001. The Second IMACS Seminar on Monte Carlo Methods.

[50] B. Sudret. Global sensitivity analysis using polynomial chaos expansions. Reliability Engineering & System Safety, 93(7):964 – 979, 2008.

[51] M. Villegas, F. Augustin, A. Gilg, A. Hmaidi, and U. Wever. Application of the Polynomial Chaos Expansion to the simulation of chemical reactors with uncertainties. Mathematics and Computers in Simulation, 82:805–817, 2012.

[52] N. Wiener. The Homogeneous Chaos. Amer. J. Math., 60:897–936, 1938.

[53] D.J. Wilkinson. Stochastic modelling for systems biology. Chapman & Hall/CRC Mathematical and Computational Biology Series. Chapman & Hall/CRC, Boca Raton, FL, 2006.

[54] D.B. Xiu and G.E. Karniadakis. The Wiener-Askey Polynomial Chaos for stochastic differential equations. SIAM J. Sci. Comput., 24:619–644, 2002.