Journal of Computational Physics 273 (2014) 374–392

Contents lists available at ScienceDirect

Journal of Computational Physics

www.elsevier.com/locate/jcp

Multi-scenario modelling of uncertainty in stochastic chemical systems

R. David Evans [b], Luis A. Ricardez-Sandoval [a,b,*]

[a] Department of Chemical Engineering, University of Waterloo, 200 University Avenue West, Waterloo, ON, N2L 3G1, Canada
[b] Waterloo Institute for Nanotechnology, University of Waterloo, 200 University Avenue West, Waterloo, ON, N2L 3G1, Canada

ARTICLE INFO

Article history:
Received 4 July 2013
Received in revised form 16 March 2014
Accepted 20 May 2014
Available online 27 May 2014

Keywords:
Uncertainty
Stochastic chemical system
Chemical Master Equation
Monte Carlo
Gene regulation

ABSTRACT

Uncertainty analysis has not been well studied at the molecular scale, despite extensive knowledge of uncertainty in macroscale systems. The ability to predict the effect of uncertainty allows for robust control of small scale systems such as nanoreactors, surface reactions, and gene toggle switches. However, it is difficult to model uncertainty in such chemical systems as they are stochastic in nature, and require a large computational cost. To address this issue, a new model of uncertainty propagation in stochastic chemical systems, based on the Chemical Master Equation, is proposed in the present study. The uncertain solution is approximated by a composite state comprised of the averaged effect of samples from the uncertain parameter distributions. This model is then used to study the effect of uncertainty on an isomerization system and a two gene regulation network called a repressilator. The results of this model show that uncertainty in stochastic systems is dependent on both the uncertain distribution, and the system under investigation.

1. Introduction

In the current age of mathematical models, uncertain terms are often introduced as a means to account for variability in process outputs that determine a system's performance. Bulk models of chemical reaction systems have been well studied and characterized [1,2], however these models can only be applied when the populations of chemical species are significantly large. As applications of chemistry approach the nanoscale, such models may become inaccurate due to random fluctuations in the position and energy of atoms and molecules causing significant stochastic (random) noise [3–5]. On such scales, chemical reactions are better modelled by the Chemical Master Equation (CME) which describes reactions as discrete probabilistic events [6,7]. In practice, the CME and its derivatives have been used to study biological and inorganic reactions such as enzyme catalysis [8], circadian rhythms [9], genetic switches [10–12], heat shock response [13], and crystallization [14].

Although small scale chemical reactions are known to be stochastic and are well studied, the effect of random parameters on such reactions has been studied very little. This is despite the fact that parameter uncertainty is an important and well-studied property [15], especially in large scale chemical reactions [16,17]. Stochastic reaction networks have had parameter-dependant properties such as the sensitivity examined [18–21], however there is a lack of literature regarding model parameter uncertainty, especially while using the CME. This lack of knowledge is potentially due to the large number

* Corresponding author. Tel.: +1 519 888 4567x38667.
E-mail address: laricardezsandoval@uwaterloo.ca (L.A. Ricardez-Sandoval).

http://dx.doi.org/10.1016/j.jcp.2014.05.028
0021-9991/

of equations required to create a CME model, which increases exponentially with the number of species and their respective populations because of a necessity to account for nearly every potential state of the system. This results in prohibitive computational costs when modelling complex systems. Algorithms for approximating uncertainty propagation in the CME are further complicated due to representing and propagating the additional information involved with the uncertain parameters, causing an additional computational cost. Hence, there is a need to develop practical and efficient methods that can describe the output variability in systems modelled with the CME due to model parameter uncertainty.

The goal of this work is to model and analyze uncertainty propagation in the CME. Potential uncertain terms in the CME were identified, and an algorithm called the Multiple Scenario Chemical Master Equation (MSCME) is proposed to model its propagation. The proposed algorithm functions by taking samples of the uncertain parameter distributions and calculating the effect of each separately before weighting back to a composite state. Two representative reaction systems are presented as case studies for the method. First, a system comprised of a simple interconversion between two species is presented which models a simple isomerization. Many reaction systems can be decomposed into a series of single step reactions of similar structure. Also, the small, finite number of states considered for this system results in small computational costs. Hence this system serves as an illustrative example of the proposed method and validation of the proper functioning of the algorithm. The main system of interest is a gene regulation system called a repressilator [12,22]. This models the bidirectional repression of two genes by introducing reaction rates dependent on the opposite gene. By understanding the effect of uncertainty in the cooperativity parameters associated with this system, it is possible to better characterize the effects of influences such as protein irregularities and inhibition on a gene regulation system.

Design of pharmaceuticals requires explicit knowledge of the ways in which a drug will react in a cellular environment, however fluctuations due to cellular processes mean that the circumstances under which a reaction will take place are not constant, and are subject to uncertainty. Biological chemical systems, including feedback controls, gene regulatory networks and protein catalyzed reactions have very small populations which further magnifies the effect of uncertainty in the cellular environment [23]. A genetic feedback system, the bi-directional inhibition of two genes, is modelled in the second case study of this work. Many external environment variables affect this system besides the gene products, such as the inhibition state proteins that catalyze the reactions. However, general models will account for the most significant of these components while leaving many unaccounted for. These unaccounted parameters introduce uncertainty into the model, as some protein or chemical might cause small but measurable changes to a reaction by acting on one of the measured components. By characterizing and modelling the effect of this uncertainty, the operating boundaries and stability of the system can be assessed more realistically.

Parameters characterizing a set of reactions are often determined experimentally or through simulation. For small systems such as catalytic nanoparticles, fluctuations result in very large uncertainties in the reaction rates [24]. While the time evolution of fine scale events such as molecular motion and structure of certain systems cannot be explicitly modelled due to prohibitive computational times, these phenomena can be described as a time-varying uncertain parameter in a coarse-scale model that can be solved in reasonable computational times. For example, a molecular dynamics model that explicitly describes the molecular motion of the reactants in a heterogeneous catalytic system cannot be used to determine the optimal operating conditions for this process because of the prohibitive computational costs associated with that calculation. Instead, a coarse-grained model, which includes time-varying uncertain parameters that account for the system's variability due to molecular motion, can be derived and used to account for that fine-scale phenomenon. This issue is also present for experimental measurements, due to the difficulty of obtaining measurements of micro and nanoscale phenomena. As nanostructured catalysts and nanoreactors become commonplace in industrial chemical engineering, the ability to predict and account for the effect of uncertainty on process outputs will become important.

The next section of this work outlines the mathematic formulation of the algorithm proposed, as well as required background on the CME. Case studies A and B (isomerization and repressilator systems, respectively) are examined to test the merit of the proposed algorithm. Computational cost and scaling of the algorithm is analyzed in context of other methods to solve the CME. Two variants of the proposed method are compared in this study. Conclusions regarding the algorithm's effectiveness and the effects of uncertainty propagation through the examined systems are stated in the conclusions.

2. Mathematical formulations

The CME establishes formalism with which to define a reaction system and solve for its evolution through time. As such, its structure was examined to define methods to approximate the effect of model parameter uncertainty on the system's outputs.

The algorithm proposed in this work was derived directly from a system of equations generated by the CME, despite faster algorithms being available that solve or approximate its solution, e.g., order reduction [12,25,26], Poisson representation [27,28], tensor based approaches [29,30], and the Kinetic Monte Carlo (KMC) [31,32]. These approaches or solution methods have been designed to exclusively solve the CME and therefore require modification if they are to account for parameter uncertainty, which would also introduce further sources of error even if the models are capable of accommodating uncertain terms. The most widely used method, the KMC, has a relatively large computation time for smaller molecular populations [31]. Similarly, introducing additional uncertain parameters to the KMC slows its convergence rate because it functions by taking the average of many stochastic trajectories to obtain a result, and adding uncertain parameters into the analysis increases the disparity between trajectories. This is further discussed in Section 4, when the computational time