

Local diffusion models for stochastic reacting systems: estimation issues in equation-free numerics

C. P. CALDERON*

Department of Chemical Engineering, Princeton University, Princeton, NJ, USA

(Received January 2007; in final form January 2007)

Extracting coarse-grained derivative information from fine scale, atomistic/stochastic simulations constitutes an important component of multiscale numerics for reacting systems. In this paper, we demonstrate the use of local parametric diffusion models; observing the output of short bursts of stochastic simulation, the drift and diffusion coefficients of such local models are obtained “on demand” by maximizing an approximation of the likelihood function. The parametric fit utilizes the likelihood expansion of Ait-Sahalia, giving closed-form expressions for the transition density of both scalar and multivariate processes. The fine scale simulations generating the data are the Gillespie stochastic simulation algorithm as well as stochastic differential equations (SDEs). The postulated *local* parametric diffusion models are SDEs with affine drift and diffusion functions; the information extracted from these (estimated) local parametric models is utilized in various “equation-free” computational tasks. We demonstrate how such local diffusion modeling can (1) contribute to linking stochastic simulation with traditional, continuum numerical methods such as fixed point and initial value problem solving algorithms as well as variational integration; (2) aid in extracting quantitative information about the invariant distribution associated with the underlying stochastic system; and (3) be used in computational model reduction motivated by singular perturbation theory.

Keywords: Equation-free computation; Local diffusion process estimation; Maximum likelihood; Stochastic reaction networks

1. Introduction

The primary interest of this paper is to enhance fine scale (atomistic, stochastic) simulations of chemically reacting systems by developing an efficient estimation strategy which uses short bursts of atomistic simulation output to estimate local “effective equations”; this information is then communicated to traditional, continuum numerical methods for further processing [37,42]. In previous studies [27,37,38,40,56] such so-called *equation-free* numerical methods appealed to regression in time in order to estimate quantities associated with effective equations “on-demand” from fine scale simulation data. Here we describe and implement a local estimation strategy that focuses on state-space estimation [31]. For a simple illustration of the two viewpoints, consider the following differential equation consisting of a “signal” (\mathbf{F}) and

a “noise” ($d\mathbf{N}_t/dt$):

$$\frac{d\mathbf{X}}{dt} = \mathbf{F}(\mathbf{X}) + \frac{d\mathbf{N}_t}{dt}. \quad (1)$$

Given one or more observation sequences $\{\mathbf{X}_t\}_{t=0}^T$, regression in time would attempt to explicitly model \mathbf{X} as a function of t (e.g. through a polynomial). A state-space estimation procedure would attempt to explicitly model \mathbf{F} as a function of \mathbf{X} given this(ese) same sequence(s).

If accurate effective equations can be obtained in closed-form for a problem, it is clearly desirable to work with these equations, because more detailed atomistic simulations such as Gillespie’s stochastic simulation algorithm (SSA) [28] or molecular dynamics (MD) [30,34,42] are often too computationally expensive to carry out in a reasonable amount of CPU time with current processor technology. Coarse-grained, effective equations

*Corresponding author. Email: ccaldero@princeton.edu

‡By noise, we mean a stochastic process which has increments that are independent and identically distributed (iid).

provide a means of usefully approximating important features of the full atomistic simulation and are typically much computationally cheaper than the full simulation. The problem with this approach is that global effective equations are usually not known *a priori*, nor easy to explicitly derive in closed-form.

In this paper, we describe and implement a computer-assisted strategy aimed at overcoming these difficulties by appealing to local maximum likelihood estimation [11,12]. In the simple case just given, we would attempt to repeatedly find *local* linear approximations of $F(\cdot)$ as opposed to a global nonlinear one (figure 1). By showing how the information contained in such local models can be integrated into a variety of classical numerical algorithms, the paper illustrates the advantages that taking this viewpoint can offer in enhancing the computer-assisted analysis of a class of complex multiscale reacting systems; the method is also applicable to other types of multiscale models, e.g. [11,13].

In what follows, we illustrate our approach using data generated by two different fine scale “inner simulators”: both SSA and stochastic differential equations (SDEs). We will make extensive use of diffusion approximations, i.e. we assume the output of all of our atomistic simulation can be adequately modeled by an SDE that is driven by a Brownian motion [25]. Diffusion approximations of SSA type simulations have been extensively studied from both theoretical [7,21] and simulation perspectives [14,18,32,55]. Many properties of limiting cases where a

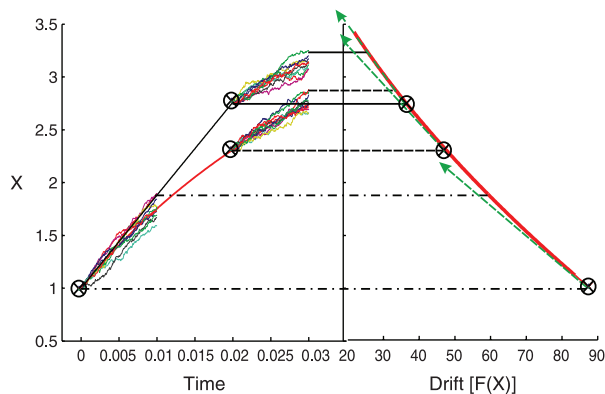


Figure 1. Local MLE. The estimation strategy employed in this paper shifts the focus from estimation in time (left) to estimation in state (right). The models used assume that a diffusion approximation of the actual local dynamics through linear (with respect to the state) drift and diffusion coefficient functions is accurate. The left panel of the figure plots three batches of trajectories corresponding to a coarse projective integration computation (see Section 4); the panel on the right plots the nonlinear drift function which governs the dynamics (in this example the noise term does not depend on the state). The three different line types connecting the two regions correspond to the range of the state values observed in the three different simulation batches. The circles in the figure correspond to the point where an ensemble of trajectories with Dirac distributed initial conditions was initialized (and where the drift was then estimated). The straight dotted arrows in the right panel show the linearization of the true nonlinear drift function with respect to the state at the “expansion points” (the circles).

“size-parameter” $\Omega \rightarrow \infty$ have been established [7,29]. The *coarse observables* [37] used here correspond to species concentrations $\equiv N_i/\Omega$, where N_i is the number of molecules of species i § and the physical interpretation of the constant Ω depends on the application (e.g. it can correspond to a fixed system volume). We note that it is not currently completely understood *quantitatively* how large Ω should be in order for diffusion approximations to be useful in general SSA type processes. In Ref. [12] a computer-assisted approach was presented which yielded some quantitative *a posteriori* answers to this problem. The final portion of this paper builds on this approach and demonstrates how the statistical tools developed in Ref. [33] can be modified to test the statistical validity of a diffusion approximation-based model reduction strategy in systems that exhibit multiple characteristic time scales. We note a difference between the diffusion approximations used here and those used in Refs. [7,29]: we do not equate the number of driving Brownian terms with the number of “elementary reaction events” [29]; each of our state components has only one Brownian term associated with it. In this paper, we simply approximate the dynamics associated with the global nonlinear jump process through a sequence of locally linear SDEs.

The remainder of this article is organized as follows: Section 2 presents the local models used and the assumptions made about the data (we estimate parameters associated with scalar as well as multivariate models); this section also reports some established properties of the estimator used in addition to reviewing some advances in mathematical statistics and finance relevant to our task. Section 3 presents the model systems studied. In Section 4 we aim at providing illustrative examples that demonstrate how the information contained in the estimated parameters associated with the simple (local) models proposed can be linked with a variety of traditional numerical methods used to study dynamical systems. In Section 5 we demonstrate how the local models can be used to approximate features associated with a system where noise plays an important role in the observed dynamics. Specifically we calculate the mode of an “almost”-invariant distribution associated with a nonlinear scalar SDE using only local models. In Section 6, we demonstrate how the approach can be used in numerical model reduction tasks. In this section we also demonstrate some goodness-of-fit test procedures that can be combined with our estimation strategy; we conclude in Section 7.

2. Local diffusion models and likelihood expansions

In what follows, we will assume *a priori* that species concentrations (N_i/Ω) of the SSA and SDE simulations based on chemical kinetic reaction schemes are good

§The N_i typically also tend to infinity as $\Omega \rightarrow \infty$.

“coarse observables” [38], and that the evolution of these observables can be adequately approximated by an effective diffusion process that possesses deterministic trend coefficients; that is, an (in principle unavailable in closed-form) SDE of the form:

$$d\mathbf{X}_t = \boldsymbol{\mu}(\mathbf{X}_t; \boldsymbol{\theta})dt + \boldsymbol{\Sigma}(\mathbf{X}_t; \boldsymbol{\theta})d\mathbf{W}_t. \quad (2)$$

Here \mathbf{X}_t is a stochastic process which is meant to model the evolution of the observable(s), \mathbf{W}_t represents a vector of standard Brownian motions, and the functions $\boldsymbol{\mu}(\cdot; \boldsymbol{\theta})$ and $\boldsymbol{\Sigma}(\cdot; \boldsymbol{\theta})$ are the drift and diffusion coefficients of the process.

In the classical parametric setup one assumes that the parameterized function families to which the drift and diffusion coefficient functions belong are known, and that the parameter vector $\boldsymbol{\theta}$ is finite dimensional. In practice, one rarely knows a class of functions which can be used to successfully describe the *global* dynamics of the observables; in the equation-free computations below, however, we simulate the true process for only relatively short bursts of time. It therefore makes sense to (locally) consider the following SDE:

$$d\mathbf{X}_t = (\mathbf{A} + \mathbf{B}(\mathbf{X}_t - \mathbf{X}_0))dt + (\mathbf{C} + \mathbf{D}(\mathbf{X}_t - \mathbf{X}_0))d\mathbf{W}_t. \quad (3)$$

where \mathbf{W}_t , \mathbf{X}_t , \mathbf{X}_0 , \mathbf{A} and $\mathbf{C} \in \mathbb{R}^d$ and \mathbf{B} and $\mathbf{D} \in \mathbb{R}^{d \times d}$. The d Brownian motions are assumed independent; the vector multiplying them, by slight abuse of notation, contains the nonzero elements of the approximation of the diagonal matrix $\boldsymbol{\Sigma}$; extending the model to accommodate correlated noise is also possible. Throughout it is assumed that the quantity $\boldsymbol{\Sigma}\boldsymbol{\Sigma}^T$ (where the superscript T denotes a matrix transpose) associated with the *true* process is positive definite[¶].

This simple model is based on the fact that we expect smooth evolution of moments of the observables; at the same time we do take into account the state dependence of the noise. Neglecting this dependence can cause bias in the estimation of the drift [11,52]. An advantage of the local model is that one does not need to calibrate a global one; in addition, the local model naturally provides derivative information at *specific points in phase space* (this statement will be clarified by a variety of examples in later sections).

The parameters of this local linear model are estimated through techniques associated with maximum likelihood estimation (MLE). The motivation for using MLE techniques stems from the fact that under certain regularity conditions [63] such estimators are asymptotically efficient (in time series sample size) as regards the variance of the estimated parameter distribution. In many circumstances the asymptotic parameter distri-

butions of maximum likelihood estimates can be obtained analytically (or numerically). This can guide one in determining how many atomistic simulation realizations are necessary to achieve a given level of accuracy. Additional (basic) information about MLE is discussed further in subsection 2.1.

The *local* estimation strategy used here does complicate analyzing properties of the estimator. This issue appears to pose a challenging problem to the statistics community. This is in part due to the difficulty in decoupling effects due to approximating the dynamics of the true process by a diffusion model, the errors introduced by the approximation of the transition density associated with the assumed diffusion model, and the uncertainty associated with finite sampling effects using nonstationary time series to calibrate the (local) models. We should explicitly state that we assume the predominant error in the applications reported is caused by sampling uncertainty caused by finite sample sizes (and not nonlinear effects caused by the local model breaking down). Investigation of the trade-off between bias and variance in this setting will be the subject of future research. Some other technical issues related to the local diffusion model approach are discussed in Ref. [11].

The key advantages that a local MLE strategy offers are that it (1) avoids the necessity of specifying global dynamics; (2) makes a wide variety of statistical inference tools available and (3) allows local model “flexibility”. We already discussed the first item; the second item is relevant to a variety of multiscale modeling applications; the statistical structure ascribed to the data by the assumed model usually allows one to easily test a variety of statistical hypothesis about the model given the data (e.g. goodness-of-fit tests are discussed in Section 6). The final item is relevant for a couple of reasons. Estimating a *local* state-space model may offer several advantages over a point derivative estimate. Many multiscale computational procedures require detailed stochastic dynamical information around a certain nominal state-space point (a.k.a. the “constraint point” \mathbf{X}_0 [4,51,62]); yet the system dynamics may significantly change when one moves slightly away from the constraint point. For a more concrete example of this type of phenomenon relevant to some MD systems, the state is often “constrained” around a point \mathbf{X}_0 by adding a harmonic potential $\Phi := k/2(\mathbf{X}_0 - \mathbf{X})^2$ to the original MD Hamiltonian [51] where k is a “spring constant” (which is typically an input parameter in the simulation). Such modified potentials are often used in both equilibrium and nonequilibrium simulations to compute the free energy associated with certain system observables. In these types of applications, the observables are often called “reaction coordinates”

[¶]In our local diffusion model approximations, it is possible to obtain a zero diffusion coefficient. The optimization routine used to find the MLE parameter did not encounter these types of problematic points in the phase space and parameter regime explored in this paper, however this possibility complicates analyzing the transition density expansions used (discussed in Section 2.2).

[4,62,51,13]. In these situations one is often interested in the state dependence of both the drift and diffusion coefficients. Our local estimation strategy readily accommodates situations where the (*effective*) drift and diffusion term vary smoothly with respect to the state by creating models that only require linear approximations of the drift and diffusion coefficient to be valid in the “ \mathbf{X}_0 -neighborhood” selected (figure 1). The “flexibility” mentioned earlier also refers to the fact that the same local modeling idea can be used to incorporate other noise terms, e.g. jump transitions. The main challenge associated with modeling with more complicated noise processes lies in determining how to approximate the transition density of the assumed model.

A large part this paper utilizes only the information contained in the drift portion of the local models. This is done because, in some systems, one knows that the system dynamics converge to a deterministic limit. When the functional form of this deterministic limit is unknown *a priori* (this can occur in kinetic Monte Carlo simulations with lateral interactions [47,48] as opposed to perfectly mixed SSA models) we could apply the same ideas presented here to estimate function values “on-demand”. For the “inner simulators” of the systems studied, it is known that when $\Omega \rightarrow \infty$ the dynamics of the concentration observables tend to ODEs whose RHS corresponds to the diffusion approximation’s drift function [21]. In Ref. [12] we illustrated how the probability integral transform can be used to quantify how large Ω needs to be for a diffusion approximation to be statistically valid; in what follows we simply assume Ω is large enough for the diffusion approximation to be meaningful.

2.1 Maximum likelihood estimation for discretely observed diffusions

We now recall a few basic facts about MLE estimation; standard references include, e.g. [31,35,63]. It is assumed throughout that the *exact* distribution associated with the parametric model admits a continuous density whose logarithm is well defined almost everywhere and is three times continuously differentiable with respect to the parameters [41].

MLE is based on maximizing the log-likelihood (\mathcal{L}_θ) with respect to the parameter vector (for our model $\theta \equiv [\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D}]$):

$$\mathcal{L}_\theta \equiv \log(f(\mathbf{X}; \theta)). \quad (4)$$

In the above equation, \mathbf{X} corresponds to a matrix of observations $\in \mathbb{R}^{d \times M}$ where d is the dimension of the state and M is the length of the time series; $f(\mathbf{X}; \theta)$ corresponds to the probability of making observation \mathbf{X} . For a single sample path of a discretely observed diffusion known to be initialized at \mathbf{x}_0 (a lower case symbol is used because the estimated local diffusion models use the symbol \mathbf{X}_0 for the

“expansion point”), $f(\mathbf{X}; \theta)$ can be evaluated as [31]:

$$f(\mathbf{X}; \theta) = \delta_{\mathbf{x}_0} \prod_{m=1}^{M-1} f(\mathbf{X}_m | \mathbf{X}_{m-1}; \theta). \quad (5)$$

In this equation $f(\mathbf{X}_m | \mathbf{X}_{m-1}; \theta)$ represents the conditional probability (transition density) of observing \mathbf{X}_m given the observation \mathbf{X}_{m-1} and θ ; $\delta_{\mathbf{x}_0}$ represents the Dirac distribution at the (known) initial state. Subscripts on time series observations mark the time index (columns of \mathbf{X}) and superscripts (see below) are used to distinguish between different sample paths. In our illustrative examples, we search for the parameter vector that is best over *all* observations (we have an ensemble of N paths of length M). In this case our expression for the log-likelihood (given the data and transition density) takes the form:

$$\mathcal{L}_\theta := \sum_{i=1}^N \sum_{m=1}^M \log(f(\mathbf{X}_m^i | \mathbf{X}_{m-1}^i; \theta)). \quad (6)$$

Assume the existence of an invertible symmetric positive definite “scaling matrix” matrix $\mathcal{F}_{(M, \theta)}$ [45] associated with the estimator; the subscripts are used to make the dependence of the scaling matrix on M and θ explicit. For the “standard” $N = 1$ case in time series analysis, under some additional regularity assumptions [35,63], one has the following limit for a *correctly specified* parametric model:

$$\mathcal{F}_{(M, \hat{\theta})}^{(1/2)}(\theta_M - \hat{\theta}) \xrightarrow{\mathbb{P}_{\hat{\theta}}} N(\mathbf{0}, \mathbf{I}). \quad (7)$$

Here $\hat{\theta}$ is the true parameter vector; θ_M represents the parameters estimated with a finite time series of length M ; $\xrightarrow{\mathbb{P}_{\hat{\theta}}}$ denotes convergence in distribution [31,63] under $\mathbb{P}_{\hat{\theta}}$ (the distribution associated with the density $f(\mathbf{X}; \hat{\theta})$); $N(\mathbf{0}, \mathbf{I})$ denotes a normal distribution with mean zero and an identity matrix for the covariance. For a correctly specified model family, $\mathcal{F}_{(M, \hat{\theta})}$ can be estimated in a variety of ways [45,64]. The appeal of MLE lies in that, asymptotically in M , the variance of the estimated parameters is the smallest that can be achieved by an estimator that satisfies the assumed regularity conditions [35,63].

2.2 Transition density expansions

Here we briefly outline the key features of the recent work of Ait-Sahalia [2,3] used in our coarse-grained computations below. The problem with estimation based on a simple model like the one given in equation (3) is that the transition density associated with the process is not known in closed form. In recent years, many attempts to approximate the transition density have appeared in the literature; some techniques depend on analytical approximations while others are simulation-based (see, e.g. [1,2,5,8,24,53]). We have used, with some success, the expansions found in Refs. [1,2,3]. High accuracy can be

obtained using this method to approximate the transition density associated with a *scalar* process; the multivariate case is discussed in Ref. [3]. The basic idea behind the scalar case, presented in Refs. [1,2], is as follows: One first transforms the process given in equation (2) into a new process [2]:

$$dY_t = \mu_Y(Y_t; \theta)dt + dW_t$$

$$Y \equiv \gamma(X; \theta) = \int^x \frac{du}{\sigma(u; \theta)} \quad (8)$$

$$\mu_Y(y; \theta) \equiv \frac{\mu(\gamma^{-1}(y; \theta); \theta)}{\sigma(\gamma^{-1}(y; \theta); \theta)} - \frac{1}{2} \frac{\partial \sigma}{\partial x}(\gamma^{-1}(y; \theta); \theta). \quad (9)$$

An additional change of variables brings the transition density of the process closer to a standard normal density $Z \equiv \Delta^{-1/2}(Y - y_0)$ where Δ is the time between observations. The transformations introduced allow the use of a Hermite basis set in order to approximate the transition density of the original process via the following series:

$$p_Z(\Delta, z|y_0; \theta) \approx \phi(z) \sum_{j=0}^K \eta_Z^{(j)}(\Delta, y_0; \theta) H_j(z) \quad (10)$$

$$\eta_Z^{(j)}(\Delta, y_0; \theta) \equiv \frac{1}{j!} \int_{-\infty}^{\infty} H_j(z) p_Z(\Delta, z|y_0; \theta) dz$$

$$:= \frac{1}{j!} \mathbb{E} \left[H_j \left(\Delta^{-1/2} (Y_{t+\Delta} - y_0) \right) | Y_t = y_0; \theta \right]. \quad (11)$$

In the above, H_j represents the j th Hermite polynomial and $\phi(\cdot)$ is the standard normal density. The coefficients needed for the approximation are obtained through the conditional moments of the process Y_t . Ait-Sahalia outlines [2] a procedure which exploits the connection between the SDE and the associated Kolmogorov equations in order to develop a closed-form expression for a Taylor series expansion of the $\eta_Z^{(j)}$ coefficients. If the coefficient functions satisfy the assumptions laid out in Ref. [2], the approximation converges to the true transition density. The approximation improves as $K \rightarrow \infty$. For fixed, finite K the approximation typically improves as $\delta t \rightarrow 0$. In numerical applications one must always deal with a finite K ; problems may arise in the truncated expansion: the approximation of the density may not normalize to unity or, worse, it may become negative (see [1,3,5] for some possible remedies).

In the multivariate case, it becomes more difficult to introduce an analog of Y_t [3]. Nonetheless, it is still possible to construct a series motivated by the methodology used in the scalar case; however, one now needs to expand in (state) space *and* time, whereas the Hermite expansion yielded a series “in time only”. Ait-Sahalia [3] outlines an approach which makes use of a recursion for calculating the coefficients of the expansion in the multivariate case. We have had some success in using these expansions, even in cases where convergence of the infinite series is not guaranteed by the conditions given in Refs. [1,2]. Using

different function classes (made computationally feasible by the extension of Bakshi and Ju [5,6]), such as sigmoidal functions for the diffusion coefficient, may help circumvent some of the convergence problems associated with the transition density applied to the types of models given in equation (3). Other computational problems associated with the expansions are discussed in Ref. [11]; as with any applications of MLE, *a priori* estimates of the range of the parameters of interest [11,45,63] can enhance the performance of the optimization algorithm. The comparison study [36] recommends the use of the expansions by Ait-Sahalia for a wide class of diffusion models. Beyond the estimation itself, these expansions can also be helpful in obtaining diagnostics that depend on knowledge of the transition density (such as goodness-of-fit tests [33]) and asymptotic error analysis [45], though the methodology we propose can utilize any reliable transition density approximation scheme.

2.3 Diagnostic tools

The probability integral transform [16] (PIT) is a powerful hypothesis testing tool which can be used in a variety of inference procedures associated with time series analysis [33]. It can be used after one selects a model for the data; the technique requires one to have the ability to evaluate the transition density associated with the assumed model. In parametric diffusion modeling applications, previous studies have shown that various transition density expansions can be reliably used in place of the exact transition density with the PIT [12,33]. The PIT is a transform which constructs a new random variable (Z_n) given an observation pair and an assumed transition density. This random variable is used to set up tests that aim at determining whether the actual dynamics observed are consistent with the assumed model. The scalar case is easiest to demonstrate (the multivariate is a fairly straightforward extension [33]) and the transformation is shown below:

$$Z_n := \int_{-\infty}^{X_n} p(X'_n | X_{n-1}; \theta) dX'_n$$

$$Z_n \sim q(Z_n) \equiv \frac{dQ(Z_n)}{dZ_n} \quad (12)$$

$$X_n \sim f(X_n | X_{n-1}) \equiv \frac{dF(X_n | X_{n-1})}{dx_n}.$$

The symbol \sim denotes that the random variable on the left of the symbol is distributed according to the (assumed continuous) density on the right. The transition density of the true process (conditional on X_{n-1}) is denoted by $f(\cdot | X_{n-1})$, the corresponding transition density assumed (or fit) by the modeler is represented by $p(\cdot | X_{n-1}; \theta)$, and $q(\cdot)$ represents the density of the PIT random variable (which implicitly depends on the true transition density through the data). Under a correctly specified model, the Z_n 's are independent and uniformly distributed on the interval $[0,1]$, independent of the transition density [16]. The transition

density of the true process ($f(X_n|X_{n-1})$) may not be adequately described by the assumed model class; in this case $q(Z_n)$ will not be a uniform density due to the discrepancy between $f(X_n|X_{n-1})$ and $p(X_n|X_{n-1};\theta)$. In fact if the two densities are not equal, the Z_n series constructed from the data and the assumed model will often exhibit time correlation. In Ref. [33] a suite of statistical tests are developed which exploit these two facts. Here we utilize the Q -test statistic; to construct it one needs to first construct the series $\{Z_n\}_{n=1}^M$ from the series $\{X_n\}_{n=0}^M$ and the estimated θ . One then uses a nonparametric estimate of the PIT random variable density in order to construct the following function:

$$\hat{g}_j(z_1, z_2) \equiv \frac{1}{M-j} \sum_{\tau=j+1}^M K_h(z_1, \hat{Z}_\tau) K_h(z_2, \hat{Z}_{\tau-j}). \quad (13)$$

Here, $K_h(\cdot)$ is the kernel used to obtain the nonparametric density estimate (see Equation (7) on pg 44 of [33]), h is the kernel bandwidth, the parameter j specifies the lag (for our applications we always set $j = 1$), and \hat{Z}_τ represents a PIT variable constructed from the observed data. One then defines the random variable

$$\hat{M}_1(j) \equiv \int_0^1 \int_0^1 (\hat{g}_j(z_1, z_2) - 1)^2 dz_1 dz_2. \quad (14)$$

Finally one obtains a test statistic from the random variable above through the relation

$$\hat{Q}(j) \equiv (h(M-j)\hat{M}_1(j) - A_h^0) / V_0^{\frac{1}{2}} \quad (15)$$

where the centering and scaling parameters (A_h^0 and V_0) are listed in equations (10) and (11) of Ref. [33]. The above random variable converges weakly to a normal asymptotically in M ; this fact can be used in a variety of goodness-of-fit tests. The major appeal (for our purposes) of the above test statistic is its applicability to nonstationary time series. Another desirable feature associated with the PIT test is that if a test is rejected, one can even construct an “ M -test statistic” in order to assist one in diagnosing what features of the assumed model are inadequate [33]. We make a minor modification of the Q -test statistic procedure in Section 6.3 to illustrate how this technique can be used to assist in coarse projective integration.

2.4 Short-burst estimation

Calibrating our local models requires us to specify the parameters M , \mathbf{X}_0 , ε , and τ . Initial conditions are drawn from a uniform distribution (of width $\varepsilon/2\mathbf{X}_0$) centered at \mathbf{X}_0 . We have not yet systematically explored the issue of prescribing an initial distribution of the *coarse observables* which aims at enhancing parameter estimation. Short bursts of atomistic simulations were run for τ time units and time series were created by making M uniformly spaced observations in this time interval (resulting in time series spaced by $\delta t \equiv \tau/M$).

3. Model systems

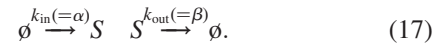
In this paper we study three model systems. Their detailed (“atomistic”) evolution rules are first presented; we then report the deterministic limiting ODE associated with each system and our reason for using each particular model. The functional form of the limiting SDE associated with a given model is determined once the reaction mechanism is prescribed (see [7,25]).

3.1 Michaelis–Menten systems (model 1)

The classical Michaelis–Menten (MM) reaction mechanism is of the form:



where S represents the substrate, E the enzyme, C the complex (substrate bound to enzyme), and P the product [49]. We introduce the following “reactions” to the scheme in order to prevent a steady state corresponding to extinction of S :



The first step is meant to model a source term, while the second is an outward flux of S (the symbol \emptyset indicates a species that is not within the boundaries of the well-mixed system). We simulate the system above using both an SDE model and the SSA. The classical reaction model proceeds until all of the substrate is converted into product. In the SSA it becomes possible to exhaust the substrate in finite time; this type of behavior is not observed in the limiting ODE description. The extra reaction steps given in equation (15) were introduced in order to keep the “steady-state” distribution significantly away from such an extinction event. After adding the new “flow terms” and taking into account mass balances, the resulting system of ODEs describing the system (valid at the infinite Ω limit) is given by:

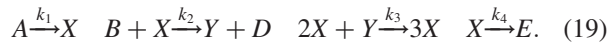
$$\begin{aligned} \frac{dS}{dt} &= -k_1(E_0 - C)S + k_{-1}(C) + \alpha - \beta S \\ \frac{dC}{dt} &= k_1(E_0 - C)S - (k_{-1} + k_2)C. \end{aligned} \quad (18)$$

The constant E_0 represents the maximum enzyme concentration that can be obtained in the system (it is assumed that E , C , and P can not flow out of the system). In this paper we carry out numerical temporal integration and fixed-point computations using noisy data; in effect we are performing equation-free computations for the *estimated* local drift function. In Section 5, we apply projective integration [27] to the estimated drift. There, both explicit and implicit projective forward Euler schemes are used for numerical integration. The parameters of the reaction scheme above were selected in order to yield both stiff as well as non-stiff parameter case problems (only two sets of

parameters are used in the MM systems studied and they are distinguished by the labels “stiff” and “non-stiff”). In the stiff parameter case, a linearization around the stable fixed point of the limiting ODE (above) resulted in a matrix where the ratio of the eigenvalues (denoted by ϵ) was ≈ 0.09 (with the largest eigenvalue having a magnitude of ≈ 213.4). In Section 6, we use the parameters in the stiff regime and couple the local estimation strategy with the ideas behind Quasi-Steady State Approximations (QSSA) [46] and projective integration.

3.2 The Brusselator (model 2)

Here, we again use the SSA and a SDE as the fine scale, “inner” atomistic simulation. The kinetic scheme for the system is as follows:



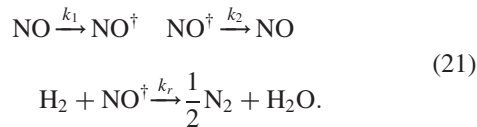
All letters above correspond to different chemical species; the concentrations of A and B are assumed to be held constant by a reservoir. After incorporating mass conservation, the infinite particle ODE limit reads [61]:

$$\begin{aligned} \frac{dX}{dt} &= k_1 A - (k_2 B + k_4) X + k_3 X^2 Y \\ \frac{dY}{dt} &= k_2 B X - k_3 X^2 Y. \end{aligned} \quad (20)$$

We tune the parameters (reported in the results section) of the system to result in a limit cycle for the corresponding limiting ODE. We use this system to illustrate how our local diffusion models can be used in order to assist with variational calculations in the presence of a periodic solution.

3.3 NO reduction (model 3)

In this model we only use a SDE as the fine scale, “inner” atomistic simulation. The kinetic scheme [59] for the system is as follows:



The scalar limiting ODE corresponding to this system is:

$$\frac{dX}{dt} = k_1(1 - X) - k_2 X - k_r(1 - X)^2 X. \quad (22)$$

In the above, X corresponds to the concentration of NO^\dagger (the superscript indicates that the species is absorbed on a catalytic surface). This model is used because of its nonlinearity; it illustrates how one can use certain equation-free ideas coupled with local diffusion models in order to determine properties of invariant distributions [15,25] associated with the process. We limit this last numerical study to scalar diffusions because an analytical expression for the invariant distribution can be determined

(and used to check our numerical results) simply by inspecting the coefficients of the SDE [15].

4. Deterministic numerical methods

In this section we illustrate the linking of local diffusion modeling with traditional numerical methods. We will show how continuum numerical analysis becomes, in effect, a protocol for *designing computational experiments* with the fine scale “inner simulator”; observation and estimation provide the quantities necessary to design “intelligent” subsequent numerical experiments. This iterative scheme (performing short bursts of fine scale computation, processing the observations, and using the results to design subsequent numerical experiments) constitutes the basis of equation-free multiscale computation. In this section, we run the atomistic simulations and estimate *all* of the parameters of a local diffusion model, but only carry out equation-free computations using the parameters associated with the drift component of the local models.

4.1 Coarse projective integration

Here we obtain local parametric models from short burst output of both the SSA and SDE versions of the modified MM equations. We will show how the local diffusion model information can be used to solve an initial value problem by using *coarse projective integration* on the estimated drift component of our surrogate models. We tune the system parameters ($k_1 = 1.0$, $k_{-1} = 1.0$, $k_2 = 1.0$, $\alpha = 1.0$, $\beta = 1.0$, $E_0 = 1.0$) to lie in a non-stiff regime of the limiting deterministic ODE; this is done so that both explicit and implicit integration schemes can be used with a reasonably sized timestep (Δt).

Before proceeding, we demonstrate the basic idea behind coarse projective integration through some simple examples. In coarse explicit Euler projective integration, one runs (one or more realizations of) an atomistic simulator and collects a sequence of observations $\{\mathbf{X}_i\}_{i=0}^T$; here the observations are spaced uniformly in time, separated by an interval (δt). In earlier formulations of projective integration, a regression in time would yield the (assumed constant) “slope” ($\mathbf{F}(\cdot)$) evaluated at a point \mathbf{X}_T associated with the burst of atomistic simulation data. This slope would then be used to update the (expected, coarse) state by making a projective step ($\mathbf{X}^{\text{exp}} = \mathbf{X}_T = \Delta t \mathbf{F}(\mathbf{X}_T)$; where $\Delta t \gg \delta t$). The procedure would then be iterated. If it is desired to use an implicit integrator, employing a least squares estimator for the time derivative can lead to a computationally costly fixed-point procedure at every projective timestep [37].

The local models we use allow one to estimate $\mathbf{F}(\cdot)$ in a neighborhood of \mathbf{X}_0 (the extent of this neighborhood is determined by the region where the SDE coefficient functions can be well approximated by linear functions). For processes and conditions under which a local model is

accurate, this allows one more flexibility in projective integration since one can, at a low computational cost, use either explicit or implicit integrators.

The formulas

$$\begin{aligned}
 \mathbf{X}_n^{\text{exp}} &= \mathbf{X}_{n-1} + \Delta t \mathbf{F}(\mathbf{X}_{n-1}) \\
 &\approx \mathbf{X}_{n-1} + \Delta t (\mathbf{A} + \mathbf{B}(\mathbf{X}_{n-1} - \mathbf{X}_0)) \\
 \mathbf{X}_n^{\text{imp}} &= \mathbf{X}_{n-1} + \Delta t \mathbf{F}(\mathbf{X}_n^{\text{imp}}) \\
 &\approx \mathbf{X}_{n-1} + \Delta t (\mathbf{A} + \mathbf{B}(\mathbf{X}_n^{\text{imp}} - \mathbf{X}_0))
 \end{aligned}
 \tag{23}$$

list the deterministic explicit and implicit Euler integrators (respectively). These can be used individually or in tandem; in the latter case they form a predictor-corrector (PC) scheme (one uses the explicit scheme to provide an initial guess for the implicit scheme [26,44]). The straight line in the left panel of figure 1 represents a prototypical predictor step (using a local MLE parameter estimate for the time derivative) for data generated using a generic SDE with state independent noise and a nonlinear drift function. The other two batches of simulations in figure 1 represent data sets that could be used to make the corrector step.

In this case of mild nonlinearity one might be able to solve a single (estimated) linear equation in the corrector step and introduce a fairly small approximation error. This would avoid estimating a sequence of local linear models as would be done in a standard corrector fixed-point iteration. We now return to the MM problem and show how the modified estimation procedure can be linked with projective integration. It should be noted that the implicit Euler method requires one to set up a Newton iteration scheme (shown in detail in the next section) at each time

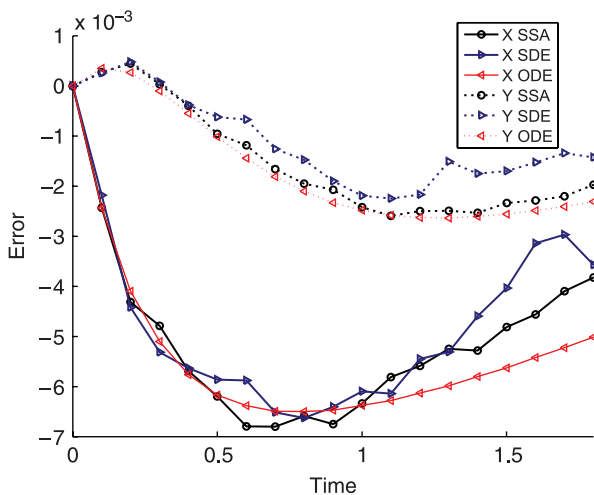


Figure 2. Implicit coarse projective Euler integration (non-stiff regime). Details: $\Omega = 2000$, $\delta t = 1 \times 10^{-3}$, $\Delta t = 0.1$, $M = 100$, $N_{\text{path}} = 500$, $\varepsilon = 0.01$, number of NR iterations = 4 per integration step. The larger magnitude error corresponds to X and the smooth curve corresponds to the truncation error associated with carrying out the deterministic explicit Euler on an ODE identical to the drift of the SDE (and close to that of the SSA for the Ω used).

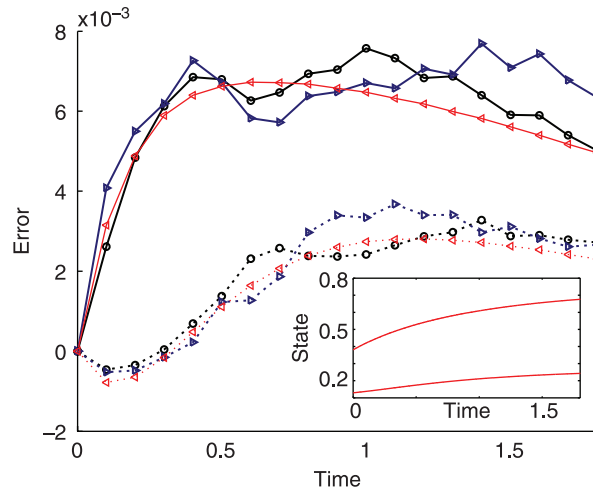


Figure 3. Explicit coarse projective Euler integration (non-stiff regime). Details: $\Omega = 2000$, $\delta t = 1 \times 10^{-3}$, $\Delta t = 0.1$, $M = 100$, $N_{\text{path}} = 500$, $\varepsilon = 0.01$. The larger magnitude error corresponds to X and the smooth curve corresponds to the truncation error associated with carrying out the deterministic explicit Euler (on the *known* drift). The symbols used are the same as in figure 2.

step. The results of carrying out such an implicit projective integration scheme are shown in figure 2. The inset in figure 3 shows the numerical integration results (carried out with MATLAB’s ode23s numerical integrator) of the *known* limit ODE. The main portion of the figure plots the difference between the explicit projective integration scheme minus the ODE solution (note that for finite Ω the SSA process may have a different drift function than the limiting ODE; the SDE process is *constructed* to have the expected limiting drift function). The results of carrying out the explicit Euler scheme (with the same step size) and “perfect” function evaluations are also plotted, in order to give one a best case scenario of accuracy *if* the parameters estimated follow the limiting ODE (table 1).

In both the implicit and explicit integration cases the accuracy is comparable in magnitude. The explicit projective integration procedure only requires the “constant” terms associated with a local model; however, adding the linear terms to the state-space model allows one to exploit a longer time series in the estimation step (assuming a fixed δt) because the derivative no longer needs to remain constant. The results from the individual Euler schemes indicate that one could potentially use a PC-type projective integrator. Most PC schemes offer a local truncation error estimate (LTE) (for example the Adams’ family of PC schemes allows for this); estimating a simple local affine model at each required point allows for more sophisticated numerical integrators, yielding schemes that have (*partial*) error estimation built in (i.e. estimation of the deterministic integrator discretization error). Incorporating estimation uncertainty is a separate issue; *if* the proposed model is statistically adequate, then information contained in estimated parameter distributions [11] may be used to approximate the effect of parameter uncertainty on integration performance [22,23,50]. We estimate the integrator

Table 1. Estimation of the Local Truncation Error (LTE) committed on a single coarse projective step (from a common initial condition). The “Measured” case corresponds to the error determined by taking a single explicit projective step and subtracting the result from an “exact” deterministic integrator (ode23s); the “Predicted” case corresponds to the error estimated by the predictor-corrector pair (the data reported is for a single PC iteration using a common initial condition).

	ODE	SDE	SSA
	X		
Measured	-2.4379×10^{-3}	-2.1817×10^{-3}	-2.4306×10^{-3}
Predicted	-2.7928×10^{-3}	-3.1228×10^{-3}	-2.5027×10^{-3}
	Y		
Measured	3.48092×10^{-4}	2.56070×10^{-4}	2.70245×10^{-4}
Predicted	5.64664×10^{-4}	3.89497×10^{-4}	3.66594×10^{-4}

discretization error by implementing the simple Adams method based on the first order Euler PC pair; consult Lambert [44] or Gear [26] for details. In this case the LTE is estimated by: $LTE_i(t_n) \approx (X_i^{imp}(t_n) - X_i^{exp}(t_n))/2$ (we define the LTE as the numerical method’s estimate minus the exact solution to the ODE and the subscripts on variables indicate vector components). Table 1 shows that the *estimated* LTEs based on the first order schemes shown in this section come close to the observed LTEs; note that the precision is limited by the noisy estimation of the system parameters and recall once more that the drift function of the diffusion approximation to the SSA is not known *exactly* for finite Ω ;

4.2 Equation-free fixed-point methods for integrators

Previous studies [47,48] have already addressed how one can use the “equation-free” framework to find steady states through the zeros of a *timestepper* (a discrete time map resulting from an integrator scheme) using a Newton–Raphson (NR) method. Here we demonstrate how the information contained in our local diffusion model can be used to carry out such a NR scheme to solve the fixed point problems arising in implicit integration using the timestepper associated with the implicit Euler projective integration (the results of this integration scheme applied to the modified MM system were already shown in the previous section). In general, the fixed point scheme is carried out by specifying the initial condition and the projective step size (Δt) and then iteratively

Table 2. Real component of (estimated) monodromy matrix eigenvalue closest to unity ($\Omega = 12 \times 10^3$). “FD” corresponds to using forward-differences, “CD” to using centered differences, and “MLE” to using local linear models in order to estimate the monodromy matrix.

Case	FD	CD	MLE
SSA	0.54074	0.90025	1.00018
SDE	0.70389	0.98679	1.05189

determining \mathbf{X}_n by solving

$$\mathbf{G}(\mathbf{X}_n) \equiv \mathbf{X}_n - (\mathbf{X}_{n-1} + \Delta t \mathbf{F}(\mathbf{X}_n)) = 0 \quad (24)$$

for $n = 1, \dots, N = \tau/\Delta t$ (where τ is the final time of the coarse projective trajectory). Solving this equation through a Newton procedure (given an initial guess $\mathbf{X}_n^{(0)}$) results in a fixed-point recursion of the form:

$$\mathbf{X}_n^{(k)} = \mathbf{X}_n^{(k-1)} - \left(\frac{\partial \mathbf{G}(\mathbf{X})}{\partial \mathbf{X}} \Big|_{\mathbf{X}=\mathbf{X}_n^{(k-1)}} \right)^{-1} \mathbf{G}(\mathbf{X}_n^{(k-1)}). \quad (25)$$

The nature of our local model allows us to approximate the RHS of an ODE (e.g. equation (18)) directly allowing us to approximate the Newton scheme above by:

$$\mathbf{X}_n^{(k)} = \mathbf{X}_n^{(k-1)} - (\mathbf{I} - \Delta t \mathbf{B})^{-1} \mathbf{G}(\mathbf{X}_n^{(k-1)}). \quad (26)$$

Here

$$\begin{aligned} \mathbf{G}(\mathbf{X}_n^{(k-1)}) &\approx \mathbf{X}_n^{(k-1)} \\ &- (\mathbf{X}_{n-1} + \Delta t (\mathbf{A} + \mathbf{B}(\mathbf{X}_n^{(k-1)} - \mathbf{X}_0))) \end{aligned} \quad (27)$$

and all of the parameters above are estimated after specifying \mathbf{X}_0 (a parameter of the local model) to be $\mathbf{X}_n^{(k-1)}$. The results of such an NR iteration scheme are shown in figure 4; the parameters ($k_1 = 150.0, k_{-1} = 0.1, k_2 = 110.0, \alpha = 20.0, \beta = 1.0, E_0 = 0.4$) of the modified MM scheme were chosen in the regime where the limiting ODE is stiff. The value of Ω was chosen to be large enough for a diffusion approximation to be physically meaningful. It is clear upon inspection of the rate of convergence of the NR scheme that the procedure using data from a genuine inner diffusion SDE comes closer to quadratic convergence (the feature that makes the classical NR method appealing). It should also be noted that this

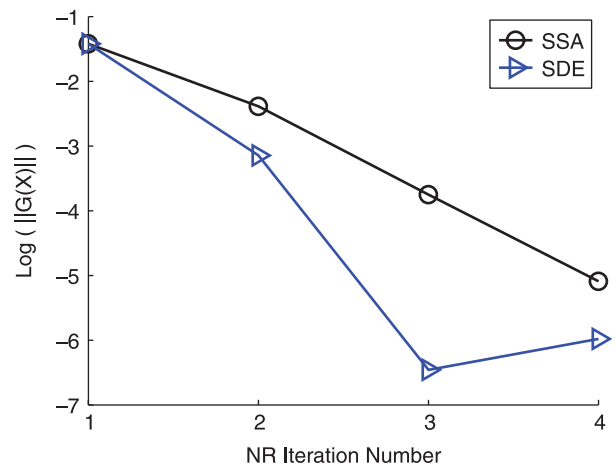


Figure 4. NR procedure to find a zero of the equation needed by each step of the implicit Euler timestepper (equation 27). A timestepper based on a single projective step is shown (system parameters are selected in the stiff regime). Numerical details: $\Omega = 2000, \delta t = 1 \times 10^{-3}, \Delta t = 0.1, M = 100, N_{\text{path}} = 500, \varepsilon = 0.01$, number of NR iterations = 4.

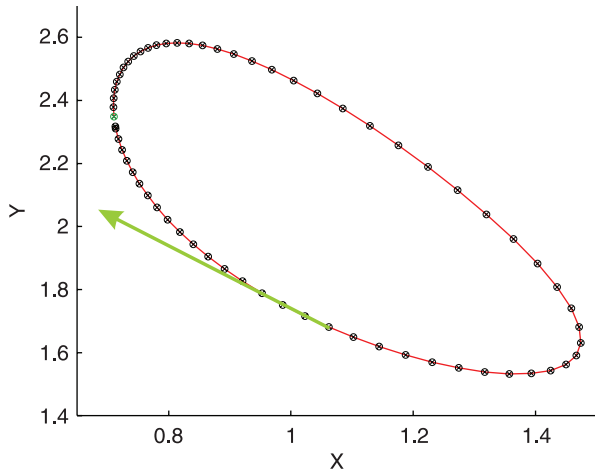


Figure 5. Phase portrait of the coarse ‘the Brusselator’ (model 2). The arrow indicates the trajectory’s time direction. The local Jacobians needed in the variational integration are obtained through estimating the “ \mathbf{B} ” parameter (equation 2) from ensembles of short burst SSA simulations initialized at the N_{grid} base points shown as circles (see text).

fixed-point scheme is limited by the sampling error associated with estimation; knowledge of the parameter distribution can assist one in designing computational experiments that achieve a specified desired accuracy. We again refer the reader to Ref. [11] for a more detailed discussion. One should contrast this equation-free Newton method with that in Ref. [47] where the Jacobians of the Newton scheme were determined using a finite difference approximation coupled with least squares estimation. With noisy data, determining the appropriate finite difference stencil when the noise is not uniform in state-space (as it is in all applications shown here) is quite nontrivial. Braatz *et al.* have developed a finite difference scheme [58] that can efficiently extract derivatives in the presence of noise; their approach requires knowledge of the noise distribution. We expect a “small” finite difference stencil around the state point to have a small signal to noise ratio. A larger stencil increases the signal to noise ratio, but the linearity of the function becomes questionable if one uses too large a stencil. After we estimate the parameters by MLE type techniques, we can easily read off the corresponding Jacobian (i.e. the linearized drift, \mathbf{B}) associated with \mathbf{X}_0 (figure 5).

4.3 Variational integration for periodic solutions

We now use SSA and SDE simulations of the Brusselator scheme as our fine scale “inner” stochastic simulator. Our purpose is to show how the simple ideas illustrated so far can be exploited to estimate the monodromy matrix [60] associated with the coarse limit cycle of the Brusselator using the parameter values $k_1 = 1.0$, $k_2 = 1.0$, $k_3 = 1.0$, $k_4 = 1.0$, $A = 1.0$, and $B = 0.4$. The monodromy matrix, $\mathbf{V}(t = T)$, of a limit cycle of period T for a *deterministic* ODE model can be evaluated by

simultaneously solving the following initial value problems, where \mathbf{X}^{IC} is a point on the limit cycle:

$$\frac{d\mathbf{X}}{dt} = \mathbf{F}(\mathbf{X}; \theta) \quad \mathbf{X}_0 = \mathbf{X}^{\text{IC}} \quad (28)$$

$$\frac{d\mathbf{V}}{dt} = \frac{\partial \mathbf{F}(\mathbf{X}; \theta)}{\partial \mathbf{X}} \cdot \mathbf{V} \quad \mathbf{V}_0 = \mathbf{I}; \quad (29)$$

its eigenvalues (also called Floquet multipliers) quantify the stability of the deterministic limit cycle; one generically lies at 1 (associated with translational invariance of the limit cycle in time) and the oscillation is stable if the remaining multipliers have norm less than 1.

In the above, the zero subscripts denote the state values at $t = 0$. When the ODEs in the above system are not explicitly available (as will be the case when we use the output of a noisy atomistic simulator to estimate temporal derivatives by using regression in time), the RHS of the second set of ODEs is sometimes estimated using a finite difference stencil [56]. In such cases, the monodromy matrix associated with the coarse limit cycle can be approximated starting at the initial conditions above and performing numerical integration of the above ODEs through some standard temporal discretization scheme. For each timestep of this integration one estimates the local “sensitivity matrix” (the local Jacobian of the integrator map), \mathbf{S}_n , by a finite difference scheme. In this formulation, \mathbf{S}_n is then used to iteratively update the variational components \mathbf{V}_n via $\mathbf{V}_n = \mathbf{S}_n \mathbf{V}_{n-1}$.

This is perhaps easiest to see by inspecting the equations defining a trajectory obtained by an explicit forward Euler scheme:

$$\begin{aligned} \mathbf{X}_1 &= \mathbf{X}_0 + \Delta t \mathbf{F}(\mathbf{X}_0) \\ &\vdots \\ \mathbf{X}_n &= \mathbf{X}_{n-1} + \Delta t \mathbf{F}(\mathbf{X}_{n-1}). \end{aligned} \quad (30)$$

The variational integration accompanying this simple integration sequence is as follows:

$$\begin{aligned} \frac{\partial \mathbf{X}_1}{\partial \mathbf{X}_0} &= \left(\mathbf{I} + \Delta t \frac{\partial \mathbf{F}(\mathbf{X})}{\partial \mathbf{X}} \Big|_{\mathbf{X}_0} \right) \frac{\partial \mathbf{X}_0}{\partial \mathbf{X}_0} = \mathbf{S}_1 \mathbf{V}_0 = \mathbf{S}_1 \mathbf{I} \\ &\vdots \\ \frac{\partial \mathbf{X}_n}{\partial \mathbf{X}_0} &= \left(\mathbf{I} + \Delta t \frac{\partial \mathbf{F}(\mathbf{X})}{\partial \mathbf{X}} \Big|_{\mathbf{X}_{n-1}} \right) \frac{\partial \mathbf{X}_{n-1}}{\partial \mathbf{X}_0} = \mathbf{S}_n \mathbf{V}_{n-1} = \mathbf{S}_n (\mathbf{S}_{n-1} \dots \mathbf{S}_1 \mathbf{I}). \end{aligned} \quad (31)$$

Note how in the above approximation of \mathbf{V} we effectively take the initial first-step sensitivity matrix (\mathbf{S}_1) and apply a series of matrix multiplications. The “left matrix multiplier” comes from finite difference information obtained around the state point associated with time step n ; this ultimately results in an approximation of the sensitivity of an ODE trajectory to the initial condition.

Coarse-grained limit cycle calculations using the projective integration formalism were first reported in Ref. [56]. In Ref. [12] we showed how a simple extension of the method presented above, coupled with

a suitably defined Poincaré section, can be used in computations that converge to a coarse limit cycle. In this section we assume that the exact limit cycle is already known. With the local diffusion model approach one does not need to resort to estimating the sensitivity matrix by finite differences. We circumvent these finite difference computations by partitioning the time (of the known) period associated with the limit cycle into 64 subintervals. We then initialize a batch of brief atomistic simulations around the state point corresponding to the beginning of a subinterval. Within each subinterval we directly estimate $\partial F(\mathbf{X})/\partial \mathbf{X} \approx \mathbf{B}$ by using both methods (local MLE and finite difference estimates). In the local MLE formulation one updates \mathbf{V} in a time interval by assuming that $\partial F(\mathbf{X})/\partial \mathbf{X}$ is constant or “frozen” (within each subinterval this quantity is estimated using local MLE). One can loosely compare this frozen approximation to a forward Euler scheme applied to the ODE in equation (29), where the left matrix is constant within each local model neighborhood. Under this type of piecewise frozen approximation, we can “exactly” integrate the variational equations in each subinterval. The quotes are used because the “exact” trajectory is corrupted by estimation error *and* the system Jacobian is not truly constant in any given subinterval.

Results of carrying this procedure out for the SDE and SSA inner simulator case with $\Omega = 2000$ are shown in figures 6 and 7, respectively. The diffusion approximation is only valid for large Ω , so the exact system parameters are not *precisely* known for the SSA case (whereas in the inner SDE simulator case, the drift is prescribed). Figure 8 demonstrates how the local MLE results obtained using SSA data appear to converge to the expected ODE limit as the system size parameter

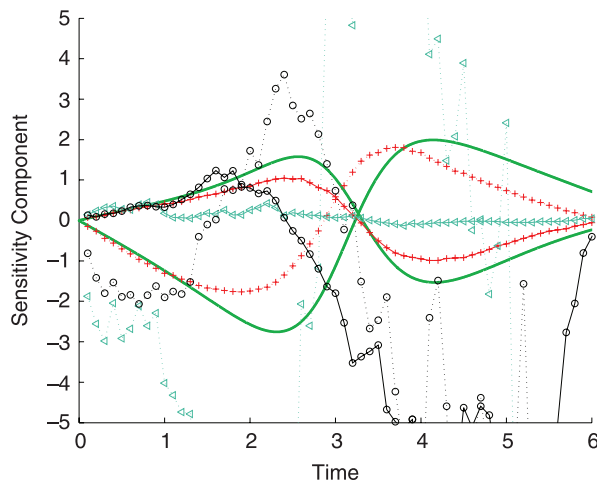


Figure 6. Evolution of the off-diagonal entries of the variational matrix using SSA as the inner simulation process. Numerical details: $N_{\text{path}} = 600$, $\Omega = 2 \times 10^3$, $\delta t = 5 \times 10^{-4}$, $\varepsilon = 0.05$, $M = 200$, $\Delta t = \tau_{\text{period}}/N_{\text{grid}} \approx 0.3$. The solid line represents the results obtained by a deterministic integration of the *known* limiting ODEs. The crosses correspond to results obtained using a local diffusion model; the circles correspond to integration using a centered difference scheme’s estimate and the triangles correspond to integration using a forward difference scheme’s estimate of the local model.

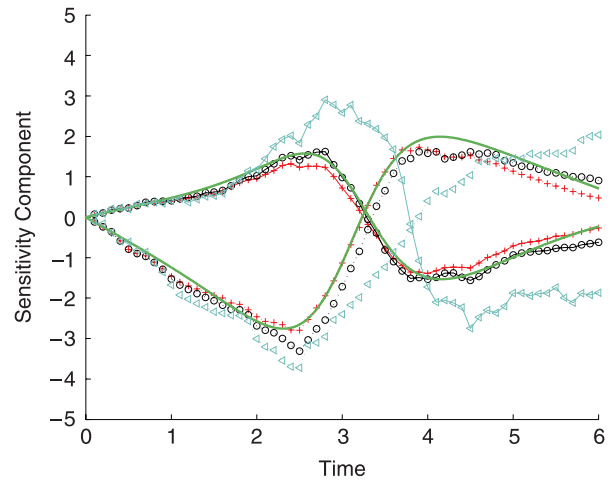


Figure 7. Evolution of off-diagonal entries of the variational matrix evolution using SDE as the inner simulation process. Numerical details: $N_{\text{path}} = 600$, $\Omega = 2 \times 10^3$, $\delta t = 5 \times 10^{-4}$, $\varepsilon = 0.05$, $M = 200$, $\Delta t = \tau_{\text{period}}/N_{\text{grid}} \approx 0.3$. For symbol details refer to figure 6.

increases (Ω is increased to 12,000). It should be noted that, despite the fact that the drift of the process appears to have practically converged to the limiting ODE equation in the local MLE estimates, the monodromy matrix estimated by finite differences is quite sensitive to the ε parameter (varied in results not reported, but fixed throughout here).

We already mentioned that for deterministic limit cycles, one eigenvalue of the monodromy matrix must be unity. In table 2 we show the real part of the largest eigenvalue of our estimated monodromy matrix. In the low noise inner SDE simulator case ($\Omega = 12,000$), MLE and the centered difference scheme with least squares estimation have comparable accuracy; however in the SSA inner simulator case the local MLE seems to outperform both the centered and forward finite difference schemes in the simulation paths observed.

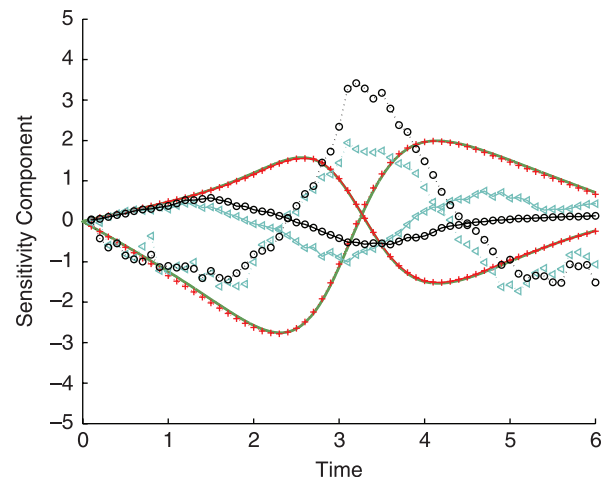


Figure 8. Off-diagonal entries of the variational matrix evolution using SSA as the inner simulation process. Numerical details $N_{\text{path}} = 600$, $\Omega = 2 \times 10^3$, $\delta t = 5 \times 10^{-4}$, $\varepsilon = 0.05$, $M = 200$, $\Delta t = \tau_{\text{period}}/N_{\text{grid}} \approx 0.3$. For symbol details refer to figure 6.

5. Numerical methods for finding extrema of distributions

Up to this point, we have illustrated the linking of local estimation with continuum numerical methods that act only on the estimated drift function of a diffusion model. For systems where noise is inherent (such as some simple stochastic models of cellular environments), the drift alone may be inadequate for characterizing the system features we are interested in. For example, suppose the invariant density associated with a given scheme is unimodal. The mode of a probability density is the point at which the density achieves a maximum value; unimodal implies that there is a unique maximum. We use the term “mode” to refer to both local and global maxima. In this section we use a model that has a diffusion coefficient that depends on the (scalar) state; in this case the mode of the density is not the same as the point at which the drift function has a zero [40,57]. In such situations we can use the information contained in our local diffusion models to locate distinguishing features (such as a mode) of the invariant density of our system.

If a scalar coarse observable can be modeled by an SDE of the form:

$$d\mathbf{X}_t = \mu(\mathbf{X}_t)dt + \sigma(\mathbf{X}_t)dW_t \quad (32)$$

and the coefficients in this equation are globally well-behaved (i.e. the drift and diffusion coefficients satisfy certain regularity conditions [15] which are satisfied, by construction, in our case) one can compute the invariant probability density using only the SDE coefficient functions via the relation [57]:

$$P^{\text{EQ}}(X) = \frac{\exp(-G(\mathbf{X}))}{C}$$

where

$$\begin{aligned} C &\equiv \int_{\Omega_p} \exp(-G(X'))dX' \\ G(X) &\equiv - \int_{X_{\text{ref}}}^X \left(\frac{\mu(X')}{D(X')} \right) dX' + \log(D(X)) \\ D(X) &\equiv (\sigma(X))^2/2. \end{aligned} \quad (33)$$

In the above, the subscript Ω_p denotes integration over the entire probability space and it is assumed that both X and $X_{\text{ref}} \in \Omega_p$ where X_{ref} is an arbitrary state point to which all “free energy differences” are taken relative from. To illustrate how local modeling can be used to approximate a global property (the equation above for the invariant density depends on the global nonlinear SDE coefficients) we consider a SDE corresponding to Model 3 and locate extremal points of the associated invariant distribution by

locating the zeros of:

$$\begin{aligned} H(X) &\equiv \frac{dG(X)}{dX} = - \left(\frac{\mu(X)}{D(X)} \right) + \frac{1}{D(X)} \frac{\partial D(X)}{\partial X} \\ &\approx \frac{-2(A + B(X - X_0))}{(C + D(X - X_0))^2} + \frac{2D}{(C + D(X - X_0))}. \end{aligned}$$

Note that all of the necessary information to make this (*approximate*) function evaluation is available upon evaluation of a single local affine diffusion model. Carrying out a Newton–Raphson iteration for the distribution mode requires higher order derivative information; this is obtained here by appealing to finite differences, although it would (in principle) be possible to *directly* estimate the derivative information in the local model at the cost of estimating more parameters in the local diffusion models.

The drift component of our illustrative model corresponds to a limiting ODE that has two stable fixed-points for the parameter values used [47]. We used an SDE as our data generating process (our inner simulator) so that we can accurately calculate the invariant distribution of the model. The function, $G(\cdot)$, that is used in determining the invariant density is referred to as “the effective free energy”. This function depends on the coverage (X) and is plotted in figure 9 for the system parameters we used. The plot indicates the existence of two distinct wells in the effective free energy surface. In the infinite time limit, the most probable value of the state is ≈ 0.99 . However, if one starts a simulation near the “other” local effective free energy minimum and evolves over a time that is long -yet much shorter than the characteristic switching time between the two wells- one can approximate the distribution of the particles “stuck” in the local equilibrium density maximum by using $G(X)$. An explicit analytic probability density expression can be found using the

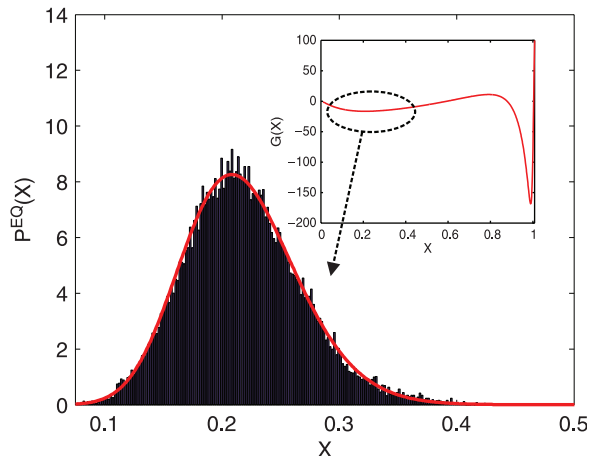


Figure 9. Empirical “almost”-invariant distribution obtained by computing 1000 sample paths for a time sufficient to observe an apparently stationary distribution. Our estimate of the distribution (solid line) is also shown. The inset shows the effective free energy (G) of the system calculated from the *known* SDE coefficients and the dotted arrow indicates the portion of the free energy surface that was used to construct the approximation.

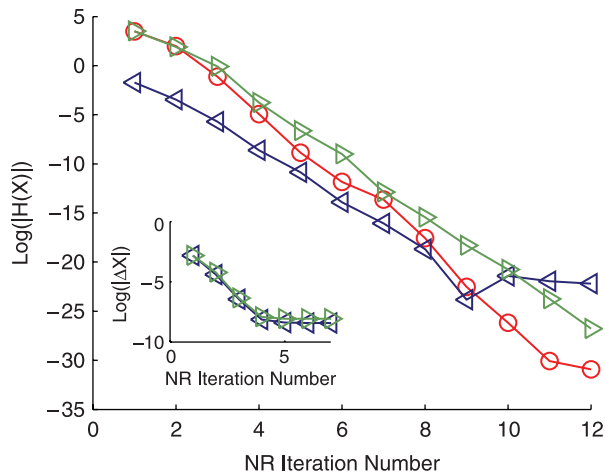


Figure 10. Application of a NR scheme to locate a local maximum of the invariant distribution. The curve with circles is the result obtained by using the exact *known* expression. The curves with triangles show the result obtained by using the local diffusion model parameter estimates (see text for an explanation of the difference between the two curves). The inset plots the difference between the NR guess of the fixed-point and the *known* zero as a function of iteration number (after roughly four iterations the update to the state stagnates).

coefficients of the known SDE (in order to focus on the mode of the PDF in the shallower well, we normalize the PDF so that it has support in the interval $X \in [0, 0.8]$). The resulting (“almost”-invariant) density is shown by the red curve in figure 9. Note that the local maxima of this (closed-form) invariant density is at $X \approx 0.208$, whereas the nearest zero of the limit ODE *drift* is at $X \approx 0.211$. The histogram plotted was obtained by running an ensemble of simulations initialized as a Dirac distribution at $X = 0.20$. The simulation evolved until the empirical density effectively ceased to change in time (quantified by monitoring the first three moments of the empirical distribution). Then 3×10^4 samples were collected and used to construct the histogram ($k_1 = 1.0$, $k_2 = 1.0 \times 10^{-2}$, $k_r = 6.0$). The result of carrying out a Newton fixed-point scheme to determine the value of the state point corresponding to the local distribution maximum highlighted in figure 9 is shown in figure 10. The three different curves correspond to the NR scheme associated with the exact (known) SDE coefficient functions (circles), those obtained by estimating *all* of the coefficient derivatives with respect to X by using finite differences (light triangles), and those obtained by estimating only the second derivatives in X by finite differences (dark triangles).

6. Model reduction

Equation free numerical methods allow us to investigate coarse-grained properties of an atomistic simulator by “wrapping” classical numerical algorithms around local models calibrated from a small set of system observables (measured over a relatively short time burst [38]). Often, in the presence of a significant time scale separation in the

effective dynamics, a complex atomistic simulation may be adequately characterized by a small set of coarse observables.

In the following sections we demonstrate a computer-assisted approach on a toy two-dimensional problem (an SDE version of the MM system studied earlier) when the stiff parameters are used for the atomistic simulator. Before proceeding to performing numerics in the stochastic case, let us recall some facts and terminology from the deterministic limit. Assume a generic ODE of the form

$$\frac{dX}{dt} = f(X, Y) \quad \frac{dY}{dt} = \frac{1}{\epsilon} g(X, Y)$$

where f and g are comparable in magnitude in their domains and the parameter ϵ is used to quantify the “stiffness” of the problem [27]; in the situation above, a system is considered stiff when $\epsilon \ll 1$. Let $r(X)$ be a function that satisfies the relation $g(X, r(X)) = 0$ everywhere in the domain of the above ODE and let $F(U) := dU/dt = f(X, r(X))$. The $r(\cdot)$ function can be used to reduce the dimension of the original problem, and the zeros of $g(X, r(X))$ define a curve in XY space. In deterministic problems, this curve (in general, surface) is sometimes referred to as the *slow manifold* [27].

In our stochastic analog of this type of system, a genuine diffusion is used as the data generating inner process. We thus focus on the errors associated with singular perturbation approximations—as opposed to errors associated with the diffusion approximation of different inner simulators. The literature contains extensive references treating these types of problems; see, for example [14, 17–19, 27, 32, 55]. In the stochastic case, we will use the term “slow manifold” to refer to the surface (analogous to the previous example) defined using only the *drift* function. We start with a simple stochastic example of such multiple time scale dynamics; we then report on “computational savings” that projective integration may offer in these types of systems, and conclude by reporting on hypothesis testing procedures that can be used to quantify whether the simplifying assumptions we make about the dynamics are statistically justified.

6.1 QSSA and estimation

It is well-known that the QSSA of stochastic processes is significantly more intricate than its deterministic counterparts [14, 18, 20, 32, 55]. In the stiff parameter regime studied, our modified MM scheme is such that the substrate (denoted in this section by X) is the “slow component” of interest and the complex (denoted here by Y) is the “fast component” whose detailed dynamics are considered uninteresting. In realistic applications, one will likely not have such a convenient partition of the slow and fast “modes” [27]; we will nevertheless assume this here to simplify the presentation. We are interested in performing numerical integration of a limiting ODE on X

given only noisy observations on X coming from a full stochastic simulation of X and Y .

We are interested in exploiting the stochastic relaxation to a slow manifold (associated with the drift functions) if and when it exists. Before continuing, we present a simulation result to explicitly highlight the behavior we are concerned with. Three batches of MM simulations were executed using multiple (1×10^4) replica trajectories starting from three different Dirac distributions of initial conditions. The X coordinate of all of the Dirac distributions was 0.71; the Y coordinate of the Dirac distribution in one of the simulation batches was initialized *on the deterministic slow manifold*. The other two batches had Dirac distributions with a Y coordinate significantly off the deterministic slow manifold. Figure 11 plots the evolution of the contours of the empirical distributions obtained from these three simulation batches.

The solid lines in figure 11 show the evolution of the (empirical) mean of the distributions associated with the three Dirac-distributed initial conditions. Inspection of the figure shows that the variance associated with the Y coordinate initially increases quickly (in all cases); subsequently the rate of change of the variance (in Y) diminishes. The variance associated with X increases much slower, at a relatively constant rate. One should also note that the dynamical evolution of the two runs corresponding to Dirac distributions initialized away from the deterministic slow manifold is slightly different during the initial short period of integration; yet once the two distributions “relax” to the neighborhood of the deterministic slow manifold, the dynamics of *all three* distributions become nearly identical. This can be confirmed by observing the two different types of approaches of the ensemble mean to the slow manifold (indicated by solid line and symbols outlined in black) and the similar final position and slope of the mean in the phase-portrait given in figure 11. It should also be noted that this common final slope and position in the phase

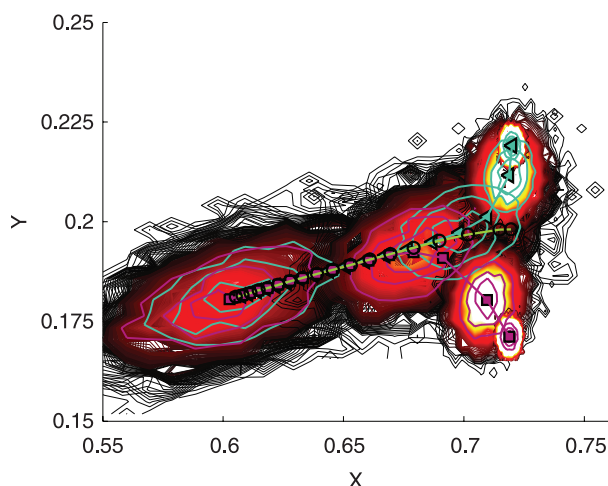


Figure 11. Phase portrait of evolution statistics for the stochastic MM system in a parameter regime where the limiting ODE dynamics are stiff (see text).

portrait is not due to approaching equilibrium: all three of the trajectory distributions are still evolving (albeit slowly) towards the single mode associated with the “coarse steady-state” (located at $\approx [0.58, 0.176]$).

In more complicated systems (such as those arising from inner MD simulations [34]) one hopes that similar properties hold, so that the system may be effectively reducible to one with a small number of coarse observables. Realistic models are typically characterized by *many* fast variables and a relatively small number of useful slow variables. Unfortunately, one rarely knows how to properly initialize the fast degrees of freedom (ideally we would distribute them on some form of slow manifold, but we usually do not have *a priori* knowledge of this surface in complex systems). The observations made in the previous example indicate that useful dynamical information can still be obtained even if one does not have knowledge of the effective slow manifold in some high-dimensional system. One can still accelerate the full simulation if one can determine the time period needed for the full simulation to “heal” [37] itself from “bad” initial guesses of the effective slow manifold. The computations we present in the next two sections aim at (1) developing estimation tools that can be used to facilitate coarse projective integration in such stiff stochastic systems and (2) determining if a lower-dimensional effective dynamical model is meaningful based only on observations of the (chosen) slow variable(s).

6.2 Computational savings of coarse projective integration

In this section we apply the explicit projective Euler method to carry out numerical integration on the drift component of the *reduced* effective system. That is, while our inner simulator produces as output the coupled stochastic evolution of X and Y , we only estimate a local effective diffusion approximation of X . The effective reduced model is assumed statistically meaningful *a priori*; the following section tests this assumption *a posteriori*. Our primary interest is in demonstrating how coarse projective integration can be used as a tool for accelerating the detailed inner atomistic simulations. Coarse projective integration is appealing in situations where the size of the projective step is large relative to the atomistic simulation time. Clearly, as one uses less atomistic data, the quality of the parameter estimates degrades due to finite size sampling effects. In this section we present simulation results that numerically demonstrate this type of trade-off.

*Meaning that the model parameter vector is known without any error.

**To facilitate what follows, all processes studied have initial distributions that correspond to Dirac distributions in the “slow” variable, all time series are

For an effective model of the X dynamics to close meaningfully, adequate averaging in Y must take place (in between observations in time, the atomistic code should sample many significantly different values of Y with X essentially fixed). The likelihood of this increases as the magnitude of an ε type parameter decreases and/or the time between observations increases (assuming that the time window is still small in relation to the slow time scale of the X evolution). In our modified (stochastic) MM system, ε corresponds to the ratio of the smallest to largest eigenvalue of the drift function linearized about a state point where the drift is zero. In the stiff regime studied here and in the next section, recall that this parameter has a value of ≈ 0.09 . The reduction is asymptotic in nature; the dynamics of the reduced process and the dynamics of the full stochastic simulation will almost always differ. However, the difference will typically become harder to measure as the magnitude of an ε type parameter decreases.

For the systems studied here, the difference between the dynamics of the *deterministic* QSSA and the effective dynamics associated with the mean of the X trajectories coming from the *full* modified MM system is small but measurable (figure 12). As a result of the small discrepancy, we use the deterministic QSSA reaction rates in order to determine the initial parameter guesses used in the ML optimization routine. We then estimate a local diffusion model using a batch of atomistic simulations starting the X coordinate in a uniform neighborhood (using $\varepsilon = 0.10$ recall that the “stiffness” parameter is represented by ε) of the specified X_0 and

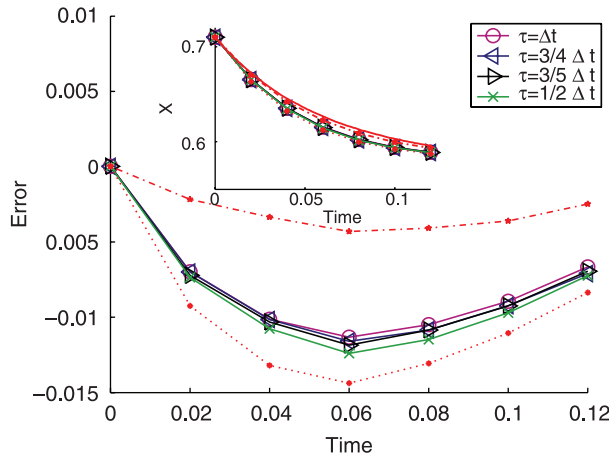


Figure 12. Numerical Model Reduction; Parameters $N_{\text{path}} = 6 \times 10^2$, $\Omega = 1 \times 10^3$, $\delta t = 5 \times 10^{-5}$, $\varepsilon = 0.1$, $\Delta t = 0.02$ (see text for discussion on size of M used). The figure shows the difference between a reduced coarse projective explicit Euler integration and a reference trajectory obtained by taking the empirical average of an ensemble of 1×10^4 replicas of the *full* system with the same initial condition using a stochastic explicit Euler [39] $\delta t = 5 \times 10^{-7}$. The dotted line shows the difference between the corresponding deterministic QSSA explicit Euler integration (same step size) and the reference trajectory. The dash-dot line shows the difference between the “exact” deterministic integration result (using MATLAB’s ode45 and the QSS function) and the reference trajectory. The inset plots the trajectories associated with each integration scheme (the solid red curve is the empirically determined reference trajectory which is considered “exact” in this application).

initializing the fast Y coordinate on the deterministic slow manifold (for each random X drawn, a corresponding Y value on the slow-manifold was determined). We evolve the atomistic simulations using a time step of $\delta t = 1 \times 10^{-5}$, recording observations every 5 steps. The full set of parameters associated with the scalar effective diffusion were obtained first using the entire time series; then the drift was re-estimated by down-sampling the time series by 4 and estimating only the drift parameters. The macroscopic step associated with the coarse projective explicit Euler scheme was $\Delta t = 0.02$ in all cases, and the size of the time series used for estimation in four (different) simulation runs was (400,300,240,200), these sizes correspond to the ratios

$$\left(\frac{\Delta t}{5\delta t}, \frac{(3/4)\Delta t}{5\delta t}, \frac{(3/5)\Delta t}{5\delta t}, \frac{(1/2)\Delta t}{5\delta t} \right).$$

The numerical error of the reduced coarse projective scheme was quantified by running a batch of 1×10^4 trajectories starting from the same initial configuration (using the same δt) for the same time as the projective integration scheme and then using the average of the batch as the “effective reference trajectory”. The errors obtained in this manner are given in figure 12; one observes a systematic decrease in error magnitude as the sample size (hence length of time series) used to generate the information used in the projective time step increases (which decreases the “savings” of the projective scheme). The systematic difference is due mainly to the fact that the initial parameter guesses were biased (we used deterministic QSSA parameter values to initialize) and all simulations had the same termination criterion for the optimization algorithm employed for parameter estimation. With a shorter estimation interval (recall we obtain a local model parameter vector using *ensembles* of time series), the likelihood function, viewed as a function of the *finite* number of observations and the parameter vector, becomes “noisier” (i.e. further from the large sample likelihood surface). Since we start the (parameter fitting) optimization routine in a neighborhood of a biased parameter estimate, the algorithm has a higher chance of getting stuck in a parameter regime near the biased initial parameter guess when smaller time series are used. Using a more sophisticated optimization routine (we used the IMSL’s Nelder–Mead simplex search) would likely reduce the computational cost.

The magnitude of the differences in the trajectories obtained using longer time intervals for estimation (less computational savings) is small in relation to the magnitude of the difference between the explicit Euler integration (using an integrator step size equal to the projective step size) of the deterministic QSSA system and the “exact trajectory” of the same QSSA system (obtained by integrating with MATLAB’s ode45 function). This suggests that the primary source of error is due to the Euler discretization in the coarse projective integration scheme, rather than to the effective model estimation error.

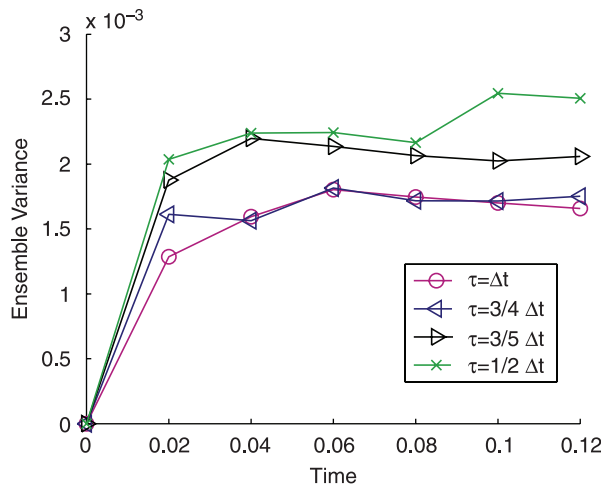


Figure 13. The procedure in figure 12 was repeated 100 times (recall that each parameter vector estimated for a projective step was the single parameter vector optimal over an *ensemble* of trajectories) and the resulting variance (as a function of time) of the reduced coarse projective integration is plotted here. It should be noted that a monotonic decrease in variance is not observed because each path contains errors associated not only with standard estimation uncertainty, but errors resulting from projective integration and artifacts of model misspecification.

Figure 13 illustrates how the variance of the projective integration procedure depends on sample size. It should be noted that this “variance” consists of errors due to an accumulation of projective integration discretization errors and errors associated with estimating a sequence of effective local diffusion models using an imperfect model and finite samples.

6.3 Hypothesis testing with nonstationary multiscale data

In this section we use a slight variant of Hong and Li’s Q-test [33] in order to determine if the dynamics observed (during the types of estimation intervals shown above) in our nonstationary time series are consistent with the assumed diffusion model. Recall the discussion at the end of Section 6.1; there it was mentioned that if one could quantify a healing time, then it might be possible to *estimate* a lower dimensional (reduced) effective model.

Our estimation-based model reduction strategy calibrates a stochastic dynamical model by only observing some slow variable(s). However, in order to have some confidence in our estimated reduced models, we need to ensure that our time series samples are significantly beyond a transient regime associated with fast time scale relaxation. If we are not beyond such a transient regime, then the reduced model we estimate would depend on the “details” of the relaxation. If there is a large time scale separation in the system under study, typically only a small portion of time series data needs to be discarded for a reduced (estimated) model to be a useful and reliable approximation. Throwing out a large quantity of “good” data is, however, also undesirable because it causes an

unnecessary increase in parameter uncertainty. Not discarding enough initial observations (where transient effects are still important) can make the model depend on the initial Y value or cause the assumption of diffusive process noise to be poor. It would be useful to determine (quantitatively) how much healing time is needed for a meaningful reduced dynamical model to be reliably estimated. This is the goal of the computations presented in the next subsection.

6.4 Windowed hypothesis tests. In Hong and Li’s Q-test [33], they appeal to asymptotic results (along a single sample path) that require the length of the time series to tend to infinity. In time series analysis situations, one typically attempts to numerically realize an infinite sample length by either (1) fixing the sampling frequency and letting the time of the terminal observation tend to infinity or (2) fixing the terminal observation time and letting the time between observations tend to zero. If a diffusion model actually generates the data, either type of limiting method can usually be used to produce the sample sizes needed.

However, in projective integration, we use short bursts of simulation data. This makes the terminal time of our observation small in relation to the actual time scale of interest. Sampling too frequently in time with multiscale signals often causes diffusion models to be invalid [11,52]. Both of these facts make the direct application of Hong and Li’s asymptotically motivated Q-test problematic.

Coincidentally, for a *correctly* specified model*, the distribution of the PIT series is known [16]; this distribution is not asymptotic in nature. So in order to overcome the difficulties mentioned in the previous paragraph and still utilize the ideas laid out in Ref. [33], we carry out a “windowed” Q-test by doing the following**:

- (i) Propose a parametric model and a specific parameter value that locally describes the process dynamics (this serves as the null hypothesis).
- (ii) Accurately simulate [39] a *batch* of genuine diffusion paths consistent with the null hypothesis. If one strongly believes that the error associated with the transition density approximation employed is negligible, one can bypass this simulation step because the distribution of the PIT series is known under the null [33].
- (iii) Divide each sample path (time series) into B sub-blocks.
- (iv) Compute the Q-test statistic (which requires one to first compute the corresponding PIT time series) within each sub-block for each time series.
- (v) Use the sub-blocks corresponding to the same point in time and empirically measure the distribution of the Q-test statistic (which will typically significantly differ from the large sample normal limit if B is large relative to the total number of observations along individual sample paths) for each sub-block. Use this

data to generate a cumulative distribution function (CDF) of the windowed test statistic for *each* of the B sub-blocks. and now,

- (vi) Obtain samples from the actual process of interest. Then specify a significance level [9] and use the CDFs in step (v) to determine the critical values. Repeat steps (iii) and (iv) with the actual process data; finally determine whether the null is rejected (within each of the B sub-blocks) for each of the *individual* sample paths being tested.

The method employed is by no means optimal. For example, we do not take parameter uncertainty into account, nor do we account for errors associated with empirically determining distributions in step (v). However the simple test used does prove useful in some respects as demonstrated in the following subsection.

6.4 Application to reduced MM stochastic dynamics.

In order to carry out step (i), we obtained local parameter estimates of an effective scalar diffusion process by local MLE using a batch of 500 time series (where X and Y were both evolved) spaced by $\delta t = 1 \times 10^{-4}$ using $M = 4000$ observations. All time series started with Y initialized on the deterministic slow manifold (all simulations in this section initialize X as a Dirac distribution at 0.71). Again, we would like to remind the reader that a *single parameter* vector was obtained which optimized the likelihood function given in equation (5). The first quarter of the time series was discarded (this provided more than enough “healing time”) when estimating the null parameters. Then steps (ii)–(v) were carried out using the local SDE implied by the parameters of the estimated (scalar) process. The null process was sampled every $\delta t = 1 \times 10^{-5}$ and the Euler–Maruyama scheme [39] was used to simulate the paths; we also set $B = 100$. Finally step (vi) was carried out using the *full* system to generate data, but this time Y was initialized as a Dirac distribution significantly away from the deterministic slow manifold (here $Y = 0.17$ was used as the initial Dirac distribution) and the process was (again) sampled using $\delta t = 1 \times 10^{-5}$. An increased sampling frequency was used in steps (ii)–(vi) because we wanted to create a significant portion of “imperfect” time series data.

We specified the significance level of all of our windowed hypothesis tests to be 15%. The results of carrying out this procedure are shown in figure 14 where one observes that after $\tau \approx 0.015$ the rejection rate (percent of the batch that had a test statistic greater than the critical value) roughly approaches that associated with a correctly specified model (i.e. the rate of rejection is consistent with that of the null). The insets plot the mean (solid line) and standard deviation (dotted lines) of the slow variable (left) and fast variable (right). The right inset shows that the point at which the rate of change of the “time derivative” of the empirical mean of the fast variable reaches a minimum magnitude roughly corresponds to the point at which

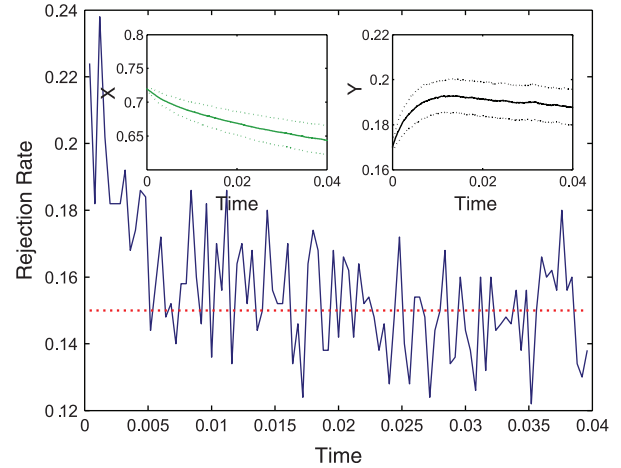


Figure 14. Rejection rate using an empirically determined critical value. The Q-statistic (using a lag of 1) [33] was constructed by partitioning each individual time series into $B = 100$ bins and calculating the test statistic in each bin; we define the rejection rate as the fraction of rejected paths (over an N_{path} simulation). The critical value was determined from the empirical distribution of this statistic (using a known scalar reference SDE model whose parameters were determined by those estimated using $X_0 = 0.71$ with Y distributed near the deterministic slow manifold (see text)). Other system parameters: $N_{\text{path}} = 5 \times 10^2$, $\Omega = 2 \times 10^3$, $\delta t = 1 \times 10^{-5}$, $\varepsilon = 0$, $M = 4000$. The solid line indicates the Type I error rate specified (this determined the one-sided critical value using our empirical reference CDF). The insets show the evolution of the first two moments of the slow (left) and fast (right) components.

reliable samples can be taken for the reduced model (as measured by the rejection rate leveling off).

For a variety of complex systems one typically does not know of an observable analogous to Y in the MM system, making it difficult to quantitatively determine when one can start collecting data for reduced model estimation. The computational tool presented here does not require knowledge of a representative fast variable. The technique only requires a rough preliminary parameter estimate of the reduced model and knowledge of a good slow observable (“reaction coordinate”). The technique can be applied to determine how much data can be used in situations where one discards a significant portion of the time series due to concerns about “healing effects”. The method can even be applied to determine which sampling rate is too frequent in situations where the true data cannot be modeled as a genuine diffusion [20,29] at small observation times. For a simple example of this type of application, see [12]. It should be noted, that taking other factors into account (like those stated at the end of the previous subsection) may slightly modify the shape of the measured rejection rate curve, but the qualitative nature of the test should not change.

7. Conclusions

We have demonstrated various computer-assisted applications where calibrating a local diffusion model with data collected from an inner fine scale stochastic/atomistic simulator can accelerate the extraction of information

from the inner simulator. The local diffusion models can be used to design “intelligent” subsequent computational experiments by linking them to deterministic numerical methods in the so-called equation-free framework. Accelerated simulation via coarse projective integration was demonstrated, as was a procedure through which a small set of local computational experiments could be used to determine the maximum of a “almost”-invariant distribution associated with an effectively scalar stochastic system. The approach also proved useful in computational model reduction tasks. The main advantages of wrapping our parametric models around the inner simulator in order to design computational experiments with it are (1) that it allows one to efficiently obtain coarse derivative information around a coarse state point without needing to constrain the inner simulation and (2) it allows one to utilize a powerful set of statistical tools for quantitatively determining how well the assumed local diffusion model fits the empirical data for the observation frequency and sample size used. The careful numerical analysis of such multiscale stochastic schemes is the subject of current research (e.g. [17,18]).

Acknowledgements

The author acknowledges I. Kevrekidis for useful discussions and suggestions. This work was partially supported by the Ford foundation NRC Fellowship, DARPA, the US Department of Energy and ACS-PRF.

Notes

¹ The standard diffusion approximation of the SSA [25,7,29] uses r Brownian noise terms, where r is the number of reaction events; each reaction event i has a “propensity” given by the function $a_i(\cdot)$ and “state change vector” ν_i [29]. For a 1-d problem, one can (without much complication) write another diffusion driven by a single Brownian term which is distributionally equivalent [25] to the diffusion process driven by r Brownian motions. The diffusion coefficient of this equivalent process is given by $(\Omega^{-1} \sum_{i=1}^r (\nu_i^2 a_i))^{\frac{1}{2}}$. In order to avoid technical complications associated with imaginary diffusion coefficients in our “Model 3”, we instead used $(\Omega^{-1} \sum_{i=1}^r (\nu_i^2 a_i^2))^{\frac{1}{2}}$ for the diffusion term (and used the “correct” limiting drift term); this actually makes the diffusion coefficient more nonlinear (which made the benchmark problem more “interesting”), but note that this *new* diffusion no longer corresponds to the limiting SDE of the SSA model.

References

- [1] Y. Ait-Sahalia. Closed-form likelihood expansions for multivariate diffusions. Technical Report (2001).
- [2] Y. Ait-Sahalia. Maximum-likelihood estimation of discretely-sampled diffusions: a closed-form approximation approach. *Econometrica*, **70**, 223 (2002).
- [3] Y. Ait-Sahalia, R. Kimmel. Estimating affine multifactor term structure models using closed-form likelihood expansions. NBER Technical Working Papers 0286, National Bureau of Economic Research, Inc., December. Available at <http://ideas.repec.org/p/nbr/nberte/0286.html> (2002).
- [4] A. Amadei, A.B.M. Linssen, H.J.C. Berendsen. Essential dynamics of proteins. *Proteins*, **17**, 412 (1993).
- [5] G. Bakshi, N. Ju. A refinement to Ait-Sahalia’s (2002) “Maximum likelihood estimation of discretely sampled diffusions: a closed-form approximation approach”. *J. Business*, **78**, 2037 (2005).
- [6] G. Bakshi, N. Ju, H. Ou-Yang. Estimation of continuous-time models with an application to equity volatility dynamics. *J. Finan. Econ.*, **82**, 227 (2006).
- [7] K. Ball, T. Kurtz, L. Popovic, G. Rempala. Asymptotic analysis of multiscale approximations to reaction networks. <http://www.citebase.org/abstract?id=oai:arXiv.org:math/0508015> (2005).
- [8] B.M. Bibby, M. Sørensen. Martingale estimation functions for discretely observed diffusion processes. *Bernoulli*, **1**, 17 (1995).
- [9] P.J. Bickel, K.J. Doksum. *Mathematical Statistics: Basic Ideas and Selected Topics*, Prentice Hall, Upper Saddle River, NJ (2001).
- [10] M. Bradnt, P. Santa-Clara. Simulated likelihood estimation of diffusions with an application to exchange rate dynamics in incomplete markets. *J. Finan. Econ.*, **63**, 161 (2002).
- [11] C.P. Calderon. Fitting effective diffusion models to data associated with a “glassy potential”: estimation, classical inference procedures and some heuristics. *SIAM MMS*, (to appear).
- [12] C.P. Calderon, G.A. Tsekouras, A. Provata, I.G. Kevrekidis. Coarse graining the Cyclic Lotka-Volterra Model: SSA and local maximum likelihood estimation. In *Proceedings of Model Reduction and Coarse-Graining Approaches for Multiscale Phenomena*, A. Gorban, N. Kazantzis, I. Kevrekidis, H. Ottinger, C. Theodoropoulos (Eds.), Workshop at U. Leicester, Leicester, UK, August 24–26, Springer-Verlag, Heidelberg (2005).
- [13] C.P. Calderon. On the use of local diffusion models in potential of mean force computations. *J. Chem. Phys.*, **126**, 084106 (2007).
- [14] Y. Cao, D.T. Gillespie, L.R. Petzold. The slow-scale stochastic simulation algorithm. *J. Chem. Phys.*, **122**, 014116 (2005).
- [15] A.S. Dalalyan, Y.A. Kutoyants. Asymptotically efficient estimation of the derivative of the invariant density. *Stat. Inference Stoch. Process.*, **6**, 89 (2003).
- [16] F.X. Diebold, T. Gunther, A. Tay. Evaluating density forecasts with applications to financial risk management. *Int. Econ. Rev.*, **39**, 863 (1998).
- [17] W.E.D. Liu, E. Vanden-Eijnden. Analysis of multiscale methods for stochastic differential equations. *Comm. Pure Appl. Math.*, **58**, 1544 (2005).
- [18] W.E.D. Liu, E. Vanden-Eijnden. Nested stochastic simulation algorithm for chemical kinetic systems with disparate rates. *J. Chem. Phys.*, **123**, 194107 (2005).
- [19] E. Vanden-Eijnden. Numerical techniques for multi-scale dynamical systems with stochastic effects. *Comm. Pure Appl. Math.*, **1**, 385 (2003).
- [20] M. El-Ansary, H. Khalil. On the interplay of singular perturbations and wide-band stochastic fluctuations. *SIAM J. Control Optim.*, **24**, 83 (1986).
- [21] S. Ethier, T. Kurtz. *Markov Processes: Characterization and Convergence*, Wiley-Interscience, New York (1986).
- [22] E. Fournié, J.L. Lasry, P.L. Lions, J. Lebuchoux, N. Touzi. Applications of Malliavin calculus to Monte Carlo methods in finance. *Finance Stoch.*, **3**, 391 (1999).
- [23] E. Fournié, J.L. Lasry, J. Lebuchoux, P.L. Lions. Applications of Malliavin calculus to Monte Carlo methods in finance ii. *Finance Stoch.*, **5**, 201 (2001).
- [24] A.R. Gallant, G. Tauchen. Which moments to match? *Econometric Theory*, **12**, 657 (1996).
- [25] C.W. Gardiner. *Handbook of Stochastic Models*, Springer-Verlag, Berlin (1985).
- [26] C.W. Gear. *Numerical Initial Value Problems in Ordinary Differential Equations*, Prentice-Hall, Upper Saddle River, NJ (1971).
- [27] C.W. Gear, T.J. Kaper, I.G. Kevrekidis, A. Zagaris. Projecting to a slow manifold: singularly perturbed systems and legacy codes. *SIAM J. Appl. Dyn. Syst.*, **4**, 711 (2005).
- [28] D.T. Gillespie. Exact stochastic simulation of coupled chemical reactions. *J. Phys. Chem.*, **81**, 2340 (1977).
- [29] D.T. Gillespie, L.R. Petzold. Improved leap-size selection for accelerated stochastic simulation. *J. Chem. Phys.*, **119**, 8229 (2003).
- [30] J. Gullingsrud, R. Braun, K. Schulten. Reconstructing potentials of mean force through time series analysis of steered molecular dynamics simulations. *J. Comp. Phys.*, **151**, 190 (1999).

- [31] J.D. Hamilton. *Time Series Analysis*, Princeton University Press, Princeton, NJ (1994).
- [32] E.L. Haseltine, J.B. Rawlings. Approximate simulation of coupled fast and slow reactions for stochastic chemical kinetics. *J. Chem. Phys.*, **117**, 6959 (2002).
- [33] Y. Hong, H. Li. Nonparametric specification testing for continuous-time models with applications to term structure of interest rates. *Rev. Finan. Stud.*, **18**, 37 (2005).
- [34] G. Hummer, I.G. Kevrekidis. Coarse molecular dynamics of a peptide fragment: free energy, kinetics, and long-time dynamics computations. *J. Chem. Phys.*, **118**, 10762 (2003).
- [35] P. Jeganathan. Some aspects of asymptotic theory with applications to time series models. *Econometric Theory*, **11**, 818 (1995).
- [36] B. Jensen, R. Poulsen. Transition densities of diffusion processes: numerical comparison of approximation techniques. *J. Derivatives*, **9**, 18 (2002).
- [37] I.G. Kevrekidis, C.W. Gear, J.M. Hyman, P.G. Kevrekidis, O. Runborg, K. Theodoropoulos. Equation-free coarse-grained multiscale computation: enabling microscopic simulators to perform system-level tasks. *Comm. Math. Sci.*, **1**, 715 (2003).
- [38] I.G. Kevrekidis, C.W. Gear, G. Hummer. Equation-free: the computer-aided analysis of complex multiscale systems. *AIChE J.*, **50**, 474 (2004).
- [39] P. Kloeden, E. Platen. *Numerical Solution of Stochastic Differential Equations*, Springer-Verlag, Berlin (1992).
- [40] D.I. Kopelevich, A.Z. Panagiotopoulos, I.G. Kevrekidis. Coarse-grained kinetic computations for rare events: application to micelle formation. *J. Chem. Phys.*, **122**, 044908 (2005).
- [41] S. Kullback, R.A. Leibler. On information and sufficiency. *Ann. Math. Stat.*, **22**, 79 (1951).
- [42] R. Kupferman, A.M. Stuart. Fitting sde models to nonlinear Kac-Zwanzig heat bath models. *Phys. D*, **199**, 279 (2004).
- [43] H. Kushner. *Weak Convergence Methods and Singularly Perturbed Stochastic Control and Filtering Problems*, Birkhauser, Boston (1990).
- [44] J.D. Lambert. *Numerical Methods for Ordinary Differential Systems: The Initial Value Problem*, Wiley, New York (1991).
- [45] L. Le Cam, G.L. Yang. *Asymptotics in Statistics: Some Basic Concepts*, Springer-Verlag, New-York (2000).
- [46] P.K. Maini, M.A. Burke, J.D. Murray. On the quasi-steady-state assumption applied to Michaelis–Menten and suicide substrate reactions with diffusion. *Phil. Trans. R. Soc.*, **337A**, 299 (1991).
- [47] A. Makeev, D. Maroudas, I.G. Kevrekidis. Coarse stability and bifurcation analysis using stochastic simulators: kinetic Monte Carlo examples. *J. Chem. Phys.*, **116**, 10083 (2002).
- [48] A. Makeev, D. Maroudas, A.Z. Panagiotopoulos, I.G. Kevrekidis. Coarse bifurcation analysis of kinetic Monte Carlo simulations: a lattice-gas model with lateral interactions. *J. Chem. Phys.*, **117**, 8229 (2002).
- [49] J.D. Murray. *Mathematical Biology I: An Introduction*, Springer-Verlag, New-York (2004).
- [50] D. Nualart. *The Malliavin Calculus and Related Topics*, Springer, New York (1995).
- [51] S. Park, K. Schulten. Calculating potentials of mean force from steered molecular dynamics simulations. *J. Chem. Phys.*, **120**, 5946 (2004).
- [52] G.A. Pavliotis, A.M. Stuart. Parameter estimation for multiscale diffusions. <http://www.citebase.org/cgi-bin/citations?id=oai:arXiv.org:math/0603668> (2006).
- [53] A.R. Pedersen. A new approach to maximum likelihood estimation for stochastic differential equations based on discrete observations. *Scand. J. Stat.*, **22**, 55 (1995).
- [54] P.E. Protter. *Stochastic Integration and Differential Equations*, Springer, Berlin (2005).
- [55] C. Rao, A. Arkin. Stochastic chemical kinetics and the quasi-steady-state assumption: application to the Gillespie algorithm. *J. Chem. Phys.*, **118**, 4999 (2003).
- [56] R. Rico-Martínez, C.W. Gear, I.G. Kevrekidis. Coarse projective kMC integration: forward/reverse initial and boundary value problems. *J. Comput. Phys.*, **196**, 474 (2004).
- [57] H. Risken. *The Fokker-Planck Equation*, Springer-Verlag, Berlin (1984).
- [58] E. Rusli, T.O. Drews, D.L. Ma, R.C. Alkire, R.D. Braatz. Robust nonlinear feedback-feedforward control of a coupled kinetic Monte Carlo finite-difference simulation. *J. Process Control*, **16**, 409 (2006).
- [59] E. Schutz, N. Hartmann, Y. Kevrekidis, R. Imbühl. Microchemical engineering of catalytic reactions. *Catal. Lett.*, **54**, 181 (1998).
- [60] R.U. Seydel. *Practical Bifurcation and Stability Analysis: From Equilibrium to Chaos*, Springer-Verlag, New York (1994).
- [61] J.M.T. Thompson, H.B. Stewart. *Nonlinear Dynamics and Chaos*, Wiley, West Sussex, UK (2002).
- [62] G.M. Torrie, J.P. Valleau. Nonphysical sampling distributions in Monte Carlo free-energy estimation: umbrella sampling. *J. Comp. Phys.*, **23**, 187 (1977).
- [63] A. van der Vaart. *Asymptotic Statistics*, Cambridge University Press, New York (1998).
- [64] H. White. Maximum likelihood estimation of misspecified models. *Econometrica*, **50**, 1 (1982).