SAND2022-14004C

Data assimilation for microstructure evolution in kinetic Monte Carlo

Anh Tran *, Yan Wang, Theron Rodgers

Abstract Modeling grain growth has been a subject of interest in computational material science, as it occurs in thermal-based processing methods such as annealing and sintering. Kinetic Monte Carlo with Potts model is often used as an integrated computational materials engineering (ICME) grain growth model and can generate high-fidelity synthetic microstructures. In this paper, we offer a data-driven stochastic calculus perspective on the kinetics of grain growth and model the microstructure evolution through the lens of stochastic differential equations, based on Langevin dynamics and Fokker-Planck equation to forecast the grain size distribution. We demonstrate that our proposed approach agrees reasonably well with the hybrid Potts-phase field model using SPPARKS in forecasting the long-term evolution of grain size distribution.

Key words: Langevin equation, Fokker-Planck, advection-diffusion, stochastic reduced-order model, data assimilation, grain growth, kinetic Monte Carlo, SPPARKS

1 Introduction

Integrated Computational Materials Engineering (ICME) models are powerful tools to study the process-structure-property linkages at multiple length and time scales. Numerous ICME models have been developed to study material systems from quantum to continuum scales. Quantifying local distributions of microstructures or configurations and mapping statistical ensembles to the measurable properties at a larger scale become an important methodology in the simulations of multiscale systems. Describing microstructures statistically and modeling their evolution have been one of the important topics in computational materials science.

Here, we propose a stochastic modeling approach to accelerate the predictions of statistical QoIs in material evolution, specifically the statistical descriptors of microstructures such as grain size distribution of polycrystalline metals. Mathematically, the propagation of statistical distributions can be modeled as continuous-time stochastic processes. We apply the Kramers-Moyal expansion to model the evolution of probability density function (pdf) of the QoI along time. When only the first and second orders of the expansion are considered, the model is simplified to the well-known Fokker-Planck equation. The Fokker-Planck equation is a deterministic partial differential equation of pdfs, which is equivalent to the stochastic differential equation usually employed to model Langevin dynamics. The dynamics of a statistical QoI is modeled by the Fokker-Planck equation as a one-dimensional continuous-time stochastic process, where the drift and diffusion coefficients in the equation are estimated and calibrated using the available time-series dataset, collected from ICME simulations for some period of time. Once calibrated, the Fokker-Planck equation can be applied to predict the dynamics of pdfs for the statistical QoIs directly and very efficiently for a much longer time period, instead of continuously relying on the actual ICME simulations. Thus the Fokker-Planck equation can be regarded as a stochastic reduced-order model (SROM) in our approach. The advantage of the proposed SROM is that the predictions of statistical QoIs in the simulated material systems can be

Anh Tran
Scientific Machine Learning, Sandia National Laboratories, Albuquerque, NM. e-mail: anhtran@sandia.gov

Yan Wang
Woodruff School of Mechanical Engineering, Georgia Institute of Technology, Atlanta, GA 30332 e-mail: yan-wang@gatech.edu

Theron Rodgers
Computational Materials and Data Science, Sandia National Laboratories, Albuquerque, NM. e-mail: trodger@sandia.gov

* Corresponding author: e-mail: anhtran@sandia.gov

significantly accelerated because the time scale used in the Fokker-Planck equation can be higher than the ones in the ICME model, whichever ICME model has been used to generate the training dataset.

The parameters of the SROM are the drift and diffusion coefficients. They need to be calibrated based on the ICME models. With well-calibrated parameters, the SROM can predict the evolution of QoIs. In this work, the distributions of QoIs from the ICME models as time series are divided into two time periods or stages. The data from the first stage are used to train and calibrate the SROM, whereas the second stage is used to test the SROM performance. In the rest of this paper, we denote $X(t)$ as a one-dimensional QoI in an ICME model, where $X(t)$ is considered as a continuous-time stochastic process and modeled using the Langevin equation. The outline of the paper is as follows. Section 2 describes the 1-dimensional Fokker-Planck equation and provides the problem statement. Section 3 describes the application in kinetic Monte Carlo with SPPARKS [23, 24] and numerical results of the proposed SROM approach. Section 4 discusses and concludes the paper.

2 The Fokker-Planck equation

In this section, the background of uncertainty propagation using the Kramers-Moyal expansion and Fokker-Planck equation is provided, as it is the backbone of our proposed methodology. The one-dimensional (1D) non-linear Langevin equation for stochastic variable $X(t)$, which is the QoI in this paper, provides the preliminary context which gives rise to the Fokker-Planck equation. Following the notation of Risken [26] and Frank [6], the 1D non-linear Langevin equation reads

$$\dot{X} = h(X,t) + g(X,t)\Gamma(t), \tag{1}$$

where the Langevin force $\Gamma(t)$ is assumed to be a Gaussian distributed, white noise term with zero mean and δ correlation function [17], i.e.

$$\langle \Gamma(t) \rangle = 0, \quad \langle \Gamma(t)\Gamma(t') \rangle = \delta(t-t'). \tag{2}$$

Here, $\langle \cdot \rangle$ denotes the expectation, $\delta(t-t')$ is the Dirac-delta function, and the intensity of noise is 1 by convention [29]. If $h(X,t) = 0$ and $g(X,t) = 1$, Equation 1 describes the Wiener process. There are two alternative ways to interpret the drift and diffusion coefficients of the Fokker-Planck equation. The first is based on Itô calculus and the second is based on Stratonovich calculus, depending on the existence of spurious or noise-induced drift [26]. For Itô calculus, $D^{(1)}(X,t) = h(X,t)$, whereas for Stratonovich calculus, $D^{(1)}(X,t) = h(X,t) + \dfrac{\partial g(X,t)}{\partial X} D^{(2)}(X,t)$. For both Itô and Stratonovich calculus, $D^{(2)}(X,t) = g^2(X,t)$. With Itô calculus, Equation 1 can be rewritten as $dX(t) = h(X,t)dt + g(X,t)dW_t$, where W_t is the Wiener process, which is the source of randomness [29]. $X(t)$ is the QoI dependent on time, also referred to as the state variable. We restrict $X(t)$ to be 1-dimensional, even though it can be generalized for higher dimensional QoIs. In the scope of this paper, the Itô calculus is used to interpret the stochastic process in Equation 1, that is,

$$dX(t) = D^{(1)}(X,t)dt + D^{(2)}(X,t)dW(t), \tag{3}$$

where $W(t)$, $t \geq 0$ is a scalar Wiener process, $D^{(1)}(X,t)$ and $D^{(2)}(X,t)$ are the drift and diffusion coefficients, respectively. The stochastic process described in Equation 1 is a Markov process with δ-correlated force; this Markov properties is destroyed if $\Gamma(t)$ is no longer δ-correlated [26].

The ordinary differential equation of pdfs based on the Kramers-Moyal expansion is a general model of the evolution of probability distributions. The Fokker-Planck equation is a special case when only the first- and second-order spatial derivatives are considered. The Kramers-Moyal expansion for the pdf $f(X,t)$ associated with stochastic variable $X(t)$ can be written as [26]

$$\frac{\partial f(X,t)}{\partial t} = \sum_{j=1}^{n} \left(\frac{-\partial}{\partial X} \right)^j D^{(j)}(X,t) f(X,t). \tag{4}$$

where n is the number of truncated terms in Kramers-Moyal expansion, $D^{(j)}(X,t)$ is the Kramers-Moyal expansion coefficient. The Pawula's theorem [26] states that the Kramers-Moyal expansion may stop either after the first term or after the second term; if it does not stop after the second term, then it must contain an infinite number of terms. With only the first two terms, the Kramers-Moyal expansion is reduced to the Fokker-Planck equation as

$$\frac{\partial f(X,t)}{\partial t} = -\frac{\partial}{\partial X}\left[D^{(1)}(X,t)f(X,t)\right] + \frac{\partial^2}{\partial X^2}\left[D^{(2)}(X,t)f(X,t)\right], \tag{5}$$

where $D^{(1)}$ and $D^{(2)}$ are the drift and diffusion coefficients, respectively, and both are spatio-temporal functions.

Assume the evolution of the stochastic variable $X(t)$, which is the QoI, can be modeled by the Fokker-Planck equation. This assumption holds if $X(t)$ obeys a Langevin equation with Gaussian δ-correlated noise, it can be shown that all coefficients other than drift and diffusion coefficients vanish, and the Kramer-Moyal expansion simply reduces to the Fokker-Planck equation (cf. [26], Section 1.2.7). The Fokker-Planck equation is sometimes referred to as the backwards or second Kolmogorov equation in literature [25]. Furthermore, for the case of one stochastic variable QoI $X(t)$ in Equation 1, it is always possible to convert the Langevin equation from a multiplicative noise force g to an additive noise force, through a transformation of variable (cf. [26], Section 3.3).

Theorem 1 ([17, 27, 13]) *Denote the n^{th} central moment as $M^{(n)}(X,t)$, then*

$$\frac{\partial}{\partial t}M^{(n)}(X,t) = n!D^{(n)}(X,t) + O(\Delta t^2) \tag{6}$$

Thus, the Kramers-Moyal coefficients can be estimated from sampling the time-series data as

$$D^{(n)}(X,t) = \lim_{\Delta t\to 0}\frac{1}{n!\Delta t}\langle[X(t+\Delta t) - X(t)]^n\rangle\Big|_{X(t)=x}. \tag{7}$$

It is noted that Theorem 1 follows from the Taylor series expansion in deriving Kramers-Moyal expansion. The evolution of mean and variance can be modeled through the two following corollaries.

Corollary 1 *The expectation of $X(t)$, denoted as $\mathbb{E}[X(t)]$,*

$$\mathbb{E}[X(t)] := \int_{-\infty}^{\infty} Xf(X,t)dX, \tag{8}$$

must satisfy

$$\frac{\partial}{\partial t}\mathbb{E}[X(t)] = \int_{-\infty}^{\infty} D^{(1)}(X,t)f(X,t)dX. \tag{9}$$

If the drift coefficient is a temporal function, only i.e. $D^{(1)}(X,t) = D^{(1)}(t)$, then

$$\frac{\partial}{\partial t}\mathbb{E}[X(t)] = D^{(1)}(t). \tag{10}$$

Corollary 2 *Assume that the drift coefficient is a temporal function, i.e. $D^{(1)}(X,t) = D^{(1)}(t)$, then the variance of $X(t)$, denoted as $\mathbb{V}[X(t)]$,*

$$\mathbb{V}[X(t)] := \int_{-\infty}^{\infty}(X - \mathbb{E}[X(t)])^2 f(X,t)dX, \tag{11}$$

must satisfy

$$\frac{\partial}{\partial t}\mathbb{V}[X(t)] = 2\int_{-\infty}^{\infty} D^{(2)}(X,t)f(X,t)dX \tag{12}$$

If the diffusion coefficient is also a temporal function, i.e. $D^{(2)}(X,t) = D^{(2)}(t)$, then

$$\frac{\partial}{\partial t}\mathbb{V}[X(t)] = 2D^{(2)}(t). \tag{13}$$

2.1 Analytical solutions of Fokker-Planck equation

Here we present three simple analytical examples that can capture a wide range of phenomenon with increasing complexity. These examples are the analytical solution of the one-dimensional Fokker-Planck equation as in Equation 5.

The first example is a Gaussian pdf with no drift and constant diffusion parameter,

$$f_1(X,t) = \frac{1}{\sqrt{4\pi Dt}} e^{-\frac{X^2}{4Dt}}. \tag{14}$$

It is easy to see that for $f_1(X,t)$, $\mathbb{E}[X(t)] = 0$, $D^{(1)} = \frac{\partial \mathbb{E}[X(t)]}{\partial t} = 0$. $\mathbb{V}[X(t)] = 2Dt$, $D^{(2)} = \frac{1}{2}\frac{\partial \mathbb{V}[X(t)]}{\partial t} = D$.

The second example is a Gaussian pdf with constant drift, where the mean is moving with a constant velocity and no diffusion,

$$f_2(X,t) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(X-\mu t)^2}{2\sigma^2}}. \tag{15}$$

In $f_2(X,t)$ case, $\mathbb{E}[X(t)] = \mu t$, $D^{(1)} = \frac{\partial \mathbb{E}[X(t)]}{\partial t} = \mu$. $\mathbb{V}[X(t)] = \sigma^2$, $D^{(2)} = \frac{1}{2}\frac{\partial \mathbb{V}[X(t)]}{\partial t} = 0$.

The third example, which is the most general among these examples, is a Gaussian pdf with constant drift, where the mean is moving with a constant velocity and constant diffusion, where the variance increases linearly with respect to time,

$$f_3(X,t) = \frac{1}{\sqrt{4\pi Dt}} e^{-\frac{(X-\mu t)^2}{4\pi Dt}}. \tag{16}$$

For $f_3(X,t)$, $\mathbb{E}[X(t)] = \mu t$, $D^{(1)} = \frac{\partial \mathbb{E}[X(t)]}{\partial t} = \mu$. $\mathbb{V}[X(t)] = 2Dt$, $D^{(2)} = \frac{1}{2}\frac{\partial \mathbb{V}[X(t)]}{\partial t} = D$.

2.2 Problem statement & Assumptions

Figure 1 provides a schematic overview of the proposed SROM method. First, the ROM needs to be trained, where the drift and diffusion coefficients are calibrated using the predicted QoIs from ICME models of materials. After test and validation with additional materials simulation data, the SROM then can predict the uncertainty propagation efficiently with a much longer time step than the time steps used in the ICME models.

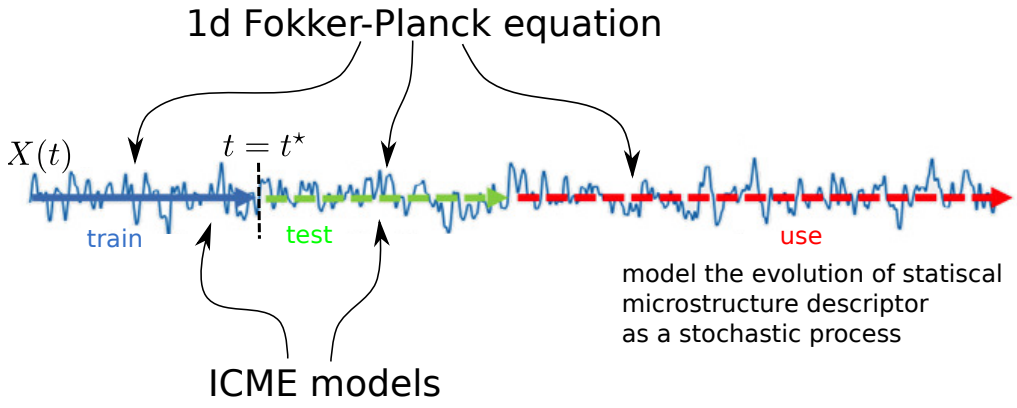


Fig. 1: The SROM models $X(t)$, which is a statistical microstructure descriptor and QoI, as a stochastic process with Langevin equation. The SROM is trained using the first part of the time-series dataset, i.e. the training dataset. After the drift and diffusion coefficients are trained, the trained ROM is validated using the second part of the time-series dataset, i.e. the testing dataset. After the SROM is trained and validated, it can be deployed to predict the pdf of QoI in the future.

Some notable work to estimate the drift-diffusion parameters from experimental and computational time-series dataset include noisy electrical circuit [9], stochastic dynamics of metal cutting [18, 15, 14], meteorological data for sea surface wind [28, 10] to name a few. Interested readers are referred to the review paper of Friedrich et al. [7] for extensive multi-disciplinary applications across many scientific and engineering fields. Here, we applied and extended the method into the field of computational materials science with applications to microstructure evolution. We assume that the noise associated with QoIs are δ-correlated, in order to preserve the Markov property of the Langevin model. Furthermore, the noise is also assumed to be independent of the QoI $X(t)$. In practice, the noise is not strictly δ-correlated, which results in Markov-Einstein time τ_{ME}, such that for sampling intervals $\tau < \tau_{ME}$, the Markov property does not hold [17]. However, there has been proofs that the Markov assumption is a valid assumption, for example, in the field of fluid mechanics with small-scale turbulence [25]. Some recent efforts are also noted in adopting Fokker-Planck equation for non-Markovian process [11, 12]. Mousavi et al.[20] and Anvari et al. [1] discussed in great details regarding the Markov-Einstein time scale τ_M – the minimum time interval over which the time series $X(t)$ can be considered as a Markov process. Friedrich et al [7] (cf. Section 2.2.5) also pointed out that the Markovian property is "usually violated by many physical systems". Furthermore, we assume that the drift and diffusion coefficients are temporal (but not spatial) functions, even though both time-independent [22] and time-dependent coefficients [19] have been studied in the literature. Also, in this paper, we restrict the QoI $X(t)$ to be 1-dimensional.

3 Normal grain growth with hybrid Potts phase-field in kinetic Monte Carlo

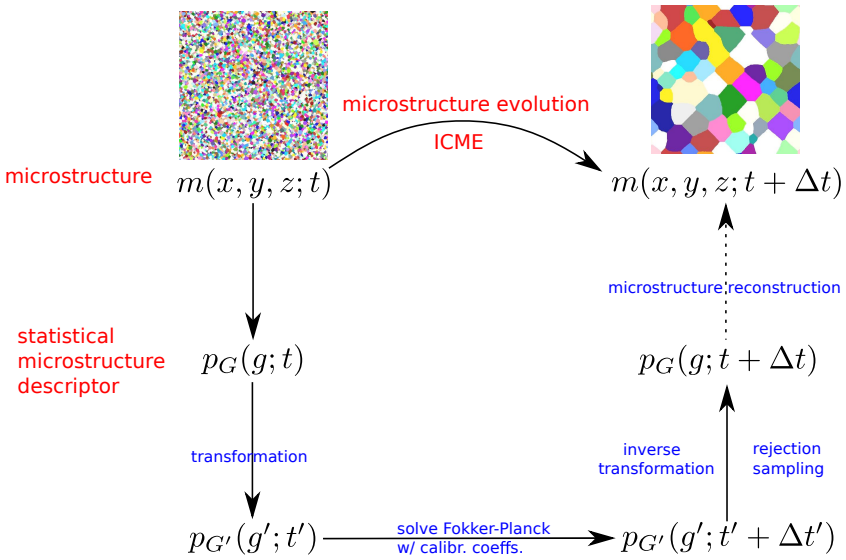


Fig. 2: Conceptual map for applying SROM to kMC. First, a training/testing SPPARKS dataset is generated. From this dataset, the grain size probability density distribution as a statistical microstructure descriptor is extracted. A truncated time-step $t^\star$ is imposed to separate training and testing datasets (cf. Figure 1). A log-log transformation is applied to both the grain size as well as the time-step. Using the (transformed) training dataset, we calibrate the drift and the diffusion coefficients in the Fokker-Planck equation, which is the SROM. With the initial probability density of the log of grain size at $\log(t^\star)$, we then solve the Fokker-Planck to forecast the probability of the log of grain size in the future. An inverse transformation is applied to return to the grain size and the original time domain, followed by rejection sampling to estimate the probability density of the grain size in time $t > t^\star$. We then compare the grain size distribution between the Fokker-Planck and the SPPARKS model. Note that using this predicted grain size distribution, we could reconstruct an ensemble of statistically equivalent microstructures. The Fokker-Planck model predictions agree very well with the SPPARKS kinetic Monte Carlo simulations.

In this example, the hybrid Potts-phase field model from Homer et al. [16] based on kinetic Monte Carlo SPPARKS framework [24] is used to investigate the evolution of the grain area during the grain growth. The hybrid Potts-phase field model is applied on a

simple two-component, two-phase system, where the bulk free energy of the system is described as [16]

$$E_v(q, C) = \lambda[(C - C_1)^2 + (C_2 - C)^2] + a(C - C_3)^2 q_\alpha + \alpha(C_4 - C)^2 q_\beta, \tag{17}$$

where $\lambda = 0.3, C_1 = 0.25, C_2 = 0.75, C_3 = 0.05, C_4 = 0.95$, and $\alpha = 0.5$. A computational domain of 5000 pixel $\times$ 5000 pixel is used to perform the 2D grain growth kMC simulation. 50 kMC simulations are performed for 20,000 Monte Carlo steps (mcs), where 40 Monte Carlo events are observed, with the last microstructure is obtained at 16,681.1 mcs. Time-step in kMC is an exponentially distributed random variable [38], which is inversely proportional with the rate constant and typically generated from the uniformly distributed seed.

Exploiting the fact that the log-normal distribution is one of the most commonly used distributions to characterize the grain size, as well as the fact that in the kinetic Monte Carlo method, time-step is an exponentially distributed random variable, we apply the log transformation to both grain size and time, i.e. $X \to \log X$, $t \to \log t$, before modeling the QoI (i.e. X after transformation) using the Fokker-Planck equation.

The drift and diffusion coefficients are calibrated by simple linear regression, which minimizes the ℓ^2 error, based on Corollaries 1 and 2. The initial and training pdfs are constructed using the kernel density estimation method with the normal kernel distribution. The selected bandwidth is optimal for the normal kernel density [2]. The Tikhonov regularization is applied to the initial pdf to reduce the chance of numerical divergent for the Fokker-Planck solver.

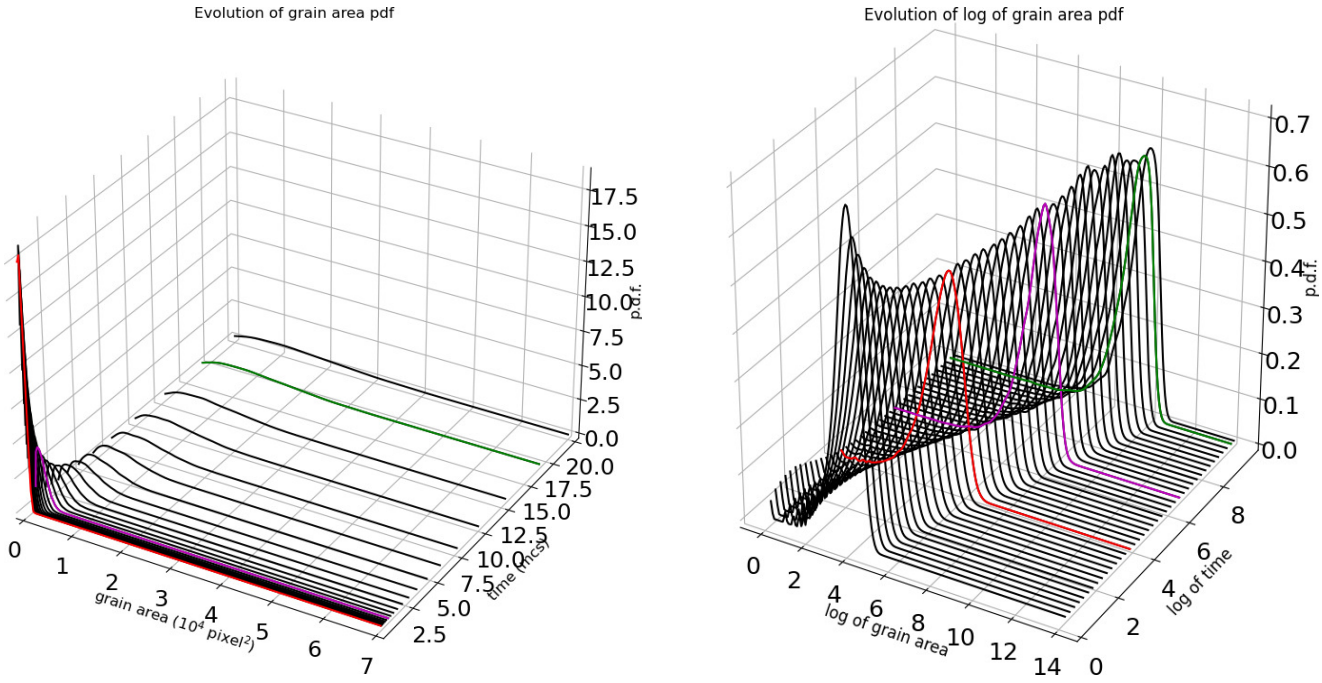
The Monte Carlo events from 46.5 mcs to 599.5 mcs are used as the training dataset, while the Monte Carlo events from 744.375 mcs to 16681.1 mcs are used as the testing dataset. Specifically, the training dataset includes snapshots at $t \in \{46.5, 60., 77.5, 100., 129.25, 166.875, 215.5, 278.375, 359.5, 464.25, 599.5\}$ mcs. The testing dataset includes snapshots at $t \in \{774.375, 1000., 1291.62, 1668.12, 2154.5, 2782.62, 3593.88, 4641.62, 5994.88, 7742.75, 10000., 12915.5, 16681.1\}$ mcs. The time $t^\star = 599.5$ mcs is used to split the training and testing datasets. By simple linear regression, we obtained the constant drift and diffusion coefficients, respectively, as $D^{(1)}(t) = 0.7320$, whereas $D^{(2)}(t) = -0.02931$.

Figure 3a shows the evolution of grain area pdfs before transformation, while Figure 3b shows the evolution of grain area pdfs after transformation. Figure 4 shows the solution of the Fokker-Planck equation after the log transformation at three different snapshots: Figure 4a at the beginning of the training dataset, Figure 4b at the end of the training dataset, and Figure 4c at the end of the testing dataset. Excellent agreement with the testing dataset is obtained.

From the solution of the Fokker-Planck equation shown in Figure 4, we apply rejection sampling algorithm to draw samples and reconstruct the pdf of the grain size, by inverting the log transformation, i.e. $X \to \exp(X)$. Figure 5 shows the comparison between the reconstructed pdfs from Fokker-Planck solution and the original pdfs from SPPARKS. It is observed that even though the Fokker-Planck solution has a longer tail distribution, in general, the last testing pdf at 16,681.1 mcs agrees relatively well with the reconstructed pdf from the Fokker-Planck solution. This demonstrates if the Fokker-Planck coefficients are well-trained, a prediction about the evolution of the microstructural descriptor using the trained ROM can be made with a good level of accuracy.

4 Discussions & Conclusion

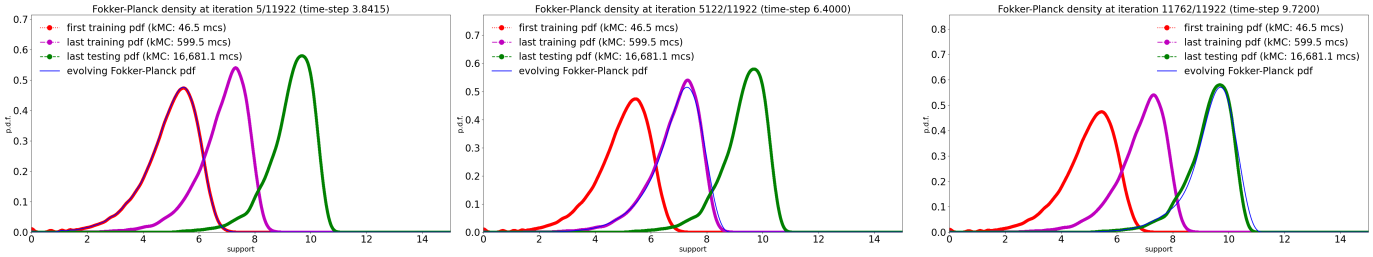
In this paper, we propose and successfully demonstrate the application of a SROM, formulated by Fokker-Planck equation, to a normal grain growth with kinetic Monte Carlo simulations using SPPARKS. Specifically, we demonstrate that extrapolating the grain size pdf in long time agrees very well with the SPPARKS simulations. The proposed SROM has also been applied to molecular dynamics and phase-field simulation, but SPPARKS remains the most successful application so far [37]. It is noteworthy that even though in this particular example, the simple linear regression is used to calibrate the drift and diffusion coefficients in ℓ^2 loss function, a more generalized black-box optimization approach, e.g. Bayesian optimization method [31, 36, 35, 32, 33, 34], could be used for a more generalized loss function. Analytically, the solutions to some microstructure descriptors are well-known in the field of materials science. In the case study of kinetic Monte Carlo for norml grain-growth problem, it is known that the average grain size (isotropic grain boundary energies and mobilities) grows as $t^{1/2}$, as described in Ng [21] and Breithaupt et al[3]. Interested readers are referred to the works of Friedrich et al. [7, 17, 27, 9, 8] for various applications of stochastic method in time-series and estimation the drift and diffusion coefficients of Fokker-Planck equations in low- and high-sampling rates. It remains questionable to us as to how the $\log t$ transformation yields a uniformly distributed random variable, even though mathematically, it is that $e^{-\Delta t} = re^{1/k}$ should be uniformly distributed, because $r \sim \mathcal{U}[0, 1]$, k is the rate constant [38], and Δt is supposed to be an exponentially distributed random variable. The issue remains an open question for future study.



(a) Evolution of grain area (before transformation) pdf in kMC simulation shows a diffusion-dominant type of Fokker-Planck equation.

(b) Evolution of log grain area (after transformation) pdf in kMC simulation shows a typical Fokker-Planck equation.

Fig. 3: Evolution of grain area pdf. before (Figure 3a) and after (Figure 3b) transformation.



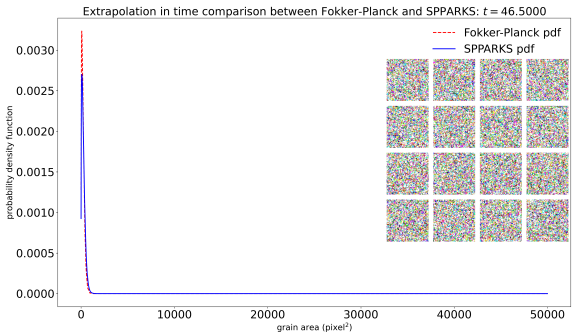
(a) 46.5 mcs

(b) 599.5 mcs

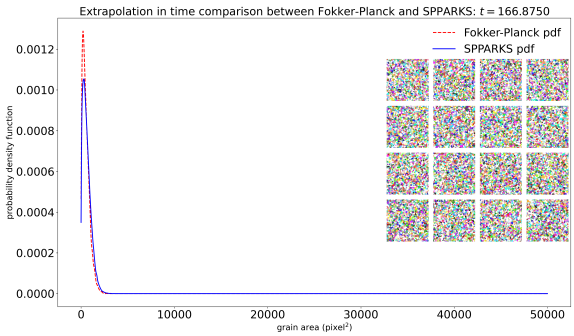
(c) 16,681.1 mcs

Fig. 4: Evolution of log of grain area distribution by (a) kinetic Monte Carlo simulations and (b) Fokker-Planck equation with calibrated coefficients at three different snapshots: beginning of training (Figure 4a), end of training and beginning of testing (Figure 4b), and end of testing (Figure 4c).

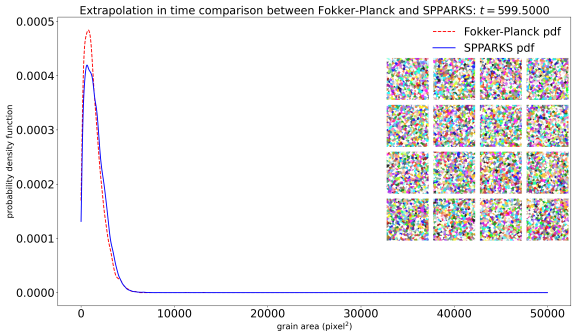
This manuscript provides a numerical evidence that the grain growth in 2d, through SPPARKS, are perhaps related to a geometric Brownian motion, where the logarithm of the grain can be modeled by a Brownian motion or Wiener process under Itô interpretation. The relationship between Langevin equation and Anderson-Kubo process, Ornstein-Uhlenbeck process, Black-Scholes process or geometric Brownian motion is fairly known (cf. Chapter 8 [30]), yet little has been brought into the materials science community. Practically, it validates the numerical implementation of SPPARKS, which agrees with the theoretical mathematical modeling results that has been long established in the literature [21, 3]. For those who wonder what this manuscript is related to data assimilation, there is a direct connection between ensemble Fokker-Planck equation and ensemble Kalman filter; more specifically, ensemble Kalman filter applies the sequential Markov chain Monte Carlo to solve the Fokker-Planck equation [5, 4], by viewing the dynamical model as a stochastic differential equation with Itô calculus [5], as done in this study.



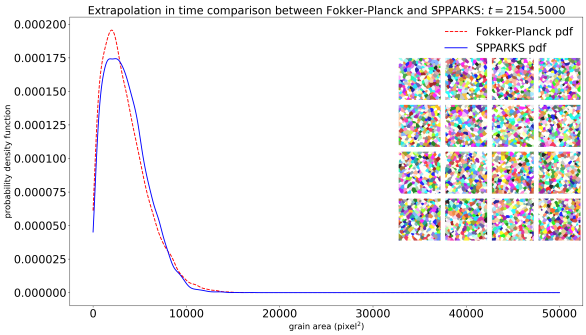
(a) 46.5 mcs



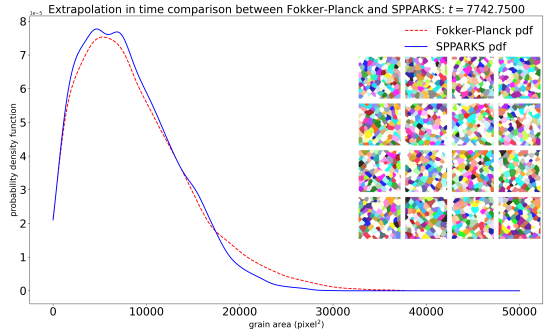
(b) 166.875 mcs



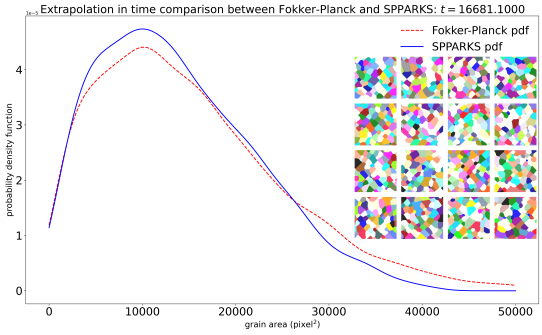
(c) 599.5 mcs



(d) 2154.5 mcs



(e) 7742.75 mcs



(f) 16,681.1 mcs

Fig. 5: Evolution of grain area distribution reconstructed by rejection sampling algorithm from the Fokker-Planck solutions.

Acknowledgment

This article has been authored by an employee of National Technology & Engineering Solutions of Sandia, LLC under Contract No. DE-NA0003525 with the U.S. Department of Energy (DOE). The employee owns all right, title and interest in and to the article and is solely responsible for its contents. The United States Government retains and the publisher, by accepting the article for publication, acknowledges that the United States Government retains a non-exclusive, paid-up, irrevocable, world-wide license to publish or reproduce the published form of this article or allow others to do so, for United States Government purposes. The DOE will provide public access to these results of federally sponsored research in accordance with the DOE Public Access Plan https://www.energy.gov/downloads/doe-public-access-plan. This paper describes objective technical results and analysis. Any

subjective views or opinions that might be expressed in the paper do not necessarily represent the views of the U.S. Department of Energy or the United States Government.

References

1. Anvari, M., Tabar, M., Peinke, J., Lehnertz, K.: Disentangling the stochastic behavior of complex time series. Scientific reports **6**(1), 1–12 (2016)
2. Bowman, A.W., Azzalini, A.: Applied smoothing techniques for data analysis: the kernel approach with S-Plus illustrations, vol. 18. OUP Oxford (1997)
3. Breithaupt, T., Hansen, L.N., Toppaladoddi, S., Katz, R.F.: The role of grain-environment heterogeneity in normal grain growth: A stochastic approach. Acta Materialia **209**, 116699 (2021)
4. Evensen, G.: Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. Journal of Geophysical Research: Oceans **99**(C5), 10143–10162 (1994)
5. Evensen, G.: The ensemble Kalman filter: Theoretical formulation and practical implementation. Ocean dynamics **53**(4), 343–367 (2003)
6. Frank, T.D.: Nonlinear Fokker-Planck equations: fundamentals and applications. Springer Science & Business Media (2005)
7. Friedrich, R., Peinke, J., Sahimi, M., Tabar, M.R.R.: Approaching complexity by stochastic methods: From biological systems to turbulence. Physics Reports **506**(5), 87–162 (2011)
8. Friedrich, R., Renner, C., Siefert, M., Peinke, J.: Comment on "Indispensable finite time corrections for Fokker-Planck equations from time series data". Physical Review Letters **89**(14), 149401 (2002)
9. Friedrich, R., Siegert, S., Peinke, J., Siefert, M., Lindemann, M., Raethjen, J., Deuschl, G., Pfister, G., et al.: Extracting model equations from experimental data. Physics Letters A **271**(3), 217–222 (2000)
10. Gille, S.T.: Statistical characterization of zonal and meridional ocean wind stress. Journal of Atmospheric and Oceanic Technology **22**(9), 1353–1372 (2005)
11. Giuggioli, L., McKetterick, T.J., Kenkre, V., Chase, M.: Fokker–Planck description for a linear delayed Langevin equation with additive Gaussian noise. Journal of Physics A: Mathematical and Theoretical **49**(38), 384002 (2016)
12. Giuggioli, L., Neu, Z.: Fokker–Planck representations of non-Markov Langevin equations: application to delayed systems. Philosophical Transactions of the Royal Society A **377**(2153), 20180131 (2019)
13. Gottschall, J., Peinke, J.: On the definition and handling of different drift and diffusion estimates. New Journal of Physics **10**(8), 083034 (2008)
14. Gradišek, J., Govekar, E., Grabec, I.: Qualitative and quantitative analysis of stochastic processes based on measured data, II: Applications to experimental data. Journal of sound and vibration **252**(3), 563–572 (2002)
15. Gradišek, J., Grabec, I., Siegert, S., Friedrich, R.: Qualitative and quantitative analysis of stochastic processes based on measured data, I: Theory and applications to synthetic data. Journal of sound and vibration **252**(3), 545–562 (2002)
16. Homer, E.R., Tikare, V., Holm, E.A.: Hybrid Potts-phase field model for coupled microstructural–compositional evolution. Computational Materials Science **69**, 414–423 (2013)
17. Honisch, C., Friedrich, R.: Estimation of Kramers-Moyal coefficients at low sampling rates. Physical Review E **83**(6), 066701 (2011)
18. Kleinhans, D., Friedrich, R., Nawroth, A., Peinke, J.: An iterative procedure for the estimation of drift and diffusion coefficients of Langevin processes. Physics Letters A **346**(1-3), 42–46 (2005)
19. Lin, W.T., Ho, C.L.: Similarity solutions of the Fokker–Planck equation with time-dependent coefficients. Annals of Physics **327**(2), 386–397 (2012)
20. Mousavi, S., Reihani, S., Anvari, G., Anvari, M., Alinezhad, H., Tabar, M.: Stochastic analysis of time series for the spatial positions of particles trapped in optical tweezers. Scientific Reports **7**(1), 1–11 (2017)
21. Ng, F.S.: Statistical mechanics of normal grain growth in one dimension: A partial integro-differential equation model. Acta Materialia **120**, 453–462 (2016)
22. Pesce, G., McDaniel, A., Hottovy, S., Wehr, J., Volpe, G.: Stratonovich-to-Itô transition in noisy systems with multiplicative feedback. Nature communications **4**(1), 1–7 (2013)
23. Plimpton, S., Thompson, A., Slepoy, A.: Stochastic parallel particle kinetic simulator. Tech. rep., Sandia National Lab.(SNL-NM), Albuquerque, NM (United States) (2008)
24. Plimpton, S., Thompson, A., Slepoy, A.: SPPARKS kinetic monte carlo simulator (2012)
25. Renner, C., Peinke, J., Friedrich, R.: Experimental indications for Markov properties of small-scale turbulence. Journal of Fluid Mechanics **433**, 383–409 (2001)
26. Risken, H.: The Fokker Planck equation, Methods of solution and application 2nd Ed. Springer Verlag, Berlin, Heidelberg (1989)
27. Siefert, M., Kittel, A., Friedrich, R., Peinke, J.: On a quantitative method to analyze dynamical and measurement noise. EPL (Europhysics Letters) **61**(4), 466 (2003)
28. Sura, P., Gille, S.T.: Interpreting wind-driven Southern Ocean variability in a stochastic framework. Journal of marine research **61**(3), 313–334 (2003)
29. Tabar, M.R.R.: The Langevin Equation and Wiener Process, pp. 39–48. Springer International Publishing, Cham (2019). DOI 10.1007/978-3-030-18472-8_5. URL https://doi.org/10.1007/978-3-030-18472-8_5
30. Tabar, R.: Analysis and data-based reconstruction of complex nonlinear dynamical systems, vol. 730. Springer (2019)
31. Tran, A., Eldred, M., Wildey, T., McCann, S., Sun, J., Visintainer, R.J.: aphBO-2GP-3B: a budgeted asynchronous parallel multi-acquisition functions for constrained Bayesian optimization on high-performing computing architecture. Structural and Multidisciplinary Optimization **65**(4), 1–45 (2022)
32. Tran, A., Mitchell, J.A., Swiler, L.P., Wildey, T.: An active-learning high-throughput microstructure calibration framework for process-structure linkage in materials informatics. Acta Materialia **194**, 80–92 (2020)

33. Tran, A., Sun, J., Furlan, J.M., Pagalthivarthi, K.V., Visintainer, R.J., Wang, Y.: pBO-2GP-3B: A batch parallel known/unknown constrained Bayesian optimization with feasibility classification and its applications in computational fluid dynamics. Computer Methods in Applied Mechanics and Engineering **347**, 827–852 (2019)
34. Tran, A., Tran, M., Wang, Y.: Constrained mixed-integer Gaussian mixture Bayesian optimization and its applications in designing fractal and auxetic metamaterials. Structural and Multidisciplinary Optimization **59**, 2131–2154 (2019)
35. Tran, A., Tranchida, J., Wildey, T., Thompson, A.P.: Multi-fidelity machine-learning with uncertainty quantification and Bayesian optimization for materials design: Application to ternary random alloys. The Journal of Chemical Physics **153**, 074705 (2020)
36. Tran, A., Wildey, T., McCann, S.: sMF-BO-2CoGP: A sequential multi-fidelity constrained Bayesian optimization for design applications. Journal of Computing and Information Science in Engineering **20**(3), 1–15 (2020)
37. Tran, A., Wildey, T., Sun, J., Liu, D., Wang, Y.: A stochastic reduced-order model for statistical microstructure descriptors evolution. Journal of Computing and Information Science in Engineering pp. 1–18 (2022)
38. Voter, A.F.: Introduction to the kinetic Monte Carlo method. In: Radiation Effects in Solids, pp. 1–23. Springer (2007)

Proofs

Here, we present a short proof for Corollary 1. Assume vanishing boundary conditions at an exponential rate of the pdf, i.e. $f(X,t) \propto e^{-X^2} \Rightarrow \lim_{X\to\pm\infty} f(X,t) = 0$.

Proof Here, we integrate by part and utilize the vanishing boundary conditions of the pdf $f(X,t) \to 0$ as $X \to \pm\infty$.

$$\begin{aligned}
\frac{\partial}{\partial t}\mathbb{E}[X(t)] &= \frac{\partial}{\partial t}\int_{-\infty}^{\infty} X f(X,t) dX = \int_{-\infty}^{\infty} X\frac{\partial f}{\partial t} dX \\
&= \int_{-\infty}^{\infty}\left[-X\frac{\partial}{\partial X}\left(D^{(1)}f\right) + X\frac{\partial^2}{\partial X^2}\left(D^{(2)}f\right)\right] dX \\
&= -\int_{-\infty}^{\infty} X\frac{\partial}{\partial X}\left(D^{(1)}f\right) dX + \int_{-\infty}^{\infty} X\frac{\partial^2}{\partial X^2}\left(D^{(2)}f\right) dX \\
&= -[XD^{(1)}f]_{-\infty}^{\infty} + \int_{-\infty}^{\infty} D^{(1)}f dX + [X\frac{\partial}{\partial X}(D^{(2)}f)]_{-\infty}^{\infty} - \int_{-\infty}^{\infty}\frac{\partial}{\partial X}(D^{(2)}f) dX \\
&= -[XD^{(1)}f]_{-\infty}^{\infty} + \int_{-\infty}^{\infty} D^{(1)}f dX + [X\frac{\partial}{\partial X}(D^{(2)}f)]_{-\infty}^{\infty} - [(D^{(2)}f]_{-\infty}^{\infty} \\
&= \int_{-\infty}^{\infty} D^{(1)}(X,t) f(X,t) dX
\end{aligned} \tag{18}$$

If the drift coefficient is a temporal function $D^{(1)}(X,t) = D^{(1)}(t)$, then the last equation simplifies to

$$\frac{\partial}{\partial t}\mathbb{E}[X(t)] = D^{(1)}(t), \tag{19}$$

as $\int_{-\infty}^{\infty} f(X,t) dX = 1$. □

Here, we present a short proof for Corollary 2 by integrating by part and utilizing the vanishing boundary conditions of $f(X,t)$.

Proof First, consider

$$
\begin{aligned}
\frac{\partial}{\partial t}\mathbb{E}[X(t)^2] &= \frac{\partial}{\partial t}\int_{-\infty}^{\infty} X^2 f(X,t)dX = \int_{-\infty}^{\infty} X^2 \frac{\partial f}{\partial t}dX = \int_{-\infty}^{\infty} X^2 \left[-\frac{\partial}{\partial X}\left(D^{(1)}f\right) + \frac{\partial^2}{\partial X^2}\left(D^{(2)}f\right)\right]dX \\
&= \int_{-\infty}^{\infty} X^2 \left[-\frac{\partial}{\partial X}\left(D^{(1)}f\right) + \frac{\partial^2}{\partial X^2}\left(D^{(2)}f\right)\right]dX \\
&= \int_{-\infty}^{\infty} -X^2 \frac{\partial}{\partial X}\left(D^{(1)}f\right)dX + \int_{-\infty}^{\infty} X^2 \frac{\partial^2}{\partial X^2}\left(D^{(2)}f\right)dX \\
&= -[X^2 D^{(1)}f]_{-\infty}^{\infty} + 2\int_{-\infty}^{\infty} XD^{(1)}f dX + [X^2 \frac{\partial}{\partial X}(D^{(2)}f)]_{-\infty}^{\infty} - 2\int_{-\infty}^{\infty} X\frac{\partial}{\partial X}(D^{(2)}f)dX \\
&= -[X^2 D^{(1)}f]_{-\infty}^{\infty} + 2\int_{-\infty}^{\infty} XD^{(1)}f dX \\
&\quad + [X^2 \frac{\partial}{\partial X}(D^{(2)}f)]_{-\infty}^{\infty} - 2[XD^{(2)}f]_{-\infty}^{\infty} + 2\int_{-\infty}^{\infty} D^{(2)}f dX \\
&= 2\int_{-\infty}^{\infty} XD^{(1)}f dX + 2\int_{-\infty}^{\infty} D^{(2)}(X,t)f(X,t)dX
\end{aligned}
\tag{20}
$$

Observe that if the drift coefficient is a temporal function $D^{(1)}(X,t) = D^{(1)}(t)$, then by Corollary 1, $\frac{\partial}{\partial t}\mathbb{E}[X(t)] = D^{(1)}(t)$. Thus,

$$
\begin{aligned}
2\int_{-\infty}^{\infty} XD^{(1)}f dX &= 2\int_{-\infty}^{\infty} XD^{(1)}(t)f(X,t)dX = 2D^{(1)}(t)\int_{-\infty}^{\infty} Xf(X,t)dX \\
&= 2D^{(1)}(t)\int_{-\infty}^{\infty} Xf(X,t)dX = 2D^{(1)}(t)\mathbb{E}[X(t)] \\
&= 2\frac{\partial \mathbb{E}[X(t)]}{\partial t}\mathbb{E}[X(t)].
\end{aligned}
\tag{21}
$$

If the diffusion coefficient is also a temporal function $D^{(2)}(X,t) = D^{(2)}(t)$, then Equation 20 becomes

$$
\frac{\partial}{\partial t}\mathbb{E}[X^2(t)] = 2\frac{\partial \mathbb{E}[X(t)]}{\partial t}\mathbb{E}[X(t)] + 2D^{(2)}(t). \tag{22}
$$

After a few algebraic manipulation, we obtain

$$
\frac{\partial}{\partial t}\mathbb{V}[X(t)] = \frac{\partial}{\partial t}(\mathbb{E}[X(t)^2] - [\mathbb{E}[X(t)]]^2) = \frac{\partial}{\partial t}\mathbb{E}[X(t)^2] - 2\frac{\partial \mathbb{E}[X(t)]}{\partial t}\mathbb{E}[X(t)] = 2D^{(2)}(t). \tag{23}
$$