

A STOCHASTIC MULTISCALE COUPLING SCHEME TO ACCOUNT FOR SAMPLING NOISE IN ATOMISTIC-TO-CONTINUUM SIMULATIONS*

MAHER SALLOUM[†], KHACHIK SARGSYAN[‡], REESE JONES[§], BERT DEBUSSCHERE[‡],
HABIB N. NAJM[‡], AND HELGI ADALSTEINSSON[¶]

Abstract. We present a methodology to assess the predictive fidelity of multiscale simulations by incorporating uncertainty in the information exchanged between the atomistic and continuum simulation components. Focusing on uncertainty due to finite sampling in molecular dynamics (MD) simulations, we present an iterative stochastic coupling algorithm that relies on Bayesian inference to build polynomial chaos expansions for the variables exchanged across the atomistic-continuum interface. We consider a simple Couette flow model where velocities are exchanged between the atomistic and continuum components. To alleviate the burden of running expensive MD simulations at every iteration, a surrogate model is constructed from which samples can be efficiently drawn as data for the Bayesian inference. Results show convergence of the coupling algorithm at a reasonable number of iterations. The uncertainty associated with the exchanged variables significantly depends on the amount of data sampled from the MD simulations and on the width of the time averaging window used in the MD simulations. Sequential Bayesian updating is also implemented in order to enhance the accuracy of the stochastic algorithm predictions.

Key words. finite sampling, Bayesian inference, coupling, molecular dynamics, polynomial chaos expansion

AMS subject classifications. 33C45, 60G99, 62F15, 65C20, 76D99, 74A25, 62P30

DOI. 10.1137/110844404

1. Introduction. Over the past two decades, there has been growing interest in modeling physical systems with phenomena that operate over a range of length and time scales. These are often encountered in fluid-solid interactions commonly found in nanodevices and microfluid systems [2, 38, 45]. To understand these phenomena, atomistic or molecular dynamics (MD) simulations have to be performed in conjunction with continuum macroscale simulations. Hence, such multiscale simulations involve two or more model components operating at different scales. In the current study, we focus on problems that involve disparity in their length scales that range between the atomistic and continuum scales.

1.1. Motivation. Consider a general atomistic-continuum multiscale problem, schematically represented in Figure 1, in which various flows of information can be

*Received by the editors August 15, 2011; accepted for publication (in revised form) February 29, 2012; published electronically May 22, 2012. This work was supported by the U.S. Department of Energy Office of Science through the Applied Mathematics program in the Office of Advanced Scientific Computing Research (ASCR). Sandia National Laboratories is a multiprogram laboratory managed and operated by Sandia Corporation, a wholly owned subsidiary of Lockheed Martin Corporation, for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-AC04-94AL85000. The U.S. Government retains a nonexclusive, royalty-free license to publish or reproduce the published form of this contribution, or allow others to do so, for U.S. Government purposes. Copyright is owned by SIAM to the extent not limited by these rights.

<http://www.siam.org/journals/mms/10-2/84440.html>

[†]Sandia National Laboratories, MS 9158, Livermore, CA 94550 (mnsallo@sandia.gov).

[‡]Sandia National Laboratories, MS 9051, Livermore, CA 94550 (ksargsy@sandia.gov, bjdebus@sandia.gov, hnnaajm@sandia.gov).

[§]Sandia National Laboratories, MS 9404, Livermore, CA 94550 (rjones@sandia.gov).

[¶]Sandia National Laboratories, MS 9158, Livermore, CA 94550 (hadalst@sandia.gov). Current address: Google, Mountain View, CA 94043.

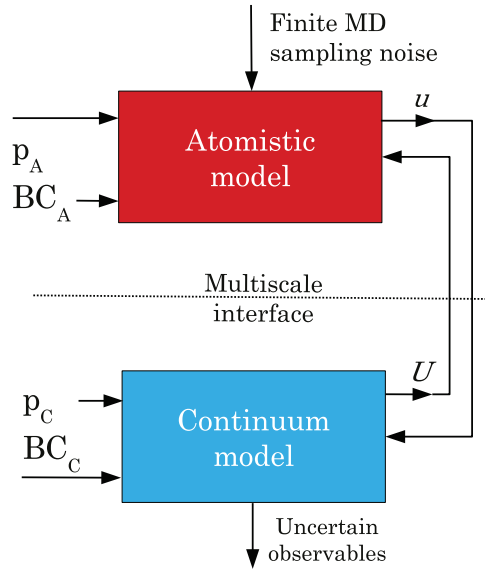


FIG. 1. Information flow in a general stochastic atomistic-continuum multiscale simulation. Inputs to the atomistic simulation are parameters P_A , boundary conditions BC_A , and a variable U extracted from the continuum simulation. Inputs to the continuum simulation are parameters P_C , boundary conditions BC_C , and a variable u extracted from the atomistic simulation. Each of the input parameters and boundary conditions can be uncertain, resulting in uncertainties in the observables that are extracted from the continuum simulation.

distinguished. The main components in the multiscale simulation framework are the continuum and atomistic models, with input parameters P_C and P_A , respectively, as well as externally imposed boundary conditions BC_C and BC_A . Each of these inputs can be uncertain. Macroscale observables extracted from the MD simulation through averaging generally have intrinsic noise due to the finite amount of MD samples available, given the high computational cost of MD simulations. The atomistic and continuum models exchange information across a multiscale interface illustrated as the dotted line in Figure 1. This interface is commonly referred to as the overlap or handshake region [13, 33, 35, 56]. The exchanged information is represented in terms of two variables u and U as shown in the figure. For the present discussion, we assume that relevant macroscale observables are to be extracted from the continuum model. The uncertainty in the parameters and boundary conditions as well as the intrinsic noise in the atomistic model contribute to the uncertainty of these macroscale observables.

With all of the above-mentioned sources of uncertainty, the problem statement addressed is the following: *based on all inputs into the simulation, what is the resulting uncertainty in the predicted values of the macroscale observables extracted from the continuum simulation?* In order to answer this question, it is crucial to first quantify the uncertainty in the coupling variables between the atomistic and continuum simulations. Hence, the objective of the present work is to calculate the coupling variables u and U and their associated uncertainty. The variable u extracted from the MD simulation is fed to the continuum component of the system while U is imposed on the MD simulation, similarly to the Schwartz alternating method [48], until convergence is reached.

1.2. Literature review. Mathematically, the information exchanged between the different system scales is passed in terms of boundary conditions [45]. O’Connell and Thompson were the first to study such a coupling procedure [42]. They considered a nanochannel flow decomposed into three main regions: atomistic, continuum, and overlap regions. The continuum region extends until the point where the Navier–Stokes equations are no longer applicable. The remaining part of the domain is handled by an MD simulation. An overlap region is responsible for exchanging the information between the continuum and atomistic regions. Similar approaches were used in subsequent studies by Hadjiconstantinou [21], Hadjiconstantinou and Patera [22], Li, Liao, and Yip [33, 34, 35], Delgado-Buscalioni and Coveney [10], Kotsalis, Walther, and Koumoutsakos [27, 28], Yasuda and Yamamoto [56], and Donev et al. [13]. For instance, Li, Liao, and Yip used the statistical maximum likelihood inference method to estimate macroscale field variables such as density, temperature, and velocity from MD simulation data while minimizing the disturbance to particle dynamics upon information exchange. In a series of papers, Alexander, Garcia, and Tartakovsky [3, 4] developed a hybrid particle-continuum algorithm for hydrodynamic problems. The algorithm is focused on a diffusion partial differential equation for the continuum model that exchanges fluctuations with a random-walk particle model.

In almost all these studies, the coupling was performed near a fluid-solid interface where the wall velocity was the main variable of interest. A common challenge encountered in these studies was the choice of the coupling variables and boundary conditions. While in some studies the velocity at the multiscale interface was both an input to and output of the MD simulation, in other studies this velocity was imposed on the continuum model and the corresponding flux (shear stress) was used as a boundary condition to the MD model. The latter technique is not straightforward since shear stress is not a fundamental quantity in MD. An attempt to impose shear stress in an MD simulation was performed by Ren [45], who distributed this stress as forces among the particles close to the wall following a predetermined function. This function assigns the amplitude of the exerted force depending on the distance of the particle to the wall. The work of Ren [45] also addressed the stability of various coupling schemes and showed the importance of choosing the coupling variables.

All of the aforementioned studies were successful in effectively exchanging the appropriate information between a continuum and an atomistic simulation. However, these studies did not incorporate uncertainty in the MD simulation parameters. Also, for situations where there is a strong separation of time and/or length scales between the continuum and particle simulations, studies in the literature did not address uncertainty due to sampling noise in a general coupled setting. This noise level is, in a central limit sense, inversely proportional to the number of particles present in the simulation and to the averaging time window width. Completely eliminating the noise through averaging over samples is not practically feasible, as it requires an infinite simulation time for a given number of particles. In this context, Hadjiconstantinou et al. [23] derived analytical expressions for the noise amplitude in stand-alone particle simulations as a function of the number of particles and the Knudsen and Mach numbers. These expressions are useful in cases of hydrodynamics phenomena where the Knudsen, Mach, and acoustic numbers are known. Assessing the uncertainties is more challenging, however, when atomistic and continuum simulations are coupled in a way that sampling noise in the particle simulation outputs also leads to uncertainties in the particle simulation inputs, through the coupling with the continuum layer. In this scenario, both the effects of uncertain inputs and sampling noise on the particle simulations need to be quantified. Hence, a general method to quantify the

uncertainty that corresponds to the intrinsic finite-sampling noise in the atomistic-continuum coupling is a crucial and, at the same time, a challenging step towards an overall uncertainty analysis of multiscale model predictions. The development of such a method is the focus of the current paper.

1.3. Approach. A flexible and robust methodology is needed to properly characterize and propagate all relevant sources of uncertainty through the full multiscale simulation. The different sources of uncertainty shown in Figure 1 can be grouped into two major categories. The first is intrinsic to the atomistic part of the simulation where finite sampling is required, resulting in a noisy output from MD simulations. The second type encompasses the external uncertainty in the parameters and/or boundary conditions in both the continuum and the atomistic parts of the simulation. In this study, we focus on the uncertainty due to the intrinsic noise in MD simulations and assume separation of length and/or time scales between the atomistic and continuum models. This implies that the continuum formulation does not resolve any molecular fluctuations, and the uncertainty in the coupling variables represents the lack of knowledge due to finite sampling for the extraction of the macroscale quantities from the particle simulations, as opposed to previous studies where the molecular fluctuations themselves have been exchanged between particle and continuum simulations [3, 4, 13]. For this purpose, we express the variables u and U (see Figure 1) exchanged between the atomistic and continuum models in terms of polynomial chaos expansions (PCEs) [19, 40, 52]. This methodology offers the flexibility of accounting for parametric uncertainty; however, that will be the subject of future work. PCEs have been extensively used in the literature to represent and propagate uncertainties in many physical systems [8, 9, 17, 18, 29, 30, 31, 53, 54, 55]. Obtaining a PCE based on sampled data is commonly performed using numerical integration employing quadrature rules [44]. However, the data derived from atomistic simulations is noisy due to the finite sampling in MD simulations. Using quadrature rules to derive a PCE in the case of a noisy function is not a straightforward exercise and can be problematic due to the lack or absence of convergence [26]. Among several other possibilities for obtaining PCEs from noisy data, we argue for and implement Bayesian inference [49] since it is suitable for handling systems with various sources of uncertainty and is robust with respect to noisy data.

Based on the above, we develop in this paper a method to quantify uncertainty from an atomistic simulation using Bayesian inference. Further, focusing on the uncertainty due to the intrinsic noise in MD simulations, we present a computationally efficient coupling algorithm that evaluates the variables exchanged between the two components in an atomistic-continuum simulation under uncertainty in terms of PCEs. We consider a simple Couette flow as the physical system model that we use to develop the stochastic multiscale coupling methodology. The Couette flow offers the advantages of being efficient to simulate, common in many fluid mechanics applications where MD simulations are required near walls with a “no-slip” boundary condition, and sufficiently complex to develop a general stochastic coupling methodology.

This paper is organized as follows. A brief background about PCEs and Bayesian inference is given in section 2. In section 3.1, the geometry of the multiscale Couette flow is described, and in section 3.2, the exchanged variables are defined. Next, the different steps required to quantify uncertainty in an atomistic simulation in terms of the PCE of the output variable are established in section 3.3, with different levels of problem complexity. Propagating uncertainty in the continuum model is described

in section 3.4, and the overall coupling algorithm is described in section 3.5. The MD simulations and other numerical methods used in the algorithm are presented in section 4. Finally, results of the stochastic multiscale coupling of the simple Couette flow are given in sections 5 and 6.

2. Background theory. In this section we introduce the notation and present relevant background material on PCEs and Bayesian inference.

2.1. PCEs. Let $(\Omega, \mathcal{F}, \mathcal{P})$ denote a probability space, where Ω is the sample space, \mathcal{F} is an appropriate σ -algebra on Ω , and \mathcal{P} is a probability measure. We consider systems with random entities, parameterized by a finite collection of real-valued independent and identically distributed (i.i.d.) random variables ξ_1, \dots, ξ_n on Ω . \mathcal{P}_ξ denotes the joint distribution function of the random vector $\xi = (\xi_1, \dots, \xi_n)^T$. Note that since the ξ_j are i.i.d., they share a common distribution function, \mathcal{P} , and consequently $\mathcal{P}_\xi(\mathbf{x}) = \prod_{j=1}^n \mathcal{P}(x_j)$ for $\mathbf{x} \in \mathbb{R}^n$. If, for example, ξ is a standard normal random variable on Ω , we write $\xi \sim \mathcal{N}(0, 1)$ and $\mathcal{P}_\xi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$ is the probability density function (PDF) of ξ . Any random variable $u \in L^2(\Omega, \mathcal{F}, \mathcal{P}_\xi)$ admits an expansion of the form

$$(1) \quad u = \sum_{k=0}^{\infty} u_k \Psi_k(\xi),$$

where the $\{\Psi_k\}_{k=0}^{\infty}$ is an orthogonal basis with respect to the density of ξ ,

$$(2) \quad \langle \Psi_k \Psi_l \rangle = \int_{\Omega} \Psi_k(\mathbf{x}) \Psi_l(\mathbf{x}) \mathcal{P}_\xi(\mathbf{x}) d\mathbf{x} = \delta_{kl} \langle \Psi_k^2 \rangle,$$

where δ_{kl} is the Kronecker delta. The expansion (1) is known as the PCE [7, 19, 24, 29, 52] of u . Particularly, in the case where $\xi_i \stackrel{\text{iid}}{\sim} \mathcal{N}(0, 1)$, the $\{\Psi_k\}_{k=0}^{\infty}$ are n -variate Hermite polynomials [1]. In practical computations, we approximate $u(\xi)$ with a truncated series,

$$(3) \quad u(\xi) \approx \sum_{k=0}^P u_k \Psi_k(\xi),$$

where P is finite and depends on the truncation strategy adopted. We consider truncations based on the total degree of the retained polynomials in the series such that P is a function of the stochastic dimension n and expansion order p according to

$$(4) \quad P + 1 = \frac{(n + p)!}{n!p!}.$$

Here p refers to the largest polynomial degree in the expansion. We denote by \mathcal{V}^p the subspace of $L^2(\Omega)$ spanned by $\{\Psi_0, \Psi_1, \dots, \Psi_P\}$. Every element of \mathcal{V}^p is a linear combination of the polynomial chaos (PC) basis $\{\Psi_k\}_{k=0}^P$. Random variables in $L^2(\Omega)$ are approximated by elements in \mathcal{V}^p . One way to derive the PC coefficients of a $u \in \mathcal{V}^p$ is by projection on the PC basis following

$$(5) \quad u_k = \frac{\langle u \Psi_k \rangle}{\langle \Psi_k^2 \rangle}, \quad k = 0, \dots, P.$$

As mentioned in the introduction, this requires numerical evaluation of the projection integrals $\langle u \Psi_k \rangle$ using quadrature rules. These are problematic in situations when the u data is noisy. Another way to compute PC coefficients is Bayesian inference, which will be introduced in the following section.

2.2. Bayesian inference. Let $\mathbf{m} \in \Omega$ denote a vector of model parameters and \mathbf{d} a vector of observable data. A forward model Γ relates the data \mathbf{d} to the parameters \mathbf{m} by $\mathbf{d} \approx \Gamma(\mathbf{m})$ [37]. We use Bayes' rule to derive the posterior PDF for the model parameters \mathbf{m} given the observed data \mathbf{d} [37, 50]:

$$(6) \quad \mathcal{P}(\mathbf{m}|\mathbf{d}) = \frac{\mathcal{P}(\mathbf{d}|\mathbf{m})\mathcal{P}(\mathbf{m})}{\int_{\Omega} \mathcal{P}(\mathbf{d}|\mathbf{m})\mathcal{P}(\mathbf{m})d\mathbf{m}}.$$

The prior $\mathcal{P}(\mathbf{m})$ and posterior $\mathcal{P}(\mathbf{m}|\mathbf{d})$ probabilities represent degrees of knowledge of \mathbf{m} *before* and *after* observing the data \mathbf{d} , respectively. The denominator in (6) is independent of \mathbf{m} and is merely a normalizing constant for the purposes of this work. The key component in (6) is the likelihood function $L(\mathbf{m}) = \mathcal{P}(\mathbf{d}|\mathbf{m})$. To construct the likelihood, one needs to make assumptions regarding the distribution of the discrepancy between the model prediction and the data. Assuming that a forward model Γ predicts N realizations of the data \mathbf{d} with a discrepancy $\boldsymbol{\epsilon} = [\epsilon_1, \epsilon_2, \dots, \epsilon_N]^T$ that has a PDF $\mathcal{P}_{\boldsymbol{\epsilon}}(\cdot)$, i.e., $\mathbf{d} = \Gamma(\mathbf{m}) + \boldsymbol{\epsilon}$, we can write the likelihood as

$$(7) \quad \mathcal{P}(\mathbf{d}|\mathbf{m}) = \mathcal{P}_{\boldsymbol{\epsilon}}(\mathbf{d} - \Gamma(\mathbf{m})).$$

Further, if $\{\epsilon_i\}_{i=1}^N$ are i.i.d. with $\epsilon_i \sim \mathcal{P}_{\epsilon}(\cdot)$ (a condition which holds in the present context, as explained when used later below), then we have

$$(8) \quad \mathcal{P}(\mathbf{d}|\mathbf{m}) = \prod_{i=1}^N \mathcal{P}_{\epsilon}(d_i - \Gamma(\mathbf{m})).$$

In this paper, we infer the PC coefficients of a variable extracted from an MD simulation (see section 3.3). Thus, the PC coefficients are model parameters in the inference procedure outlined in this section. Starting from prior knowledge of the PC coefficients and given data \mathbf{d} in the likelihood function, the posterior distribution calculated using (6) includes better knowledge of the PC coefficients.

3. Mathematical formulation.

3.1. Model geometry. Consider the atomistic-continuum steady Couette flow schematically depicted in Figure 2. The typical configuration for Couette flow, where two walls separated by a fluid region are moving in opposite directions at a velocity V , is shown in Figure 2(a). The regions close to the walls (depicted in red) are the ones where MD simulations should be performed in order to capture wall-slip phenomena that cannot be resolved in fully continuum simulations. By virtue of symmetry, we can restrict our study to the lower half of the continuum domain. Using Galilean invariance, we then can map this domain into the situation where the lower wall is stationary while the fluid at the centerline is moving at a velocity V . This system, which is shown in Figure 2(b) and (c), is the target of this study. The system contains a discrete overlap region shown as a dashed red area, where coupling information is communicated between the continuum and atomistic simulations. The distance δ from the wall to the overlap region needs to be chosen such that all continuum and atomistic governing equations are applicable in the overlap region. Figure 2(c) is a magnified view of the MD simulation domain where the bottom wall is stationary and the top wall is moving at velocity $2U$ such that the velocity at $y = h_{MD}$ is equal to U , where U is the fluid velocity at $y = h_{MD}$ in Figure 2(b). Further details about the MD simulation are given in section 4.1.

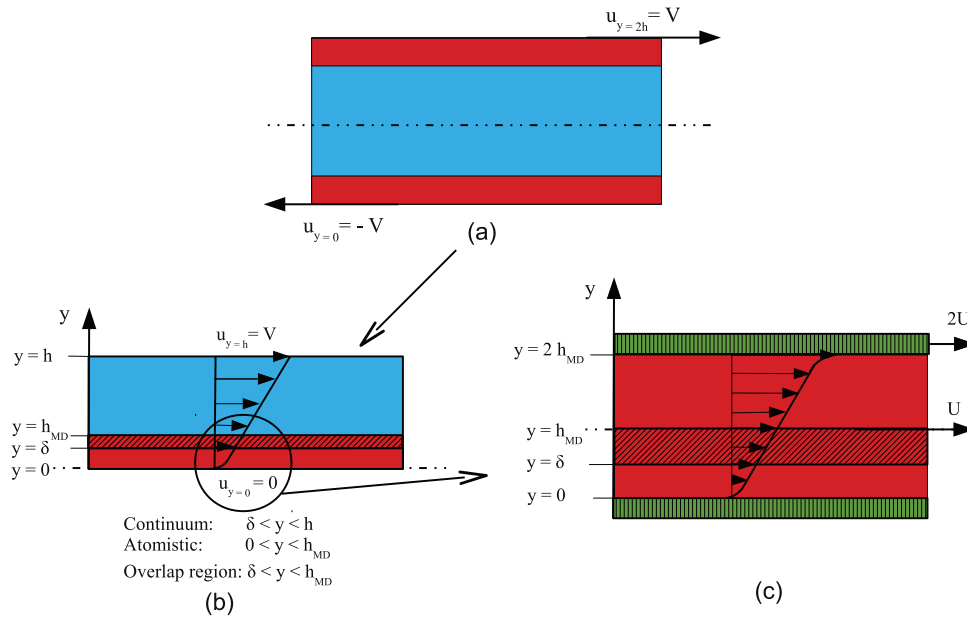


FIG. 2. A schematic showing different components of the continuum-atomistic simulation. The red color (dark shading) signifies the atomistic portion of the simulation, while blue (light shading) represents the continuum portion. The schematics show (a) a general symmetric Couette flow multiscale geometry where V is the continuum wall velocity, (b) a transformed half-geometry where the atomistic simulation is next to the stationary wall; the dashed red (dark shaded) area is an overlap region between atomistic and continuum, and (c) an atomistic Couette flow simulation performed next to the wall where U is the velocity imposed on the MD region in (b); the green area (above and below with vertical lines) denotes the region in the atomistic simulation that models the walls.

3.2. Variable exchange between the atomistic and continuum simulations. In the setting of the Couette flow in Figure 2(b), the key variables bridging the atomistic and continuum simulations are either shear stress (F) or a velocity (V). There are four different coupling schemes based on the classification of Ren [45].

1. FV scheme: the shear stress is extracted from the continuum simulation and imposed as a boundary condition on the atomistic simulation. In exchange, an averaged velocity is extracted from the atomistic simulation at a sample box centered at $y = \delta$ and imposed on the continuum at $y = \delta$.
2. VV scheme: a velocity is extracted from the continuum simulation at $y = h_{MD}$ and imposed as a boundary condition on the atomistic simulation at $y = h_{MD}$. In exchange, an updated velocity is extracted from the atomistic simulation at $y = \delta$ and imposed on the continuum at $y = \delta$ (see Figure 3).
3. VF scheme: a velocity is extracted from the continuum simulation at $y = \delta$ and imposed as a boundary condition on the atomistic simulation at $y = \delta$. In exchange, shear stress is calculated from the atomistic simulation across $y = \delta$ and imposed on the continuum.
4. FF scheme: the shear stress is exchanged back and forth between the continuum and atomistic regions.

Since shear stress, unlike velocity, is not a fundamental quantity in MD and therefore difficult to control directly, the FV and FF schemes that impose shear stress

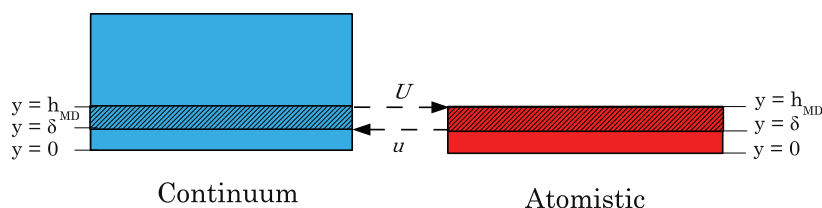


FIG. 3. A schematic showing the locations of variable exchange between the atomistic and continuum simulations in a VV coupling scheme.

on the MD system were eliminated from consideration. Stability considerations with the VF scheme further discussed in section 3.5 led us to use the VV coupling scheme in subsequent multiscale simulations. The resulting exchange of variables between the atomistic and the continuum simulations is visualized in Figures 3 and 4.

As mentioned in the introduction, the output u of an atomistic simulation is an uncertain quantity. This uncertainty is visualized and quantified in sections 5.1 and 5.2. Thus, imposing u on the continuum simulation (see Figure 4) incurs an uncertain output U , which in turn is imposed on the MD simulation. Hence, in a coupled setting, the outputs u and U of the atomistic and continuum simulations, respectively, are uncertain quantities. These coupling variables are defined on the macroscale continuum level. For instance, u in Figures 3 and 4 is not raw MD velocity data; it is the uncertain macroscale velocity estimated from the MD simulation velocity data. For each value of U imposed on the atomistic model, different replica MD simulations are performed [2] to sample over the many degrees of freedom for the initial particle positions and velocities. When the MD simulation reaches a statistical steady state, the resulting averaged velocity output still has noise with a magnitude depending on the length of the averaging time window. We represent u and U in the form of PCEs and denote by \mathbf{u} and \mathbf{U} the vectors containing the PC coefficients of the variables.

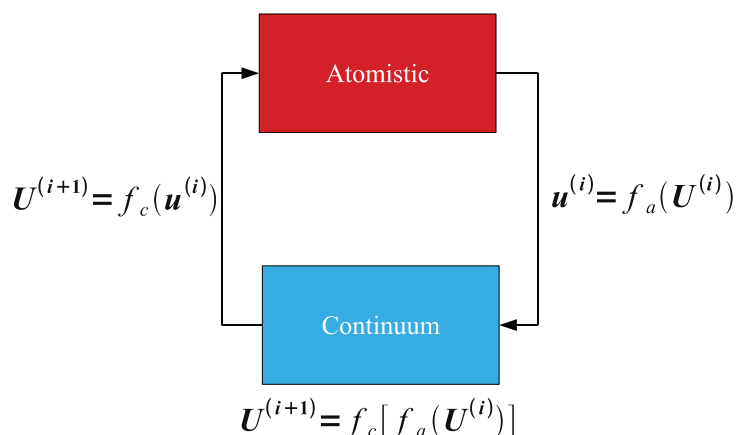


FIG. 4. A schematic showing the exchange of the variables in terms of PC vectors in a VV coupling scheme between the atomistic and continuum simulations. The superscript index (i) indicates the current coupling iteration.

The solution $\{\mathbf{u}, \mathbf{U}\}$ is then determined by solving the following system of equations:

$$(9) \quad \begin{aligned} \mathbf{u} &= f_a(\mathbf{U}), \\ \mathbf{U} &= f_c(\mathbf{u}), \end{aligned}$$

where f_a and f_c are functions that relate the PC vectors \mathbf{u} and \mathbf{U} at the atomistic and continuum levels, respectively. We discuss the details of f_a and f_c in sections 3.3 and 3.4. In section 3.5, we describe an iterative scheme to solve the system (9) illustrated in Figure 4.

3.3. Quantifying intrinsic uncertainty in the atomistic simulations. In this section we describe how to quantify the intrinsic uncertainty in the output u of the MD simulation. We consider different cases by order of complexity. We first quantify in section 3.3.1 the uncertainty in the MD simulation on which a deterministic velocity U is imposed. In section 3.3.2, we describe a method to propagate an uncertain velocity $U(\xi)$ through the MD simulation and to quantify the output uncertainty. In section 3.3.3, we extend this method to the situation where the atomistic simulation is coupled to a continuum model. In all the cases addressed here, the uncertainty is due to the finite MD sampling in the atomistic simulation. We express the uncertainty in our output variable in the form of a PCE, and we use Bayesian inference to determine its PC coefficients. The major steps of this procedure are illustrated in Figure 5.

3.3.1. Atomistic simulation with deterministic input. For a deterministic input velocity U , we set up N_r replica MD simulations. From each replica simulation, we gather N_t samples of the output velocity v obtained when the simulation reaches a statistical steady state. These samples are not MD velocities but short-time averages of the raw MD velocity. As such, we obtain $N = N_r N_t$ independent noisy samples $\{v_j\}_{j=1}^{N_r N_t}$ of u as illustrated in Figure 5(top). A straightforward method for computing the long-term averaged velocity u is simply averaging over the replicas and samples $\{v_j\}_{j=1}^{N_r N_t}$. However, this method does not allow the estimation of the uncertainty in the long-term averaged velocity. Therefore we use these samples as data \mathbf{d} in a Bayesian inference process to determine the posterior distribution on the output scalar velocity u .

We assume a Gaussian noise model that relates u to the short-time averaged simulation samples v_j such that

$$(10) \quad v_j = u + \sigma \eta_j,$$

where the i.i.d. standard normals $\{\eta_j\}$ represent the noise in the short-term MD sample averages, an assumption justified based on the central limit theorem (CLT). The variance σ^2 of the discrepancy is expected to decrease inversely proportional to the number of samples N and will be inferred as a *hyperparameter* [37, 39] along with u . Bayes' rule is then written as

$$(11) \quad \mathcal{P}(u, \sigma^2 | \mathbf{d}) \propto \mathcal{P}(\mathbf{d} | u, \sigma^2) \mathcal{P}(u, \sigma^2),$$

where \mathbf{d} is the velocity data v_j obtained from the MD simulations. Note that the inference relies on a constant σ noise model since the discrepancies are due to finite MD sampling which is largely independent of the macroscale input velocity for a constant temperature MD simulation and a fixed MD system size. The likelihood

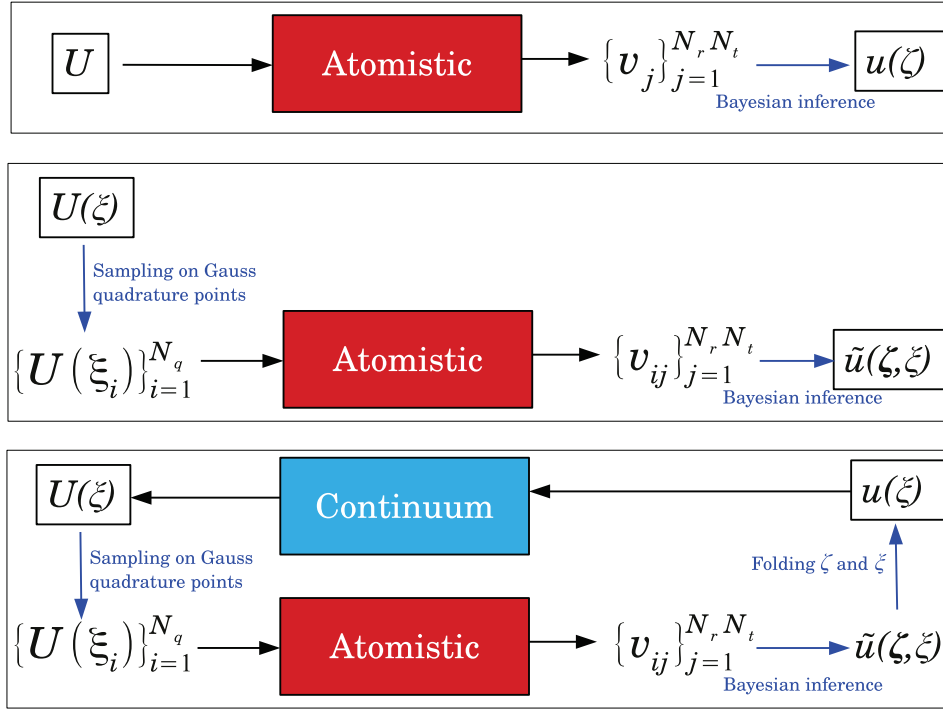


FIG. 5. A schematic showing the steps taken to evaluate the atomistic simulation output u as a function of U . Top: U is deterministic, and $u(\zeta)$ is inferred based on the MD data v_j . Middle: U is uncertain; it is sampled on Gauss quadrature points, and atomistic simulations are performed for each sample. $\tilde{u}(\xi, \zeta)$ is inferred based on the MD data v_{ij} . Bottom: In addition to the steps in the middle row, the uncertainties are folded into one stochastic dimension to produce the PCE of the output u .

function, based on (10), is written as

$$(12) \quad \mathcal{P}(\mathbf{d}|u, \sigma^2) = (2\pi)^{-N/2} (\sigma^2)^{-N/2} \exp\left(-\frac{\boldsymbol{\epsilon}^T \boldsymbol{\epsilon}}{2\sigma^2}\right),$$

where $N = N_r N_t$ and the elements of the discrepancy vector $\boldsymbol{\epsilon}$ are given by

$$(13) \quad \begin{aligned} \epsilon_j &= \sigma \eta_j \\ &= v_j - u. \end{aligned}$$

We assume independent priors for u and σ^2 , i.e., $\mathcal{P}(u, \sigma^2) = \mathcal{P}(u)\mathcal{P}(\sigma^2)$, and assign to u an improper uniform prior [50] on $[-\infty, +\infty]$. For the hyperparameter, σ^2 , leveraging the fact that there is a complete a priori ignorance about its value except that it cannot be negative, we assume a Jeffreys prior:

$$(14) \quad \mathcal{P}(\sigma^2) = \frac{1}{\sigma^2}.$$

If the likelihood (12) incorporates a large amount of data, then the joint prior has a minimal role in the resulting posterior. Conversely, if the likelihood function brings only a small amount of data, the distribution of the resulting posterior is comparable to the prior. The effect of the amount of data will be discussed in section 5.

Based on the chosen priors and the Gaussian noise model, the corresponding posterior is a normal-scaled inverse Gamma (NSIG) distribution [12] such that

$$(15) \quad \mathcal{P}(u, \sigma^2 | \mathbf{d}) = \frac{\sqrt{\nu}}{\sqrt{2\pi\sigma^2}} \frac{b^a}{\Gamma(a)} \left(\frac{1}{\sigma^2}\right)^{a+1} \exp\left(-\frac{2b + \nu(\mu - u)^2}{2\sigma^2}\right).$$

The distribution parameters are given by

$$(16) \quad \begin{aligned} \nu &= N, \\ \mu &= \frac{1}{N} \sum_{j=1}^N v_j, \\ a &= (N-1)/2, \\ b &= \frac{1}{2} \left(\sum_{j=1}^N (v_j)^2 - N\mu^2 \right). \end{aligned}$$

We are interested primarily in the posterior over u irrespective of the corresponding values of σ^2 . This posterior over u is obtained by marginalizing (11) over σ^2 :

$$(17) \quad \mathcal{P}(u | \mathbf{d}) = \int_0^\infty \mathcal{P}(u, \sigma^2 | \mathbf{d}) d\sigma^2,$$

which is a Student-t distribution [20, 47] $\mathcal{S}(\gamma, \mu, \mathbf{S})$ given by

$$(18) \quad \mathcal{P}(u | \mathbf{d}) = \frac{\Gamma[(\gamma+1)/2]}{\Gamma(\gamma/2) \sqrt{\gamma\pi\mathbf{S}} \left[1 + \frac{(u-\mu)^2}{\gamma\mathbf{S}}\right]^{(\gamma+1)/2}},$$

where $\mathbf{S} = 2b/(\gamma N)$, and the number of degrees of freedom γ , mean μ , and variance Σ are given by

$$(19) \quad \begin{aligned} \gamma &= N-1, \\ \mu &= \frac{1}{N} \sum_{j=1}^N v_j, \\ \Sigma &= \frac{\gamma}{\gamma-2} \mathbf{S} = \frac{2b}{N(\gamma-2)} \end{aligned}$$

and u is given by

$$(20) \quad u = \mu + \sqrt{\mathbf{S}}\zeta,$$

where ζ follows the distribution $\mathcal{S}(\gamma, 0, 1)$. For large values of N , the Student-t distribution $\mathcal{S}(\gamma, 0, 1)$ approaches the normal distribution $\mathcal{N}(0, 1)$ such that, in this limit, $\mathcal{P}(u | \mathbf{d})$ is close to a Gaussian.

3.3.2. Atomistic simulation with uncertain input. Consider an uncertain input velocity to the atomistic simulation, represented by a PCE $U(\xi)$. In order to propagate this uncertainty through the atomistic simulation, N_q samples $\{U(\xi_i)\}_{i=1}^{N_q}$ are first drawn from the distribution of $U(\xi)$. The choice of these samples is not restricted to specific locations; for convenience, we select the samples at the Gauss

quadrature points. For each of these samples $U(\xi_i)$, N_r replica simulations are set up. From each replica simulation, we gather N_t samples of the statistical steady state velocity v . Hence, for each $U(\xi_i)$, we obtain $N_r N_t$ uncorrelated and independent samples $\{v_{ij}\}_{j=1}^{N_r N_t}$ as illustrated in Figure 5(middle). These samples are used as data \mathbf{d} in the Bayesian inference process, described in this section, to estimate the PC coefficients of the output variable \tilde{u} , which is the long-term average velocity inferred from the MD simulation. This quantity has uncertainty due to both the uncertainty in the input velocity $U(\xi)$ and the MD sampling noise. Instead of computing \tilde{u} by simply averaging the MD velocity samples over time and replicas, we infer the PC coefficients for the representation of \tilde{u} using the data obtained from MD replicas and samples as illustrated in Figure 6. For a given input $U(\xi_i)$, the velocity v obtained from MD simulations exhibits a spread in its value due to the finite-sampling noise as discussed previously.

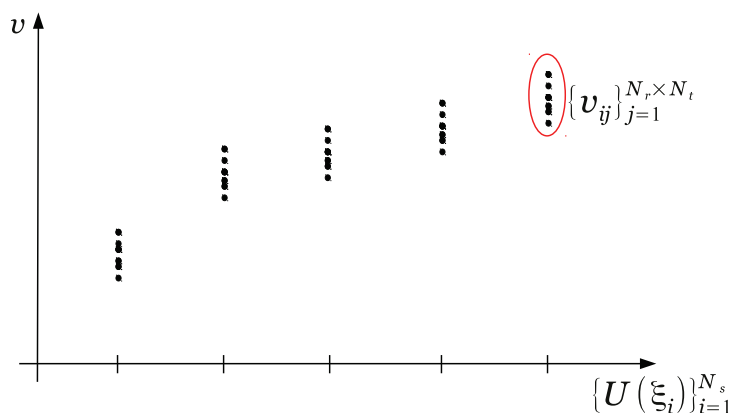


FIG. 6. A schematic showing samples v_{ij} obtained from MD simulations as a function of the input velocity $U(\xi_i)$.

We represent the random variable \tilde{u} with a PCE as

$$(21) \quad \tilde{u}(\xi) = \sum_{k=0}^P \tilde{u}_k \Psi_k(\xi).$$

Similarly to section 3.3.1, we assume a Gaussian noise model that relates \tilde{u} in (21) to the actual simulation results v_{ij} :

$$(22) \quad v_{ij} = \sum_{k=0}^P \tilde{u}_k \Psi_k(\xi_i) + \sigma \eta_{ij}.$$

Here also, $\{\eta_{ij}\}$ are i.i.d. standard normals and represent the noise in the short-term MD sample averages. The variance σ^2 will also be inferred as a hyperparameter along with the PC coefficient vector $\tilde{\mathbf{u}} = \{\tilde{u}_k\}_{k=0}^P \in \mathbb{R}^{P+1}$. Bayes' rule is then written as

$$(23) \quad \mathcal{P}(\tilde{\mathbf{u}}, \sigma^2 | \mathbf{d}) \propto \mathcal{P}(\mathbf{d} | \tilde{\mathbf{u}}, \sigma^2) \mathcal{P}(\tilde{\mathbf{u}}, \sigma^2),$$

where \mathbf{d} is the velocity data v_{ij} obtained from the MD simulations. The likelihood function $\mathcal{P}(\mathbf{d} | \tilde{\mathbf{u}}, \sigma^2)$ is given by (12), where, in this case, $N = N_q N_r N_t$ and the vector

ϵ is given by

$$(24) \quad \epsilon = \mathbf{d} - \mathbf{Q}\tilde{\mathbf{u}}.$$

Elements of the vector $\mathbf{d} \in \mathbb{R}^N$ and the matrix $\mathbf{Q} \in \mathbb{R}^{N \times (P+1)}$ are given by

$$(25) \quad \begin{aligned} d_m &= v_{ij}, \\ Q_{mk} &= \Psi_k(\xi_i), \\ m &= j + (i-1)N_r N_t, \\ i &= 1, \dots, N_q, \\ j &= 1, \dots, N_r N_t, \\ k &= 0, \dots, P. \end{aligned}$$

As in the previous section, we assign to the PC coefficients an improper uniform prior [50] on $[-\infty, +\infty]$ due to the complete prior ignorance of their values. For the necessarily positive but otherwise unconstrained hyperparameter σ^2 , the corresponding posterior is a multivariate NSIG distribution [12] in this case such that

$$(26) \quad \mathcal{P}(\tilde{\mathbf{u}}, \sigma^2 | \mathbf{d}) \propto (\sigma^2)^{-a-3/2-P/2} \exp\left(-\frac{2b + (\tilde{\mathbf{u}} - \boldsymbol{\lambda})^T \underline{\nu}(\tilde{\mathbf{u}} - \boldsymbol{\lambda})}{2\sigma^2}\right).$$

The distribution parameters are given by

$$(27) \quad \begin{aligned} \underline{\nu} &= \mathbf{Q}^T \mathbf{Q}, \\ \boldsymbol{\lambda} &= \underline{\nu}^{-1} \mathbf{Q}^T \mathbf{d}, \\ a &= [N - (P+1)]/2, \\ b &= \frac{1}{2} (\mathbf{d}^T \mathbf{d} - \mathbf{d}^T \mathbf{Q} \boldsymbol{\lambda}). \end{aligned}$$

The marginal posterior on $\tilde{\mathbf{u}}$ is a multivariate Student-t distribution [20, 47] $\mathcal{S}(\gamma, \boldsymbol{\lambda}, \underline{\Sigma})$:

$$(28) \quad \mathcal{P}(\tilde{\mathbf{u}} | \mathbf{d}) = \frac{\Gamma[(\gamma + P + 1)/2]}{\Gamma(\gamma/2)(\gamma\pi)^{(P+1)/2} |\underline{\Sigma}|^{1/2} \left[1 + \frac{1}{\gamma}(\tilde{\mathbf{u}} - \boldsymbol{\lambda})^T \underline{\Sigma}^{-1}(\tilde{\mathbf{u}} - \boldsymbol{\lambda})\right]^{(\gamma+P+1)/2}},$$

where $\underline{\Sigma} = 2b\underline{\nu}^{-1}/\gamma$, and the number of degrees of freedom γ , mean vector $\boldsymbol{\lambda} \in \mathbb{R}^{P+1}$, and covariance matrix $\underline{\Sigma} \in \mathbb{R}^{(P+1) \times (P+1)}$ are given by

$$(29) \quad \begin{aligned} \gamma &= N - (P + 1), \\ \boldsymbol{\lambda} &= \underline{\nu}^{-1} \mathbf{Q}^T \mathbf{d}, \\ \underline{\Sigma} &= \frac{\gamma}{\gamma - 2} \underline{\Sigma} = \frac{2b}{\gamma - 2} \underline{\nu}^{-1}. \end{aligned}$$

This expression captures two sources of uncertainty. The first source is inherited from the uncertainty in $U(\xi)$ that was imposed on the atomistic model, while the second source of uncertainty is the one reflected in the posterior of each of the PC coefficients of $\tilde{\mathbf{u}}$. This distribution represents uncertainty due to the finite amount of MD data that was available for the Bayesian inference procedure.

The marginal posterior of the inferred PC coefficient vector $\tilde{\mathbf{u}}$ is a multivariate Student-t distribution with the mean vector $\boldsymbol{\lambda}$ and a covariance matrix $\underline{\Sigma}$ (see (29)).

Thus, we can express $\tilde{\mathbf{u}}$ as [15, 51]

$$(30) \quad \tilde{\mathbf{u}} = \begin{pmatrix} \tilde{u}_0 \\ \tilde{u}_1 \\ \vdots \\ \tilde{u}_P \end{pmatrix} = \boldsymbol{\lambda} + \underline{\Lambda} \boldsymbol{\zeta} \sim \mathcal{S}(\gamma, \boldsymbol{\lambda}, \underline{\Sigma}),$$

where $\underline{\Lambda} \in \mathbb{R}^{(P+1) \times (P+1)}$ is the lower-triangular Cholesky factor of the scale matrix $\underline{\Sigma} = \underline{\Lambda} \underline{\Lambda}^T$, and the vector $\boldsymbol{\zeta} \in \mathbb{R}^{P+1}$ comprises i.i.d. random variables distributed according to $\mathcal{S}(\gamma, 0, 1)$. Furthermore, the linear combination of Student-t variables \tilde{u}_k given by

$$(31) \quad \begin{aligned} \tilde{u} &= \sum_{k=0}^P \tilde{u}_k \Psi_k(\xi) \\ &= \boldsymbol{\Psi}(\xi)^T (\boldsymbol{\lambda} + \underline{\Lambda} \boldsymbol{\zeta}) \end{aligned}$$

forms an uncertain response surface in terms of a Student-t process such that

$$(32) \quad \tilde{u} \sim \mathcal{S}(\gamma, \boldsymbol{\Psi}(\xi)^T \boldsymbol{\lambda}, \boldsymbol{\Psi}(\xi)^T \underline{\Sigma} \boldsymbol{\Psi}(\xi)),$$

where $\boldsymbol{\Psi} = [\Psi_0, \dots, \Psi_P]^T$.

3.3.3. Atomistic simulation with stochastic input coupled to continuum. In this case, the input U to the atomistic model is uncertain as illustrated in Figure 5(bottom). Hence we can apply the same Bayesian inference machinery of the previous section to propagate this uncertainty in the atomistic model. Bayesian inference results in a distribution for each of the inferred PC coefficients of \tilde{u} after marginalizing over the hyperparameter σ^2 . Since the atomistic and continuum models are coupled, solving for the variables u and U requires iteration between the two models as discussed at the end of section 3.2. Hence, it is useful to have a well-defined PCE for the uncertain atomistic model output u such that it can be fed to the continuum model. In principle, the two sources of uncertainty in \tilde{u} are of the same nature since the stochastic dimension ξ in this case is also the result of MD finite-sampling noise from a previous stochastic coupling iteration. Here, we propose a series of transformations that enable “folding” these uncertainties into one stochastic dimension resulting in a representation of u with a single degree of freedom, i.e., with a one-dimensional PCE that has deterministic coefficients. The task is to represent the random variable \tilde{u} in terms of a PCE with a single degree of freedom ξ' such that

$$(33) \quad \tilde{u} \approx u = \sum_{k=0}^P u_k \Psi_k(\xi').$$

We propose using a Galerkin projection of $\tilde{u}(\xi, \boldsymbol{\zeta})$ onto the basis in ξ' . The expression of \tilde{u} in (31) allows sampling with very little computational cost. Indeed, this requires only sampling Student-t and Gaussian distributions. Using kernel density estimation (KDE) [6], we build the cumulative distribution function (CDF), $F(\cdot)$, of the obtained samples. Let $\Phi(\cdot)$ be the CDF of the standard normal distribution. By construction of the CDF, $F(\tilde{u})$ and $\Phi(\xi')$ are uniformly distributed random variables. Therefore \tilde{u} and $F^{-1}(\Phi(\xi'))$ have the same distribution, allowing the mapping of \tilde{u} to the dimension ξ' and the construction of a PCE for u following (33). The deterministic

coefficients u_k can be computed by Galerkin projection using quadrature according to

$$(34) \quad u_k = \frac{\langle F^{-1}(\Phi(\xi'))\Psi_k(\xi') \rangle}{\langle \Psi_k(\xi')^2 \rangle}.$$

To avoid accumulating new notation with every iteration, we will replace ξ' by ξ , since the two are the same in terms of distribution, and write

$$(35) \quad u = \sum_{k=0}^P u_k \Psi_k(\xi).$$

The representation (35) has a single degree of freedom and, by construction, is equivalent in distribution to (31), which involves two degrees of freedom.

3.4. Propagating uncertainty in the continuum simulation. The PCE of the velocity obtained from the MD simulations is imposed on the continuum system as shown in Figure 3. We formulate the continuum Couette flow mathematical model based on Figure 2(b). We assume a steady, laminar, Newtonian flow, in which case the Navier–Stokes equations reduce to

$$(36) \quad \frac{d^2 U}{dy^2} = 0 \quad \text{for } \delta \leq y \leq h.$$

Based on section 3.1, the boundary conditions are set such that the velocity extracted from the MD simulation is imposed at $y = \delta$ in the continuum simulation. The boundary conditions become

$$(37) \quad U(y = \delta) = u,$$

$$(38) \quad U(y = h) = V.$$

Finally the analytical solution to the problem is

$$(39) \quad U(y) = V + \frac{(V - u)(y - h)}{h - \delta}.$$

In the absence of any uncertainty in the continuum model parameters, we propagate the uncertainty present in u intrusively in (39). We would like to obtain a PCE for U of the form

$$(40) \quad U(y, \xi) = \sum_{k=0}^P U_k(y) \Psi_k(\xi).$$

By substituting the PCE of u (35) in (39), we derive the PC coefficients of $U_k(y)$ as

$$(41) \quad U_k(y) = \begin{cases} V + \frac{(V - u_k)(y - h)}{h - \delta} & \text{for } k = 0, \\ \frac{-u_k(y - h)}{h - \delta} & \text{for } 0 < k \leq P. \end{cases}$$

Finally, we extract the velocity U at $y = h_{MD}$ (see Figures 2 and 3) to obtain

$$(42) \quad U_k = \begin{cases} V + \frac{(V - u_k)(h_{MD} - h)}{h - \delta} & \text{for } k = 0, \\ \frac{-u_k(h_{MD} - h)}{h - \delta} & \text{for } 0 < k \leq P. \end{cases}$$

3.5. Coupling procedure and convergence study. The simulation is initiated by an initial guess of \mathbf{U} , denoted by $\mathbf{U}^{(0)}$, that is imposed on the atomistic simulation. The obtained output $\mathbf{u}^{(0)}$ is imposed on the continuum region (see Figure 2(b)), which generates a new value of \mathbf{U} denoted by $\mathbf{U}^{(1)}$. This process is repeated until convergence in u and U .

Based on the nonlinear system (9) in section 3.2, we relate the values of \mathbf{U} between two consecutive iterations (see Figure 4 and (9)) by

$$(43) \quad \mathbf{U}^{(i+1)} = f_c[f_a(\mathbf{U}^{(i)})].$$

The fixed-point iterative equation (43) converges under the shrinking condition [11, 41]:

$$(44) \quad \left| \frac{d\mathbf{U}^{(i+1)}}{d\mathbf{U}^{(i)}} \right| < 1.$$

In the general case where \mathbf{U} and \mathbf{u} are vectors containing PC coefficients, the above derivative is a Jacobian matrix. In this case, the eigenvalues of the matrix should be bounded by the unit circle to guarantee stability and convergence [11]. Computing this Jacobian matrix involves calculating the derivative of f_a , which is not a straightforward exercise since, as discussed in section 3.3, f_a involves complicated steps with no analytical derivatives available. However, if we consider the deterministic (no sampling noise) case with scalar u and U , we can compute the left-hand side of (44) based on (43) as

$$(45) \quad \left| \frac{dU^{(i+1)}}{dU^{(i)}} \right| = f'_c f'_a.$$

Further, consider a “toy” analytical model for f_a of the MD simulation as the one that directly maps the input velocity U to the extracted velocity u without any uncertainty and without the sampling and Bayesian inference described in section 3.3. For the VV scheme, a reasonable choice for such a model is a linear relationship between u and U , given the linear nature of the Couette flow (see Table 1). We also compute f_c based on the analytical solution of the Couette flow described in the previous section. This allows us to find out whether or not condition (44) is met in this particular case. Table 1 shows the relationships f_a , f_c and their associated convergence condition for the different coupling schemes described in the beginning of this section. We note that the convergence of the coupling scheme strongly depends on the nature of the exchanged variables based on the analysis of Ren [45]. It was found that the FF scheme is always unstable regardless of the flow regime and the frequency of data exchange between continuum and atomistic. However, the VV and FV cases were stable for both static and time-dependent MD. The VF scheme is stable if the frequency of data exchange is above a certain threshold. Overall, the FV scheme was found to be the superior scheme with better convergence and stability properties, as well as smaller statistical errors.

In this paper, velocities are exchanged between the atomistic and continuum models. Hence, the corresponding VV scheme is convergent and stable when these velocities are deterministic according to Table 1. While these results do not guarantee that the VV scheme will be convergent for the fixed-point iteration scheme in the case when u and U are uncertain, we have found the VV scheme to show good convergence properties in practice for the uncertain case as well.

TABLE 1

The relationships $\{f_a, f_c\}$ between scalar input/output coupling variables for a “toy” atomistic model and the convergence condition (44) for different coupling schemes.

Scheme	“Toy” f_a	f_c	Convergence condition (44)
FV	$u = \tau \frac{\delta}{\mu}$	$\tau = (V - u) \frac{\mu}{h - \delta}$	$\frac{\delta}{h - \delta} \ll 1$
VV	$u = U \frac{\delta}{h_{MD}}$	$U = V + \frac{(V - u)(h_{MD} - h)}{h - \delta}$	$\frac{\delta(h - h_{MD})}{h_{MD}(h - \delta)} < 1$
VF	$\tau = u \frac{\mu}{\delta}$	$u = V - \tau \frac{h - \delta}{\mu}$	$\frac{h - \delta}{\delta} \gg 1$

4. MD numerical implementation. In this section, we cover the numerical implementation of the MD computations used in the multiscale simulations. First, we present details of the atomistic simulation of the Couette flow in section 4.1. Since atomistic simulations are too expensive to be performed upon each coupling iteration, we present in section 4.2 a method to infer a numerical surrogate for the MD simulation and analyze its accuracy.

4.1. MD simulations. The MD simulation geometry suggested in Figure 2(c) consists of a stationary and a moving wall. When performing the MD simulation, it is advisable for temperature control that the center of mass of the simulation box remain fixed. Hence, for the actual MD simulations, we perform a shift in the velocity U such that the walls are moving in opposite directions at the same velocity U as depicted in Figure 7. We consider a three-dimensional domain of size $31 \times 87 \times 31$ Å in the x , y , and z directions, respectively. The size in the motion direction x and in z was chosen to be big enough such that particle motions are not affected by the periodic boundary conditions in these two directions.

The domain is occupied by particles with mass 18 g/mol (representing water molecules) at a density 1000 Kg/m³ (corresponding to 2729 particles for the selected domain dimensions). The soft interparticle interaction is modeled by the Lennard-Jones (LJ) pairwise potential ϕ_{ij} [32]. For particles i and j separated by a distance r , ϕ_{ij} is given by

$$(46) \quad \phi_{ij} = 4\phi_0 \left[\left(\frac{\rho}{r} \right)^{12} - \left(\frac{\rho}{r} \right)^6 \right].$$

We set the LJ parameters for water to $\phi_0 = 0.152$ Kcal/mole and $\rho = 3.15$ Å [5]. The boundary conditions are periodic in the x and z directions. The moving walls in the Couette flow are modeled with particles in the regions W_{bottom} and W_{top} , where $-12 \leq y \leq 0$ Å and $63 \leq y \leq 75$ Å, respectively, such that $h_{MD} = 31.5$ Å. The particles in the boundary regions have the same properties as those in the flow domain. To simulate moving walls, we fix these particles to lie on a regular lattice and have a prescribed velocity vector $\{U, 0, 0\}$ in W_{top} and $\{-U, 0, 0\}$ in W_{bottom} for the whole simulation period.

MD computations were performed with LAMMPS [43] with a time step of $\Delta t = 1$ fs for a constant temperature $T = 298$ K, using a velocity Verlet time integration of the Newtonian equations and a simple velocity rescale temperature control performed every 100 steps [5, 43]. The simulation is divided into two main parts. The simulation is first run for 10000 steps until the velocity output reaches a statistical steady state. Next, velocity samples are collected until the noise amplitude due to finite sampling is reduced below a target cutoff. While the initial particle positions are on a regular grid in the simulation box, the initial velocities are assigned randomly. For each set of

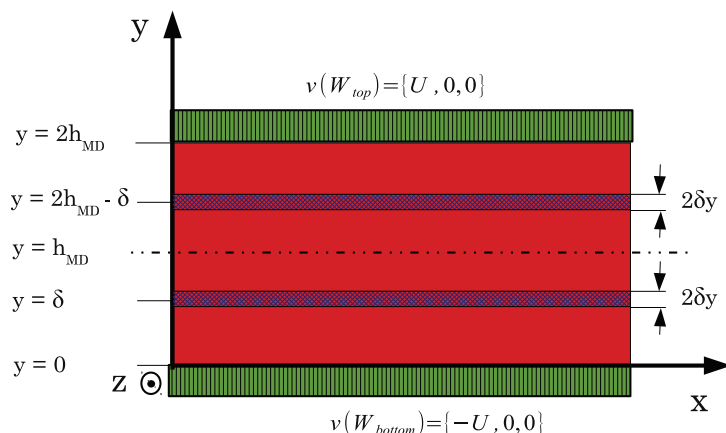


FIG. 7. A schematic showing the MD simulation domain. The hatched areas at $y = \delta$ and $y = 2h_{MD} - \delta$ are the regions where the velocity is extracted. The green area (above and below with vertical lines) denotes the regions W_{bottom} and W_{top} that simulate the moving walls.

inputs we perform different replica simulations to sample over the degrees of freedom in the initial conditions.

For the VV coupling scheme, we extract the velocity at $y = \delta$ (see Figures 2(c) and 7). We choose $\delta = 20.5 \text{ \AA}$, which is far enough from the wall such that the velocity profile is linear, making the continuum laws applicable in this location as discussed in the beginning of section 3.1. This value of δ is the same for all the considered cases in this paper. We evaluate the velocity first by averaging particle velocities over the x and z directions in the region limited by $y = \delta \pm \delta y$ with $\delta y = 2 \text{ \AA}$, illustrated by the hatched area in Figure 7. The average velocity is evaluated at every MD simulation time step. The noise amplitude of this raw velocity is found to be equal to 33.31 m/s, which is in good agreement with the expression of Hadjiconstantinou et al. [23, eq. 1], which predicts an amplitude of 35.72 m/s based on our MD simulation parameters. We then employ a moving time average with a window width of t_w to determine an ensemble average for the short-time averaged velocity.

The extracted velocities as a function of MD simulation time are shown in Figure 8(top). For all the considered averaging time windows, the mean amplitudes of the velocities at $y = 2h_{MD} - \delta$ and $y = \delta$ are found to be comparable to each other, as expected from the symmetry of the domain around $y = h_{MD}$. Extracting the velocity at $y = 2h_{MD} - \delta$ essentially provides a free replica of the simulation for $y = \delta$. We take advantage of this free replica in all the simulations. The velocity at $y = h_{MD}$ fluctuates around zero also as expected. The amplitude of the velocity noise decreases with increasing the averaging time window. The noise in the velocity evaluation is observed in the PDF of the collected velocities v in Figure 8(bottom).

4.2. Surrogate for the atomistic simulations. The coupling formulation described in sections 3.2, 3.3.3, and 3.5 requires multiple iterations in a multiscale setting. In each iteration, multiple MD simulations need to be performed. Each MD simulation of the box described in the previous section requires about 6 hours of computation on an 8 core machine running at a CPU speed of 2.8 GHz. Hence, performing a multiscale simulation under uncertainty becomes a significantly expensive exercise. We alleviate the computational cost by constructing a surrogate for the MD

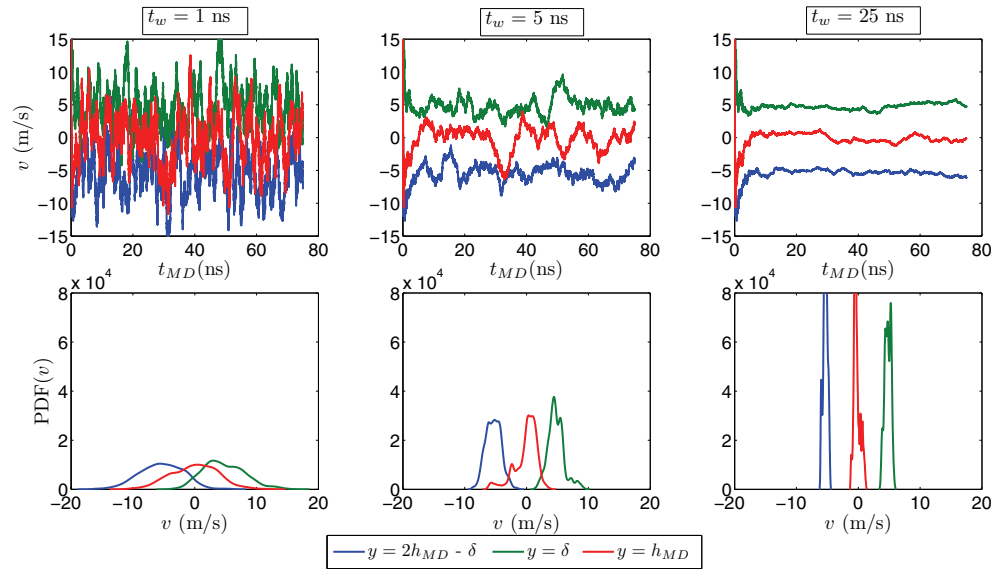


FIG. 8. Plots showing short-time averaged velocities extracted from the Couette flow MD simulation (see Figure 7) for $U = 10$ m/s and different moving time averaging window widths, as indicated. The top row shows the output velocities as a function of the MD simulation time t_{MD} . The bottom row shows the PDFs of the MD time sequence of output velocities v . Plots are generated for output velocities extracted at $y = 2h_{MD} - \delta$, $y = h_{MD}$, and $y = \delta$, as indicated.

simulation output. We infer this surrogate based on precomputed MD data obtained at different input velocities U_s .

4.2.1. Inference of the surrogate model. For a given application, one can select the range of the uncertain input variable and parameters of the MD simulation and set up MD simulations for the different realizations of the input variables and parameters that fall in the selected range. For the Couette flow simulation that is the focus of this paper, we pick $N_s = 10$ values for the input velocity $U_{s,\min} \leq U_s \leq U_{s,\max}$ m/s. $U_{s,\min}$ and $U_{s,\max}$ are chosen to be about half and double the estimated value of U , respectively. For each input velocity, we draw N_{MD} sample values for the extracted velocity v short-time averaged over a time window t_w . Samples of this short-time averaged MD velocity data are depicted as black dots in Figure 9.

We employ the Bayesian inference construction discussed in section 3.3.2 to infer a response surface of this MD data. Due to the observed nearly linear relationship between v and U , we model the expression of the output velocity v as [15, 51]

$$(47) \quad v = v_0 + v_1 \left(\frac{U - \bar{U}_s}{U_{s,\max} - U_{s,\min}} \right) + \sigma\eta,$$

where $\bar{U}_s = (U_{s,\min} + U_{s,\max})/2$ and $\sigma\eta$ reflects the noise present in the short-time averaged MD data. The distribution of η is modeled as standard normal, based on the CLT, with σ being the standard deviation of the velocity data. Since thermal fluctuations, which are largely independent of the input velocity U , are the source of this sampling noise, we can assume σ to be constant over the range of input velocities. For $U = 0$, i.e., stationary walls in the MD Couette flow simulation, we expect the mean velocity $\bar{v}(0)$ to be equal to zero over the MD simulation box. Thus, one could

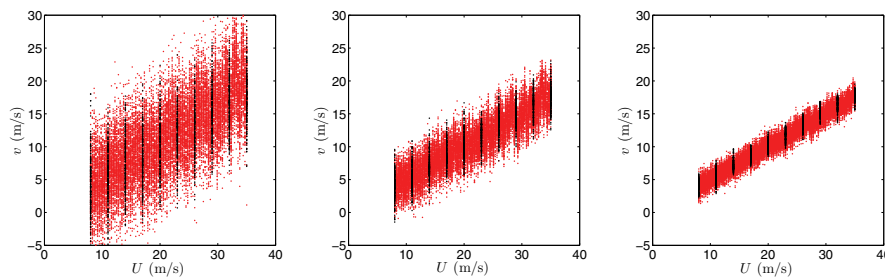


FIG. 9. Plots showing results of the MD simulation surrogate. The black dots represent sampled short-time averaged velocity data obtained from MD simulations for the selected range of the input velocity ($U_{s,\min} = 8$ m/s and $U_{s,\max} = 35$ m/s), while the red dots are the samples of the output velocity v computed by the surrogate model. Plots are generated for $N_{MD} = 200$ and for different noise levels, i.e., different MD time averaging windows, as indicated. Dot colors are distinguishable only in the online version.

infer only v_1 and σ . However, in this work, we infer all three variables v_0 , v_1 , and σ for a more general methodology. Note that while we can use the same inference procedure here as in section 3.3.2, a key difference is that here we infer a surrogate model for the short-time averaged data, whereas before we focused on the long-term averaged MD outputs. We do not marginalize over σ here, and therefore we obtain a bivariate NSIG distribution instead of a bivariate Student-t distribution. We draw samples from the surrogate model by sampling the bivariate NSIG distribution of $\{v_0, v_1, \sigma^2\}$ whose parameters $(\lambda_s, \underline{\nu}_s, a_s, b_s)$ can be computed with a data vector \mathbf{d}_s and a matrix \mathbf{Q}_s (not shown) whose values are based on the MD data and their trend (47) (the black dots in Figure 9) similarly to the expressions in (25)–(27).

The parameters $(\lambda_s, \underline{\nu}_s, a_s, b_s)$ are tabulated for a given MD time averaging window width t_w to be used in the stochastic coupling algorithm. The forward model given by these parameters enables drawing samples from the posterior predictive distribution [16, 36] of v , serving as a substitute for the full MD model since (47) includes the spread of data due to short-time averaging. Thus, rather than performing N_r MD simulations upon each atomistic-continuum coupling iteration and drawing N_t MD data samples from each simulation, we draw $M = N_r N_t$ samples v_s from the surrogate by [12]

1. sampling $\eta \sim \mathcal{N}(0, 1)$,
2. sampling σ from an inverse Gamma distribution with parameters a_s and b_s ,
3. sampling $\{v_0, v_1\}$ from a bivariate normal PDF with parameters $(\lambda_s, \sigma^2 \underline{\nu}_s^{-1})$,
4. applying (47) to calculate v_s .

4.2.2. Accuracy of the surrogate model. The stochastic coupling algorithm described in section 3 is designed to estimate u , the long-term mean of the MD output velocity data. However, when using the surrogate model, the coupling algorithm estimates the mean of the data drawn from the surrogate. This data should match the original raw MD data within a certain degree of accuracy. In this section, we study this accuracy, focusing on the width of the joint marginal posterior on $\{v_0, v_1\}$ which depends on the amount of data used to generate the surrogate model. This posterior follows a bivariate Student-t distribution $\mathcal{S}(\gamma_s, \lambda_s, \underline{S}_s)$ with degrees of freedom γ_s ,

mean λ_s , and scale matrix \underline{S}_s given by

$$(48) \quad \begin{aligned} \gamma_s &= N_s N_{MD} - 2, \\ \lambda_s &= \nu_s^{-1} \underline{Q}_s^T \underline{d}_s, \\ \underline{S}_s &= \frac{2b_s}{\gamma_s} \nu_s^{-1}, \end{aligned}$$

where \underline{d}_s , \underline{Q}_s , and b_s (see (27)) are computed based on the short-time averaged MD velocity data. We can therefore write $\{v_0, v_1\}$ as

$$(49) \quad \begin{pmatrix} v_0 \\ v_1 \end{pmatrix} = \bar{\mathbf{v}} + \underline{\Theta} \begin{pmatrix} \xi_0 \\ \xi_1 \end{pmatrix},$$

where $\bar{\mathbf{v}} = \{\bar{v}_0, \bar{v}_1\} = \lambda_s$, $\underline{\Theta}$ is the lower-triangular Cholesky factor of the scale matrix $\underline{S}_s = \underline{\Theta} \underline{\Theta}^T$, and ξ_0 and ξ_1 are i.i.d. Student-t random variables $\mathcal{S}(\gamma_s, 0, 1)$.

The covariance matrix $\underline{\Sigma}_s = \gamma_s \underline{S}_s / (\gamma_s - 2)$ is closely related to the posterior width and determines the uncertainty in the estimation of the inferred variables due to the limited amount of available data. We would like to quantify the posterior width along all dimensions; hence we define an overall measure of the posterior width as the square root of the trace of $\underline{\Sigma}_s$:

$$(50) \quad \kappa \equiv \sqrt{\text{Tr}(\underline{\Sigma}_s)}.$$

Since κ measures the uncertainty in the surrogate parameters $\{v_0, v_1\}$, it also indicates the accuracy of the surrogate, i.e., how close the average of the values of v_s drawn from the surrogate is to the short-time means of the original MD data. We study the effect of the number of MD data samples N_{MD} and the time averaging window width t_w on the accuracy.

In Figure 10(top row), we plot the obtained values of $\{\bar{v}_0, \bar{v}_1, \bar{\sigma}\}$ ($\bar{\sigma} = \sqrt{b_s/(a_s - 1)}$ [51]) and κ as a function of N_{MD} and t_w . As N_{MD} and t_w increase, \bar{v}_0 and \bar{v}_1 approach a constant value, indicating a more accurate solution. Concurrently, the noise amplitude $\bar{\sigma}$ decreases as a function of the time averaging window width t_w since increasing t_w reduces the intrinsic variabilities of an averaged property extracted from an MD simulation. Note that by computing \bar{v}_0 and \bar{v}_1 using (48)–(49), and setting $U = 0$ in (47), we obtain $\bar{v}(0) \approx 0.3$ m/s, which is quite close to the expected value of $\bar{v}(0) = 0$, especially compared to the input velocities $8 \leq U_s \leq 35$.

In Figure 10(bottom row), we plot the posterior width measure κ as a function of N_{MD} and t_w . As expected, κ decreases with increased MD data, and the logarithmic slope of this decrease is very close to the value of 0.5 expected from the CLT and consistent with the expression derived by Hadjiconstantinou et al. [23, eqs. 3–4]. A similar trend can be observed in the decrease of κ with t_w . This decreasing trend is not surprising since increasing t_w leads to a decrease in $\bar{\sigma}$. In fact, $\bar{\sigma}$ is the original scatter in the MD data as illustrated in Figure 9, and it can be used to gauge the uncertainty measured by κ in the posterior for v_0 and v_1 . Thus, we aim to minimize $\kappa/\bar{\sigma}$ such that the uncertainty in $\{v_0, v_1\}$ is small relative to the original MD data noise amplitude. We plot this ratio together with κ in Figure 10(bottom row). The plot suggests that $\kappa/\bar{\sigma}$ is not significantly affected by t_w but strongly depends on N_{MD} . We therefore opt to infer the surrogate with $N_{MD} = 200$ short-time averaged MD data samples, which guarantees a $\kappa/\bar{\sigma} < 5\%$.

For the range of U considered in our case, $8 \leq U \leq 35$ m/s, we plot values of v drawn from a surrogate inferred for $N_{MD} = 200$ in Figure 9 (red dots). The results show good agreement between the MD data and the data drawn from the surrogate.

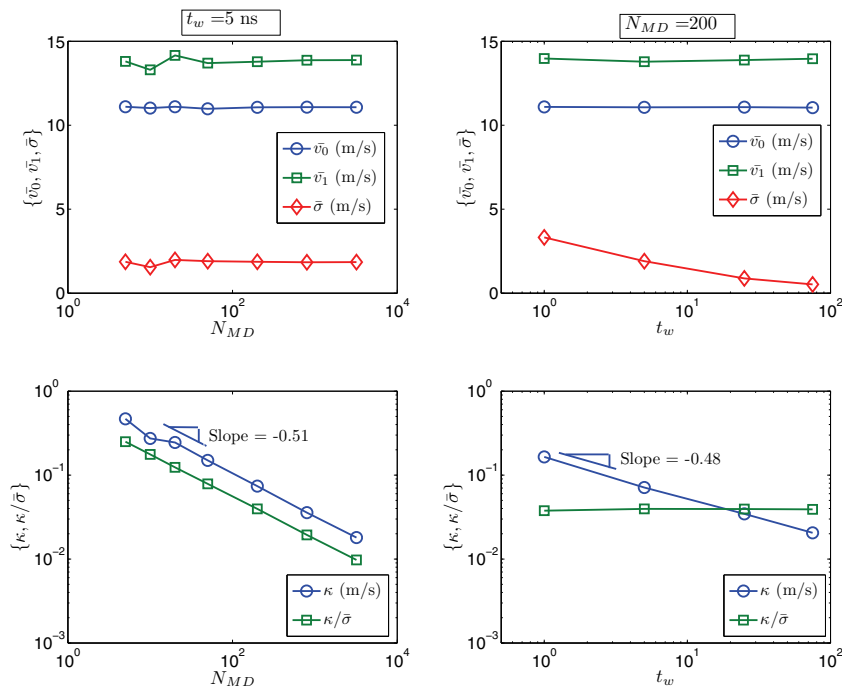


FIG. 10. Plots showing the variation of the inferred variables $\{\bar{v}_0, \bar{v}_1, \bar{\sigma}\}$ (top row) as a function of the number of short-time averaged MD data samples N_{MD} for $t_w = 5$ ns (left column) and the time averaging window width t_w for $N_{MD} = 200$ (right column). Also plotted are the measures of the inferred posterior width $\{\kappa, \kappa/\bar{\sigma}\}$ (bottom row).

5. Application to coupled atomistic-continuum simulations of the Couette flow. In this section, we apply the methodology described in the previous sections to the multiscale Couette flow model. We report results for increasing order of complexity of the stochastic multiscale problem following the organization of section 3.3. We first consider in section 5.1 the atomistic model alone, from which we infer the uncertain output velocity u based on a deterministic velocity input U . In section 5.2, we quantify the uncertainty in the atomistic model output u based on an uncertain velocity input $U(\xi)$. Finally, in section 5.3, we consider the fully coupled stochastic atomistic-continuum multiscale simulation and iterate between both models until convergence. In section 5.4.1, we explore the effect of sequential Bayesian updating (SBU) on the simulation outputs. In all of these cases, we study the effect of the time averaging window width t_w and the amount of data M drawn from the MD surrogate model on the uncertainty. Finally, we study in section 5.4.2 the effectiveness of the surrogate MD model as a function of the number of coupling iterations when SBU is invoked.

5.1. Atomistic simulation with deterministic input. We consider the MD simulation illustrated in Figure 7 on which a deterministic velocity $U = 10$ m/s is imposed. As shown in the plots of Figure 8, the output velocity at $y = \delta$ is noisy due to finite MD sampling. We implement the Bayesian inference approach described in section 3.3.1 to infer the output velocity u and the noise σ in the short-time averaged

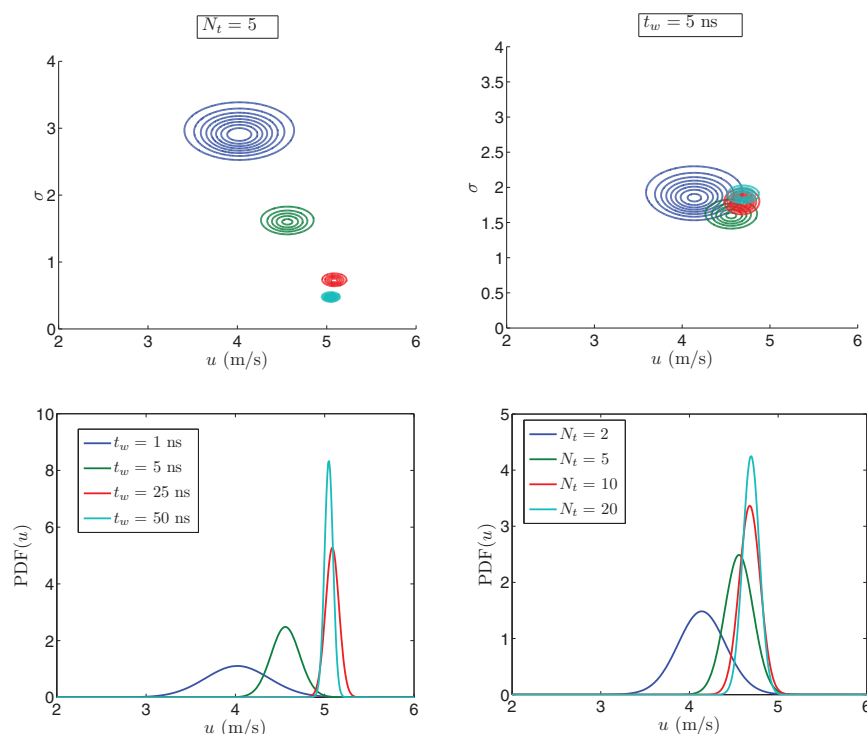


FIG. 11. Plots showing (top) the joint posterior on $\{u, \sigma\}$ given by NSIG distributions and (bottom) the PDF of u after marginalizing over σ^2 given by Student- t distributions when u is extracted from an atomistic simulation with a deterministic input $U = 10$ m/s. Shown are results generated for $N_r = 4$, $N_t = 5$, and different values of t_w (left) and results generated for $t_w = 5$ ns and different values of N_t (right), as indicated.

velocity data.

For different time averaging window widths t_w and the same number $N_t = 5$ of samples drawn from the $N_r = 4$ replica MD simulation, the resulting posterior is plotted in Figure 11(left) before (top) and after (bottom) marginalizing over σ^2 . When t_w is increased, the mean value of σ decreases as expected, while the mean value of the inferred velocity u approaches a fixed value. Moreover, the posterior width also decreases because of the noise reduction as discussed in section 4.2.2. When the number N_t of samples drawn from the MD simulations is increased, the distributions for u and σ narrow down around the true values of the long-term averaged MD output velocity and the MD noise level, respectively, as depicted in Figure 11(right).

The uncertainty due to the finite amount of data in the Bayesian inference is reflected in the posterior widths shown in Figure 11 for the different cases considered. In fact, increasing the amount of data in the inference overcomes the ignorance about the inferred quantities that was reflected in the selection of priors.

5.2. Atomistic simulation with uncertain input. A more complicated case study is performed in this section where the atomistic problem has the same geometry and configuration as the previous case but with an uncertain input velocity $U = 10 \pm 1$ m/s. We represent U as a PCE with one stochastic dimension $n = 1$ and an expansion order $p = 1$. Hence we write $U(\xi) = U_0 + U_1\xi$, where ξ has a normal

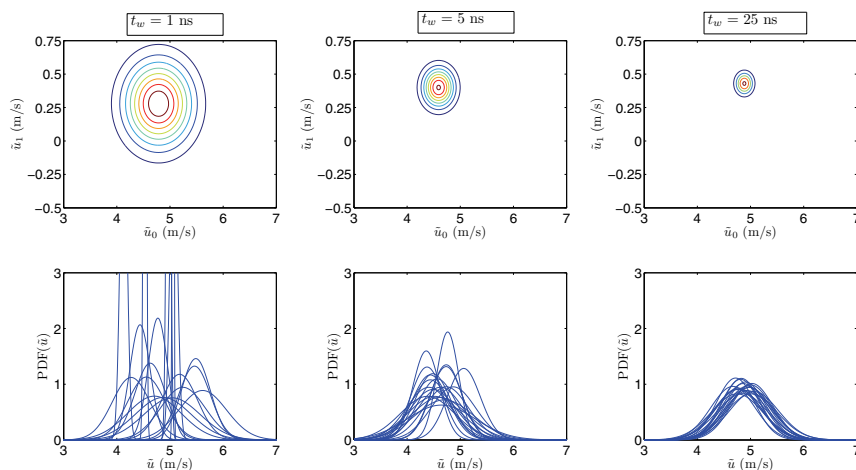


FIG. 12. Plots showing (top row) the joint posterior on $\{\tilde{u}_0, \tilde{u}_1\}$ after marginalizing over σ^2 given by Student- t distributions and (bottom row) the Gaussian PDFs of $\tilde{u}(\xi)$ where different lines indicate different values of ζ in (30) when u is extracted from an atomistic simulation with a stochastic input $U = 10 \pm 1$ m/s. Shown are results generated for $N_q = 5$, $N_r = 4$, $N_t = 5$, and different values of the averaging time window width t_w , as indicated.

distribution. Following the Bayesian inference approach developed in this study, we sample $U(\xi)$ over $N_q = 5$ Gauss–Hermite quadrature points (see section 3.3.2) and infer \tilde{u} based on data drawn from the MD short-time averaged samples.

In Figure 12, we report results of the inference of \tilde{u} for different t_w . After marginalizing over σ^2 , the joint posterior over $\{\tilde{u}_0, \tilde{u}_1\}$ in the top row of Figure 12 exhibits a decrease in its posterior width when t_w is increased for the same reasons discussed in section 4.2.2. There are two sources of uncertainty in this case as described in section 3.3.2. The first one is present in $U(\xi)$, which is imposed on the MD simulation, and the second source is due to the finite amount of data from the MD simulations. In the bottom row of Figure 12, we plot PDFs of $\tilde{u}(\xi)$ for different values of ζ in (30). For short averaging time windows (left column in Figure 12), the PC coefficients of $\tilde{u}(\xi)$ are not well known, leading to strong variability in the PDF of $\tilde{u}(\xi)$. With more averaging, the PCE of $\tilde{u}(\xi)$ is better defined.

5.3. Atomistic simulation with uncertain input coupled to continuum.

In this section, we consider the fully coupled stochastic multiscale system illustrated in Figure 4, where we set the height of the continuum domain to $h = 31.5$ nm and impose a continuum wall velocity of $V = 200$ m/s. The steady state continuum flow field is computed. This configuration has some overlap in length scales between the continuum and particle simulations but a strong separation in time scales. For water, our chosen velocity results in a Reynolds number approximately equal to 6 such that the flow falls in the laminar regime [14] and the linear velocity profile for the continuum solution is valid. The near-wall velocity for the coupled simulation falls within the range used to infer the MD surrogate model. The geometry and configuration of the atomistic simulation is given in Figure 7, and the iterative coupling scheme is illustrated in Figure 4 and Figure 5(bottom). We assume a first order PCE for both u and U ($P + 1 = 2$) and start the coupling algorithm with a deterministic value of $U = 10$ m/s ($\mathbf{U}^{(0)} = \{10, 0\}$). In the first iteration, the propagation of

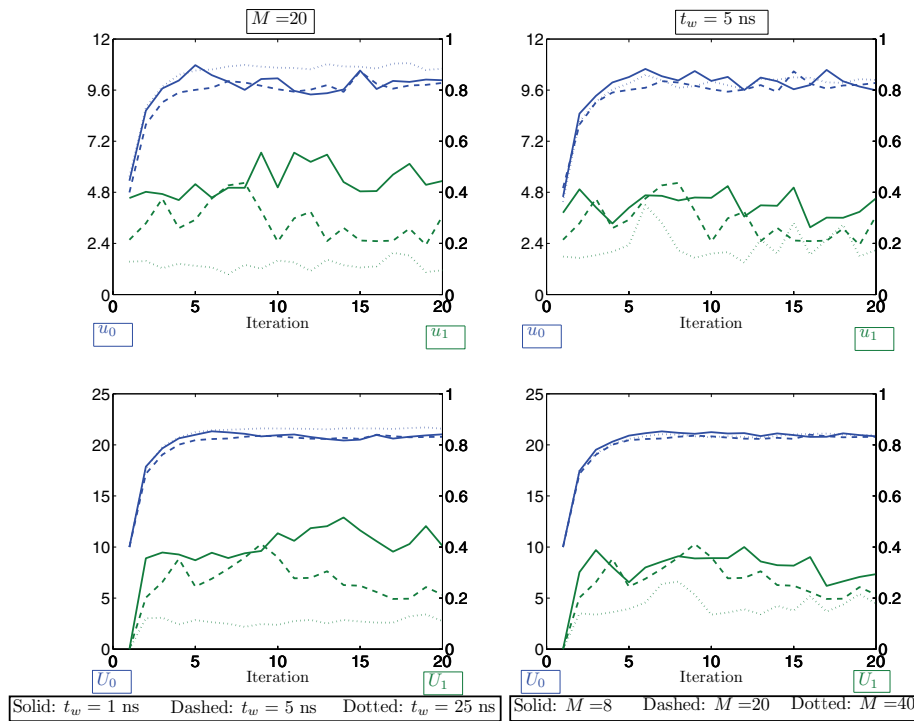


FIG. 13. Plots showing the variables u (top) and U (bottom) exchanged between the atomistic and continuum components of a stochastic multiscale simulation. The mean u_0 (resp., U_0) and standard deviation u_1 (resp., U_1) are reported for a continuum velocity $V = 200$ m/s as a function of the number of iterations for $N_q = 5$ and different values of M and t_w , as indicated.

U in the MD simulation is performed based on the formulation and test case of sections 3.3.1 and 5.1, respectively. In subsequent iterations, U becomes uncertain and is propagated in the atomistic model following the method in section 3.3.2. The data used to build the likelihood function (12) is drawn from the MD surrogate model in this case instead of running new MD simulations. Furthermore, “folding” the two sources of uncertainties is required as discussed in section 3.3.3 in order to obtain a one-dimensional PCE of u , which is propagated in the continuum simulation using the analytical solution (42) to calculate a new PCE for U .

Figure 13 shows at each coupling iteration the predicted mean and standard deviation of the variables u and U exchanged between the atomistic and continuum components. The mean of both variables converges and reaches a statistical steady state after about 10 iterations. The observed fluctuations in subsequent iterations reflect the lack of knowledge of the variables due to the finite amount of data available to update the same prior (improper prior for the PC coefficients and Jeffreys prior for σ^2) at every iteration. Increasing either t_w or M lowers the amplitude of these fluctuations by including additional data in the inference and thereby overrules the priors. Increasing either t_w or M also lowers the standard deviation of the predicted variables as expected and as discussed previously.

5.4. Convergence acceleration.

5.4.1. Sequential Bayesian updating. With the iterative nature of the coupled system, one can effectively incorporate information from all previous iterations by picking more informative priors at every step. We employ sequential Bayesian updating (SBU) [25, 46] to take into account additional data in the inference as the system evolves, resulting in less noisy inference of the computed variables. SBU is suitable in the stochastic coupling algorithm since the prediction of the PCEs of the variables u and U evolves and invokes new data sampled from the MD simulation surrogate at each iteration. We implement SBU by replacing the prior in each iteration by the posterior obtained from the previous iteration. The joint prior at an iteration i becomes

$$(51) \quad \mathcal{P}(\tilde{\mathbf{u}}, \sigma^2)^{(i)} = \mathcal{P}(\tilde{\mathbf{u}}, \sigma^2 | \mathbf{d})^{(i-1)}.$$

This joint prior, being the posterior from the previous iteration, is a multivariate NSIG distribution as shown in section 3.3.2. We can prove that the updated marginalized posterior is also a multivariate Student-t distribution. In principle, the SBU should be started when the evolving process reaches a statistical steady state in order to reduce the noise in the variable of interest by providing more sampling data. Assuming that SBU is initiated at the second iteration, the degrees of freedom, mean, and covariance matrix at an iteration i are given by

$$(52) \quad \begin{aligned} \gamma^{(i)} &= (i-1)N - (P+1), \\ \boldsymbol{\lambda}^{(i)} &= \nu^{-1} \underline{\mathbf{Q}}^T \mathbf{d}'^{(i)}, \\ \underline{\Sigma}^{(i)} &= \frac{2b'^{(i)}}{\gamma^{(i)} - 2} \underline{\nu}^{-1}. \end{aligned}$$

Note that in this section we are using the superscript with parentheses to denote the index of an iteration. The vector \mathbf{d}' and the scalar b' at an iteration $i > 1$ are calculated as

$$(53) \quad \begin{aligned} \mathbf{d}'^{(i)} &= \frac{1}{i-1} \sum_{j=i_{SBU}}^i \mathbf{d}^{(j)}, \\ b'^{(i)} &= -\frac{1}{2} \left(\mathbf{d}'^T \underline{\mathbf{Q}} \boldsymbol{\lambda}^{(i)} - \frac{1}{i-1} \sum_{j=i_{SBU}}^i \mathbf{d}^{(j)T} \mathbf{d}^{(j)} \right). \end{aligned}$$

In Figure 14, we plot the inferred posterior over $\tilde{\mathbf{u}}$ after marginalizing over σ^2 , the PDF of u , and the obtained PDF of U after propagation in the continuum simulation. The plot shows these results after 10 iterations with and without SBU. The results suggest that the prediction of u and U is significantly less noisy when SBU is used. By choosing the prior of the inferred variable at iteration i as the posterior obtained at iteration $i-1$, the accumulated knowledge of the PCE of u up to that point is maintained for iteration i and enhanced with the inclusion of new data through a new Bayesian update step using likelihood function (12). This implies that SBU has an effect similar to increasing the number of samples M drawn from the MD surrogate model, as discussed in section 5.1. However, more samples M require more expensive computations (see (27)). Hence, SBU mitigates the computational cost while enhancing the prediction of u and U since their posterior width is decreasing monotonically with every iteration.

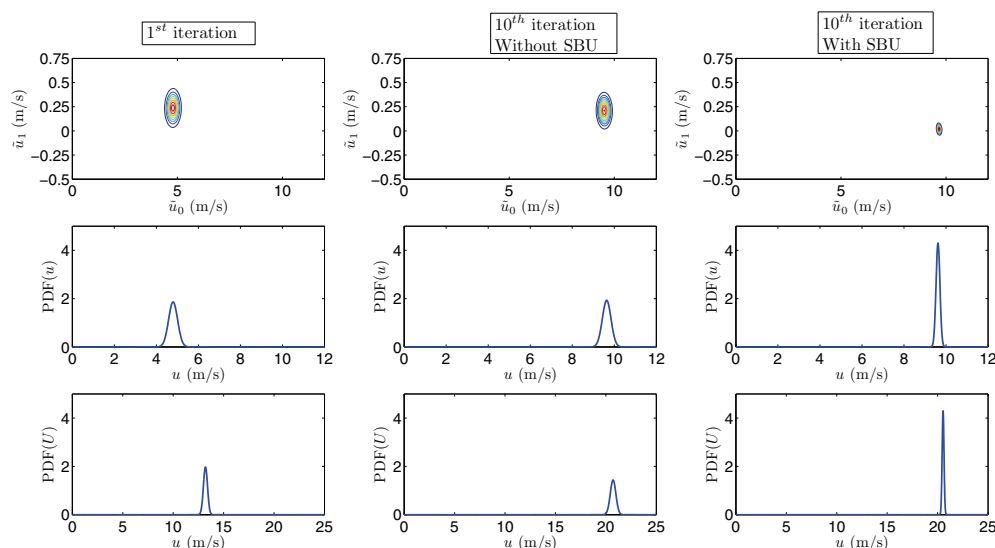


FIG. 14. Plots showing (top row) the Student- t joint posterior on \tilde{u} after marginalizing over σ^2 , (middle row) the PDF of u after “folding” the two sources of uncertainty, and (bottom row) the PDF of U after propagating u in the continuum simulation. Results are reported for $N_q = 5$, $M = 20$, and $t_w = 5$ ns. Shown are results obtained at the 1st iteration and the 10th iteration, with and without SBU, as indicated.

Similarly to section 5.3, we report in Figure 15 the predicted mean and standard deviation of the variables u and U exchanged between the atomistic and continuum components at different iterations. When SBU is used, the fluctuations in the mean of u and U are suppressed. Relative to the case when SBU is not invoked (see Figure 13), the standard deviations u_1 and U_1 are significantly reduced in consecutive iterations through the inclusion of new data that decreases the uncertainty. For given t_w and M , the additional data included in the inference with SBU results in an accuracy similar to the case without SBU when larger t_w and M are used. This can be observed by looking at the trends of the standard deviation u_1 in Figures 13 and 15. After 20 coupling iterations, the coefficients of variation (COVs) for U and u are about 0.5 and 0.25 %, respectively.

5.4.2. Limitation of SBU. The accuracy of the surrogate with respect to the original MD simulations should be taken into consideration when implementing SBU in the stochastic coupling algorithm. Based on section 5.4.1, the uncertainty in the estimated u is reduced by SBU with each iteration, and eventually it will approach the accuracy of the surrogate model itself, as characterized by the matrix $\underline{\Theta}$ discussed in section 4.2.2. At that point, iterations should be ceased, as further use of SBU would generate a u value that is ostensibly more accurate than the surrogate, which is not realistic.

The accuracy of the MD simulation surrogate model is determined by how well it can express the mean of the short-time averaged velocity extracted from the MD simulations. This accuracy is closely related to the amount of short-time averaged velocity data used to infer the surrogate. Hence, we study in this section the limit beyond which the accuracy brought by SBU approaches the intrinsic accuracy of the surrogate model described in section 4.2.2. We can compute the uncertainty in

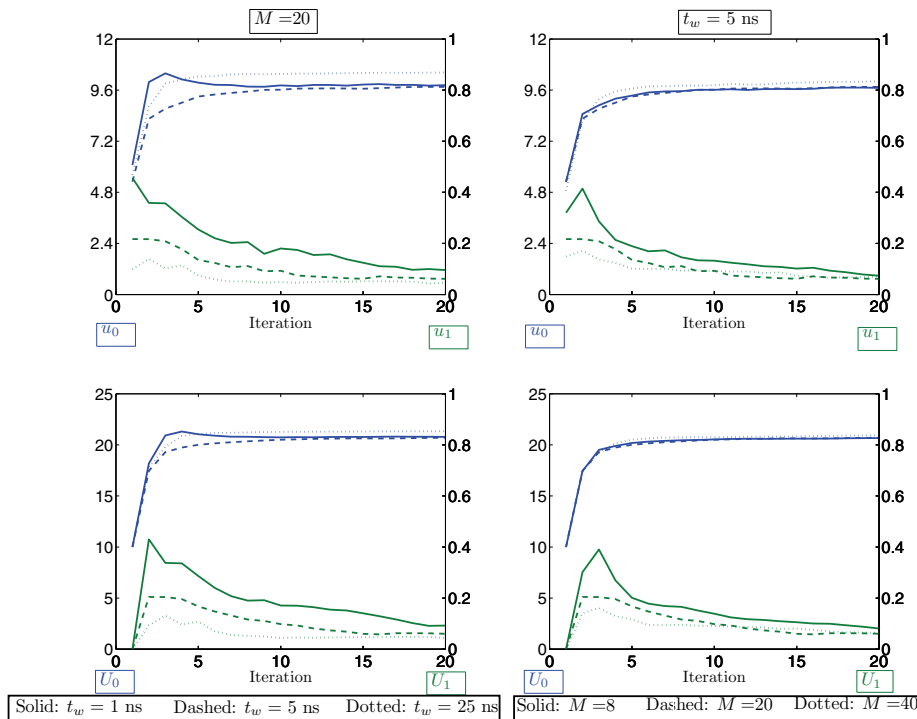


FIG. 15. Plots showing the variables u (top) and U (bottom) exchanged between the atomistic and continuum components of a stochastic multiscale simulation. The mean u_0 (resp., U_0) and standard deviation u_1 (resp., U_1) are reported for a continuum velocity $V = 200$ m/s as a function of the number of iterations when SBU is implemented for $N_q = 5$ and different values of M and t_w , as indicated.

the velocity mean (over η) \bar{v} evaluated using the surrogate model by combining (47) and (49) and focusing on the posterior over $\{v_0, v_1\}$. We obtain

$$(54) \quad \bar{v} = (\bar{v}_0 + \bar{v}_1 U_0^*) + (\theta_{11} + \theta_{21} U_0^*) \xi_0 + \theta_{22} \xi_1 U_0^*,$$

where $U_0^* = \frac{U_0 - \bar{U}_s}{U_{s,\max} - U_{s,\min}}$ and U_0 is the mean of the velocity U evaluated at any given iteration. The last two terms represent the uncertainty in \bar{v} . Thus, defining χ as the standard deviation of $\bar{v}(\xi_0, \xi_1)$, we have

$$(55) \quad \chi = \sqrt{(\theta_{11} + \theta_{21} U_0^*)^2 + \theta_{22}^2 U_0^{*2}}.$$

As seen in Figure 15, U_0 converges to a constant value with an increasing number of iterations. Hence χ , being solely dependent on U_0 , also converges to a constant for a given surrogate. On the other hand, the uncertainty in the velocity u estimated by the stochastic coupling algorithm is given by u_1 in our case when a first order expansion is used. In Figure 16(solid line), we plot χ/u_1 as a function of the coupling iterations. For all values N_{MD} of MD data points used to generate the surrogate model, the ratio χ/u_1 increases as a function of the number of iterations. This is because χ quickly approaches a constant value and u_1 decreases (see Figure 15). When the threshold $\chi/u_1 = 1$ is reached, the accuracy of the inferred coupling velocity is of the same order

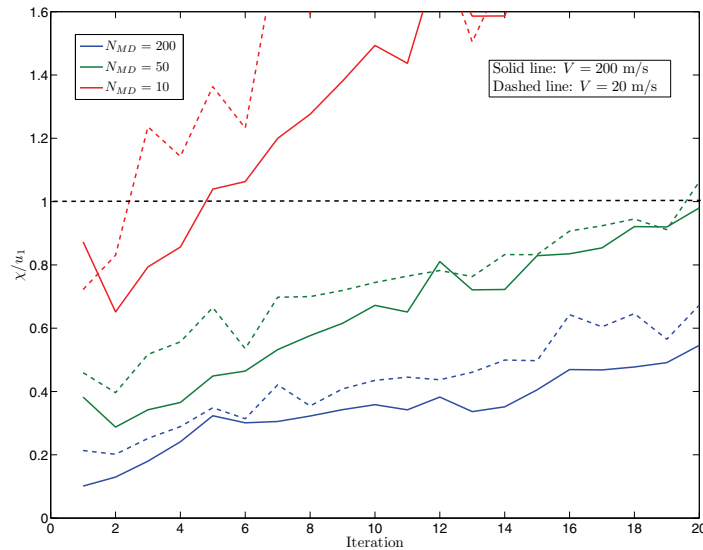


FIG. 16. A plot showing the ratio of the uncertainty χ in estimating the mean of the velocity v evaluated by the surrogate and the uncertainty u_1 in the computed velocity u . Plotted are ratios χ/u_1 obtained for a coupled stochastic atomistic-to-continuum Couette flow when continuum wall velocities $V = 200$ m/s (solid line) and $V = 20$ m/s (dashed line) are imposed on the simulation at subsequent coupling iterations with $N_q = 5$, $M = 20$, and $t_w = 5$ ns for surrogate models generated with different numbers N_{MD} of MD data points, as indicated.

as the surrogate model. This means that beyond this point, SBU does not offer any additional advantage towards estimating the mean u of the original MD data. For the case with $N_{MD} = 200$, which was used to produce all results in this paper, χ/u_1 remains below the accuracy threshold even if more than 20 iterations are performed. However, if a “poor” surrogate were to be used (e.g., $N_{MD} = 10$), this threshold would be reached at a very early stage.

6. Application to a low-speed Couette flow. Even though the continuum velocity $V = 200$ m/s considered in the previous section corresponds to a laminar flow, it may be too high in many practical applications. In this section we report results of an atomistic simulation coupled to continuum with a continuum wall velocity $V = 20$ m/s, an order of magnitude lower than the value considered in the previous section. This continuum velocity corresponds to a velocity U imposed on the MD simulation of about 2 m/s for the coupled simulation. We consider the same MD box size, number of LJ particles, and simulation parameters as previously, resulting in a similar MD noise level. Thus, for the same averaging time window width, the signal to noise ratio in the atomistic-to-continuum coupling variables will be much lower. Hence, we also explore in this section the capability of the stochastic coupling algorithm to handle low signal to noise ratios.

As before, we apply the VV coupling scheme (see section 3.2), Bayesian inference of the MD simulation output variable (see sections 3.2 and 3.3.3), inference of a surrogate model of the MD simulations (see section 4.2), and SBU (see section 5.4.1).

For the surrogate, we choose $N_s = 10$ values for the input velocity $0 \leq U_s \leq 4.5$

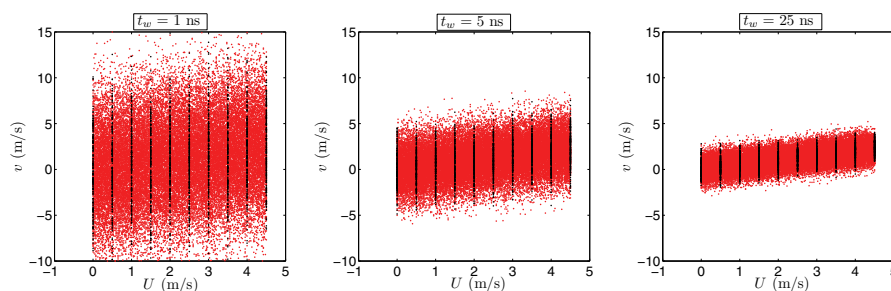


FIG. 17. Plots showing results of the MD simulation surrogate. The black dots represent short-time averaged velocity data obtained from MD simulations for the selected range of the input velocity $0 \leq U_s \leq 4.5$ m/s, while the red dots are the samples of the output velocity v computed by the surrogate model. Plots are generated for $N_{MD} = 200$ and for different noise levels, i.e., different MD time averaging windows, as indicated. Dot colors are distinguishable only in the online version.

m/s and perform the corresponding MD simulations. Similarly to Figure 9, we plot in Figure 17 MD data and data drawn from the surrogate in the given velocity range. Again, we notice a very good agreement between the MD data (black dots) and the data drawn from the surrogate (red dots).

With increasing N_{MD} and t_w , \bar{v}_0 and \bar{v}_1 quickly reach a steady value as depicted in Figure 18(top row). Note that after inferring v_0 and v_1 and setting $U = 0$, $\bar{v}(0) \approx 0.03$ m/s is very small compared to the considered velocity input range $0 \leq U_s \leq 4.5$ m/s, also as expected. The measure κ (see section 4.2.2) of the posterior width decreases with N_{MD} and t_w consistently with the CLT. For $N_{MD} = 200$, $\kappa/\bar{\sigma} < 5\%$, and hence we infer the surrogate model with this value of N_{MD} and use it in the stochastic coupling simulations.

We study the response of the fully coupled stochastic atomistic-to-continuum simulation with the fixed-point VV iteration scheme described in section 3.5. Even though the mean and standard deviation of the coupling variables u and U quickly approach a statistical steady state as shown in Figure 19, the strong fluctuations observed in subsequent iterations indicate higher uncertainty levels. This suggests that SBU is needed for this case in order to smooth out the fluctuations in the mean response and to reduce the uncertainty in the predictions of the output variables.

Figure 20 shows the response of the coupled system when SBU is used. As expected, smoother curves are obtained for the mean of u and U while the standard deviation decreases with the iterations due to the accumulation of data brought by SBU. The stochastic coupling algorithm developed in this work is therefore effective for this case of lower continuum input velocity, where finite MD sampling noise is more important.

It is important to define an appropriate convergence or stopping criterion for the coupling algorithm for the variables of interest, which are in this case the velocities u and U exchanged between the atomistic and continuum models. A simple convergence criterion is when the relative change in the mean U_0 is below a certain threshold. Or, one can keep iterating until the standard deviation (uncertainty) is reduced below a desired value. However, regardless of the convergence and/or stopping criterion used, it is a prerequisite that $\chi/u_1 \leq 1$. This ratio, defined in section 5.4.2, indicates how close the accuracy brought by SBU is to the surrogate accuracy. For $V = 20$ m/s, χ/u_1 is plotted in Figure 16(dashed line) for different values of N_{MD} . This plot suggests that for $N_{MD} = 200$ and after 20 iterations, SBU does not yet reach the

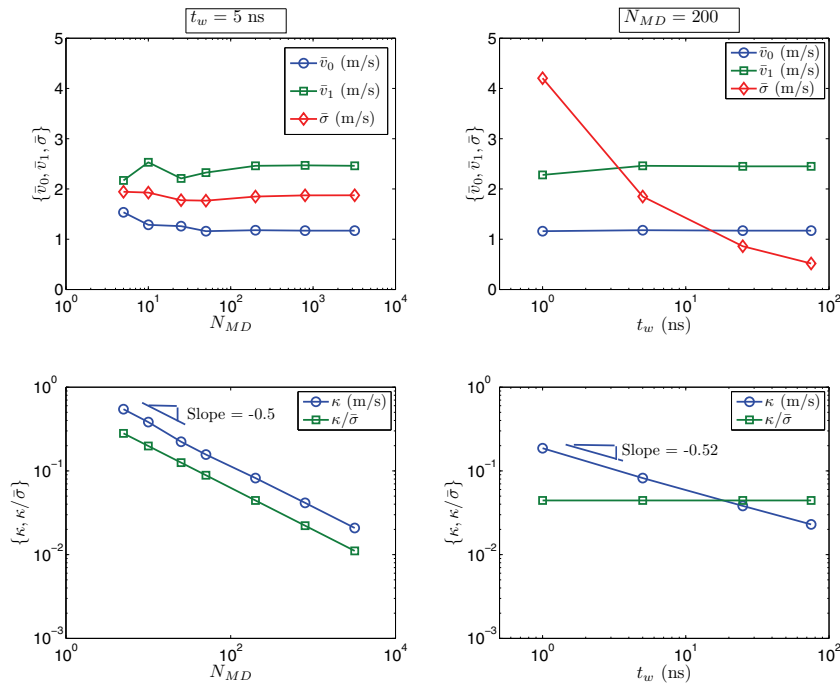


FIG. 18. Plots showing the variation of the inferred variables $\{\bar{v}_0, \bar{v}_1, \bar{\sigma}\}$ (top row) as a function of the number of short-time averaged MD data N_{MD} for $t_w = 5$ ns (left column) and the time averaging window width t_w (right column) for $N_{MD} = 200$. Also plotted are the measures of the inferred posterior width $\{\kappa, \kappa/\bar{\sigma}\}$ (bottom row).

surrogate accuracy threshold, and that more iterations could be performed in order to further reduce the uncertainty in u and U .

After 20 coupling iterations, the COV for U is about 5%. While this is higher than the ratio found when $V = 200$ m/s (see Figure 15), it is still an acceptably low COV, indicating that the stochastic coupling algorithm with SBU accurately estimates the variables u and U exchanged between the atomistic and continuum components of the multiscale simulation. Moreover, if a higher accuracy is desired, the current approach readily informs the additional MD simulations and data requirements needed to achieve this.

7. Conclusion. In this paper we have presented a systematic approach to quantifying uncertainties from atomistic simulations using Bayesian inference with a focus on uncertainty due to the sampling noise in MD simulations. This uncertainty extraction formalism is applicable to any type of particle simulation and without any prior knowledge of its physical parameters. An iterative stochastic coupling algorithm that evaluates the variables exchanged in multiscale simulations with strong separation in time and/or length scales has been developed. The posterior over the inferred variables was derived analytically. The uncertainties due to finite MD sampling at successive iterations have been “folded” into one stochastic dimension. Thus, the uncertainty in the variable extracted from the atomistic component was represented in terms of a PCE. This methodology can be extended to extract a set of continuum

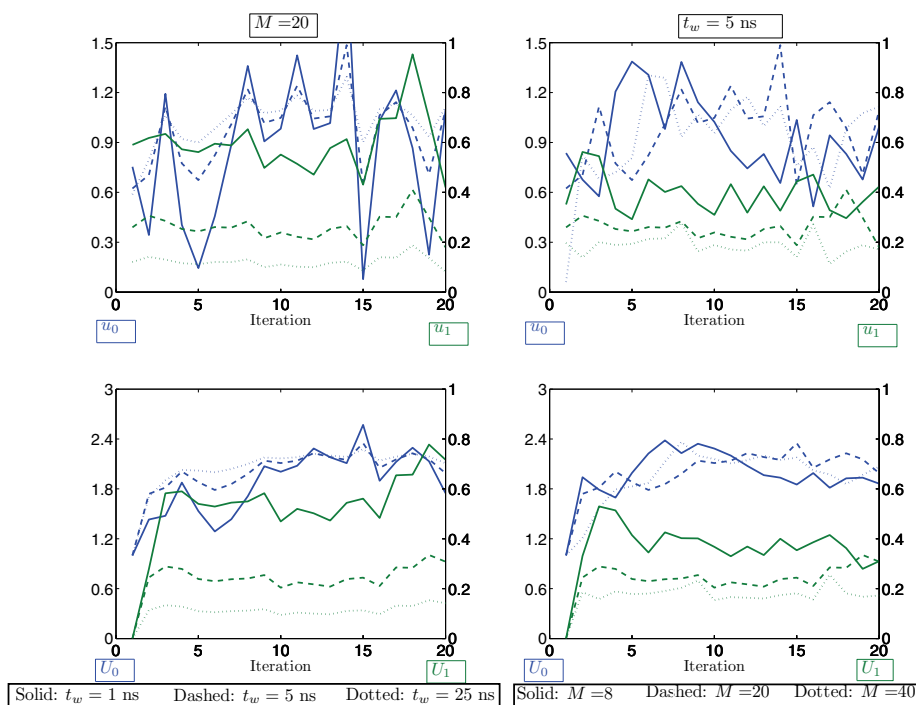


FIG. 19. Plots showing the variables u (top) and U (bottom) exchanged between the atomistic and continuum components of a stochastic multiscale simulation. The mean u_0 (resp., U_0) and standard deviation u_1 (resp., U_1) are reported for a continuum velocity $V = 20$ m/s as a function of the number of iterations for $N_q = 5$ and different values of M and t_w , as indicated.

observables instead of one variable or to account for multiple sources of uncertainty. The number of stochastic dimensions in the underlying PCE will increase accordingly and can incorporate possible dependencies between the output observables. Handling more dimensions is a subject currently under investigation by our research group. The burden of the MD simulations required at each iteration was significantly alleviated by inferring a surrogate model that accurately reflects MD results. Major findings of this study can be summarized as follows:

1. The surrogate model produces data that mimic original MD simulation data within a given degree of accuracy. This accuracy is governed by the amount of data used to infer the surrogate model parameters. The uncertainty in these parameters should be small compared to the amplitude of the noise due to MD finite sampling in order to have a high fidelity surrogate.
2. The stochastic coupling algorithm convergence and stability properties depend on the exchanged variables. For the simple Couette flow considered, a velocity-velocity fixed-point iteration scheme converges in a reasonable number of iterations.
3. Increasing the number of samples drawn from the surrogate and/or increasing the MD time averaging window improves the accuracy of the coupling variables in the atomistic-continuum simulation by narrowing the width of inferred posteriors and thereby decreasing the uncertainty in inferred quantities.

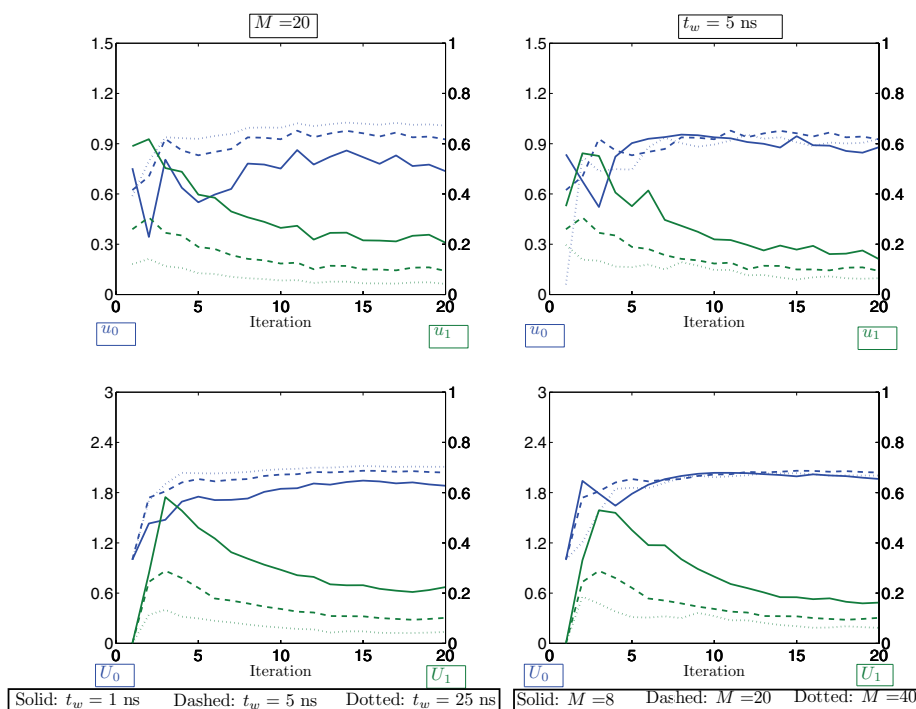


FIG. 20. Plots showing the variables u (top) and U (bottom) exchanged between the atomistic and continuum components of a stochastic multiscale simulation. The mean u_0 (resp., U_0) and standard deviation u_1 (resp., U_1) are reported for a continuum velocity $V = 20$ m/s as a function of the number of iterations when SBU is implemented for $N_q = 5$ and different values of M and t_w , as indicated.

4. When SBU is used, the stochastic atomistic-to-continuum coupling algorithm provides less noisy estimates of the exchanged variables even with a high noise to mean ratio in the data drawn from MD simulations.
5. The additional accuracy brought by SBU has to be checked against the original accuracy of the MD surrogate model at every coupling iteration, in order to detect the threshold when SBU cannot provide additional information to predict the mean of the original MD data. A high accuracy surrogate is required, in order to delay reaching this threshold until a satisfactory solution is obtained.

REFERENCES

- [1] M. ABRAMOWITZ AND I. A. STEGUN, *Handbook of Mathematical Functions*, Dover, New York, 1970.
- [2] H. ADALSTEINSSON, B. J. DEBUSSCHERE, K. R. LONG, AND H. N. NAJM, *Components for atomistic-to-continuum multiscale modeling of flow in micro- and nanofluidic systems*, Sci. Program., 16 (2008), pp. 297–313.
- [3] F. J. ALEXANDER, A. L. GARCIA, AND D. M. TARTAKOVSKY, *Algorithm refinement for stochastic partial equations: I. Linear diffusion*, J. Comput. Phys., 182 (2002), pp. 47–66.
- [4] F. J. ALEXANDER, A. L. GARCIA, AND D. M. TARTAKOVSKY, *Algorithm refinement for stochastic partial equations: II. Correlated systems*, J. Comput. Phys., 207 (2005), pp. 769–787.
- [5] M. P. ALLEN AND D. J. TIDLESLEY, *Computer Simulation of Liquids*, Oxford Sci. Publ., Oxford

- University Press, New York, 1989.
- [6] A. W. BOWMAN AND A. AZZALINI, *Applied Smoothing Techniques for Data Analysis*, Oxford University Press, New York, 1997.
 - [7] R. H. CAMERON AND W. T. MARTIN, *The orthogonal development of nonlinear functionals in series of Fourier-Hermite functionals*, Ann. of Math. (2), 48 (1947), pp. 385–392.
 - [8] B. J. DEBUSSCHERE, H. N. NAJM, A. MATTA, O. M. KNIO, R. G. GHANEM, AND O. P. LE MAÎTRE, *Protein labeling reactions in electrochemical microchannel flow: Numerical simulation and uncertainty propagation*, Phys. Fluids, 15 (2003), pp. 2238–2250.
 - [9] B. J. DEBUSSCHERE, H. N. NAJM, P. P. PÉBAY, O. M. KNIO, R. G. GHANEM, AND O. P. LE MAÎTRE, *Numerical challenges in the use of polynomial chaos representations for stochastic processes*, SIAM J. Sci. Comput., 26 (2004), pp. 698–719.
 - [10] R. DELGADO-BUSCALIONI AND P. V. COVENEY, *Continuum-particle hybrid coupling for mass, momentum, and energy transfers in unsteady fluid flow*, Phys. Rev. E (3), 67 (2003), 046704.
 - [11] J. E. DENNIS, JR., AND R. B. SCHNABEL, *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*, SIAM, Philadelphia, 1996.
 - [12] F. DOMINICI, G. PARMIGIANI, AND M. CLYDE, *Conjugate analysis of multivariate normal data with incomplete observations*, Canad. J. Statist., 28 (2000), pp. 533–550.
 - [13] A. DONEV, J. B. BELL, A. L. GARCIA, AND B. J. ALDER, *A hybrid particle-continuum method for hydrodynamics of complex fluids*, Multiscale Model. Simul., 8 (2010), pp. 871–911.
 - [14] H.-S. DOU AND B. C. KHOO, *Investigation of turbulent transition in plane Couette flows using energy gradient method*, Adv. Appl. Math. Mech., 3 (2011), pp. 165–180.
 - [15] A. GELMAN, J. B. CARLIN, H. S. STERN, AND D. B. RUBIN, *Bayesian Data Analysis*, Chapman and Hall/CRC, Boca Raton, FL, 2004.
 - [16] A. GELMAN, X. L. MENG, AND H. STERN, *Posterior predictive assessment of model fitness via realized discrepancies*, Statist. Sinica, 6 (1996), pp. 733–807.
 - [17] R. GHANEM, *Probabilistic characterization of transport in heterogeneous media*, Comput. Methods Appl. Mech. Engrg., 158 (1998), pp. 199–220.
 - [18] R. GHANEM, *Ingredients for a general purpose stochastic finite element formulation*, Comput. Methods Appl. Mech. Engrg., 168 (1999), pp. 19–34.
 - [19] R. G. GHANEM AND P. D. SPANOS, *Stochastic Finite Elements: A Spectral Approach*, Springer-Verlag, New York, 1991.
 - [20] G. GRIMMETT AND D. STIRZAKER, *Probability and Random Processes*, Oxford University Press, New York, 2011.
 - [21] N. HADJICONSTANTINO, *Hybrid atomistic-continuum formulations and the moving contact-line problems*, J. Comput. Phys., 154 (1999), pp. 245–265.
 - [22] N. HADJICONSTANTINO AND A. T. PATERA, *Heterogeneous atomistic-continuum representation for dense fluid systems*, Internat. J. Modern Phys. C, 8 (1997), pp. 967–976.
 - [23] N. G. HADJICONSTANTINO, A. L. GARCIA, M. Z. BAZANT, AND G. HE, *Statistical error in particle simulations of hydrodynamic phenomena*, J. Comput. Phys., 187 (2003), pp. 274–297.
 - [24] S. JANSON, *Gaussian Hilbert Spaces*, Cambridge University Press, Cambridge, UK, 1997.
 - [25] J. KAIPIO AND E. SOMERSALO, *Statistical and Computational Inverse Problems*, Springer-Verlag, New York, 2005.
 - [26] D. KIM, B. J. DEBUSSCHERE, AND H. N. NAJM, *Spectral methods for parametric sensitivity in stochastic dynamical systems*, Biophys. J., 92 (2007), pp. 379–393.
 - [27] E. M. KOTSALIS, J. H. WALTHER, AND P. KOUMOUTSAKOS, *Control of density fluctuations in atomistic-continuum simulations of dense liquids*, Phys. Rev. E (3), 76 (2007), 016709.
 - [28] E. M. KOTSALIS, J. H. WALTHER, AND P. KOUMOUTSAKOS, *Hybrid atomistic-continuum method for the simulation of dense fluid flows*, J. Comput. Phys., 205 (2005), pp. 373–390.
 - [29] O. P. LE MAÎTRE AND O. M. KNIO, *Spectral Methods for Uncertainty Quantification with Applications to Computational Fluid Dynamics*, Springer-Verlag, New York, 2010.
 - [30] O. P. LE MAÎTRE, O. M. KNIO, H. N. NAJM, AND R. G. GHANEM, *A stochastic projection method for fluid flow. I. Basic formulation*, J. Comput. Phys., 173 (2001), pp. 481–511.
 - [31] O. P. LE MAÎTRE, M. T. REAGAN, H. N. NAJM, R. G. GHANEM, AND O. M. KNIO, *A stochastic projection method for fluid flow. II. Random process*, J. Comput. Phys., 181 (2002), pp. 9–44.
 - [32] A. R. LEACH, *Molecular Modelling: Principles and Applications*, Pearson, Harlow, UK, 2001.
 - [33] J. LI, D. LIAO, AND S. YIP, *Coupling continuum to molecular dynamics simulation reflecting particle method and the field estimator*, Phys. Rev. E (3), 57 (1998), pp. 7259–7267.
 - [34] J. LI, D. LIAO, AND S. YIP, *Imposing field boundary conditions in MD simulation of fluids: Optimal particle controller and buffer zone feedback*, in Materials Research Society Sym-

- posium Proceedings, 1998, pp. 473–478.
- [35] J. LI, D. LIAO, AND S. YIP, *Nearly exact solution for coupled continuum/MD fluid simulation*, J. Comput. Aided Mater. Des., 6 (1999), pp. 95–102.
 - [36] S. M. LYNCH AND B. WESTERN, *Bayesian posterior predictive checks for complex models*, Sociol. Meth. Res., 32 (2004), pp. 301–335.
 - [37] Y. M. MARZOUK, H. N. NAJM, AND L. A. RAHNA, *Stochastic spectral methods for efficient Bayesian solution of inverse problems*, J. Comput. Phys., 224 (2007), pp. 560–586.
 - [38] K. M. MOHAMED AND A. A. MOHAMAD, *A review of the development of hybrid atomistic-continuum methods for dense fluids*, Microfluid. Nanofluid., 8 (2010), pp. 283–302.
 - [39] A. MOHAMMAD-DJAFARI, *Bayesian inference for inverse problems*, in Bayesian Inference and Maximum Entropy Methods in Science and Engineering, AIP Conf. Proc. 617, American Institute of Physics, Melville, NY, 2002, pp. 477–496.
 - [40] H. N. NAJM, *Uncertainty quantification and polynomial chaos techniques in computational fluid dynamics*, in Annual Review of Fluid Mechanics, Volume 41, Annual Reviews, Palo Alto, CA, 2009, pp. 35–52.
 - [41] J. NOCEDAL AND S. J. WRIGHT, *Numerical Optimization*, Springer-Verlag, New York, 1999.
 - [42] S. T. O'CONNELL AND P. A. THOMPSON, *Molecular dynamics-continuum hybrid computations: A tool for studying complex fluid flows*, Phys. Rev. E (3), 52 (1992), pp. R5792–R5795.
 - [43] S. J. PLIMPTON, *Fast parallel algorithms for short-range molecular dynamics*, J. Comput. Phys., 117 (1995), pp. 1–19.
 - [44] M. T. REAGAN, H. N. NAJM, R. G. GHANEM, AND O. M. KNIO, *Uncertainty quantification in reacting flow simulations through non-intrusive spectral projection*, Combust. Flame, 132 (2003), pp. 545–555.
 - [45] W. REN, *Analytical and numerical study of coupled atomistic-continuum methods for fluids*, J. Comput. Phys., 227 (2007), pp. 1353–1371.
 - [46] C. RIGGELSEN, *Approximation Methods for Efficient Learning of Bayesian Networks*, IOS Press, Amsterdam, 2008.
 - [47] D. ROWE, *Multivariate Bayesian Statistics: Models for Source Separation and Signal Unmixing*, Chapman and Hall/CRC, Boca Raton, FL, 2003.
 - [48] H. A. SCHWARTZ, *Über einen grenzübergang durch alternierendes verfahren*, Vierteljahrsschrift der Naturforschenden Gesellschaft in Zürich, 15 (1870), pp. 272–286.
 - [49] D. S. SIVIA AND C. J. CARLILE, *Molecular-spectroscopy and Bayesian spectral-analysis—how many lines are there*, J. Chem. Phys., 96 (1992), pp. 170–178.
 - [50] D. S. SIVIA AND J. SKILLING, *Data Analysis: A Bayesian Tutorial*, 2nd ed., Oxford Sci. Publ., Oxford University Press, New York, 2006.
 - [51] W. N. VENABLES AND B. D. RIPLEY, *Modern Applied Statistics with S*, Springer-Verlag, New York, 2002.
 - [52] S. WIENER, *The homogeneous chaos*, J. Comput. Phys., 224 (2007), pp. 560–586.
 - [53] D. XIU AND G. E. KARNIADAKIS, *Modeling uncertainty in steady state diffusion problems via generalized polynomial chaos*, Comput. Methods Appl. Mech. Engrg., 191 (2002), pp. 4927–4948.
 - [54] D. XIU AND G. E. KARNIADAKIS, *The Wiener–Askey polynomial chaos for stochastic differential equations*, SIAM J. Sci. Comput., 24 (2002), pp. 619–644.
 - [55] D. XIU, D. LUCOR, C.-H. SU, AND G. E. KARNIADAKIS, *Stochastic modeling of flow-structure interactions using generalized polynomial chaos*, ASME J. Fluids Eng., 124 (2002), pp. 51–59.
 - [56] S. YASUDA AND R. YAMAMOTO, *A model for hybrid simulations of molecular dynamics and computational fluid dynamics*, Phys. Fluids, 20 (2008), 113101.