

Quantifying Sampling Noise and Parametric Uncertainty in Atomistic-to-Continuum Simulations using Surrogate Models

Maher Salloum¹, Khachik Sargsyan², Reese Jones³,
Habib N. Najm², Bert Debusschere².

September 29, 2014

¹ Sandia National Laboratories, 7011 East Ave., MS 9158, Livermore, CA 94550, USA

² Sandia National Laboratories, 7011 East Ave., MS 9051, Livermore, CA 94550, USA

³ Sandia National Laboratories, 7011 East Ave., MS 9404, Livermore, CA 94550, USA

{mnsallo, ksargsy, rjones, hnnajm, bjdebus}@sandia.gov

Contents

1	Introduction	3
2	Problem Setup	7
2.1	Model Geometry	7
2.2	MD Numerical Implementation	9
2.3	Variable Exchange between the Atomistic and Continuum Simulations . . .	11

3	Mathematical Formulation	14
3.1	Polynomial Chaos Expansions	14
3.2	Atomistic Component: Inference of a Polynomial Response Surface	16
3.3	Continuum Component: Linear Relationship	21
4	Application to Coupled Atomistic-Continuum Couette Flow with Sampling Noise	22
5	Application to Coupled Atomistic-Continuum Couette Flow with Sampling Noise and Parametric Uncertainty	27
5.1	Solution Method	29
5.1.1	Intrusive Spectral Projection (ISP)	29
5.1.2	Non-Intrusive Spectral Projection (NISP)	30
5.2	Results	31
5.2.1	Convergence Study	34
5.2.2	Sensitivity Analysis	37
6	Conclusion	38
7	Acknowledgements	40
8	References	40

Abstract

We present a methodology to assess the predictive fidelity of multiscale simulations by incorporating uncertainty in the information exchanged between the components of an atomistic-to-continuum simulation. We account for both the uncertainty due to finite sampling in molecular dynamics (MD) simulations and the uncertainty in the physical parameters of the model. Using Bayesian inference, we represent the expensive atomistic component by a surrogate model that relates the long-term output of the atomistic simulation to its uncertain inputs. We then present algorithms

to solve for the variables exchanged across the atomistic-continuum interface in terms of Polynomial Chaos Expansions (PCEs). We consider a simple Couette flow where velocities are exchanged between the atomistic and continuum components, while accounting for uncertainty in the atomistic model parameters and the continuum boundary conditions. Results show convergence of the coupling algorithm at a reasonable number of iterations. The uncertainty in the obtained 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.

1 Introduction

A wide variety of devices, such as physical nanodevices and micro-fluid systems, achieve their function through phenomena that operate on a range of time and length scales. To properly resolve the key phenomena on all relevant scales, multiscale methods are required [1–7], often coupling different physical models across scales, such as atomistic models (*e.g.* Molecular Dynamics (MD)) and continuum macroscale formulations [8–10].

While recent years have brought tremendous progress in the field of multiscale modeling [1–7], a key requirement for predictive simulations is to also quantify the uncertainty in the multiscale simulation results. Uncertainty in multiscale simulations stems from input parameter, initial condition, and boundary condition uncertainties; from modelling assumptions on each resolved scale level; but also from modelling assumptions in the coupling between the different scales [11–14]. Uncertainty Quantification (UQ) in multiscale simulations is an active research topic, with many approaches being developed for the various types of multiscale coupling [11–13, 15, 16]. Most of the current research appears to focus on UQ on individual scale levels, or on one-way propagation of uncertainty across scales. More research is needed on requisite two-way coupling of uncertainty across scales, especially when different physical models are involved for each scale [17].

Recent work on uncertainty quantification in coupled multiphysics models has focused on dimensionality reduction in the information shared at the interface between different model components [18, 19]. More specific to coupled atomistic to continuum simulations, we have studied two-way coupling between uncertainties across the scale interface. In [14], we studied the effect of finite sampling on the atomistic level, which leads to uncertainty in the macroscale quantities that are extracted from MD simulations and passed on to the continuum level. The current paper extends this formulation to also account for parametric uncertainties on both the atomistic and continuum levels.

The class of atomistic-continuum multiscale problems studied in this work is schematically represented in Figure 1, along with the various flows of information between the scales. 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. A further source of uncertainty is the fact that macroscale (*i.e.* continuum-level) observables extracted from the MD simulation through averaging generally have uncertainty due to the finite amount of MD samples available, given the relatively high computational cost of MD simulations [14]. The atomistic and continuum models exchange information across a multiscale interface indicated by the dotted line in Figure 1. The exchanged information is conceptually represented in terms of two variables u and v . The variable v extracted from the MD simulation is fed to the continuum component of the system while u is imposed on the MD simulation. For the present discussion, we assume that the relevant outputs of the overall, coupled atomistic-continuum simulation, are macroscale observables, extracted from the continuum model. The uncertainty in the parameters and boundary conditions on all scale levels, as well as the sampling noise in macroscale information extracted from the atomistic model contribute to the uncertainty of these macroscale observables.

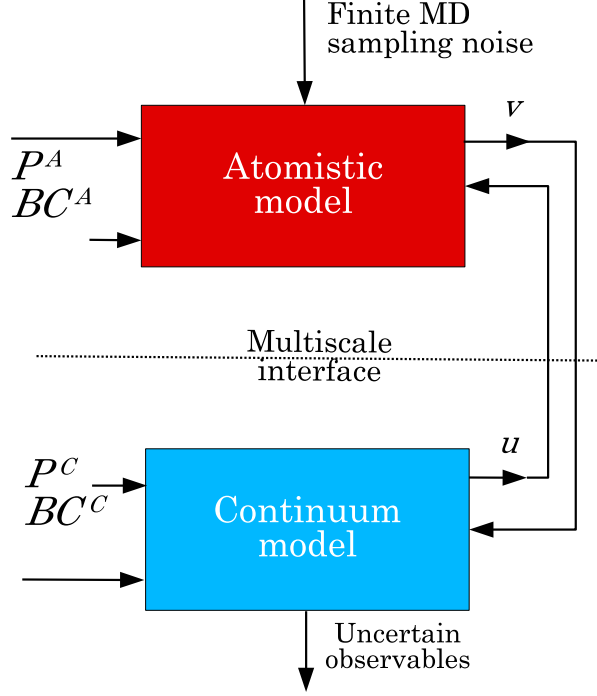


Figure 1: Information flow in a general stochastic atomistic-continuum multiscale simulation. Inputs to the atomistic simulation are parameters P^A , boundary conditions BC^A and a variable u extracted from the continuum simulation. Inputs to the continuum simulation are parameters P^C , boundary conditions BC^C and a variable v 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.

Given the above mentioned sources of uncertainty, the main question addressed in this paper is how to quantify the resulting uncertainty in the predicted values of the macroscale observables extracted from the continuum simulation. To do so, it is crucial to first quantify the uncertainty in the coupling variables between the atomistic and continuum simulations. Hence, a key aspect of the present work is to calculate the coupling variables u and v and their associated uncertainty.

The current work focuses on cases where there is a strong separation between the relevant time and/or length scales of the resolved atomistic and continuum phenomena.

Mathematically, the information exchanged between the different system scales is therefore passed in terms of boundary conditions imposed on the atomistic and continuum components of the simulation [8]. Such a coupling procedure has been studied by different research groups [20–28]. In almost all of these studies, the coupling was performed near a fluid-solid interface where the wall velocity was the main variable of interest.

The particular case studied in this work is a near-wall flow resolved with an atomistic model coupled to a continuum model for the flow further away from the wall. The continuum flow model operates on a time scale much longer than that of the natural variability of the atomistic velocities, such that the continuum model is deterministic in nature. However, extraction of macroscale observables from the atomistic model with finite sampling generates uncertainty in those observables [13], introducing uncertainty to the overall multiscale simulation.

In previous work [14], we quantified this sampling related uncertainty in coupling velocities that are extracted from atomistic simulations of a Couette flow problem. This uncertainty was then propagated to the continuum model, which sets the input parameters for the atomistic model. The uncertainty caused by the finite sampling therefore also creates uncertainty in the atomistic model inputs, and this two-way coupling leads to a fixed point iteration on the uncertain coupling variables. Polynomial Chaos Expansions (PCEs) [29–31] were used to spectrally represent the uncertain coupling variables, and a Bayesian inference approach [32] was used to determine the PCE coefficients of the macroscale velocities extracted from the atomistic simulations. A surrogate model for the atomistic simulation was used to speed up the fixed point iterations on the PCE coefficients of the coupling variables.

In the current work, the surrogate model approach is improved to streamline the coupling process, and the improved efficiency allows the incorporation of parametric uncertainty in the atomistic and continuum model parameters along with the existing uncertainty due to finite sampling on the atomistic level. As in [14], a steady-state Couette

flow problem is chosen to develop and illustrate the approach, but the approach can be generalized to more involved geometries and/or time-dependent problems.

By accounting for parametric and sampling uncertainties in a consistent way across all scales, the approach developed in this paper allows the assessment of the uncertainty in the outputs of coupled atomistic-to-continuum simulations, for a given amount of computational resources. The representation of this uncertainty with PCEs as a function of the different sources of uncertainty also shows what the dominant sources of uncertainty in the outputs of interest are, such that if more confident predictions are desired, one can determine whether this requires more computational resources (more samples), or better determined input parameters.

The paper is organized as follows. Section 2 outlines the geometry of the atomistic-to-continuum problem considered in this paper, along with an overview of the MD approach used to simulate the atomistic flow behavior, and the exchange of information between the atomistic and continuum components in the simulation. Section 3 covers the mathematical foundation for representing uncertainty and constructing surrogate models for the atomistic and continuum components. Sections 4 and 5 then detail, respectively, how these surrogate models are used to obtain the coupling variables in a case with sampling noise only, and in a case where sampling noise is combined with parametric uncertainty. We then discuss the results and conclude in Section 6.

2 Problem Setup

2.1 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 w , 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 can not 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 w . 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 a constant deterministic velocity $2u^A$ such that the velocity at $y = h_{MD}$ is equal to u^A , where u^A is the fluid velocity at $y = h_{MD}$ in Figure 2(b). Further details about the MD simulation are given in [14].

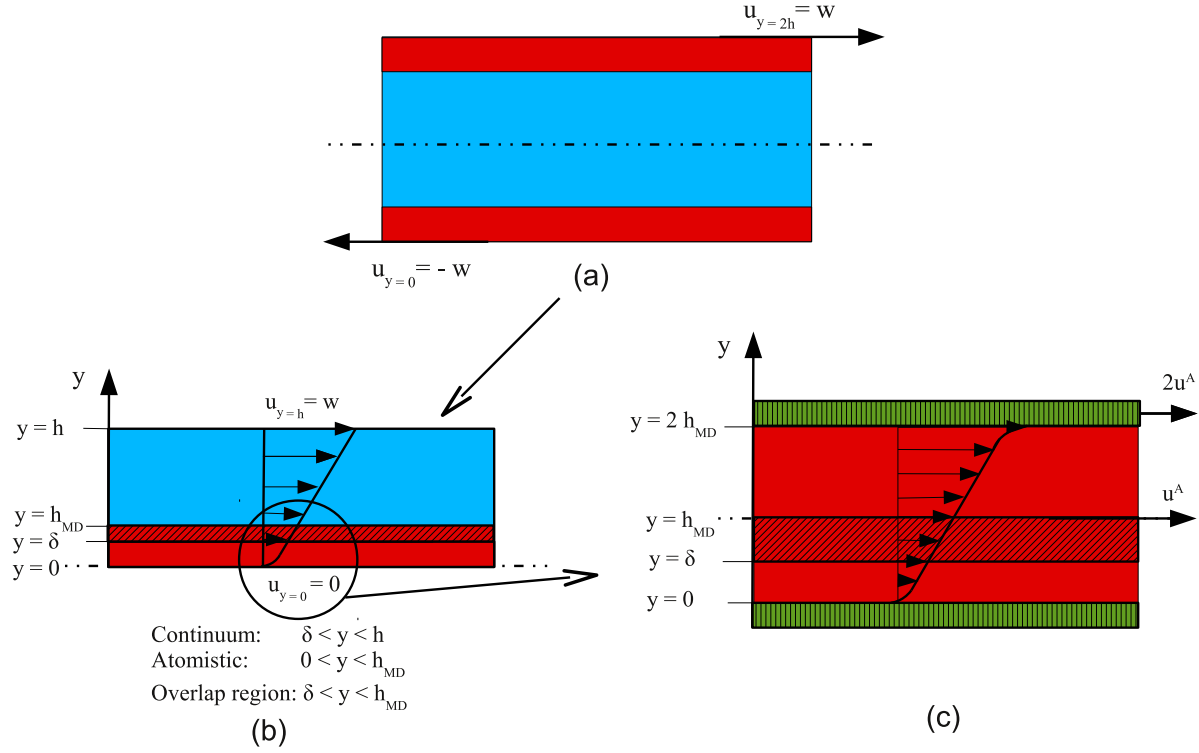


Figure 2: A schematic showing different components of the continuum-atomistic simulation. The red color signifies the atomistic portion of the simulation while blue represents the continuum portion. The schematics show: (a) a general symmetric Couette flow multiscale geometry where w is the continuum wall velocity, (b) a transformed half-geometry where the atomistic simulation is next to the stationary wall; the dashed red area is an overlap region between atomistic and continuum, and (c) an atomistic Couette flow simulation performed next to the wall where u^A is the constant deterministic velocity imposed on the MD region in (b); the green area denotes the region in the atomistic simulation that models the walls.

2.2 MD Numerical Implementation

The MD simulation geometry suggested in Figure 2(c) consists of a stationary and a moving wall. We consider a three-dimensional domain occupied by particles representing water molecules. The soft inter-particle interaction is modeled by the Lennard-Jones (LJ)

pairwise potential ϕ_{ij} [33]. For particles i and j separated by a distance r , ϕ_{ij} is given by:

$$\phi_{ij} = 4\phi_0 \left[\left(\frac{\sigma}{r} \right)^{12} - \left(\frac{\sigma}{r} \right)^6 \right]. \quad (1)$$

We set the nominal values of the LJ parameters for water to $\phi_0 = 0.152$ Kcal/mole and $\sigma = 3.15$ Å [34]. We simulate moving walls by fixing the velocity of particles lying around at $y = h_{MD}$ to a prescribed velocity vector $\{u^A, 0, 0\}$ for the whole simulation period.

MD computations were performed with LAMMPS [35] at 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 [34, 35] to maintain the steady state in the presence of mechanical energy input from the walls.

We extract the velocity at $y = \delta$ (see Figure 2(c)). We choose $\delta = 20.5$ Å, 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 2.1. Details on the extraction of the velocity can be found in [14]. This value of δ is the same for all the cases considered in this paper. 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. Velocity samples are collected until the noise amplitude in the mean 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 inputs we perform different replica simulations to sample over the degrees of freedom in the initial conditions.

Extracted velocities as a function of MD simulation time are shown in Figure 3. For all the considered averaging time windows, the amplitude of the velocity noise decreases with increasing averaging time window.

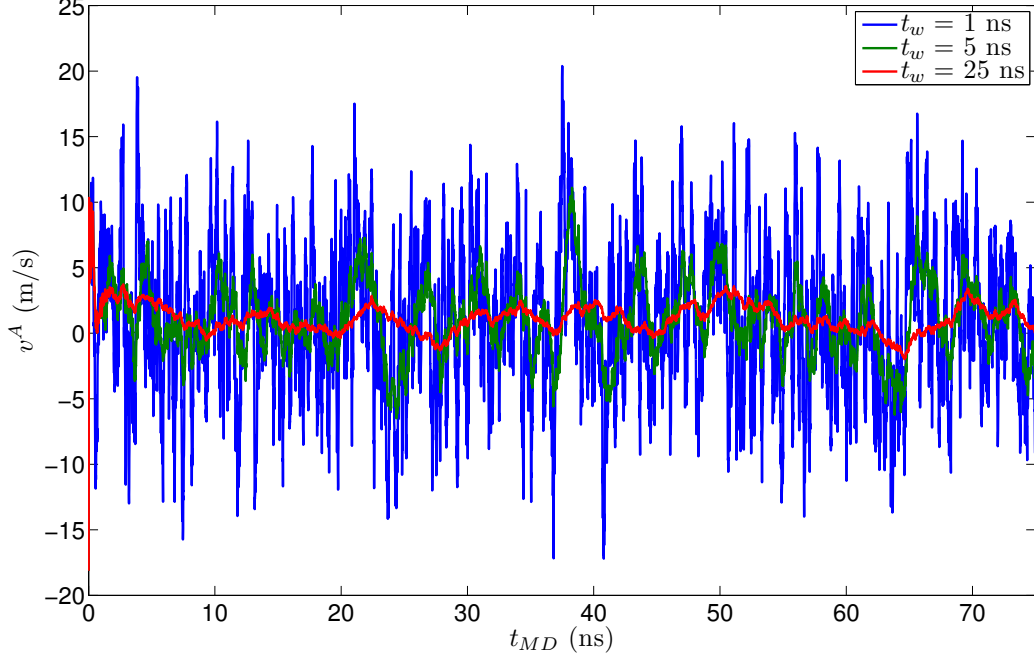


Figure 3: Plot showing short-time averaged velocities extracted from the Couette flow MD simulation (see Figure 2) for $u^A = 10$ m/s and different moving time averaging window widths, as indicated.

2.3 Variable Exchange between the Atomistic and Continuum Simulations

In the setting of the Couette flow in Figure 2(b), the key variables coupling the atomistic and continuum simulations are either shear stress (F) or velocity (V). Hence there are four different coupling schemes based on the classification of Ren [8]. Since shear stress, unlike velocity, is not a fundamental quantity in MD and therefore difficult to control directly in MD simulations, the schemes involving shear stress were not considered in this work. The stability considerations discussed in [14] led us to use the Velocity-Velocity (VV) coupling scheme where 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 4). The resulting exchange of variables

between the atomistic and the continuum simulations is visualized in Figures 4 and 5.

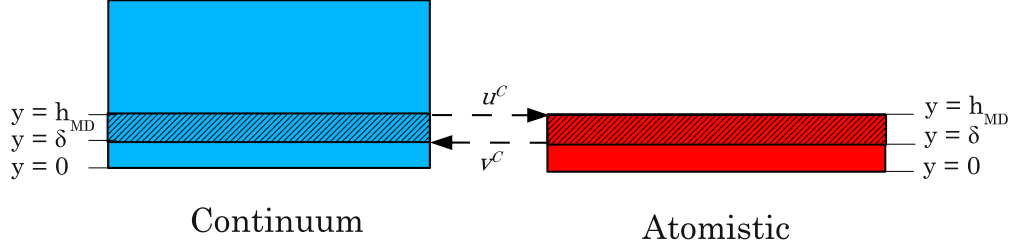


Figure 4: A schematic showing the locations of variable exchange between the atomistic and continuum simulations in a VV coupling scheme.

In these figures, the superscript C is used to signify that a particular variable is defined on the macroscale continuum level, whereas the superscript A indicates a quantity on the atomistic level.

As mentioned in the introduction, and as studied in [14], the velocity v^C is inferred from short-term averaged noisy samples of the atomistic velocity v^A , and as such, the output v^C from the atomistic simulation is treated as an uncertain quantity on the continuum level. Section 3.2 quantifies this uncertainty as a function of the time-averaging window and the number of MD velocity samples used.

Due to the uncertain input v^C to the continuum simulation (see Figure 5), the output u^C from the continuum simulation will also have uncertainty (see Section 3.3). Hence, in this two-way coupled setting, the outputs v^C and u^C of the atomistic and continuum simulations, respectively, are both uncertain quantities, and uncertainty information is passed in both directions across the interface.

The uncertainty in u^C is propagated through the atomistic model by sampling specific values u_i^A from the distribution of u^C , and imposing those values u_i^A as deterministic boundary conditions on the atomistic model. An appropriate ensemble of replica MD simulations is performed [9] to sample over the many degrees of freedom for the initial particle positions and velocities.

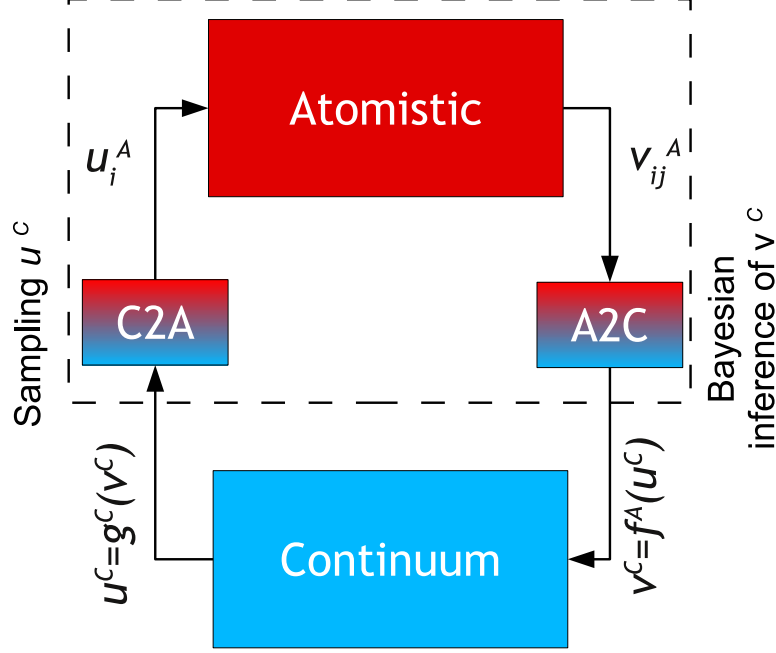


Figure 5: A schematic showing the exchange of the variables in a VV coupling scheme between the atomistic and continuum simulations. The superscript index (i) indicates the current coupling iteration.

To determine the coupling velocities u and v along with their uncertainties, an iterative scheme was developed [14] for the case where the atomic sampling noise was the only source of uncertainty in the system. The fixed point iteration scheme looped through the information flow depicted in Figure 5 until the coupling velocities converged. As this iteration required multiple runs of the atomistic model for slightly different input conditions u^A , a surrogate model was developed for the output velocities v^A of the atomistic model. This surrogate model greatly reduced the computational cost associated with the iterative scheme.

In the current work, this approach is improved by constructing a surrogate not just for the atomistic component, but for the ensemble of the *continuum to atomistic* (C2A), the atomistic, and the *atomistic to continuum* (A2C) components; *i.e.* for all components inside the box with the dashed outline in Figure 5. In doing so, the exchange of information between the atomistic and continuum models can be written in terms of variables that all

reside on the continuum level, v^C and u^C , which need to satisfy:

$$\begin{aligned} v^C &= f^A(u^C) \\ u^C &= g^C(v^C) \end{aligned} \tag{2}$$

where f^A is the surrogate model for the ensemble of atomistic components, and g^C is the surrogate for the continuum model.

The next Section 3 details how the surrogate model for the atomistic system f^A is constructed in 3.2, and the surrogate for the continuum model g^C in 3.3. The uncertainties in the exchange variables v^C and u^C are represented with Polynomial Chaos coefficients, which are introduced in Section 3.1.

3 Mathematical Formulation

3.1 Polynomial Chaos Expansions

Polynomial Chaos Expansions (PCEs) are spectral representations of random variables in terms of polynomial functions of standard random variables multiplied by deterministic coefficients. Under some generally mild assumptions [36], any finite-variance random variable u can be represented as a PCE as follows:

$$u = \sum_{k=0}^{\infty} u_k \psi_k(\xi), \tag{3}$$

for a wide class of standard random variables ξ and standard polynomials $\psi_k(\xi)$ that are orthogonal with respect to the probability distribution function of ξ . Two of the most commonly used expansions for continuous random variables are Gauss-Hermite (Hermite polynomials as functions of standard normal random variables) and Legendre-Uniform (Legendre polynomials as functions of uniform random variables) PCEs. A multivariate

generalization of the PCE is written as

$$x = \sum_{k=0}^{\infty} x_k \Psi_k(\boldsymbol{\xi}), \quad (4)$$

where x is a scalar quantity depending on a vector of standard variables $\boldsymbol{\xi} = (\xi_1, \xi_2, \dots, \xi_n)$, and $\Psi_k(\boldsymbol{\xi})$ are multivariate standard polynomials defined as $\Psi_k(\boldsymbol{\xi}) = \psi_{p_1}(\xi_1) \cdots \psi_{p_n}(\xi_n)$ according to a chosen enumeration of these polynomials for some ordering in the set of vectors of component-wise orders $(p_1, \dots, p_n)$, called multi-indices. Note that, in the simplest setting, the *stochastic dimension*, *i.e.* the size of the vector $\boldsymbol{\xi}$, is the same as the number of input parameters n . In practice, when the input parameters are independent, one can decouple the form (4) and write n independent PC expansions for each uncertain input.

$$x_i = \sum_{k=0}^{\infty} x_{i,k} \psi_k(\xi_i). \quad (5)$$

While the univariate PC expansions are usually truncated at some finite order p , the multivariate expansion (4) is typically truncated according to the total degree of the retained polynomials, *i.e.* $p_1 + \dots + p_n \leq p$, where p is a predefined value for the maximal degree. The number of terms will then be $P + 1 = (n + p)!/n!p!$.

After characterizing uncertain variables with PCEs, one can efficiently propagate their associated uncertainty through a set of model equations. One approach to do this is to substitute the PCE variables directly into the model equations and perform all operations (*e.g.* multiplication, square root) directly on the PC expansions. As such, this method is referred to as the Intrusive Spectral Projection (ISP) approach. Numerical operations on PCEs are described in detail in the work of Debusschere *et al.* [37]. Two operations that will be commonly used in this work are multiplications and square roots.

The product z of two PCEs x and y is computed as

$$z = x \otimes y \quad (6)$$

$$z_k = \frac{1}{\langle \Psi_k^2 \rangle} \sum_{i=0}^P \sum_{j=0}^P x_i y_j \langle \Psi_i \Psi_j \Psi_k \rangle \quad (7)$$

We compute the square root of a PCE using the equality $\sqrt{x} = e^{\frac{1}{2} \log x}$, i.e. by first computing its natural logarithm, multiplying the outcome by 1/2 then applying the exponential function to the result. For more details, see the work of Debusschere *et al.* [37].

Another way to derive the PC coefficients of the solution of a set of equations is to project the solution directly onto the PC basis. For example, to get the PC coefficients of a model output x , one writes

$$x_k = \frac{\langle x \Psi_k \rangle}{\langle \Psi_k^2 \rangle}, \quad k = 0, \dots, P \quad (8)$$

This requires numerical evaluation of the projection integrals $\langle x \Psi_k \rangle$ using quadrature rules, relying on samples of x for specific values of the uncertain model inputs. This method is referred to as non-intrusive spectral projection (NISP) [38], which will also be used in this study. More details about PCEs and their numerical implementations are found in [38].

3.2 Atomistic Component: Inference of a Polynomial Response Surface

To enable rapid evaluation of the output velocities of the atomistic model, for various values of its inputs, a cheaper-to-evaluate *surrogate* model is created for the atomistic model in this work. Mathematically, consider the atomistic model as a *forward* function $f(\cdot)$, relating input $\mathbf{x}$ to output $y = f(\mathbf{x})$. In general, the input $\mathbf{x}$ consists of n input parameters $\mathbf{x} = (x_1, x_2, \dots, x_n)$. Our goal is to construct the surrogate model as a re-

sponse surface $\tilde{f}(\mathbf{x}; \mathbf{m})$ that approximates the forward function $f(\cdot)$ within a hypercube $a_i \leq x_i \leq b_i$, for $i = 1, 2, \dots, n$, i.e. $f(\mathbf{x}) \approx \tilde{f}(\mathbf{x}; \mathbf{m})$. The vector of the response surface parameters $\mathbf{m}$ is the object of inference. In this paper, a polynomial model is employed for a response surface construction. First, the input parameters $x_i \in [a_i, b_i]$ are scaled to $\lambda_i \in [-1, 1]$

$$x_i = \frac{a_i + b_i}{2} + \frac{b_i - a_i}{2} \lambda_i \text{ for } i = 1, 2, \dots, n, \quad (9)$$

where $[a_i, b_i]$ is the input range for the parameter x_i .

The goal is to find a polynomial approximation $\tilde{f}(\mathbf{x}, \mathbf{m})$ for the forward function $f(\mathbf{x})$, i.e.

$$f(\mathbf{x}) \approx \tilde{f}(\mathbf{x}; \mathbf{m}) = \sum_{\alpha} m_{\alpha} \lambda(\mathbf{x})^{\alpha}, \quad (10)$$

where $\lambda(\mathbf{x})$ is the linear relationship derived from (9). Here, we used the notation $\lambda^{\alpha} = \prod_i \lambda_i^{\alpha_i}$ for the monomial corresponding to a multi-index vector α . The polynomial expansion (10) is truncated based on the total degree of the retained monomials, i.e. $\|\alpha\|_1 = \alpha_1 + \alpha_2 + \dots + \alpha_n \leq p$, where p is the order of the expansion, chosen *a priori*, leading to the number of terms $K = (n + p)!/n!p!$.

We will employ Bayesian inference to find the polynomial coefficients $\mathbf{m} = \{m_{\alpha}\}_{\|\alpha\|_1 \leq p}$ in the expansion (10). Let $\mathbf{y}$ be a vector of observable data, or a set of *training* evaluations of the forward model f at points $\mathbf{x}_i$, i.e. $y_i = f(\mathbf{x}_i)$ for $i = 1, \dots, N$. Bayes' rule writes the posterior PDF for the model parameters $\mathbf{m}$ given the observed function evaluations $\mathbf{y}$ [39, 40]:

$$\mathcal{P}(\mathbf{m}|\mathbf{y}) \propto \mathcal{P}(\mathbf{y}|\mathbf{m})\mathcal{P}(\mathbf{m}) \quad (11)$$

The prior $\mathcal{P}(\mathbf{m})$ and posterior $\mathcal{P}(\mathbf{m}|\mathbf{y})$ probabilities represent degrees of knowledge of $\mathbf{m}$ *before* and *after* obtaining the data $\mathbf{y}$, respectively. The key component in (11) is the likelihood function $L(\mathbf{m}) = \mathcal{P}(\mathbf{y}|\mathbf{m})$. To construct the likelihood, we assume a Gaussian i.i.d. discrepancy between the polynomial model prediction $\tilde{f}(\mathbf{x}; \mathbf{m})$ and the forward model $f(\mathbf{x})$ at the parameter values of interest $\mathbf{x}_i$. Specifically, the components

of the discrepancy vector $\boldsymbol{\epsilon} = [\epsilon_1, \epsilon_2, \dots, \epsilon_N]^T$ are defined as

$$\epsilon_i = y_i - \tilde{f}(\mathbf{x}_i; \mathbf{m}) \quad (12)$$

are assumed to be independent and correspond to the PDF $\pi_s(z) = \exp(-z^2/2s^2)/\sqrt{2\pi s^2}$, leading to a likelihood

$$L(\mathbf{m}; s) = \prod_{i=1}^N \pi_s(y_i - \tilde{f}(\mathbf{x}_i; \mathbf{m})) \quad (13)$$

Note that we have introduced the *hyperparameter* s that is the standard deviation of the error model for the discrepancy. In the absence of any knowledge of the magnitude of it, this parameter is also an object of inference, therefore, the Bayes' formula should be viewed as

$$\mathcal{P}(\mathbf{m}, s | \mathbf{y}) \propto L(\mathbf{m}; s) \mathcal{P}(\mathbf{m}) \mathcal{P}(s). \quad (14)$$

Since we are primarily interested in the parameters $\mathbf{m}$, we will take the marginal posterior distribution of $\mathbf{m}$ only:

$$\mathcal{P}(\mathbf{m} | \mathbf{y}) = \int_s \mathcal{P}(\mathbf{m}, s | \mathbf{y}) ds. \quad (15)$$

Independent uniform priors are taken for the polynomial coefficients, $\mathcal{P}(\mathbf{m}) = \text{const}$, while a Jeffrey's prior is assumed for the positive parameter s , i.e. $\mathcal{P}(s) \sim 1/s$. With such priors, given that the response surface (10) is linear in the parameters $\mathbf{m}$, as well as independent Gaussian assumptions on the likelihood, one can exactly solve for the posterior distribution $\mathcal{P}(\mathbf{m}, s | \mathbf{y})$ as well as marginalize over s to obtain [14]:

$$\mathcal{P}(\mathbf{m} | \mathbf{y}) \propto \left(1 + \frac{1}{\gamma} (\mathbf{m} - \boldsymbol{\mu})^T \mathbf{V}^{-1} (\mathbf{m} - \boldsymbol{\mu}) \right)^{-\frac{\gamma+K}{2}} \quad (16)$$

which is a multivariate Student-t distribution

$$\mathbf{m} | \mathbf{y} \sim \mathcal{MST}(\boldsymbol{\mu}, \mathbf{V}, \gamma) \quad (17)$$

with mean

$$\boldsymbol{\mu} = (\mathbf{Q}^T \mathbf{Q})^{-1} \mathbf{Q}^T \mathbf{y}, \quad (18)$$

scale matrix¹

$$\mathbf{V} = \frac{\mathbf{y}^T \mathbf{y} - \mathbf{y}^T \mathbf{Q} (\mathbf{Q}^T \mathbf{Q})^{-1} \mathbf{Q}^T \mathbf{y}}{N - K - 1} (\mathbf{Q}^T \mathbf{Q})^{-1} \quad (19)$$

and degrees of freedom

$$\gamma = N - K - 1. \quad (20)$$

To this end, we note that when the degrees of freedom is large, i.e. $\gamma \gg 1$, this multivariate Student-t distribution converges to a multivariate normal distribution with mean $\boldsymbol{\mu}$ and covariance $\mathbf{V}$. In Eqs. (18) and (19), the *measurement matrix* $\mathbf{Q} \in \mathbb{R}^{N \times K}$ is defined as the matrix of evaluations of the monomials $\boldsymbol{\lambda}^\alpha$ in (10) at the training $\mathbf{x}$ -values. That is, $\mathbf{Q}_{nk} = \boldsymbol{\lambda}(\mathbf{x}_n)^{\alpha(k)}$, where $\alpha(k)$ is the k -th multi-index in the ordering of the monomial terms in (10). With a multivariate Student-t distribution as polynomial coefficients, the polynomial approximation becomes a Student-t process. The marginal distribution of this process at each fixed value of $\mathbf{x}$ is a Student-t random variable

$$\tilde{f}(\mathbf{x}) \sim \mathcal{ST}(\mathbf{q}(\mathbf{x})^T \boldsymbol{\mu}, \mathbf{q}(\mathbf{x})^T \mathbf{V} \mathbf{q}(\mathbf{x}), \gamma), \quad (21)$$

where the vector $\mathbf{q}(\mathbf{x}) \in \mathbb{R}^K$ is defined as $\mathbf{q}(\mathbf{x})_k = \boldsymbol{\lambda}(\mathbf{x})^{\alpha(k)}$.

Therefore, one can write

$$\tilde{f}(\mathbf{x}) = \mathbf{q}(\mathbf{x})^T \boldsymbol{\mu} + \zeta \sqrt{\mathbf{q}(\mathbf{x})^T \mathbf{V} \mathbf{q}(\mathbf{x})}, \quad (22)$$

where ζ is a standard Student-t random variable, $\zeta \sim \mathcal{ST}(0, 1, \gamma)$.

In the present context, $\tilde{f}(\mathbf{x})$ is a response surface for the continuum wall velocity v^C , which is the key output of the atomistic model. While the atomistic model has many input parameters and boundary conditions, the response surface is constructed as

¹For completeness, we note that the covariance is proportional to the scale matrix $\boldsymbol{\Sigma} = \frac{\gamma}{\gamma-2} \mathbf{V}$.

a function of only the inputs that have variability. In the case without external sources of uncertainty, as discussed in section 4, the only input is simply the continuum input velocity u^C . Therefore $\mathbf{x} = \{u^C\}$ and the atomistic surrogate becomes

$$v^C = f^A(u^C, \zeta). \quad (23)$$

To infer this surrogate model, a set of short-term averaged values of the atomistic velocity v^A , sampled at different values of the input u^C is used as data. Due to the finite amount of sampling over the raw atomistic velocities used to extract those data points, large amounts of noise may be present. Nevertheless, since the data consists of a short-term averaged MD simulation outputs, the noise can approximately be characterized as Gaussian due to the Central Limit Theorem [14]. The other component of the discrepancy between the response surface and data is the error associated with the accuracy of the response surface itself. We assume, however, that the polynomial response surface order is chosen large enough – but without having to overfit – to render this generally correlated discrepancy component negligibly small. The outcome of the Bayesian inference of the response surface relies on the standard Student-t random variable ζ in (22) and (23) that corresponds to the variability in the posterior distribution, and represents the uncertainty due to the sampling noise as well as due to a possible model discrepancy between the true atomistic model and the response surface. Again, with large enough order in the polynomial response surface, one can safely assume that the sampling noise is the only contributor to the variability described by ζ .

For the case studied in section 5, the Lennard-Jones parameter σ corresponding to the soft sphere radius in the MD force model is assumed to be uncertain. The atomistic surrogate model needs to cover that external source of uncertainty as well, such that $\mathbf{x} = \{u^C, \sigma\}$ and the surrogate model becomes

$$v^C = f^A(u^C, \sigma, \zeta) \quad (24)$$

In this case, the data used in the inference process consists of short-term averaged atomistic velocities v^A , sampled over a range of values for u^C and σ .

3.3 Continuum Component: Linear Relationship

The velocity obtained from the MD simulations, or its polynomial approximation, is imposed on the continuum system as shown in Figure 4. We formulate the continuum Couette flow mathematical model based on Figure 2(b). Assuming a steady, laminar, Newtonian flow, the Navier-Stokes equations reduce to:

$$\frac{d^2u}{dy^2} = 0, \quad \text{for } \delta \leq y \leq h \quad (25)$$

Based on Section 2.1, 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:

$$u(y = \delta) = v^C \quad (26)$$

$$u(y = h) = w, \quad (27)$$

leading to the linear relationship for the velocity extracted at $y = h_{MD}$,

$$u(y = h_{MD}) = u^C = w + \beta(w - v^C) = g_C(v^C, w), \quad (28)$$

where $\beta = \frac{h_{MD} - h}{h - \delta}$.

4 Application to Coupled Atomistic-Continuum Couette Flow with Sampling Noise

In this section, we consider the case of coupled atomistic-continuum Couette flow with uncertainty due to sampling noise, similarly to our previous study [14]. This case enables comparing the results of the current continuum level surrogate model approach to our previous atomistic level stochastic coupling approach. We assume deterministic values for all the parameters in the coupled system, namely, h , w , ϕ_0 and σ . The only source of uncertainty therefore is due to finite sampling in the extraction of macroscale variables from the MD simulation. Hence, this uncertainty only exists in the expression of f^A in (2). Here, the response surface construction for the atomistic component described in Section 3.2 is carried out for one input only, $\mathbf{x} = \{u^C\}$, and the coupled system of equations reduces to:

$$v^C = \mathbf{q}[u^C(\zeta)]^T \boldsymbol{\mu} + \zeta \sqrt{\mathbf{q}[u^C(\zeta)]^T \mathbf{V} \mathbf{q}[u^C(\zeta)]}, \quad (29)$$

$$u^C = w + \beta(w - v^C) \quad (30)$$

We rely on $N_{MD} = 200$ short-time MD data averaged over a window $t_w = 5$ ns for each of the five training values of the input $u_i^C \in \{0.0, 1.125, 2.25, 3.375, 4.5\}$ m/s. These five values are equally spaced and chosen around a nominal value of u^C assumed to be equal to 2.25 m/s. Due to the linear nature of the Couette flow, we assume a linear relationship (i.e. an exact, linear response surface) between u^C and v^C that gives:

$$\mathbf{q}(u^C) = \{1, u^C\}^T. \quad (31)$$

for $\sigma = 3.15 \text{ \AA}$, $\boldsymbol{\mu}$ and $\mathbf{V}$ can be derived according to Section 3.2 as:

$$\begin{aligned}\boldsymbol{\mu} &= (-0.056, 0.55)^T \text{m/s} \\ \mathbf{V} &= \begin{pmatrix} 1.17 \times 10^{-5} & -2.08 \times 10^{-6} \\ -2.08 \times 10^{-6} & 1.84 \times 10^{-6} \end{pmatrix} \text{m}^2/\text{s}^2\end{aligned}\tag{32}$$

Given the relative simplicity of this case, we start by providing a graphical solution of the system formed by Eqs. (29) and (30). Plotting v^C versus u^C according to Eq. (29) leads to a family of curves parametrized by $\zeta \sim \mathcal{ST}(0, 1, \gamma)$ as discussed previously. Note that since the degrees of freedom is large $\gamma = 190 \gg 1$ in the Student-t parameter definition (20), we can approximate ζ by a normal distribution $\mathcal{N}(0, 1)$. On the other hand, plotting Eq. (30) leads to a unique curve. The solution of the system (29) and (30) is therefore a 1D interval of pairs (u^C, v^C) rather than a single point. This is visualized in Figure 6 for two different deterministic values of the LJ parameter σ in the MD simulation. The plots show that σ significantly affects the slope of the atomistic response curves. This is not surprising because in Eq. (1), the interaction energy between the LJ particles exhibits a 12^{th} order dependence on σ . The change in the interaction energy incurred by σ affects the fluid viscosity, hence the flow field inside the MD domain.

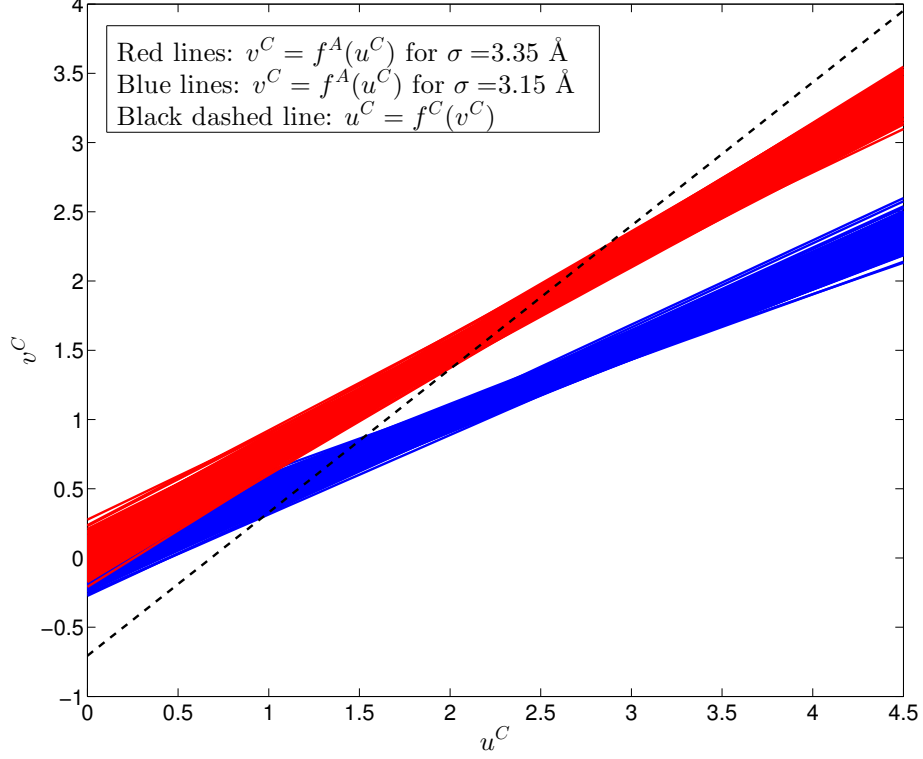


Figure 6: Plot showing: (solid lines) different realizations of the atomistic response surface sampled from Eq. 29 for two different values of the parameter σ , $N_{MD} = 200$ and $t_w = 5$ ns, and (dashed line) the continuum response surface (37) for a deterministic $w = w_0 = 20$ m/s.

The uncertain solution of the system is given by the intersection of the continuum response curve (black dashed line) and the atomistic response curves in red and blue for $\sigma = 3.35$ Å and $\sigma = 3.15$ Å, respectively. To compute this intersection formally, we assume a first order PCE for both u^C and v^C . The velocity u^C is therefore written as:

$$u^C = u_0 + u_1\zeta \quad (33)$$

Using the ISP approach outlined in Section 3.1, the PC coefficients for u^C and v^C are obtained as follows. Starting from an initial guess of the PC coefficients u_0 and u_1 , we proceed with a fixed-point iterator that computes v^C from Eq. (29) and updates u^C

using Eq. (30). This iterative procedure, summarized in Algorithm 1, is repeated until convergence of the PC coefficients. At each iteration, multiplication and square root operations take place on PCEs, as discussed in Section 3.1.

Algorithm 1 solves a system of equations in two uncertain variables expressed as PCEs using ISP in the case of sampling noise with no parametric uncertainty.

$u_0^{C,0} = u_0, u_{k>0}^{C,0} = 0$	{Initial guess of the PC coefficients of u^C }
$i = 0, i_{max} = 500, e_r = 100$	{Maximum number of iterations and initial error}
$\epsilon = 10^{-4}$	{Relative tolerance}
while $e_r > \epsilon$ and $i < i_{max}$ do	
Compute $m = \zeta \sqrt{\mathbf{q}^{T,i}(u^{C,i}) \mathbf{V} \mathbf{q}^i(u^{C,i})}$	
Compute the PCE of $v^{C,i} = \mathbf{q}^{i,T} \boldsymbol{\mu} + m$ using Eq. (29)	
Compute an updated PCE $u^{C,i+1}$ using Eq. (30)	
$e_r = \max_{0 \leq k \leq 1} 1 - u_k^{C,i+1} / u_k^{C,i} $	
$i = i + 1$	
end while	

The convergence of the solution is depicted in Figure 7 as a function of the number of iterations. Unlike the atomistic level stochastic coupling scheme [14], the current response surface approach produces smooth convergence curves with better defined convergence criteria.

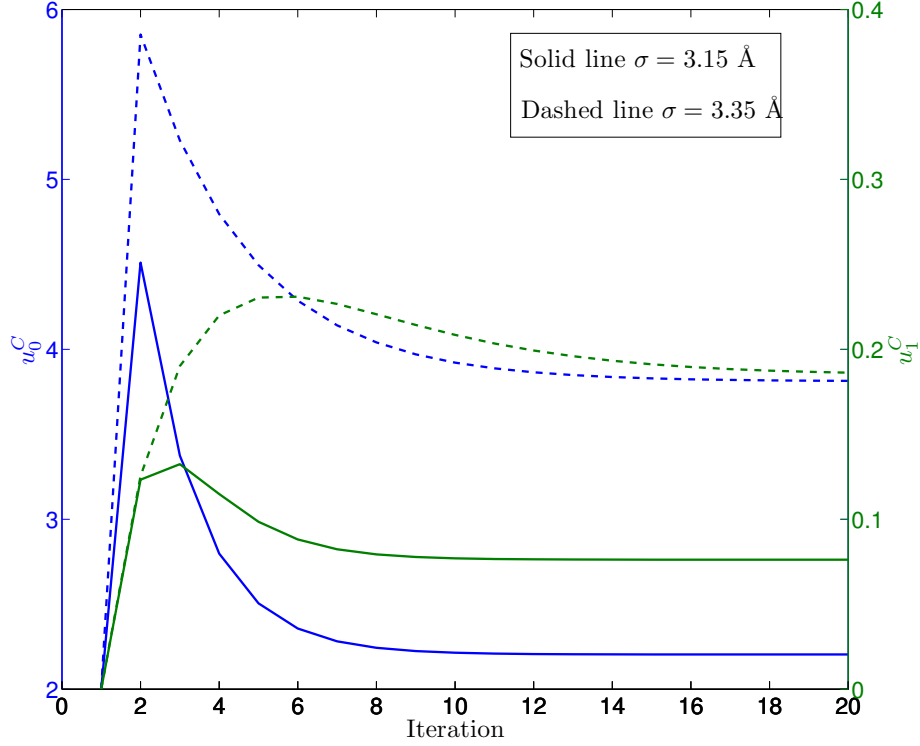


Figure 7: Plot showing the convergence of the PC coefficients of u^C with only uncertainty due to sampling noise and for two different values of the parameter σ , as indicated. Results are obtained with an atomistic response surface computed using $N_{MD} = 200$ and $t_w = 5$ ns. The ISP approach in Algorithm 1 was used to solve Eqs. (29) and (30).

The computed uncertain velocity u^C is reported in Figure 8 in terms of its mean and standard deviation. v^C can easily be obtained once u^C is computed using Eq. (30). Also shown in Figure 8 are the results obtained using the atomistic level stochastic coupling scheme [14]. The standard deviation of u^C decreases with the time averaging window t_w , as expected, and increases with σ as discussed previously. The results of the response curve approach are in good agreement with the stochastic coupling approach for all σ and t_w .

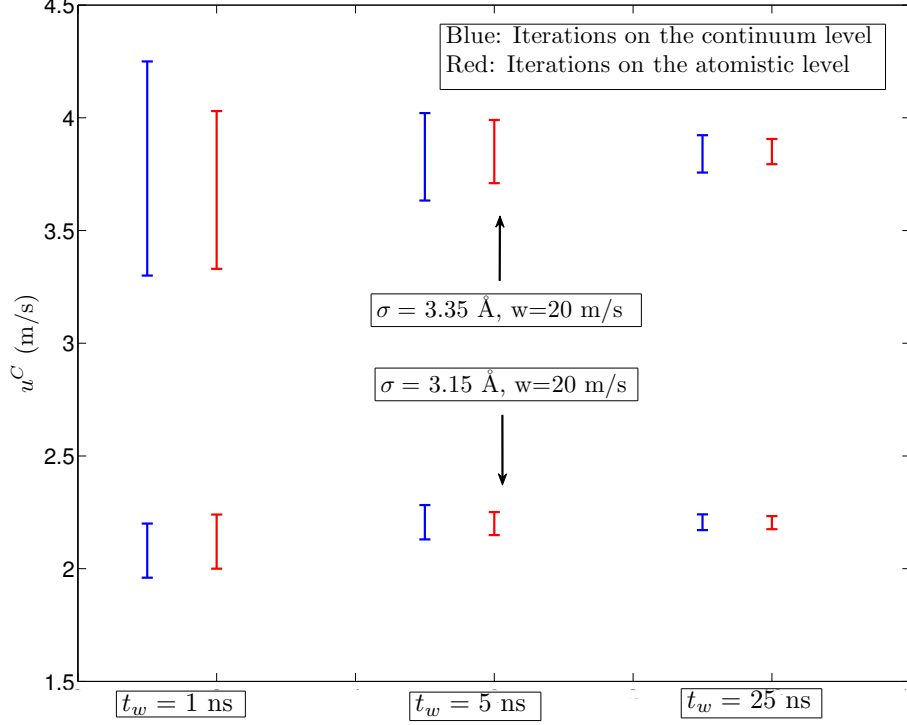


Figure 8: Plot showing the mean and standard deviation of the predicted continuum scale velocity u^C considering only uncertainty due to sampling noise for two different values of the parameter σ , as indicated. Results are obtained based on the same MD data for different time averaging windows t_w , as indicated, using the atomistic level algorithm (red) [14] and the continuum level response surface iterative scheme (blue).

5 Application to Coupled Atomistic-Continuum Couette Flow with Sampling Noise and Parametric Uncertainty

In this section, we provide a generalization of the coupling procedure described in the previous section such that we account for parametric uncertainty in addition to sampling noise. Similar to Section 4, coupling two systems is obtained by solving two equations of

the forms obtained in Sections 3.2 and 3.3. We focus on our current case of Couette flow to describe our solution method. However, the algorithms we develop are also applicable to systems with different numbers of uncertain parameters whether in the atomistic or the continuum components. We assume two uncertain parameters $\sigma(\xi_1)$ (in the atomistic component) and $w(\xi_2)$ (in the continuum component) in addition to the sampling noise $\zeta = \xi_3$. As such, the variables ξ_i for $i = 1, 2, 3$ encapsulate all the uncertain components of the coupled system. The main quantities of interest, u^C and v^C , can therefore be written as PCEs

$$v^C = \sum_{k=0}^P v_k \Psi_k(\xi_1, \xi_2, \xi_3), \quad u^C = \sum_{k=0}^P u_k \Psi_k(\xi_1, \xi_2, \xi_3). \quad (34)$$

Our goal is to find their PC coefficients u_k and v_k for $k = 0, \dots, P$, given the relations

$$\begin{aligned} v^C &= f_A(u^C, \sigma, \zeta), \\ u^C &= g_C(v^C, w). \end{aligned}$$

Thus, we would like to solve:

$$v^C(\boldsymbol{\xi}) = \mathbf{q}[u^C(\boldsymbol{\xi}), \sigma(\xi_1)]^T \boldsymbol{\mu} + \xi_3 \sqrt{\mathbf{q}[u^C(\boldsymbol{\xi}), \sigma(\xi_1)]^T \mathbf{A} \mathbf{A}^T \mathbf{q}[u^C(\boldsymbol{\xi}), \sigma(\xi_1)]} \quad (35)$$

$$= \mathbf{q}[u^C(\boldsymbol{\xi}), \sigma(\xi_1)]^T \boldsymbol{\mu} + \xi_3 \|\mathbf{A}^T \mathbf{q}[u^C(\boldsymbol{\xi}), \sigma(\xi_1)]\|_2$$

$$u^C(\boldsymbol{\xi}) = g^C(v^C(\boldsymbol{\xi}), w(\xi_2)) \quad (36)$$

for the uncertain variables u^C and v^C expressed as PCEs where $\boldsymbol{\xi} = (\xi_1, \xi_2, \xi_3)$, $\mathbf{A}$ is the Cholesky decomposition of the matrix $\mathbf{V}$ ($\mathbf{V} = \mathbf{A} \mathbf{A}^T$) and $\|\cdot\|_2$ denotes the L_2 norm.

Note that since the continuum relationship is linear in both w and v^C , one can write an analytical relationship between the PC coefficients of u^C , v^C , and w :

$$u_k = w_k + (w_k - v_k) \frac{h_{MD} - h}{h - \delta}, \text{ for } 0 < k < P. \quad (37)$$

Unlike the case of sampling noise only, the intersection of the atomistic and continuum

response surface cannot be easily visualized. We rely on computations to solve for the uncertain u^C and v^C .

5.1 Solution Method

Given the form of the system of equations (35) and (36), we use a fixed point iterations algorithm for the solution. Starting with an initial guess of one of the variables, the algorithm allows the exchange of this variable between the two equations to obtain a more accurate value at each iteration until convergence (see Figure 4). Since the variables involved in the solution are uncertain and expressed as PCEs, we devise two approaches. In the first one, the mathematical operations (addition, multiplication, etc.) on the PCEs take place intrusively at each iteration. The second one follows the NISP approach, in which the PCE of a variable is sampled and the solution for each sample is obtained separately using a fixed point iterator. The samples are then used to evaluate the integrals in Eq. (8) to recover the PCE of the solution.

5.1.1 Intrusive Spectral Projection (ISP)

The procedure used to solve for the PCEs of u^C and v^C with an ISP approach is summarized in Algorithm 2. We assume a deterministic initial value of u^C *i.e.* the PC coefficients u_k^C are equal to zero for all $k > 0$. A common entity to compute at each iteration is the vector of PCEs $\mathbf{q}(u^C)$ that describes the atomistic response surface. This involves multiple Galerkin operations over PCEs namely powers of u^C and products with σ . Many of these operations are repetitive thus we devise a recursive scheme to minimize the computational effort. Another special operation involved in this approach is the square root of a PCE required to compute the second part of the right-hand side of Eq. (35). All of these operations are performed as discussed in Section 3.1. These operations are repeated until the maximum relative error between two consecutive sets of PC coefficients of u^C is below a given tolerance. Numerical issues usually arise at large uncertainty magnitudes

when computing the logarithm of a PCE. A suitable choice of the PCE basis and order can alleviate these challenges and results in a converging algorithm.

Algorithm 2 solves a system of equations in two uncertain variables expressed as PCEs using ISP in the case of both sampling noise and parametric uncertainty.

$u_0^{C,0} = u_0, u_{k>0}^{C,0} = 0$	{Initial guess of the PC coefficients of u^C }
$i = 0, i_{max} = 500, e_r = 100$	{Maximum number of iterations and initial error}
$\epsilon = 10^{-4}$	{Relative tolerance}

while $e_r > \epsilon$ and $i < i_{max}$ **do**

Compute $\mathbf{q}^i(u^{C,i})$

Compute $\|\mathbf{A}^T \mathbf{q}^i\|_2$

Compute the PCE of $v^{C,i} = \mathbf{q}^{i,T} \boldsymbol{\mu} + \xi_3 \|\mathbf{A}^T \mathbf{q}^i\|_2$ using Eq. 35

Compute an updated PCE $u^{C,i+1}$ using Eq. 36

$e_r = \max_{0 \leq k \leq P} |1 - u_k^{C,i+1} / u_k^{C,i}|$

$i = i + 1$

end while

5.1.2 Non-Intrusive Spectral Projection (NISP)

In the NISP approach, no Galerkin operations are performed. Instead, the PCE of u^C is sampled at each iteration over $M = q^n$ deterministic Gauss quadrature points where n is the number of stochastic dimensions. For each deterministic value of the parameters σ_j , w_j and ζ_j , the system in Eqs. (35) and (36) is solved for u_j^C and v_j^C . These latter are then plugged in Eq. (8) to build the PCEs of u^C and v^C . More details about this NISP approach are given in Algorithm 3.

Algorithm 3 solves a system of equations in two uncertain variables expressed as PCEs using NISP in the case of both sampling noise and parametric uncertainty.

```

 $i_{max} = 500$                                      {Maximum number of iterations}
 $\epsilon = 10^{-4}$                                    {Relative tolerance}
 $M = q^n, q \geq p + 1$                              { $M$  is the total number of quadrature points}
for  $j = 1$  to  $M$  do
     $u_j^{C,0} = u_0$                                    {Initial guess of  $u_j$ }
     $e_r = 100$                                        {Initial error}
     $i = 0$ 
    while  $e_r > \epsilon$  or  $i < i_{max}$  do
        Compute  $v_j^{C,i}$  using Eq. (35)
        Compute an updated  $u_j^{C,i+1}$  using Eq. (36)
         $e_r = |1 - u_j^{C,i+1}/u_j^{C,i}|$ 
         $i = i + 1$ 
    end while
end for
Build the PCEs of  $u^C$  and  $v^C$  using NISP, i.e. Eq. (8)

```

5.2 Results

We consider the case of the atomistic-continuum Couette flow where parametric uncertainty exists in addition to the uncertainty due to finite sampling noise. For instance, we consider the top wall velocity w in the continuum sub-model and the LJ parameter σ in the atomistic sub-model to be uncertain. In this set of results we also rely of $N_{MD} = 200$ short-time MD data averaged at $t_w = 5$ ns. The atomistic response surface $v^C = f^A(u^C, \sigma, \zeta)$ is given by Eq. (22) for the response surface input $\mathbf{x} = (u^C, \sigma)$, i.e.

$$v^C = \mathbf{q}(u^C, \sigma)^T \boldsymbol{\mu} + \zeta \sqrt{\mathbf{q}(u^C, \sigma)^T \mathbf{V} \mathbf{q}(u^C, \sigma)} \quad (38)$$

Based on the uncertainty in σ and given its substantial effect on v^C as shown in Section 4, we assume a fourth order dependence of u^C on σ . Concurrently, we assume a linear dependence of v^C on u^C due to the linear nature of the Couette flow. Hence the polynomial basis for the response surface construction is given by, assuming a fourth order expansion in σ ,

$$\mathbf{q}[u^C(\boldsymbol{\xi}), \sigma(\xi_1)] = \{1, \sigma, \sigma^2, \sigma^3, \sigma^4, u^C, u^C \sigma, u^C \sigma^2, u^C \sigma^3, u^C \sigma^4\}^T \quad (39)$$

We infer values of $\boldsymbol{\mu}$ and $\mathbf{V}$ based on training data obtained from different MD simulations for different values of σ and u^C . We pick 7 and 5 equidistant values in the σ and u^C dimensions, respectively, thus performing a total of 35 MD simulations. These values are denoted by the black dots in Figure 9 where we plot the deterministic part of the f^A surface.

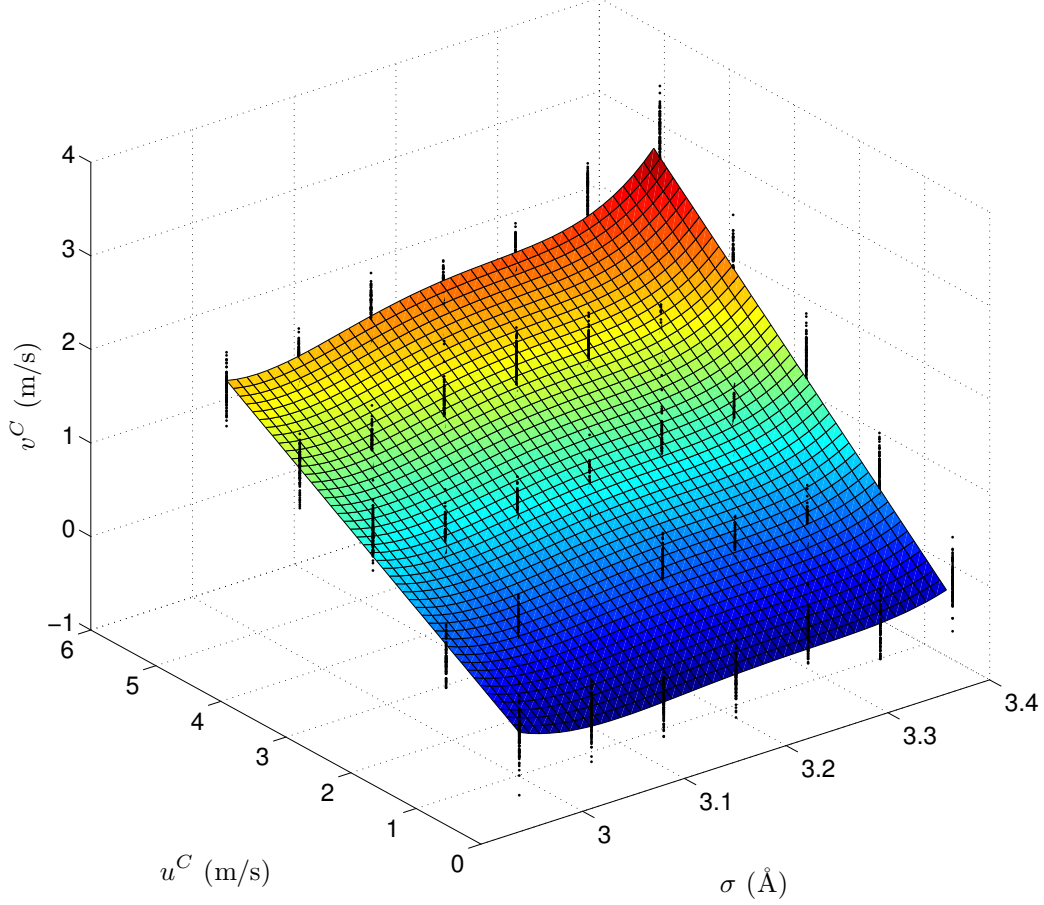


Figure 9: Surface plot of the atomistic mean response surface f^A in the presence of uncertainty in the LJ parameter σ . The black dots denote the variability of the surface due to finite sampling. The plot was generated from $N_{MD} = 200$ short-time data averaged with $t_w = 5$ ns for each sample input $\{u^C, \sigma\}$.

On the continuum side, the response surface $u^C = g^C(v^C)$ is given by Eq. (28). We assume the linear PC expressions for the uncertain inputs σ, w, ζ as given in Eq. (40). and seek PC coefficients for the quantities of interest u^C and v^C given in Eq. (34). In the current context, the uncertain input variables are cast as first-order Hermite-PC expansions in the following way (in the brackets we noted the assumed values of the PC

coefficients):

$$\begin{aligned}
\sigma &= \sigma_0 + \sigma_1 \xi_1 & [= 3.15 + 0.074 \xi_1 \text{ (Å)}] \\
w &= w_0 + w_1 \xi_2 & [= 20 + \xi_2 \text{ (m/s)}] \\
\zeta &= \xi_3,
\end{aligned} \tag{40}$$

where ξ_1, ξ_2, ξ_3 are the stochastic inputs to the problem and are standard normal. Note that, while ζ is a standard Student-t random variable, it is well-approximated here by a standard normal as discussed in Section 4.

5.2.1 Convergence Study

We apply the algorithms and methods developed in Section 5.1 to the coupled atomistic-continuum 1D laminar Couette flow problem described previously and verify the convergence in the PDF of the coupled model output in terms of the PCE order.

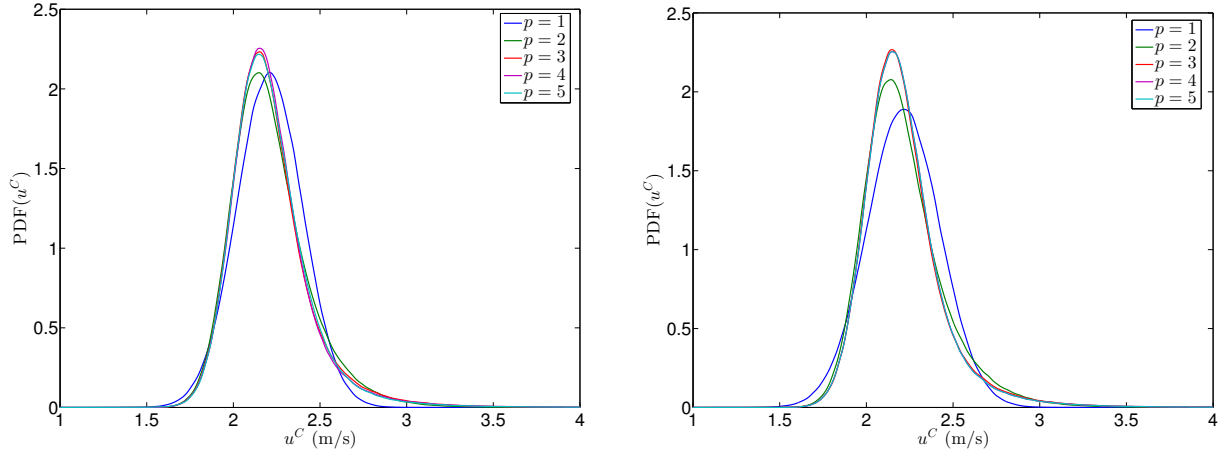


Figure 10: Plots showing the PDFs of u^C in the case of parametric uncertainty and sampling noise for different PCE orders of u^C , as indicated. Plotted are results obtained using the ISP (left) and the (NISP) approaches in the fixed point iterative scheme.

We solve for the PC coefficients u_k^C and v_k^C using the ISP and the NISP fixed iterator approaches described earlier. In the NISP approach, we rely on $M = (p + 1)^n$ Gauss-

Hermite quadrature points to build the PCEs of u^C and v^C where p is given the PCE order and n the number of stochastic dimensions. We draw a large number of samples from these PCEs and using Kernel Density Estimation [41], we construct their equivalent PDFs. These are reported in Figure 10 for u^C for different expansion orders. The results of the ISP and NISP approaches are in very good agreement. The plots imply that an expansion order $p = 3$ is required to accurately quantify the overall uncertainty in u^C . This is expected due to the assumed high order dependence of the v^C velocity on σ (see Eq. 39).

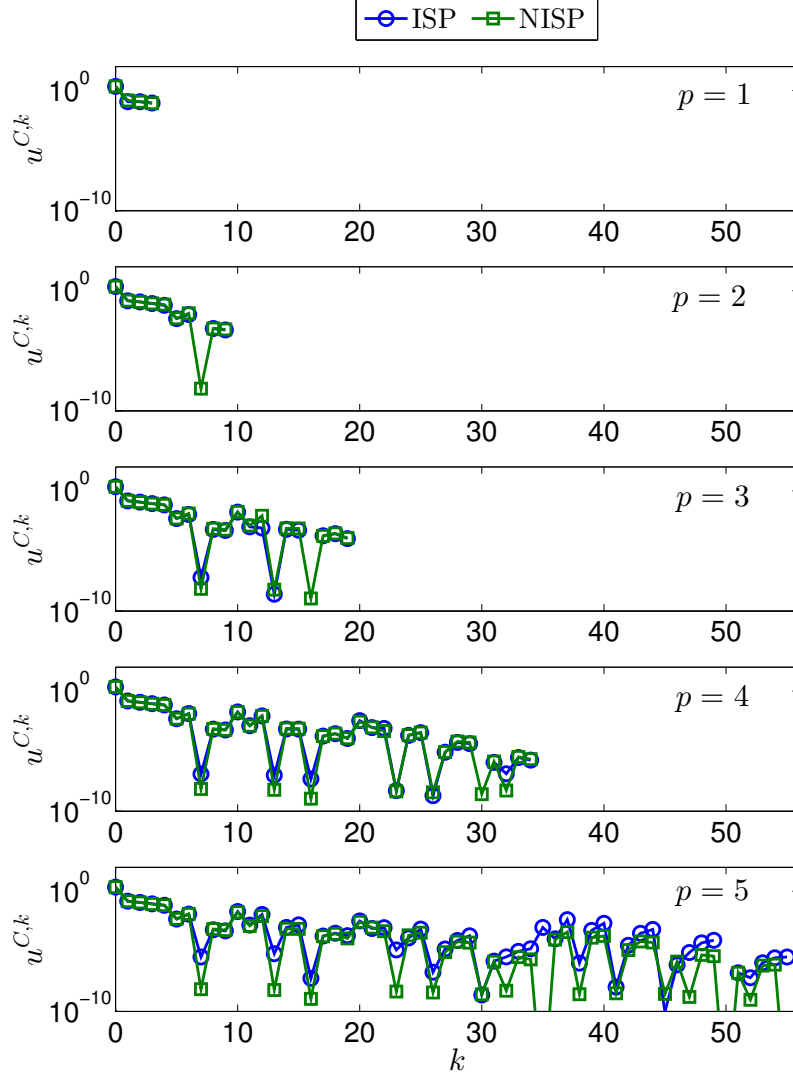


Figure 11: Plots showing the PC coefficients of u^C in the case of parametric uncertainty and sampling noise for different PCE orders of u^C , as indicated. Plotted are results obtained using the ISP (blue) and the NISP (green) approaches in the fixed point iterative scheme.

The computed PC coefficients are shown in Figure 11. Again, a very good agreement between the ISP and NISP approaches is observed. For $p \geq 2$, we notice that the PC coefficients that correspond to the PC modes with higher orders in ξ_2 have negligible amplitude. The number of such PC coefficients increases with p . This is not surprising due to the linear nature of the Couette flow problem where both u^C and v^C vary linearly

with w . Thus, a first order expansion in the $w \sim \xi_2$ dimension is sufficient in the PC representation of u^C and v^C as a function of w .

A performance study (not shown) indicates that in this response surface coupling method, the ISP approach requires more expensive computations than the NISP approach. This is due to the computational overhead required by the Galerkin operations described in Section 5.1.1 upon each fixed point iteration. This overhead quickly increases with the expansion order p since more iterations are required. On the other hand, the computational cost per sample in the NISP approach is negligible thanks to the surrogate expressions f^A and g^C that replace the expensive MD and finite element simulations with simple polynomial relationships between the quantities of interest and the system parameters.

5.2.2 Sensitivity Analysis

In engineering applications, it is often useful to understand the individual effects of the uncertain parameters on a quantity of interest. This latter can either be the u^C or v^C velocity. Such exercise is referred to as sensitivity analysis which, in the PCE context, is performed by the evaluation of the total sensitivity indices derived by Crestaux *et.al.* [42, 43]. Sensitivity index measures the fraction of output variance that can be attributed to the particular input uncertainty. In other words, the higher the sensitivity index, the stronger the contribution of the uncertainty of that parameter on the quantity of interest.

Table 1: The total sensitivity indices of u^C and v^C on the uncertain entities present in the model. Results are computed from their PC representations of u^C and v^C for $p=3$ using the parametric uncertainty level in Eq. 40 where $N_{MD} = 200$ short-time MD data averaged at $t_w=5$ ns were used.

	σ	w	ζ
u^C	0.606	0.254	0.140
v^C	0.745	0.083	0.172

The total sensitivity indices are shown in Table 1 for u^C and v^C . Even though the level of uncertainty in σ is less than 2.5% (see Eq. 40), it contributes more than 60% to the uncertainty in u^C and v^C . As shown previously in Section 4, small variations in σ result in huge variations in u^C and v^C . Thus, in order to reduce the uncertainty in the results in this problem, attention should be mainly focused on reducing the uncertainty in σ . This can be accomplished through more accurate predictions of this parameter by using more computational data and/or more experimental measurements, as described by Rizzi *et al.* [44, 45]

6 Conclusion

This paper addresses the coupling between atomistic and continuum solvers in multiscale simulations. In particular, an approach was developed to account for uncertainty due to model parameters on the continuum and the atomistic level, as well as uncertainty due to the use of a finite number of sample realizations for the extraction of macroscale quantities from atomistic simulations.

For computational efficiency, our approach constructs surrogate models or response surfaces for the coupling variables at the interface between the atomistic and continuum components, as a function of the uncertain parameters in the multiscale model. To build this surrogate model for the atomistic component, an ensemble of atomistic simulations are performed, for sampled values of the atomistic inputs. We use Bayesian inference to infer the atomistic surrogate model from short-time averaged data extracted from these simulations. The resulting posterior uncertainty in this inferred surrogate model represents the uncertainty due to the finite amount of sampling from the atomistic simulations.

The intersection of the surrogate models associated with the continuum and the atomistic components yields the values for the coupling variables. An intrusive and a non-intrusive approach are introduced to perform this intersection, while accounting for the

uncertainty in the model parameters as well as atomistic surrogate model.

The approach was demonstrated on two canonical Couette flow cases. In the first case, only noise due to finite sampling was considered, similar to our prior work [14]. The current surrogate model approach was shown to obtain a much smoother convergence than in our prior work [14], with the same accuracy. The ability of the current approach to also handle parametric uncertainty was demonstrated on a second case, where uncertainty in the atomistic force field models as well as in the continuum driving velocity was considered. The surrogate model approach performed well in this case, with the non-intrusive approach being more stable than the intrusive approach. The formulation also allows for sensitivity analysis to determine the dominant sources of uncertainty in the coupling variables. In the current case, the main source of uncertainty was shown to be the atomistic force field parameter, with the sampling noise making only a minor contribution.

Note that, in the present work, the steady state sampling of mean velocity of a region of the molecular model of the fluid is coupled to classical Stokes flow. Other authors, notably Donev *et al.* [28], couple a microscale model of the fluid to Landau- Lifshitz Navier-Stokes equations where the mean and physical fluctuations upon the mean due to the small length scales of the problem under consideration are both treated at the continuum level. A distinction between the finite sampling noise of the estimate of the time-averaged velocity in this work and the statistical hydrodynamic fluctuations at small length-scales is that the former is reducible (by taking more samples) and the latter is intrinsic, physical and aleatory. We leave a treatment of coupling through the mean and variance of the velocities of the finite sample box for future work, as well as the treatment of unsteady flow where the averaging time window will be tied to the (estimated) time-scale of the macroscopic process.

Overall, the surrogate model approach developed in this paper allows the quantitative assessment of the effects of sampling noise as well as other uncertainties on the predictive fidelity of atomistic to continuum simulations. If better confidence in the model predic-

tions is desired, this approach also allows one to determine whether additional atomistic simulation data is needed to reduce the sampling noise, or if more information is needed to better quantify the input parameters.

7 Acknowledgements

This material is based upon work supported by the U.S. Department of Energy, Office of Science, Office of Advanced Scientific Computing Research. 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 No. DE-AC04-94AL85000.

8 References

- [1] Nasr M Ghoniem, Esteban P Busso, Nicholas Kioussis, and Hanchen Huang. Multiscale modelling of nanomechanics and micromechanics: an overview. *Philosophical magazine*, 83(31-34):3475–3528, 2003. [3](#)
- [2] P Ladeveze. Multiscale modelling and computational strategies for composites. *International Journal for Numerical Methods in Engineering*, 60(1):233–253, 2004. [3](#)
- [3] Dimitri D Vvedensky. Multiscale modelling of nanostructures. *Journal of Physics: Condensed Matter*, 16(50):R1537, 2004. [3](#)

- [4] Jacob Fish. Bridging the scales in nano engineering and science. *Journal of Nanoparticle Research*, 8(5):577–594, September 2006. [3](#)
- [5] P Kanouté, DP Boso, JL Chaboche, and BA Schrefler. Multiscale methods for composites: a review. *Archives of Computational Methods in Engineering*, 16(1):31–75, 2009. [3](#)
- [6] MF Horstemeyer. Multiscale modeling: a review. In *Practical aspects of computational chemistry*, pages 87–135. Springer, 2010. [3](#)
- [7] JA Elliott. Novel approaches to multiscale modelling in materials science. *International Materials Reviews*, 56(4):207–225, 2011. [3](#)
- [8] W. Ren. Analytical and numerical study of coupled atomistic-continuum methods for fluids. *Journal of Computational Physics*, 227:1353–1371, 2007. [3](#), [6](#), [11](#)
- [9] H. Adalsteinsson, B. J. Debuschere, K. R. Long, and H. N. Najm. Components for atomistic-to-continuum multiscale modeling of flow in micro- and nanofluidic systems. *Scientific Programming*, 16:297–313, 2008. [3](#), [12](#)
- [10] K.M. Mohamed and A.A. Mohamad. A review of the development of hybrid atomistic-continuum methods for dense fluids. *Microfluid Nanofluid*, 8:283–302, 2010. [3](#)
- [11] Nicolas G Hadjiconstantinou and Anthony T Patera. Heterogeneous atomistic-continuum representations for dense fluid systems. *International Journal of Modern Physics C*, 8(04):967–976, 1997. [3](#)
- [12] J Tinsley Oden and Kumar Vemaganti. Adaptive hierarchical modeling of heterogeneous structures. *Physica D: Nonlinear Phenomena*, 133(1):404–415, 1999. [3](#)

- [13] Dimitris Drikakis and Nikolaos Asproulis. Quantification of Computational Uncertainty for Molecular and Continuum Methods in Thermo-Fluid Sciences. *Applied Mechanics Reviews*, 64(4):040801, 2011. [3](#), [6](#)
- [14] M. Salloum, K. Sargsyan, H.N. Najm, B. Debusschere, R. Jones, and H. Adalsteinsson. A stochastic multiscale coupling scheme to account for sampling noise in atomistic-to-continuum simulations. *SIAM. Multi. Model. Simu.*, 10(2):550584, 2012. [3](#), [4](#), [6](#), [8](#), [10](#), [11](#), [12](#), [13](#), [18](#), [20](#), [22](#), [25](#), [26](#), [27](#), [39](#)
- [15] Jacob Fish and Wei Wu. A nonintrusive stochastic multiscale solver. *International Journal for Numerical Methods in Engineering*, 88(9):862–879, April 2011. [3](#)
- [16] Yu Zou and Ioannis G Kevrekidis. Uncertainty Quantification for Atomistic Reaction Models: An Equation-Free Stochastic Simulation Algorithm Example. *Multiscale Modeling & Simulation*, 6(4):1217–1233, January 2008. [3](#)
- [17] Wing Kam Liu, Ashfaq Adnan, Adrian M Kopacz, Michelle Hallikainen, Dean Ho, Robert Lam, Jessica Lee, Ted Belytschko, George Schatz, Yonhua Tzeng, Young-Jin Kim, Seunghyun Baik, Moon Ki Kim, Taesung Kim, Junghoon Lee, Eung-Soo Hwang, Seyoung Im, Eiji Ōsawa, Amanda Barnard, Huan-Cheng Chang, Chia-Ching Chang, and Eugenio Oñate. Design of Nanodiamond Based Drug Delivery Patch for Cancer Therapeutics and Imaging Applications. In Dean Ho, editor, *Nanodiamonds: Applications in Biology and Nanoscale Medicine*, pages 249–284. Springer US, Boston, MA, November 2009. [3](#)
- [18] M Arnst, R Ghanem, E Phipps, and J Red-Horse. Dimension reduction in stochastic modeling of coupled problems. *International Journal for Numerical Methods in Engineering*, 92(11):940–968, June 2012. [4](#)
- [19] M Arnst, R Ghanem, E Phipps, and J Red-Horse. Measure transformation and efficient quadrature in reduced-dimensional stochastic modeling of coupled problems.

International Journal for Numerical Methods in Engineering, 92(12):1044–1080, June 2012. [4](#)

- [20] S.T. OConnell and P.A. Thompson. Molecular dynamics-continuum hybrid computations: a tool for studying complex fluid flows. *Physical Review E*, 52:R5792–R5795, 1992. [6](#)
- [21] J. Li, D. Liao, and S. Yip. Coupling continuum to molecular dynamics simulation reflecting particle method and the field estimator. *Physical Review E*, 57:7259–7267, 1998. [6](#)
- [22] J. Li, D. Liao, and S. Yip. Nearly exact solution for coupled continuum/MD fluid simulation. *Journal of Computer-Aided Materials Design*, 6:95–102, 1999. [6](#)
- [23] N. Hadjiconstantinou. Hybrid atomistic-continuum formulations and the moving contact-line problems. *Journal of Computational Physics*, 154:245–265, 1999. [6](#)
- [24] R. Delgado-Buscalioni and P. V. Coveney. Continuum-particle hybrid coupling for mass, momentum, and energy transfers in insteady fluid flow. *Physical Review E*, 67:046704, 2003.
- [25] E. M. Kotsalis, J. H. Walther, and P. Koumoutsakos. Hybrid atomistic-continuum method for the simulation of dense fluid flows. *J. Comput. Phys.*, 205(1):373 – 390, 2005. [6](#)
- [26] E.M. Kotsalis, J.H. Walther, and P. Koumoutsakos. Control of density fluctuations in atomistic-continuum simulations of dense liquids. *Phys. Rev. E*, 76(1):016709, 2007. [6](#)
- [27] S. Yasudaa and R. Yamamoto. A model for hybrid simulations of molecular dynamics and computational fluid dynamics. *Physics of Fluids*, 20:113101, 2008. [6](#)

- [28] A. Donev, J. B. Bell, A. L. Garcia, and B. J. Alder. A hybrid particle-continuum method for hydrodynamics of complex fluids. *SIAM Multiscale Modeling and Simulation*, 8:871–911, 2010. [6](#), [39](#)
- [29] S. Wiener. The Homogeneous Chaos. *J. Comp. Phys.*, 224:560–586, 2007. [6](#)
- [30] R.G. Ghanem and P.D. Spanos. *Stochastic Finite Elements: A Spectral Approach*. Springer Verlag, New York, 1991. [6](#)
- [31] H.N. Najm. Uncertainty Quantification and Polynomial Chaos Techniques in Computational Fluid Dynamics. *Annual Review of Fluid Mechanics*, 41(1):35–52, 2009. [6](#)
- [32] D.S. Sivia. *Data Analysis: A Bayesian Tutorial*. Oxford Science, 1996. [6](#)
- [33] A.R. Leach. *Molecular Modelling: Principles and Applications*. Pearson, 2001. [10](#)
- [34] M.P. Allen and D.J. Tildesley. *Computer Simulation of Liquids*. Oxford Science Publications, 1989. [10](#)
- [35] S.J. Plimpton. Fast parallel algorithms for short-range molecular dynamics. *J. Comp. Phys.*, 117:1–19, 1995. <http://lammps.sandia.gov>. [10](#)
- [36] O.G. Ernst, A. Mugler, H.-J. Starkloff, and E. Ullmann. On The Convergence of Generalized Polynomial Chaos Expansions. *ESAIM: Mathematical Modelling and Numerical Analysis*, 46:317–339, 2012. [14](#)
- [37] B.J. Debuschere, 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. Comp.*, 26(2):698–719, 2004. [15](#), [16](#)
- [38] O.P. Le Maître and O.M. Knio. *Spectral Methods for Uncertainty Quantification with Applications to Computational Fluid Dynamics*. Springer, 2010. [16](#)

- [39] Y.M. Marzouk, H.N. Najm, and L.A. Rahna. Stochastic spectral methods for efficient bayesian solution of inverse problems. *J. Comp. Phys.*, 224:560–586, 2007. [17](#)
- [40] D.S. Sivia and J. Skilling. *Data Analysis, A Bayesian Tutorial*. Oxford Science, second edition, 2006. [17](#)
- [41] A.W. Bowman and A. Azzalini. *Applied Smoothing Techniques for Data Analysis*. New York: Oxford University Press, 1997. [35](#)
- [42] T. Crestaux, O.P. Le Maître, and J.M. Martinez. Polynomial chaos expansion for sensitivity analysis. *Reliability. Eng. Sys. Safety.*, 94(7):1161–1172, 2009. [37](#)
- [43] A. Alexanderian, J. Winokur, I. Sraj, A. Srinivasan, M. Iskandarani, W.C. Thacker, and O.M. Knio. Global sensitivity analysis in an ocean general circulation model: a sparse spectral projection approach. *Comp. Geosc.*, 2(3):757–778, 2012. [37](#)
- [44] F. Rizzi, O.M. Knio, H.N. Najm, B.J. Debusschere, K.M. Sargsyan, M. Salloum, and H. Adalsteinsson. Uncertainty Quantification in MD Simulations. Part I: Stochastic reformulation of the forward problem. *SIAM J. Multiscale Model. Simul.*, 10(4):1428–1459, 2012. [dx.doi.org/10.1137/110853169](https://doi.org/10.1137/110853169). [38](#)
- [45] F. Rizzi, O.M. Knio, H.N. Najm, B.J. Debusschere, K.M. Sargsyan, M. Salloum, and H. Adalsteinsson. Uncertainty Quantification in MD Simulations. Part II: Inference of force-field parameters. *SIAM J. Multiscale Model. Simul.*, 10(4):1460–1492, 2012. [dx.doi.org/10.1137/110853170](https://doi.org/10.1137/110853170). [38](#)