# Data Driven, Predictive Molecular Dynamics for Nanoscale Flow Simulations under Uncertainty

Panagiotis Angelikopoulos,[†] Costas Papadimitriou,[‡] and Petros Koumoutsakos*,[†]

[†]Chair of Computational Science, ETH Zürich, Clausiusstrasse 33, CH-8092, Zürich, Switzerland
[‡]Department of Mechanical Engineering, University of Thessaly, GR-38334 Volos, Greece

S *Supporting Information*

ABSTRACT: For over five decades, molecular dynamics (MD) simulations have helped to elucidate critical mechanisms in a broad range of physiological systems and technological innovations. MD simulations are synergetic with experiments, relying on measurements to calibrate their parameters and probing "what if scenarios" for systems that are difficult to investigate experimentally. However, in certain systems, such as nanofluidics, the results of experiments and MD simulations differ by several orders of magnitude. This discrepancy may be attributed to the spatiotemporal scales and structural information accessible by experiments and simulations. Furthermore, MD simulations rely on parameters that are often calibrated semiempirically, while the effects of their computational implementation on their predictive capabilities have only been sporadically probed. In this work, we show that experimental and MD investigations can be consolidated through a rigorous uncertainty quantification framework. We employ a Bayesian probabilistic framework for large scale MD simulations of graphitic nanostructures in aqueous environments. We assess the uncertainties in the MD predictions for quantities of interest regarding wetting behavior and hydrophobicity. We focus on three representative systems: water wetting of graphene, the aggregation of fullerenes in aqueous solution, and the water transport across carbon nanotubes. We demonstrate that the dominant mode of calibrating MD potentials in nanoscale fluid mechanics, through single values of water contact angle on graphene, leads to large uncertainties and fallible quantitative predictions. We demonstrate that the use of additional experimental data reduces uncertainty, improves the predictive accuracy of MD models, and consolidates the results of experiments and simulations.
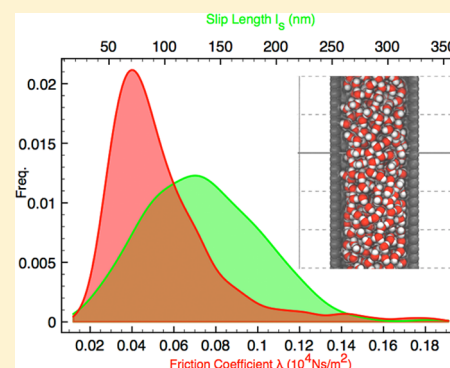
INTRODUCTION

Molecular dynamics (MD) is a powerful computational tool for investigating atomistic phenomena such as water transport in nanotubes and aquaporins as well as the structure and development of materials.[1−3] In recent years, several algorithmic developments and the availability of massively parallel computer architectures has extended the applicability and time scales that can be addressed by MD simulations.[4] Despite these extensions, a critical aspect of MD simulations is that the physics of the system being simulated hinges on the selected interaction potentials and the simulation protocol.

The parameters of MD simulation potentials are often obtained through calibration by related experiments or quantum mechanics simulations. However, when these potentials are implemented for simulating systems beyond their "calibration envelope", the interpretation and validity of the simulated predictions are questionable. The aforementioned calibration envelope thus deals with a question at the core of MD simulations: Given a calibration of an MD force-field model based on experimental observations of a quantity A, how accurately can the MD model predict any other quantity B of either the same or different nature? The quantities A and B can be of structural nature, such as radial distribution fitting of X-ray diffraction data; statistical thermodynamic nature, such as

free energies; or kinematic nature, such as transport coefficient. Such questions can be answered through systematic uncertainty quantification (UQ) and calibration of the devised potentials. So far, limited efforts[5−8] have been dedicated to this direction.

Here, we perform a systematic UQ analysis of MD potentials using as a case study nanoscale flows. We consider MD simulations involving water−graphene interactions quantified by Lennard-Jones (LJ) potentials between carbon and oxygen atoms. Force-field parametrizations based on water contact angles (WCAs) have resulted in substantially different water transport rates through carbon nanotubes (CNTs),[9] while rates reported by experiments and simulations may differ by 3 orders of magnitude. MD simulations have also reported carbon nanotube hydrophobicity to vary from strongly hydrophobic to hydrophilic behavior.[10] In other works, tweaking the potential truncation scheme leads to spurious behavior such as spontaneous filling of CNTs by water molecules[11−13]

Werder et al.[14] used MD simulations to identify the relation of the WCA to the model parameter ϵ_{C-O}^{LJ}, the well depth of the LJ potential. On the basis of the simulation predictions, they

fitted a linear relationship of the WCA to the ϵ_{C-O}^{LJ}. This linear meta-model was used to calibrate the LJ-carbon oxygen parameter ϵ_{C-O}^{LJ} to match an experimental value available for graphite WCA. The protocol was also specific, as this study was for a fixed truncation length $r_{cut} = 1.0$ nm, and with a shift electrostatics truncation scheme. In ref 15, combinations of the interaction parameters and truncation lengths were presented for use in graphite−water interaction simulations. The existence of different sets of parameters ϵ_{C-O}^{LJ} and r_{cut} that all predict accurately the mean value of the calibrated macroscopic water contact angle was also noted. However, these studies, among others, focused on minimizing the discrepancy between the mean value of the WCA predicted by the MD simulations and the mean experimental value, neglecting the effect of computational and experimental uncertainties on predictions of any output quantities of interest (QoI's).

The literature contains literally thousands of related works based on similar strategy calibrations for other materials than carbon, as in refs 9 and 16−18. Examples of QoI's predicted from WCA calibrated MD models include the self-diffusion coefficient $(D_{C_{70}})$ of C_{70} fullerenes in aqueous solutions[19,20] or the movement of isolated water molecules encaged inside a single fullerene molecule.[21]

In this work, the importance of quantifying and propagating uncertainty in MD simulations is investigated. A Bayesian uncertainty quantification and propagation (UQ+P) framework[22−27] is used to quantify and calibrate uncertainties based on available measurements from system or components tests, as well as propagate uncertainties in MD simulations to predict the uncertainties in output QoI.

The focus of the present study is to evaluate the quantitative predictive envelope of water−carbon potentials using as case study WCA calibration. The uncertainty in the carbon−oxygen force-field parameter and the cutoff radius value used to truncate pairwise molecular interactions is estimated using two different system setups for graphite models. The uncertainty is propagated to QoI's related to thermodynamic and kinematic properties of aqueous solutions of fullerenes and water transport across CNTs, illustrating the importance of data-driven UQ of MD models for robust performance evaluation of nanofluidics systems.

■ BAYESIAN UNCERTAINTY QUANTIFICATION AND PROPAGATION

UQ and Model Class Selection in MD. We consider a parametrized class $\mathcal{MD}$ of molecular model structures characterizing water−carbon interactions for MD simulations. Each model class is characterized by the functional forms of the interaction potentials (e.g., LJ potential), the electrostatic Coulombic potential, and the truncation schemes (e.g., cutoff, particle mesh Ewald). In this work, two different model classes for the WCA simulations are employed: The first model class, $\mathcal{MD}_{SG}$, regards simulation of the WCA ϕ between a water droplet and three vertically stacked atomistically resolved graphene sheets resembling the simulation model of ref 14. The second MD model class, $\mathcal{MD}_{Steele}$, assumes an infinite stack of graphene sheets constructed from the same carbon atom building blocks as $\mathcal{MD}_{SG}$, with their effective potential describing their aggregate interaction with water molecules given by the Steele 10-4-3 potential[28] (more details in the Supporting Information).

Let $\theta \in \mathbf{R}^n$ be a set of MD functional model parameters (e.g., bonded and non-bonded force-field parameters or the twin or global cutoff radius) in a model class $\mathcal{MD}$ to be estimated using experimental data. MD involve a larger number of candidate parameters for calibration. In this work, we fix all bonded terms, partial charges, and functional forms and only consider two parameters to be calibrated: the cross interaction strength ϵ_{C-O}^{LJ} and the electrostatic and van der Waals truncation length r_{cut}.

The present analysis is limited to uncertainties in these two parameters due to their relative importance in the MD simulations mentioned in ref 14 and in the specific nanoscale systems analyzed in this work. Specifically, refs 8, 9, and 29 indicated a very strong sensitivity of the water flow inside carbon nanotubes to the cross interaction strength ϵ_{C-O}^{LJ} parameter. The cutoff is also an important parameter with considerably varying values selected by users to trade-off computational effort versus accuracy in MD simulations. As a result, the cutoff values vary significantly among different software packages, not only due to the template input files but also implicitly because of the different truncation schemes used even when full particle mesh Ewald is employed.

The values of the model parameters θ are considered to be uncertain, and a probability distribution function (PDF) quantifies their plausibility. An a priori PDF $\pi(\theta|\mathcal{MD})$ is assigned to the model parameters incorporating prior information that is usually subjective based on previous knowledge, experience, physical, or computational limitations. Specifically, prior to the use of measurements, information about the importance of the cutoff values for determining the precision of simulations and the adverse consequences of high cutoff values in the computational cost should be taken into account in assigning the prior PDF for the cutoff parameter. For example, an upper bound on the support of the prior PDF can avoid values that will yield excessive computational effort.

Bayesian inference is used to update the PDF of θ based on measurements available either at molecular component or system level. For this, let $D \equiv \hat{y} = \{\hat{y}_r, r = 1, ..., n_{\hat{y}}\} \in \mathbf{R}^{n_{\hat{y}}}$ be a set of experimental observations (data), where $n_{\hat{y}}$ is the number of observations performed under different thermodynamic states. The prediction error $e = \hat{y} - g(\theta|\mathcal{MD}) \in \mathbf{R}^{n_{\hat{y}}}$ is introduced to characterize the discrepancy between the experimental observations $\hat{y}$ and the $\mathcal{MD}$ model predictions $g(\theta|\mathcal{MD})$ obtained from a particular value of the model parameters θ given the model class $\mathcal{MD}$. The predictions $g(\theta|\mathcal{MD})$ here can be idealized as, e.g., an entire simulation package that performs the simulation given the functional forms of the molecular interactions, their parameters, and geometries. The prediction error can be decomposed into three contributions $e = e^{exp} + e^{ens} + e^{m}$ accounting for the measurement (experimental) error e^{exp}, the computational error e^{ens} due to the sampling error arising from the finite time size and ensemble choice of our MD simulations, and the modeling error e^{m} due to the inadequacy of the selected $\mathcal{MD}$ to represent exactly the actual behavior of the nanofluidics system.

Assuming zero-mean normally distributed error terms and independent components, justified by the maximum entropy principle,[30] one has that $e^{exp} \sim N(0, \Sigma^{exp})$, $e^{ens} \sim N(0, \Sigma^{ens})$, and $e^{m} \sim N(0, \Sigma^{m})$, with diagonal covariance matrices $\Sigma^{exp} = \text{diag}(\nu_r^2 \hat{y}_r^2)$, $\Sigma^{ens} = \text{diag}(\sigma_r^2 \hat{y}_r^2)$, and $\Sigma^{m} = s^2 \text{diag}(\hat{y}_r^2)$, where ν_r^2 is the normalized variance of the rth thermodynamic state observation, σ_r^2 are the normalized variances measuring the variability in the MD predictions for the corresponding

measured quantities, and s^2 is the normalized variance parameter of the model error. Normalization of the variance parameters is done with respect to $\hat{y}_r$. The variance parameters σ_r^2 are known constants that could be assigned at each simulation of the MD system for a particular value of θ so that the MD simulation terminates after the variance of the response quantity of interest falls below the $\sigma_r^2 \hat{y}_r^2$ value. Finally, it should be noted that nonzero mean Gaussian models could also be readily introduced in the proposed framework to handle systematic biases in the computational or model predictions. However, in the present work, zero-mean Gaussian models were considered adequate as no systematic biases were observed.

Following the Bayesian framework, the updated distribution $f(\theta|D, \mathcal{MD})$ of the model parameters θ, given the data D and $\mathcal{MD}$, is obtained in the form[8]

$$f(\theta|D, \mathcal{MD}) = \frac{|\Sigma|^{-1/2}}{(2\pi)^{N/2}} \exp\left[-\frac{1}{2}J(\theta; \hat{y})\right]\frac{\pi(\theta|\mathcal{MD})}{f(D|\mathcal{MD})}$$

(1)

where

$$J(\theta; \hat{y}) = [\hat{y} - g(\theta|\mathcal{MD})]^T \Sigma^{-1}[\hat{y} - g(\theta|\mathcal{MD})]$$

(2)

is the weighted measure of fit between the MD model predictions and the measured data, $\Sigma = \text{diag}[(\nu_r^2 + \sigma_r^2 + s^2)\hat{y}_r^2]$ is the covariance of the prediction error, $f(D|\mathcal{MD})$ is the *evidence* of the model class $\mathcal{MD}$, selected such that the posterior PDF $f(\theta|D, \mathcal{MD})$ integrates to one, and |.| denotes determinant.

The problem of selecting between two alternative molecular model classes $\mathcal{MD}_i$ and $\mathcal{MD}_j$ based on the observed data D is addressed using the Bayes factor

$$K_{ij} = \frac{f(D|\mathcal{MD}_i)}{f(D|\mathcal{MD}_j)}$$

(3)

given by the ratio of the evidence of the two model classes. In addition, a number of μ competing classes $\mathcal{MD}_1, ..., \mathcal{MD}_\mu$ are ranked, given the data D, based on their probability

$$Pr(\mathcal{MD}_i|D) = \frac{f(D|\mathcal{MD}_i)Pr(\mathcal{MD}_i)}{f(D)}$$

(4)

where $Pr(\mathcal{MD}_i)$ is the prior probability of the model class $\mathcal{MD}_i$ and $f(D)$ is a normalization constant selected such that the sum of the posterior probabilities of all model classes is equal to 1. The most probable model class is selected as the one that maximizes $Pr(\mathcal{MD}_i|D)$ over i.

The high performance computing framework introduced in ref 8 to speed up Bayesian UQ+P computations based on the transitional Markov chain Monte Carlo (TMCMC) algorithm[31] is used for generating samples from the posterior PDF of the model parameters as well as evaluating the evidence of $\mathcal{MD}$ (more details in the Supporting Information).

Uncertainty Propagation for Robust Predictions. Let $f(q|\theta, \mathcal{MD})$ be the conditional PDF of an output QoI q (e.g., diffusion coefficients, free energies, or any other kinematic, structural, or thermodynamic observable in a MD simulation) given the model parameters θ and $\mathcal{MD}$. The conditional PDF $f(q|\theta, \mathcal{MD})$ follows a Gaussian distribution with mean $\mu_q(\theta; \mathcal{MD})$ and variance $\sigma_q^2(\theta; \mathcal{MD})$ computed for an output quantity q directly from the MD simulations for a given value of the model parameters θ.

Robust posterior predictions of q, taking into account the uncertainties in the model parameters of $\mathcal{MD}$ given the measured data D, are quantified by the posterior PDF[32]

$$f(q|D, \mathcal{MD}) = \int_\Theta f(q|\theta, \mathcal{MD})f(\theta|D, \mathcal{MD})\,\mathrm{d}\theta$$

(5)

representing an average of the conditional PDF weighted by the posterior PDF $f(\theta|D, \mathcal{MD})$ of the model parameters. Using the samples $\theta^{(i)}$, $i = 1, ..., N$ drawn from the posterior PDF $f(\theta|D, \mathcal{MD})$, the integral above is approximated by a sample estimate in terms of the conditional PDF $f(q|\theta^{(i)}, \mathcal{MD})$ (Supporting Information). Similar sample estimates are available (Supporting Information) for the posterior robust mean and variance of the output QoI q in terms of $\mu_q(\theta^{(i)}; \mathcal{MD})$ and $\sigma_q^2(\theta^{(i)}; \mathcal{MD})$. Robust predictions can also take into account the competing model classes in eq 4 by averaging their predictions using eq S8 in the Supporting Information.

■ RESULTS

Cutoff Radius Dictates MD Predictions. We first explore the effect of the cutoff radius on MD predictions by considering the potential of mean force (PMF) for two C_{70} fullerenes in aqueous solution. The predictions are based on the model $\mathcal{MD}_{SG}$ with the value of ϵ_{C-O}^{LJ} chosen to be $\epsilon_{C-O}^{LJ} = 0.439$ as proposed in ref 15 based on a WCA calibration. Figure 1 shows
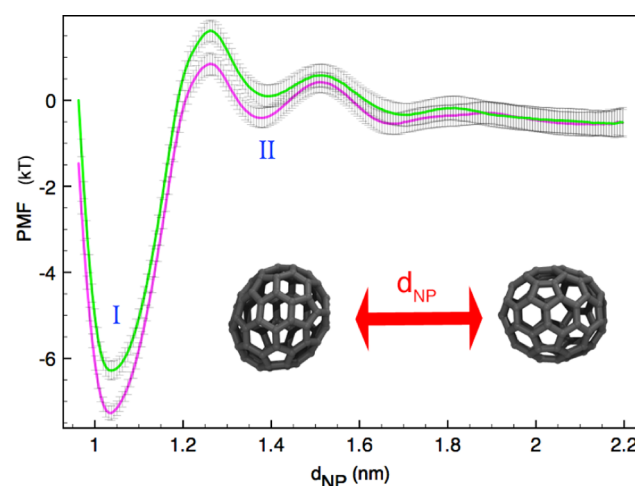
Figure 1. Potential of mean force prediction for two different cutoff radii $r_{cut} = 0.9$ nm and $r_{cut} = 1.35$ nm. Colored curves (green or magenta) indicate mean values of PMF, while gray error bars indicate computational uncertainty from bootstrapping estimates.

the predicted PMF for two different cutoff values $r_{cut} = 0.9$ nm and $r_{cut} = 1.35$ nm (see also the Supporting Information). For each cutoff value, the mean (solid curve) and the fluctuations within one standard deviation (error bars) of the predictions of the mean PMF are shown in the figure. The fluctuations are due to the ensemble error e^{ens}. Predictions from the 0.9 nm (green line) and 1.35 nm (magenta line) cutoff values differ significantly, especially for small intermolecular separations, showcasing the importance in the selection of the cutoff value for carrying out reliable predictions using MD models. The selection of the cutoff values, like other force-field parameters, should be guided from experimental data so that the estimated model parameter values are consistent with observations.

UQ+P. Full MD model simulations are used to quantify uncertainties in the MD model parameters ϵ_{C-O}^{LJ} and r_{cut}, as well

as propagate these uncertainties to predict output QoI's based on the experimental uncertainty in the WCA reported in Table 1 for $T = 298.15$ K, as well as on the computational uncertainties.

Table 1. Data Set for the WCA for Three Different Temperatures[a]

$\mathcal{T}$ (K)	$\hat{y}_r \equiv \phi$ (deg)	$\tilde{v}_r$ (deg)	ref
298.15	92.7	2.3	ref 33
328	88.5	2.7	ref 34
358	86.8	2.1	ref 34

[a]The third column $\hat{v}_r$ refers to the reported measurement uncertainties for each temperature case.

(a). One Parameter Case. The parameter ϵ_{C-O}^{LJ} is first estimated using the data. The estimation is performed for both molecular model classes $\mathcal{MD}_{SG}$ and $\mathcal{MD}_{Steele}$ given a fixed value of the cutoff radius r_{cut}. In order to explore the effect of the cutoff value in UQ+P, the following two values of the cutoff are considered: $r_{cut} = 0.9$ nm and $r_{cut} = 1.35$ nm. Thus, a total of four model classes are introduced, denoted by $\mathcal{MD}_{SG}^{0.9}$, $\mathcal{MD}_{SG}^{1.35}$, $\mathcal{MD}_{Steele}^{0.9}$, and $\mathcal{MD}_{Steele}^{1.35}$, corresponding to the underlining MD model class $\mathcal{MD}_{SG}$ or $\mathcal{MD}_{Steele}$ and the cutoff values $r_{cut} = 0.9$ or 1.35 nm.

The distributions of ϵ_{C-O}^{LJ} are shown in Figure 2a. All four model classes entail uncertainty in the estimate of ϵ_{C-O}^{LJ}, with the optimal value of ϵ_{C-O}^{LJ} and the extent of uncertainty depending on the underlying model and the cutoff value considered. The supports of the two posterior PDFs (magenta curves) in Figure 2a for ϵ_{C-O}^{LJ} regarding $\mathcal{MD}_{SG}^{0.9}$ and $\mathcal{MD}_{SG}^{1.35}$ yield a small overlap at the tails of the distributions. Thus, the uncertainty in ϵ_{C-O}^{LJ} identified by the calibration of the same underlying model depends on the values of the cutoff radius. This implies that one should not use the values of ϵ_{C-O}^{LJ} calibrated with one specific cutoff to make predictions from the same model using another cutoff value, as such predictions will not be anymore consistent with the experimental data. Differences in the PDF supports are also observed for $\mathcal{MD}_{Steele}^{0.9}$ and $\mathcal{MD}_{Steele}^{1.35}$, albeit less pronounced. This is due to the fact that the r_{cut} influence is considerably smaller for the model $\mathcal{MD}_{Steele}$, since only the SPC/E non-bonded interactions are affected from the truncation length change.

On the basis of the results in Figure 2a, the extent of uncertainty (see also Table 2) measured by the coefficient of

Table 2. Mean Values and Coefficients of Variation of the Posterior Distribution of the Model Parameter ϵ_{C-O}^{LJ}, along with the LogLikelihood of Each Model Class

model class	r_{cut} (nm)	$\overline{\epsilon}_{LJ}^{C-O}$ (kJ/mol)	$u_{\epsilon_{C-O}^{LJ}}$	LogEvidence
$\mathcal{MD}_{SG}^{0.9}$	0.9	0.4828	2.84%	1.84
$\mathcal{MD}_{SG}^{1.35}$	1.35	0.4281	3.23%	0.83
$\mathcal{MD}_{Steele}^{0.9}$	0.9	0.4628	3.47%	1.42
$\mathcal{MD}_{Steele}^{1.35}$	1.35	0.4892	3.94%	2.24

variation (COV) $u_{\epsilon_{C-O}^{LJ}}$ ranges from 2.8 to 3.9% of the mean value. The uncertainty $u_{\epsilon_{C-O}^{LJ}}$ found for $\mathcal{MD}_{Steele}^{0.9}$ differs by 50% from the one of $\mathcal{MD}_{SG}^{0.9}$, demonstrating that the uncertainty in the parameter of the LJ potential greatly depends on the underlying MD model class used for nanofluidics simulations.

We propagate the identified ϵ_{C-O}^{LJ} values in order to assess the uncertainty in the predicted WCA. The posterior PDFs of the WCA are shown in Figure 2b and are compared with the uncertainties in the experimental value of the WCA assumed to be Gaussian distributed (see also Table 1). Both baseline models $\mathcal{MD}_{SG}$ and $\mathcal{MD}_{Steele}$, for both cutoff values, provide accurate robust predictions of the WCA. The small deviations in the predictions observed between the four model classes are due to the inclusion of the computational uncertainty in generating the TMCMC samples and the use of a finite number of such samples.

Bayesian model class selection allows one to pick the most probable cutoff among the two used in this study by comparing the evidence of each model class. Specifically, on the basis of the log evidence of the model classes $\mathcal{MD}_{SG}^{0.9}$ and $\mathcal{MD}_{SG}^{1.35}$ given in Table 2, the most probable cutoff is 0.9 nm corresponding to a slightly higher LogEvidence value. The Bayes factor is 1.52, indicating a higher preference of $\mathcal{MD}_{SG}^{0.9}$ in relation to $\mathcal{MD}_{SG}^{1.35}$ based on the data. Robust predictions considering the information from both cutoff values $r_{cut} = 0.9$ nm and $r_{cut} = 1.35$ nm can be obtained from the Bayesian framework by averaging the predictions of both $\mathcal{MD}_{SG}^{0.9}$ and $\mathcal{MD}_{SG}^{1.35}$ using equation S8 in the Supporting Information, where $Pr(\mathcal{MD}_{SG}^{0.9}|D) = 0.73$ and $Pr(\mathcal{MD}_{SG}^{1.35}|D) = 0.27$ for equal model class priors.
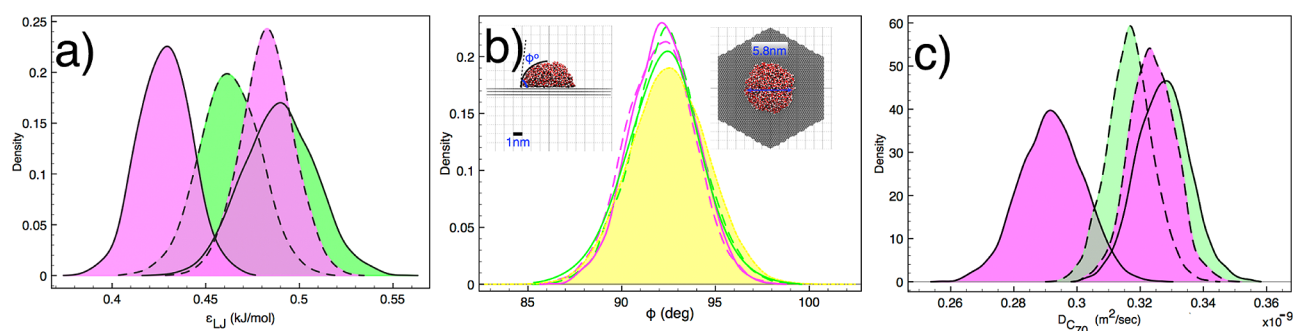

Figure 2. (a) Parameter uncertainty estimation: posterior distributions of the model parameter ϵ_{C-O}^{LJ} for model classes $\mathcal{MD}_{SG}^{0.9}$, $\mathcal{MD}_{SG}^{1.35}$, $\mathcal{MD}_{Steele}^{0.9}$, and $\mathcal{MD}_{Steele}^{1.35}$. (Inset) Snapshot of the system. (b) Uncertainty propagation: posterior distribution of the water contact angle predicted by the four model classes. PDF quantifying the experimental uncertainty is depicted in yellow color. (c) Uncertainty propagation: posterior distribution of the diffusion coefficient of a fullerene molecule in water. Dashed lines indicate $r_{cut} = 1.35$ nm, where solid lines indicate $r_{cut} = 0.9$ nm. Magenta color indicates $\mathcal{MD}_{SG}$, whereas green color indicates $\mathcal{MD}_{Steele}$.

The uncertainties identified for the parameter ϵ_{C-O}^{LJ} are next propagated to robustly predict the diffusion coefficient $D_{C_{70}}$ of a C_{70} fullerene in water. The diffusion coefficients reported herein are with the standard mean square displacement approach (see the Supporting Information) and do not include any correction term for periodic boundary conditions for system size,[35] but the uncertainties are expected to be of the same order even with the additive correction term. A theoretical estimate for $D_{C_{70}}$ via the Stokes−Einstein equation yields a value of $0.20 \cdot 10^{-9}$ m^2/s. The PDFs of the $D_{C_{70}}$, quantifying the uncertainty in the values of $D_{C_{70}}$ predicted by each one of the four model classes, are shown in Figure 2c. The results suggest that prediction of the diffusion coefficient of C_{70}, that has not been used for calibration of the model parameters, greatly depends on the underlying MD model and the cutoff value used. The 95% confidence intervals are $[0.263, 0.314]$ and $[0.304, 0.344]$ m^2/s$\cdot 10^{-9}$ for the two model classes $\mathcal{MD}_{SG}^{0.9}$ and $\mathcal{MD}_{SG}^{1.35}$, respectively, resulting in nonoverlapping supports. The extent of uncertainty of $D_{C_{70}}$ is 3.23 and 1.62% for $\mathcal{MD}_{SG}^{0.9}$ and $\mathcal{MD}_{SG}^{1.35}$, respectively. It is worth noting that robust predictions that take into account the corresponding probabilities $Pr(\mathcal{MD}_{SG}^{0.9}|D) = 0.73$ and $Pr(\mathcal{MD}_{SG}^{1.35}|D) = 0.27$ of each model class given the data can be readily obtained as a weighted average of the distribution predicted in Figure 2c from the two model classes. Similar results related to the predictions $D_{C_{70}}$ from the model class $\mathcal{MD}_{Steele}$ are obtained, although the effect of the cutoff value is less pronounced as observed in Figure 2c.

(b). Two Parameter Case. The Bayesian calibration is next performed by letting also r_{cut} be a free parameter to be estimated from the data. The two distinct two-parameter model classes are now denoted by $\mathcal{MD}_{SG}$ and $\mathcal{MD}_{Steele}$. Uniform prior PDFs are used for both parameters. The values of 0.8 and 1.5 nm are used as bounds for the support of r_{cut}, representing the range of commonly selected values.

The TMCMC samples drawn from the posterior distribution $f(\epsilon_{C-O}^{LJ}, r_{cut}|D, \mathcal{MD})$ of the two model parameters along with the estimated posterior probability functions and the marginal distributions of the two model parameters are presented as a 2 × 2 plot matrix in Figure 3 for $\mathcal{MD}_{SG}$ and Figure SI4 (Supporting Information) for $\mathcal{MD}_{Steele}$, and results also are summarized in Table 3. From the projections in the $(\epsilon_{C-O}^{LJ}, r_{cut})$ space shown in Figure 3, one can identify that the importance region of high probability volume of the posterior PDF is directed along a one-dimensional manifold inside the $(\epsilon_{C-O}^{LJ}, r_{cut})$ space, indicating that uncertainties in these model parameters are strongly correlated. Along this one-dimensional manifold, ϵ_{C-O}^{LJ} and r_{cut} are equally probable, while the posterior PDF decays more rapidly in the directions perpendicular to this manifold. The support of the posterior PDF extends along this one-dimensional manifold as far as the bounds of the uniform prior distribution of r_{cut}, signifying that the set of two parameters is unidentifiable based on available WCA measurement at a single temperature. Comparing the marginal posterior PDFs for the LJ parameter ϵ_{C-O}^{LJ} in Figure 3 with the one in Figure 2a, we note that the uncertainty in ϵ_{C-O}^{LJ} is significantly higher for the two-parameter model classes.

In addition, on the basis of the LogEvidence reported in Table 3, the Bayes factor between model classes $\mathcal{MD}_{SG}$ and $\mathcal{MD}_{Steele}$ is $f(D|\mathcal{MD}_{SG})/f(D|\mathcal{MD}_{Steele}) = 1.8$. Thus, on the basis of the experimental WCA, the model class $\mathcal{MD}_{SG}$ has
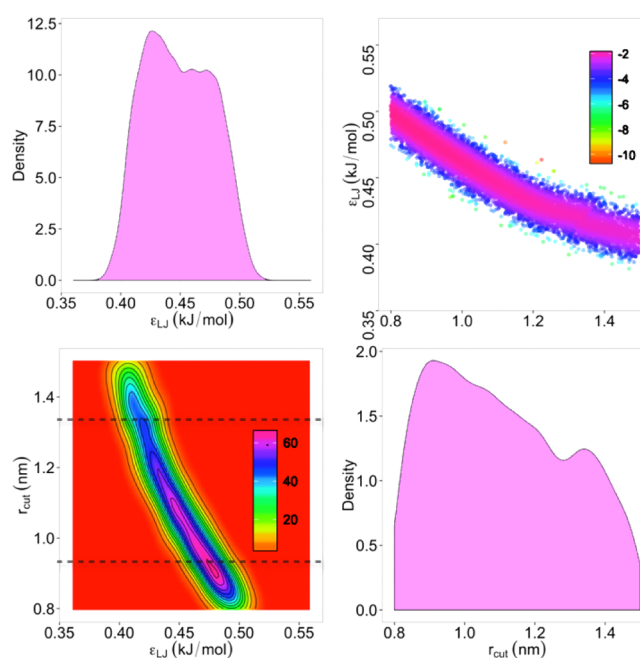
Figure 3. Top right: TMCMC samples of the posterior distributions $f(\epsilon_{C-O}^{LJ}, r_{cut}|D, \mathcal{MD}_{SG})$ in the 2-d parameter space (color bars indicate the LogLikelihood of the samples). Diagonal: Marginal distributions of the model parameters. Bottom left: Reconstructed posterior PDF (color bars indicate posterior PDF values).

slightly higher preference than the model class $\mathcal{MD}_{Steele}$ to represent the experimental data.

Different results for the posterior PDFs of the WCA are shown in Figure 4a for both model classes and compared against the experimental WCA uncertainty. The continuous spectrum of feasible values of the model parameters along the one-dimensional manifold shown in Figure 3 can accurately predict the experimental data and the associated uncertainty, similarly to the case of the one-parameter model classes.

The posterior PDF $f(D_{C_{70}}|D, \mathcal{MD}_{SG})$ quantifying the uncertainties in the self-diffusion coefficient of a C_{70} fullerene in water given the WCA experimental data for the most probable model class $\mathcal{MD}_{SG}$ is presented in Figure 4b (magenta), and it is an average of the predictions from all possible cutoff values in the neighborhood of the one-dimensional manifold, weighted by the plausibility of each value of r_{cut}. The COV of $D_{C_{70}}$ is 6.31%, and the extent of uncertainty defined by the 0.95 percentile values is contained within the interval $[0.27, 0.34]$ m^2/s$\cdot 10^{-9}$. Comparing the size of the supports of the PDFs in Figure 4b, it can be seen that the COV $u_{D_{C_{70}}}$ predicted by the two-parameter model class $\mathcal{MD}_{SG}$ is significantly higher than the one in the diffusion coefficient predicted by the one-parameter model classes (Figure 2c).

We conclude that uncertainty propagation using the two-parameter model class yields quite accurate predictions of the WCA uncertainty which are comparable to the ones predicted by the one-parameter model classes $\mathcal{MD}_{SG}^{0.9}$ and $\mathcal{MD}_{SG}^{1.35}$. However, similarly to the one-dimensional case, the corresponding predictions of output QoI that have not been used for calibration differ substantially. The additional parameter introduces a greater diversity of parameter combinations that match the calibration data, which lead to a greater diversity of predictions (and thus uncertainty) for observables that are not

Table 3. Most Probable Values and Coefficients of Variation of the Posterior Marginal Distributions of the Model Parameters ϵ_{C-O}^{LJ} and r_{cut}

model class	$\mathcal{T}$ (K)	ϵ_{C-O}^{LJ} (kJ/mol)	$u_{\epsilon_{C-O}^{LJ}}$	r_{cut} (nm)	$u_{r_{cut}}$	LogEvidence
$\mathcal{MD}_{SG}$	298.15	0.4770	6.21%	1.23	18.42%	−0.68
$\mathcal{MD}_{Steele}$	298.15	0.4487	3.94%	1.10	17.03%	−1.27
$\mathcal{MD}_{SG}$	[298.15, 328, 358]	0.5073	2.62%	0.96	14.1%	4.84
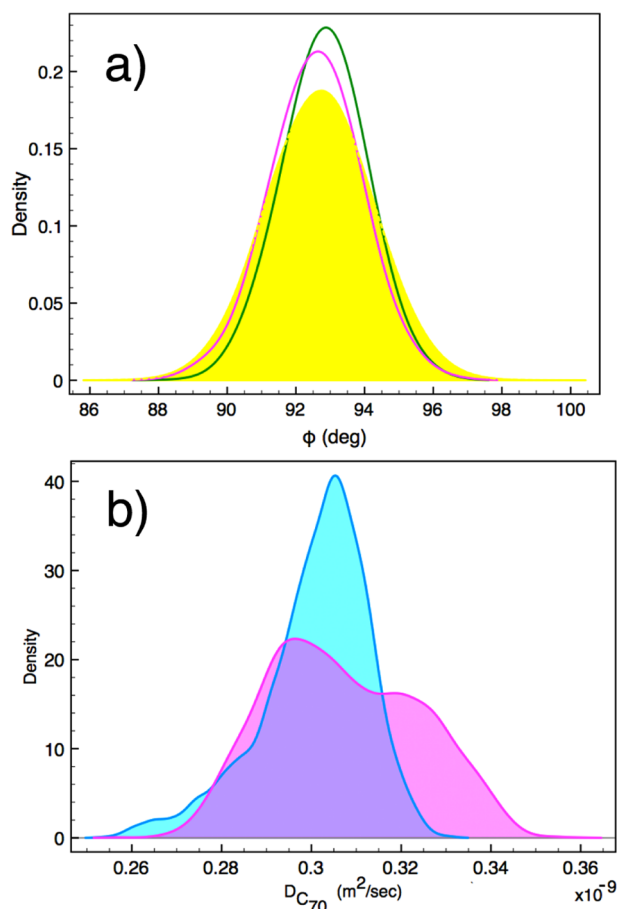
Figure 4. (a) Uncertainty propagation: posterior distribution of the water contact angle predicted from the model classes $\mathcal{MD}_{SG}$ (magenta curve) and $\mathcal{MD}_{Steele}$ (green curve). PDF quantifying the experimental uncertainty is depicted in yellow color. (b) Magenta and blue: propagation to the self-diffusion coefficient using $\mathcal{MD}_{SG}$ and experimental data from one temperature (T = 298.15 K) and three temperatures (T = 298.15, 328, and 358 K), respectively.

in the calibration. For the diffusion coefficient, the model class $\mathcal{MD}_{SG}$ predicts significantly higher uncertainty than the model classes $\mathcal{MD}_{SG}^{0.9}$ and $\mathcal{MD}_{SG}^{1.35}$. This is attributed to the fact that, in the calibration process, the parameter r_{cut} was left free to be estimated from the data. The available measurements in this case are not informative enough to reduce the uncertainties and provide more accurate estimates of the parameters, confined in a smaller region in the parameter space.

It is worth mentioning that considering more uncertain parameters in the proposed analysis, such as partial charges and σ_{C-O}^{LJ}, without using additional data, will cause an increase in the size and dimension of the uncertainty manifold in the resulting higher dimensional parameter space, extending the unidentifiability in the parameters and increasing the uncertainty bounds in the predictions of unobserved QoI.

It should be pointed out that a variety of truncation schemes can be used to model the non-covalent forces. Quantitatively, the uncertainty in the parameters will depend on the choice of the truncation schemes. However, it is not expected to get significant qualitative differences. The proposed Bayesian model class selection method can also be used to compare different truncation schemes and select the best one among alternative potential truncation schemes.[36] Furthermore, the uncertain parameter set can be augmented to include the healing length of truncation schemes[37] as a free parameter to be estimated from the data. In this case, the MD model will have more freedom to fit the data, extending further the size of uncertainties in all parameters and increasing the uncertainties in predictions of unobserved QoI.

■ ROBUST PREDICTIONS FOR NANOSCALE FLOWS

We demonstrate that UQ has important ramifications for applications in nanoscale fluid mechanics. We make robust predictions for the PMF between two fullerenes C_{70} in aqueous solution, as well as their mean breakup and dissociation time by propagating the uncertainties in ϵ_{C-O}^{LJ} and r_{cut} using $\mathcal{MD}_{SG}$ and $\mathcal{MD}_{Steele}$. These estimates are essential for targeted drug delivery[38,39] and colloidal dispersion stability.

Robust predictions of PMF are shown in Figure 5a for both $\mathcal{MD}_{SG}$ and $\mathcal{MD}_{Steele}$. Moving from larger to smaller interparticle distances, uncertainty in the PMF increases substantially. These uncertainties are due to the parametric uncertainties and should be distinguished from the computational uncertainties traditionally presented in the literature via bootstrapping e^{ens} from an isolated PMF evaluation at a fixed ϵ_{C-O}^{LJ}, shown in the inlet of Figure 5a for two different parameter sets (ϵ_{C-O}^{LJ}, r_{cut}).

Uncertainties in the estimates for the C_{70} association and dissociation rates obtained using Kramers theory are presented in Figure 5b for $\mathcal{MD}_{SG}$ (see the Supporting Information[40]). We estimate a mean aggregation time $\bar{k}_{agg}$ = 4.6 min and COV 14%, with an overall uncertainty ranging in the 95% confidence interval [3.24, 7.81] min, relevant to medicinal applications.[41,42] Similarly, the uncertainty in the thermally induced breakup mean time for two aggregated fullerenes is estimated $\bar{k}_{break}$ = 5.1 h with a COV of 22% and an overall uncertainty of [2.78, 7.31] h. The large uncertainties are attributed to the exponential dependence of the Kramers rates to the PMF well. The predictions of the uncertainties in both aggregation and breakup rates are quite sizable, unsuitable for firm predictions in applications such as drug delivery. An alternative way to estimate these rates could be found in the recent work of Morrone et al.[43] The dependence of these estimates though also hinge on the PMF and will thus still mitigate large uncertainty to the predictions.

Are similar uncertainties expected for other systems? One such system regards the friction and the water flow inside CNT arrays.[29,44] To illustrate this, consider the friction coefficient λ between carbon and water, as given by a Green−Kubo
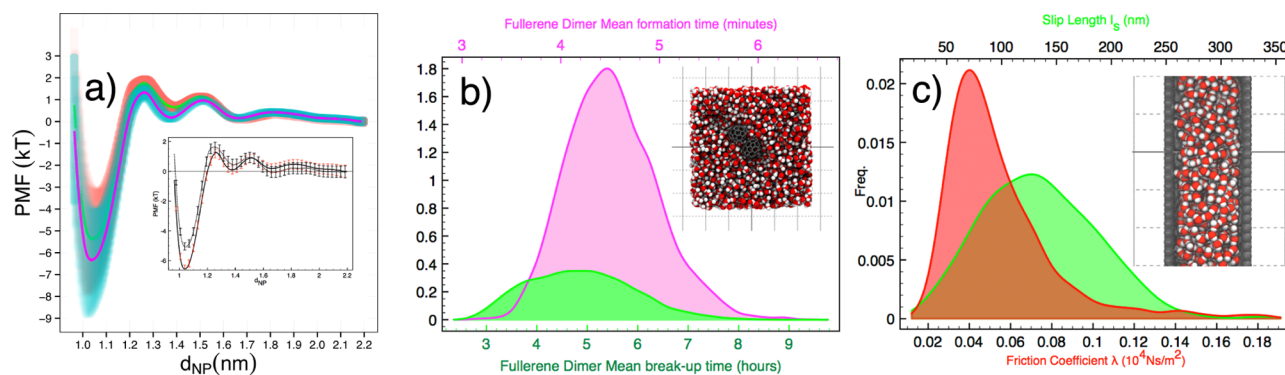
Figure 5. (a) Uncertainty propagation to the potential of mean force. Red and blue shaded areas indicate the combined parametric and ensemble uncertainty. The free energy reference point is set at the maximum measured separation. (b) Uncertainty propagation to the measurement of fullerene dimer formation (magenta-lower axis) and disassociation (green-upper x-axis) in aqueous solution. (Inlet) Snapshot of the system. (c) Robust predictions of the water−carbon friction coefficient (red-lower x-axis) and slip length (green-upper x-axis) in a typical nanopore system. (Inset) Snapshot of the system.

expression for an equilibrium flow free simulation (see the Supporting Information). The uncertainty in the two parameters ϵ_{C-O}^{LJ} and r_{cut} induces uncertainties both in the nanofriction coefficient as well as the slip length l_s being inversely related to the nanofriction coefficient and proportional to the water viscosity in the nanochannel. Robust estimates for these QoI's are given in Figure 5c by propagating the parameter uncertainties using $\mathcal{MD}_{SG}$.

Large uncertainties are observed with COV $u_\lambda = 51\%$ and COV $u_{l_s} = 67\%$. The 95% confidence intervals are [0.014, 0.114] for λ and [13, 237] for l_s. It is important to note here that the values of l_s have been reported to vary from 50 to 170 nm (see, e.g., refs 29 and 45 and refs therein). The present results attribute this volatility to the highly nonlinear dependence of these QoI's to the uncertain parameters of the investigated force fields. Note that the e^{ens} computational error contributions associated with the integrated autocorrelation function evident in the Green−Kubo formalism are relatively high, as the exact time to stop integrating is not easy to be correctly determined, requiring heavy averaging. However, we chose to neglect here this contribution (estimated to introduce a 10% extra uncertainty via the e^{ens}) to highlight the generic sensitivity of such calculations to the two parameters discussed herein. Thus, even assuming negligible computational errors e^{ens}, one would still compute widely varying slip lengths due to uncertainties in the model parameters arising from insufficient experimental data.

■ REDUCTION OF UNCERTAINTY

To alleviate the unidentifiability problem and reduce the uncertainty in the model parameters ϵ_{C-O}^{LJ} and r_{cut}, one needs to consider additional experimental data to provide extra information for estimating the parameters, hoping to confine the uncertainty in smaller region in the parameter space. To investigate the effect of additional test data on the estimation of parameter uncertainty, two additional experimental data values for the WCA are considered[34] for two different temperatures (see Table 1) and the Bayesian calibration of the two parameters ϵ_{C-O}^{LJ} and r_{cut} of the model class $\mathcal{MD}_{SG}$ is repeated. The results for the parameter estimation are shown in Figure 6 and in Table 3. Comparing Figure 6 with Figure 3, it is clear that the size of the support of the posterior PDF of the model parameters is reduced for the case where additional data are
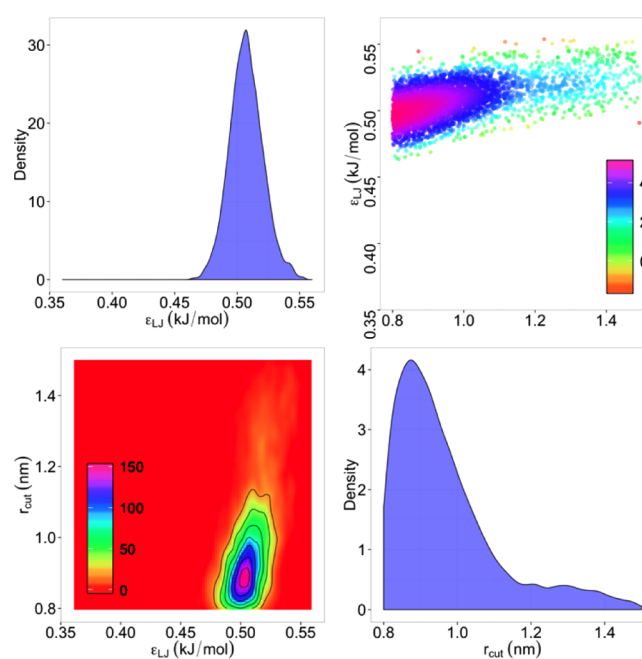
Figure 6. Posterior PDF $f(\epsilon_{C-O}^{LJ}, r_{cut}|D, \mathcal{MD}_{SG})$. Color code and subplots as in Figure 3.

used to quantify parameter uncertainties. The reduction of uncertainty is approximately 17.2% for the LJ parameter ϵ_{C-O}^{LJ} and 33.5% for the cutoff radius r_{cut}. In particular, the uncertainty in the cutoff radius is significantly reduced. The most plausible cutoff radius with higher probability based on the experimental WCA data is confined in the region [0.8, 1.1] nm. This preference for low truncation values is counterintuitive. However, one is reminded that the LJ potentials are empirical approximations to reality and the cutoff radius is one more free parameter of the MD model. This addresses a frequent misconception in molecular simulations suggesting a priori that larger cutoffs provide more accurate predictions. Using $\mathcal{MD}_{SG}$ calibrated against the three experimental WCA, the parametric uncertainty is propagated to $D_{C_{70}}$ in Figure 4b.

This COV is 3.12%, significantly reduced due to the reduction of the parameter uncertainties, and is smaller than the COV 6.21% predicted from the same model class based on a single WCA measurement.

DISCUSSION AND CONCLUSIONS

We demonstrate that a Bayesian uncertainty quantification framework is essential in order to bridge the results of experiments and MD simulations for atomistic phenomena and in particular nanoscale flows. In such systems, it is not uncommon to observe orders of magnitude differences in the reported physical quantities of interest. We provide a systematic basis on the assessment of such differences and the predictability of MD simulations for nanoscale fluid mechanics.

We demonstrate our approach, by revisiting the most common calibration methodology in nanofluidics, namely, estimating the well depth parameter of the LJ interaction potential between SPC/E water and carbon based on WCA measurements. Our approach takes into account the experimental uncertainty of WCA and the computational uncertainty involved in predictions carried out under various MD models. We find that the identified uncertainty in the LJ parameters depends on the cutoff radius used to truncate pairwise molecular interactions and the underlying model selected to predict the WCA. In addition to the well depth parameter, the uncertainty in the cutoff radius was also identified and propagated through MD models for making robust predictions of the dynamical properties of aqueous fullerene solutions and water−CNT systems. Such predictions clearly highlighted certain weaknesses of the force-field calibration method based on a single thermodynamic state of the WCA measurement. QoI such as mean times for fullerene aggregation, PMFs, and slip length for water flowing inside CNTs, exhibiting a highly nonlinear dependence on the MD model parameters, yield large uncertainties that render MD predictions insufficient for making firm assessment of the performance of the analyzed system.

The main uncertainty source in the predictions is attributed to the limited number of experimental data in relation to the number of calibrated parameters. WCA observations at a single temperature are insufficient to yield an identifiable estimate of the two model parameters. It was demonstrated that the uncertainties in the parameters and thus the uncertainties in the predictions of output QoI can be reduced by including in the calibration procedure more experimental data from MD component or system tests. The analysis in this work assumed fixed the values of a number of potential parameters that were not included in the calibration procedure. Given sufficient information from measured data, one could certainly consider additional parameters such as partial charges of water, carbon bond lengths, or computational parameters such as time step. Ways to treat such unidentifiability in the WCA paradigm include as an example the usage of surface tension measurements as additional information. It is expected that unidentifiability will be present, to a certain extent, when more parameters are included in the parameter set, to be identified from the data.

The increased popularity of MD tools dictates a closer attention to a robust reparametrization of the popular force fields before any quantitative study could be trusted in large scale applications. The proposed Bayesian framework can serve the purpose of managing uncertainties in MD simulations, integrating system or component tests for making predictions consistent with available experimental data. The quality of these data certainly affects MD model selection, uncertainty quantification, and prediction accuracy. Thus, future calibration studies must come hand in hand with accurate experimental studies at a wide range of thermodynamic states, helping achieve accurate, transferable parameter sets for reliable, robust quantitative predictions.

ASSOCIATED CONTENT

Ⓢ Supporting Information

Further description of systems used, simulation protocol, and methods. This material is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Author
*E-mail: petros@ethz.ch.

Notes
The authors declare no competing financial interest.

ACKNOWLEDGMENTS

REFERENCES

(1) Yu, J.; Yool, A. J.; Schulten, K.; Tajkhorshid, E. Mechanism of Gating and Ion Conductivity of a Possible Tetrameric Pore in Aquaporin-1. *Structure* **2006**, *14*, 1411−1423.

(2) Schlick, T.; Skeel, R. D.; Brunger, A. T.; Kalé, L. V.; Board, J.; John, A.; Hermans, J.; Schulten, K. Algorithmic Challenges in Computational Molecular Biophysics. *J. Comput. Phys.* **1999**, *151*, 9−48.

(3) Sagui, C.; Darden, T. Molecular dynamics simulations of biomolecules: Long-range electrostatic effects. *Annu. Rev. Biophys. Biomol. Struct.* **1999**, *28*, 155−179.

(4) Shaw, D. E.; Maragakis, P.; Lindorff-Larsen, K.; Piana, S.; Dror, R. O.; Eastwood, M. P.; Bank, J. A.; Jumper, J. M.; Salmon, J. K.; Shan, Y.; Wriggers, W.; et al. Atomic-Level Characterization of the Structural Dynamics of Proteins. *Science* **2010**, *330*, 341−346.

(5) Rizzi, F.; Najm, H.; Debusschere, B.; Sargsyan, K.; Salloum, M.; Adalsteinsson, H.; Knio, O. Uncertainty Quantification in MD Simulations. Part II: Bayesian Inference of Force-Field Parameters. *Multiscale Model. Simul.* **2012**, *10*, 1460−1492.

(6) Cailliez, F.; Pernot, P. Statistical approaches to forcefield calibration and prediction uncertainty in molecular simulation. *J. Chem. Phys.* **2011**, *134*, 054124−054138.

(7) Frederiksen, S. L.; Jacobsen, K. W.; Brown, K. S.; Sethna, J. P. Bayesian Ensemble Approach to Error Estimation of Interatomic Potentials. *Phys. Rev. Lett.* **2004**, *93*, 165501−165505.

(8) Angelikopoulos, P.; Papadimitriou, C.; Koumoutsakos, P. Bayesian Uncertainty Quantication and propagation in molecular dynamics simulations: a high performance computing framework. *J. Chem. Phys.* **2012**, *137*, 144103−144121.

(9) Melillo, M.; Zhu, F.; Snyder, M. A.; Mittal, J. Water Transport through Nanotubes with Varying Interaction Strength between Tube Wall and Water. *J. Phys. Chem. Lett.* **2011**, *2*, 2978−2983.

(10) Walther, J. H.; Jaffe, R. L.; Kotsalis, E. M.; Werder, T.; Halicioglu, T.; Koumoutsakos, P. Hydrophobic hydration of C60 and carbon nanotubes in water. *Carbon* **2004**, *42*, 1185−1194.

(11) Hummer, G.; Rasaiah, J. C.; Noworyta, J. P. Water conduction through the hydrophobic channel of a carbon nanotube. *Nature* **2001**, *414*, 188−190.

(12) Zimmerli, U.; Gonnet, P. G.; Walther, J. H.; Koumoutsakos, P. Curvature induced L-defects in water conduction in carbon nanotubes. *Nano Lett.* **2005**, *5*, 1017−1022.

(13) Bonthuis, D. J.; Rinne, K. F.; Falk, K.; Kaplan, C. N.; Horinek, D.; Berker, A. N.; Bocquet, L.; Netz, R. R. Theory and simulations of water ow through carbon nanotubes: prospects and pitfalls. *J. Phys.: Condens. Matter* **2011**, *23*, 184110−184120.

(14) Werder, T.; Walther, J. H.; Jaffe, R. L.; Halicioglu, T.; Koumoutsakos, P. On theWater- Graphite Interaction for Use in MD Simulations of Graphite and Carbon Nanotubes. *J. Phys. Chem. B* **2003**, *107*, 1345−1352.

(15) Jaffe, R. L.; Gonnet, P.; Werder, T.; Walther, J. H.; Koumoutsakos, P. Water−Carbon Interactions 2: Calibration of Potentials using Contact Angle Data for Different Interaction Models. *Mol. Simul.* **2004**, *30*, 205−216.

(16) Cruz-Chu, E. R.; Aksimentiev, A.; Schulten, K. WaterSilica Force Field for Simulating Nanodevices. *J. Phys. Chem. B* **2006**, *110*, 21497−21508.

(17) Ingebrigtsen, T.; Toxvaerd, S. Contact Angles of Lennard-Jones Liquids and Droplets on Planar Surfaces. *J. Phys. Chem. C* **2007**, *111*, 8518−8523.

(18) Shi, B.; Dhir, V. K. Molecular dynamics simulation of the contact angle of liquids on solid surfaces. *J. Chem. Phys.* **2009**, *130*, 034705−034710.

(19) Bezmel'nitsyn, V. N.; Eletskii, A. V.; Okun', M. V. Fullerenes in solutions. *Phys.-Usp.* **1998**, *41*, 1091−1101.

(20) Fang, K.-C.; Weng, C.-I. Molecular dynamics simulations of structural features and diffusion properties of fullerene-in-water suspensions. *J. Colloid Interface Sci.* **2008**, *318*, 188−194.

(21) Kurotobi, K.; Murata, Y. A Single Molecule of Water Encapsulated in Fullerene C60. *Science* **2011**, *333*, 613−616.

(22) Beck, J.; Yuen, K.-V. Model Selection Using Response Measurements: Bayesian Probabilistic Approach. *J. Eng. Mech.* **2004**, *130*, 192−203.

(23) Yuen, K.-V. *Bayesian Methods for Structural Dynamics and Civil Engineering*; Wiley-VCH Verlag: Singapore, 2010.

(24) Cheung, S. H.; Oliver, T. A.; Prudencio, E. E.; Prudhomme, S.; Moser, R. D. Bayesian uncertainty analysis with applications to turbulence modeling. *Reliab. Eng. Sys. Saf.* **2011**, *96*, 1137−1149.

(25) Congedo, P. M.; Colonna, P.; Corre, C.; Witteveen, J. A. S.; Iaccarino, G. Backward uncertainty propagation method in ow problems: Application to the prediction of rarefaction shock waves. *Comput. Methods Appl. Sci. Eng.* **2012**, *213-216*, 314−326.

(26) Ma, X.; Zabaras, N. A stochastic mixed finite element heterogeneous multiscale method for ow in porous media. *J. Comput. Phys.* **2011**, *230*, 4696−4722.

(27) Oden, J. T.; Hawkins, A.; Prudhomme, S. General Diffuse Interface Theories and an approach to predictive tumor growth modeling. *Math. Models Methods Appl. Sci.* **2010**, *20*, 477−517.

(28) Steele, W. A. The interaction of rare gas atoms with graphitized carbon black. *J. Phys. Chem.* **1978**, *82*, 817−821.

(29) Walther, J. H.; Ritos, K.; Cruz-Chu, E. R.; Megaridis, C. M.; Koumoutsakos, P. Barriers to Superfast Water Transport in Carbon Nanotube Membranes. *Nano Lett.* **2013**, *13*, 1910−1914.

(30) Jeffreys, H. *Theory of Probability*, 3rd ed.; Clarendon Press: Oxford, 1961.

(31) Ching, J.; Chen, Y. C. Transitional markov chain monte carlo method for Bayesian model updating, model class selection, and model averaging. *J. Eng. Mech.* **2007**, *133*, 816−832.

(32) Papadimitriou, C.; Beck, J. L.; Katafygiotis, L. S. Updating robust reliability using structural test data. *Probab. Eng. Mech.* **2001**, *16*, 103−113.

(33) Shin, Y. J.; Wang, Y.; Huang, H.; Kalon, G.; Wee, A. T. S.; Shen, Z.; Bhatia, C. S.; Yang, H. Surface-Energy Engineering of Graphene. *Langmuir* **2010**, *26*, 3798−3802.

(34) Osborne, K. L., III. Temperature-Dependence of the Contact Angle of Water on Graphite, Silicon, and Gold. Master Thesis, Worcester Polytechnic Institute, 2009.

(35) Yeh, I.-C.; Hummer, G. System-Size Dependence of Diffusion Coefficients and Viscosities from Molecular Dynamics Simulations with Periodic Boundary Conditions. *J. Phys. Chem. B* **2004**, *108*, 15873−15879.

(36) Piana, S.; Lindorff-Larsen, K.; Dirks, R. M.; Salmon, J. K.; Dror, R. O.; Shaw, D. E. Evaluating the Effects of Cutoffs and Treatment of Long-range Electrostatics in Protein Folding Simulations. *PLoS One* **2012**, *7*, e39918.

(37) van der Spoel, D.; van Maaren, P. J. The Origin of Layer Structure Artifacts in Simulations of Liquid Water. *J. Chem. Theory Comput.* **2005**, *2*, 1−11.

(38) Koumoutsakos, P.; Pivkin, I.; Milde, F. The Fluid Mechanics of Cancer and Its Therapy. *Annu. Rev. Fluid Mech.* **2013**, *45*, 325−355.

(39) Kraszewski, S.; Tarek, M.; Treptow, W.; Ramseyer, C. Affinity of C60 Neat Fullerenes with Membrane Proteins: A Computational Study on Potassium Channels. *ACS Nano* **2010**, *4*, 4158−4164.

(40) Ford, I. J.; Harris, S. A. Molecular cluster decay viewed as escape from a potential of mean force. *J. Chem. Phys.* **2004**, *120*, 4428−4440.

(41) Suh, J.; Dawson, M.; Hanes, J. Real-time multiple-particle tracking: applications to drug and gene delivery. *Adv. Drug Delivery Rev.* **2005**, *57*, 63−78.

(42) Baish, J. W.; Stylianopoulos, T.; Lanning, R. M.; Kamoun, W. S.; Fukumura, D.; Munn, L. L.; Jain, R. K. Scaling rules for diffusive drug delivery in tumor and normal tissues. *Proc. Natl. Acad. Sci. U.S.A.* **2011**, *108*, 1799−1803.

(43) Morrone, J. A.; Li, J.; Berne, B. J. Interplay between Hydrodynamics and the Free Energy Surface in the Assembly of Nanoscale Hydrophobes. *J. Phys. Chem. B* **2011**, *116*, 378−389.

(44) Falk, K.; Sedlmeier, F.; Joly, L.; Netz, R. R.; Bocquet, L. Molecular Origin of Fast Water Transport in Carbon Nanotube Membranes: Superlubricity versus Curvature Dependent Friction. *Nano Lett.* **2010**, *10*, 4067−4073.

(45) Liu, Y.; Wang, Q. Transport behavior of water confined in carbon nanotubes. *Phys. Rev. B* **2005**, *72*, 085420−085424.

14816

dx.doi.org/10.1021/jp4084713 | *J. Phys. Chem. B* 2013, 117, 14808−14816