Research

CrossMark
click for updates

Author for correspondence:
P. Koumoutsakos
e-mail: petros@ethz.ch

A hierarchical Bayesian framework for force field selection in molecular dynamics simulations

S. Wu[1], P. Angelikopoulos[1], C. Papadimitriou[2], R. Moser[3] and P. Koumoutsakos[1]

[1]Professorship for Computational Science, Clausiusstrasse 33, ETH-Zurich 8092, Switzerland
[2]Department of Mechanical Engineering, University of Thessaly, Leoforos Athinon, Pedion Areos, Volos 38334, Greece
[3]Institute of Computational Engineering and Science, UT-Austin, 201 East 24th Street, Stop C0200, Austin, TX 78712-1229, USA

We present a hierarchical Bayesian framework for the selection of force fields in molecular dynamics (MD) simulations. The framework associates the variability of the optimal parameters of the MD potentials under different environmental conditions with the corresponding variability in experimental data. The high computational cost associated with the hierarchical Bayesian framework is reduced by orders of magnitude through a parallelized Transitional Markov Chain Monte Carlo method combined with the Laplace Asymptotic Approximation. The suitability of the hierarchical approach is demonstrated by performing MD simulations with prescribed parameters to obtain data for transport coefficients under different conditions, which are then used to infer and evaluate the parameters of the MD model. We demonstrate the selection of MD models based on experimental data and verify that the hierarchical model can accurately quantify the uncertainty across experiments; improve the posterior probability density function estimation of the parameters, thus, improve predictions on future experiments; identify the most plausible force field to describe the underlying structure of a given dataset. The framework and associated software are applicable to a wide range of nanoscale simulations associated with experimental data with a hierarchical structure.

THE ROYAL SOCIETY
PUBLISHING

1. Introduction

The exploration of nanoscale phenomena requires a synergy of experiments and simulations. The study of water transport in carbon nanotube membranes is a striking example that helps showcase this need. The first experiments [1,2] reported flow rates that were 3–5 orders of magnitude larger than those estimated by continuum hydrodynamics and attracted significant interest from the community. Simulations in short carbon nanotubes (CNTs) [3,4] appear to confirm these fast flow rates, in particular when extrapolated to larger lengths used in the experiments, while conflicting with simulations in periodic domains [5]. More recent simulations [6] in CNT membranes of $2\,\mu\text{m}$, matching the thickness of the membranes used in the earlier experiments, confirmed the results of the periodic simulations. Recent experimental results [7,8] are also consistent with the results of these simulations. A recent review [9] shows also that earlier experimental results [10] differed with those reported in [2] by orders of magnitude while they were consistent with the results of the periodic simulations. These discrepancies showcase the challenges in performing experiments and simulations for nanoscale fluid flows and the need for synergetic studies. Novel experimental techniques [11,12] show the potential to detect molecular events inside CNTs and as such have the potential to reassess flow transport in CNTs. At the same time a number of combined experimental and simulation studies are beginning to emerge [13].

We argue that the Bayesian framework is a key methodology in bridging the experimental data and the molecular dynamics (MD) simulations [14,15]. The calibration of the force fields in MD simulations is a critical aspect of their predictive capabilities. A number of widely used force fields [16,17] have been developed over the years that have provided us with unique insight in processes ranging from protein folding and crystallization to the development of new materials [18].

The calibration of force fields in MD simulations often relies on experimental data that exhibit a special structure. This structure refers to measurements of the same physical quantity (such as viscosity, thermal conductivity, etc.), using different measurement techniques or under variable environmental conditions. One such example is the measurement of the thermal conductivity in nanofluidics [19] obtained from identical samples by 30 organizations worldwide, reporting a variance of less than 10% from the sample average. For some other quantities, the variance may be larger. Several factors can contribute to such discrepancies, such as the choice of the force field and its calibration, computational errors and experimental uncertainties or physical factors such as the hydrophilicity of the interior of nanochannels [20].

Data with such a hierarchical structure are observed in various fields, including social science studies, biochemical or material testing experiments [21]. Usually model regression techniques lump the data from different sources and perform statistical analysis on the pooled dataset [14,15]. Alternatively, one may group the data in different ways and perform in-group/cross-group statistical analysis [22]. Arguably these approaches do not capture the inherent structure of the data and may not be informative to models due to the variation of the type and level of uncertainties across the experiments. In turn, multilevel regression methods have been developed to address this type of problem [22].

Here, we propose to use a hierarchical Bayesian approach, to handle structured experimental data to calibrate and select force fields in MD simulations. This approach has been shown to provide a rigorous foundation for parameter inference and model selection [23,24] in structural dynamics [25–27]. In recent years, applications of the hierarchical Bayesian models have been expanded to a broad range of problems. Examples include the use of a hierarchical Bayesian model to study human decisions under risk in psychology [28,29], and dynamic causal modelling [30] as it is used in neuroimaging [31].

A major concern in using a hierarchical Bayesian analysis is the computational cost associated with evaluating the evidence that is used for model selection. Computing the evidence requires the evaluation of a number of nested multi-dimensional integrals, with the cost scaling as $\mathcal{O}(Np)$, where N is the presumed levels of hierarchy and p the number of parameters. Hence, for models with expensive function evaluations, such as in MD, the cost can be prohibitive. In this study,

we obtain an efficient approximation of the evidence by combining the Laplace Asymptotic Approximation (LAA) method [32] and the Transitional Markov Chain Monte Carlo (TMCMC) algorithm [33] implemented within the Π4U parallel framework [34]. The proposed scheme reduces the computational cost by orders of magnitude in our MD applications. In this paper, we compare two hierarchical and two single-level Bayesian analyses for MD simulations and analyse their benefits and drawbacks. The analysis is performed using the diffusion property of argon as obtained from MD simulations with varying environmental parameters and experimental data for the viscosity of gaseous krypton.

The remainder of this paper begins with a summary of Bayesian inference and model selection for single-level and hierarchical models (§2). Then, the utility of hierarchical models for inferring MD model parameters from experimental measurements of transport coefficients is explored by using MD simulations under different simulation conditions and with varying model parameters to 'measure' transport coefficients, which are then used to infer the model MD model parameters. In this way, the approach can be evaluated with reference to the known parameters. The details of the MD simulations are described in §3, while the results from inferring the parameters are discussed in §4. The application of the same approach to data from physical experiments is discussed in §5, and finally concluding remarks are provided in §6.

2. Bayesian inference and model selection for hierarchical models

Let $f(x, \boldsymbol{\theta})$ be a given deterministic function of a physical system, where x is the input variable and $\boldsymbol{\theta}$ is the parameters of the function. The probabilistic model based on f can be expressed as

$$y = f(x, \boldsymbol{\theta}) + \epsilon_y, \tag{2.1}$$

where the output y is the sum of an uncertain error term ϵ_y and the function f. A typical choice of probability density function (PDF) for ϵ_y is a Gaussian distribution $N(\epsilon_y | 0, \sigma_y)$ with zero mean and σ_y standard deviation. Given some observed dataset $\mathcal{D}$, one can infer the posterior PDF of $\boldsymbol{\theta}$ after specifying a prior PDF of $\boldsymbol{\theta}$, which represents the knowledge about the model parameters before any data are observed.

Bayesian models can employ several levels of probability distributions to express their parameters. For example, in the context of Gaussian distributions, samples can be drawn with a certain mean and variance but in turn the mean and variance can have probability distributions themselves, defined recursively. Here, we furthermore distinguish between *single-level* and *hierarchical models* depending on the assumed structure of the data (figure 1). In single-level models, all the data are pooled together and then a Bayesian analysis is performed to identify the model parameters. In hierarchical models, the data are first grouped in multiple subsets and the Bayesian analysis is performed in an intra- and intergroup fashion.

(a) Single-level Bayesian model

We denote a single-level model as $\mathcal{M}_S$. The posterior PDF for the model parameters ($p(\boldsymbol{\theta}|\mathcal{D}, \mathcal{M}_S)$) is obtained using Bayes' theorem:

$$p(\boldsymbol{\theta}|\mathcal{D}, \mathcal{M}_S) = \frac{p(\mathcal{D}|\boldsymbol{\theta}, \mathcal{M}_S) p(\boldsymbol{\theta}|\mathcal{M}_S)}{p(\mathcal{D}|\mathcal{M}_S)}, \tag{2.2}$$

where $p(\mathcal{D}|\boldsymbol{\theta}, \mathcal{M}_S)$ is the likelihood, $p(\boldsymbol{\theta}|\mathcal{M}_S)$ is the prior PDF and $p(\mathcal{D}|\mathcal{M}_S)$ is the evidence that will be used for model selection if multiple models are available [23,24]. Note that the evidence is simply a normalizing constant for the posterior PDF that is calculated by

$$p(\mathcal{D}|\mathcal{M}_S) = \int p(\mathcal{D}|\boldsymbol{\theta}, \mathcal{M}_S) p(\boldsymbol{\theta}|\mathcal{M}_S) \, \mathrm{d}\boldsymbol{\theta}. \tag{2.3}$$

rsta.royalsocietypublishing.org *Phil. Trans. R. Soc. A* **374**: 20150032
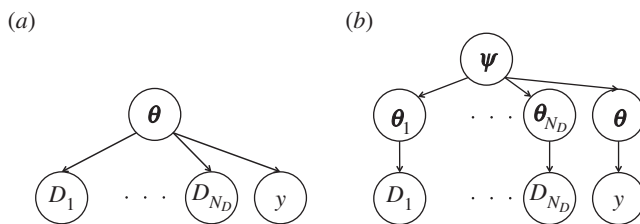
4

Figure 1. Bayesian networks for two different models of data structure. (*a*) Single-level model and (*b*) hierarchical model.

Then, a robust Bayesian approach predicts y given any input x based on the posterior PDF:

$$p(y|x, \mathcal{D}, \mathcal{M}_S) = \int p(y|x, \boldsymbol{\theta}, \mathcal{M}_S) p(\boldsymbol{\theta}|\mathcal{D}, \mathcal{M}_S)\, d\boldsymbol{\theta}. \tag{2.4}$$

In $\mathcal{M}_S$, each observation is assumed to be independent given the values of the model parameters. Hence for $\mathcal{D} = \{(\hat{x}_i, \hat{y}_i)|i = 1, \ldots, N_d\}$, this implies that $p(\mathcal{D}|\boldsymbol{\theta}, \mathcal{M}_S) = \prod_{i=1}^{N_d} p(\hat{y}_i|\hat{x}_i, \boldsymbol{\theta}, \mathcal{M}_S)$.

(b) Hierarchical Bayesian model

In a hierarchical model, the parameters $\boldsymbol{\theta}$ are characterized by hyperparameters $\boldsymbol{\psi}$ that have their own probability distribution. In the current application, hierarchical models are used to quantify the uncertainty due to the variations across different experiments through the hyperparameters $\boldsymbol{\psi}$, and all other uncertainties are accounted for through the error term ϵ_y. In this case, the hierarchical model allows the structure of the data to be exploited, which leads to a different procedure for performing Bayesian inference.

We assume that the data are obtained from $N_\mathcal{D}$ experiments, i.e. $\mathcal{D} = \{D_i|i = 1, \ldots, N_\mathcal{D}\}$ with $D_i = \{(\hat{x}_{j,i}, \hat{y}_{j,i})|j = 1, \ldots, N_{D_i}\}$. Hence, the total number of observations is $N_d = \sum_{i=1}^{N_\mathcal{D}} N_{D_i}$. The hierarchical model ($\mathcal{M}_H$) assumes that there are different values for the parameters $\boldsymbol{\theta}_i$ appropriate for each experiment D_i, and the variability of the parameters among experiments is described by distributions controlled by the hyperparameters $\boldsymbol{\psi}$ and their respective prior. Figure 1*b* shows the mapping between the model parameters and the structure of the data as a Bayesian network.

Under this model, the posterior PDF of the parameters $\boldsymbol{\theta}$ used to predict y in equation (2.4) is

$$p(\boldsymbol{\theta}|\mathcal{D}, \mathcal{M}_H) = \int p(\boldsymbol{\theta}|\boldsymbol{\psi}, \mathcal{M}_H) p(\boldsymbol{\psi}|\mathcal{D}, \mathcal{M}_H)\, d\boldsymbol{\psi}, \tag{2.5}$$

where $p(\boldsymbol{\theta}|\boldsymbol{\psi}, \mathcal{M}_H)$ is the prior model of $\boldsymbol{\theta}$. The posterior PDF of the hyperparameters $\boldsymbol{\psi}$ ($p(\boldsymbol{\psi}|\mathcal{D}, \mathcal{M}_H)$) can also be calculated using Bayes' theorem:

$$p(\boldsymbol{\psi}|\mathcal{D}, \mathcal{M}_H) = \frac{p(\mathcal{D}|\boldsymbol{\psi}, \mathcal{M}_H) p(\boldsymbol{\psi}|\mathcal{M}_H)}{p(\mathcal{D}|\mathcal{M}_H)}. \tag{2.6}$$

Here, $p(\boldsymbol{\psi}|\mathcal{M}_H)$ is the prior of $\boldsymbol{\psi}$, $p(\mathcal{D}|\mathcal{M}_H)$ is the evidence of this model $\mathcal{M}_H$. The likelihood of $\boldsymbol{\psi}$ ($p(\mathcal{D}|\boldsymbol{\psi}, \mathcal{M}_H)$) can be calculated from

$$
\begin{aligned}
p(\mathcal{D}|\boldsymbol{\psi}, \mathcal{M}_H) &= \prod_{i=1}^{N_\mathcal{D}} p(D_i|\boldsymbol{\psi}, \mathcal{M}_H) \\
&= \prod_{i=1}^{N_\mathcal{D}} \int p(D_i|\boldsymbol{\theta}_i, \mathcal{M}_H) p(\boldsymbol{\theta}_i|\boldsymbol{\psi}, \mathcal{M}_H)\, d\boldsymbol{\theta}_i \\
&= \prod_{i=1}^{N_\mathcal{D}} \int \left(\prod_{j=1}^{N_{D_i}} p(\hat{y}_{j,i}|\hat{x}_{j,i}, \boldsymbol{\theta}_i, \mathcal{M}_H) \right) p(\boldsymbol{\theta}_i|\boldsymbol{\psi}, \mathcal{M}_H)\, d\boldsymbol{\theta}_i.
\end{aligned} \tag{2.7}
$$

Table 1. Bayesian inference comparison between a single-level and a hierarchical models.

	single-level ($k = S$)	hierarchical ($k = H$)
posterior $p(\boldsymbol{\theta}\|\mathcal{D}, \mathcal{M}_k) =$	$\dfrac{p(\mathcal{D}\|\boldsymbol{\theta}, \mathcal{M}_S)p(\boldsymbol{\theta}\|\mathcal{M}_S)}{p(\mathcal{D}\|\mathcal{M}_S)}$	$\int p(\boldsymbol{\theta}\|\boldsymbol{\psi}, \mathcal{M}_H)p(\boldsymbol{\psi}\|\mathcal{D}, \mathcal{M}_H)\,\mathrm{d}\boldsymbol{\psi}$
evidence $p(\mathcal{D}\|\mathcal{M}_k) =$	$\int p(\mathcal{D}\|\boldsymbol{\theta}, \mathcal{M}_S)p(\boldsymbol{\theta}\|\mathcal{M}_S)\,\mathrm{d}\boldsymbol{\theta}$	$\int p(\mathcal{D}\|\boldsymbol{\psi}, \mathcal{M}_H)p(\boldsymbol{\psi}\|\mathcal{M}_H)\,\mathrm{d}\boldsymbol{\psi}$

Here, the first line is deduced based on the assumed hierarchical model that datasets are independent when $\boldsymbol{\psi}$ is given, and the third line assumes that data points in a dataset D_i are independent when $\boldsymbol{\theta}_i$ is given. Again, the evidence in equation (2.6) is a normalizing constant that can be calculated from

$$p(\mathcal{D}|\mathcal{M}_H) = \int p(\mathcal{D}|\boldsymbol{\psi}, \mathcal{M}_H)p(\boldsymbol{\psi}|\mathcal{M}_H)\,\mathrm{d}\boldsymbol{\psi}. \tag{2.8}$$

Table 1 summarizes a comparison between the single-level and the hierarchical models.

(c) Efficient approximate model selection

Predictions from multiple models, calibrated from a given dataset $\mathcal{D}$, can be made by

— summing up predictions from each model weighted with its posterior, i.e. hyper-robust prediction [35] or
— using the most probable model.

In either case, for each model ($\mathcal{M}_k$) the evaluation of the posterior probability $P(\mathcal{M}_k|\mathcal{D})$ is needed. Using Bayes' theorem

$$P(\mathcal{M}_k|\mathcal{D}) = \frac{p(\mathcal{D}|\mathcal{M}_k)P(\mathcal{M}_k)}{p(\mathcal{D})} \propto p(\mathcal{D}|\mathcal{M}_k)P(\mathcal{M}_k). \tag{2.9}$$

Equal prior probability is often assigned for all models $P(\mathcal{M}_k)$ so that we can write $P(\mathcal{M}_k|\mathcal{D}) \propto p(\mathcal{D}|\mathcal{M}_k)$, implying that model selection can be performed using the evidence of each model.

The calculation of the evidence requires integration over all model parameters, implying a high computational cost. This is a major challenge for Bayesian model selection especially when the function f is nonlinear with respect to the parameters $\boldsymbol{\theta}$. The computational complexity increases for hierarchical models as they imply a nested integration (equations (2.7) and (2.8)). For complex problems, such as the MD simulations presented herein, there are no analytical expressions available. A Monte Carlo Simulation (MCS) is needed to estimate $p(\mathcal{D}|\mathcal{M}_H)$ and for each MCS sample of $\boldsymbol{\psi}$, another MCS is needed to estimate $p(D_i|\boldsymbol{\psi}, \mathcal{M}_H)$ for all $i = 1, \ldots, N_D$. This nested MCS leads to an exponential increase of the number of total function evaluations. The computational cost is intractable even when the function f can be evaluated efficiently. For example, assuming f can be evaluated in 1 s and 10^4 samples are adequate for a single integral estimate, then a nested integral requires 10^8 samples, thus over 3 years, to complete one evidence estimation. Moreover, the hyperparameter space is often in a much higher dimension introducing an additional exponential increase in sample size using MCS. The application of efficient sampling methods, such as MCMCs, can alleviate but not resolve the inherently high computational demand of hierarchical Bayesian methods.

In this paper, we use the LAA method to approximate the integrals for evaluating $p(D_i|\boldsymbol{\psi}, \mathcal{M}_H)$ in equation (2.7) (see §2e for more details). Then, we use the TMCMC algorithm implemented in $\Pi4U$ [34] to draw samples that approximate the posterior PDF of $\boldsymbol{\psi}$ and estimate the evidence in equation (2.8). The combined method gives an efficient approximation to Bayesian model selection involving a hierarchical model. Note that we pick the TMCMC sampling method over other Markov Chain Monte Carlo (MCMC) methods because of three reasons: (i) it is an inherently

parallel algorithm; (ii) it is based on the idea of simulated annealing, which makes it robust to different kind of non-unimodal distributions; and (iii) the evidence is obtained as a by-product of the algorithm. Ching & Chen [33] and Hadjidoukas *et al.* [34] provide detail explanations to these points.

(d) Approximate robust posterior prediction

In the following, we limit our attention to only hierarchical models and for simplicity of notation, the term $\mathcal{M}_H$ is omitted in all the PDFs described below. The robust posterior prediction for a hierarchical model can be expressed as

$$p(y|x, \mathcal{D}) = \int p(y|x, \boldsymbol{\psi}) p(\boldsymbol{\psi}|\mathcal{D}) \, d\boldsymbol{\psi},$$

$$\text{where } p(y|x, \boldsymbol{\psi}) = \int p(y|x, \boldsymbol{\theta}) p(\boldsymbol{\theta}|\boldsymbol{\psi}) \, d\boldsymbol{\theta}. \tag{2.10}$$

The PDF $p(y|x, \boldsymbol{\psi})$ is estimated using LAA (denoted as $\tilde{p}_{LAA}(y|x, \boldsymbol{\psi})$), and $p(\boldsymbol{\psi}|\mathcal{D})$ is approximated with TMCMC samples, i.e. $p(\boldsymbol{\psi}|\mathcal{D}) = (1/N) \sum_{i=1}^{N} \delta(\boldsymbol{\psi} - \boldsymbol{\psi}^{(i)})$, where $\boldsymbol{\psi}^{(i)}$ is the ith sample from TMCMC and N is the total number of samples. The robust prediction is given by

$$p(y|x, \mathcal{D}) \approx \frac{1}{N} \sum_{i=1}^{N} \tilde{p}_{LAA}(y|x, \boldsymbol{\psi}^{(i)}). \tag{2.11}$$

In many cases, one may have particular interest in the posterior mean $E[y|x, \mathcal{D}]$ and variance $\text{Var}[y|x, \mathcal{D}]$ of the output, which can be then estimated as

$$E[y^j|x, \mathcal{D}] = \int y^j p(y|x, \mathcal{D}) \, dy$$

$$\approx \int y^j \frac{1}{N} \sum_{i=1}^{N} \tilde{p}_{LAA}(y|x, \boldsymbol{\psi}^{(i)}) \, dy$$

$$= \frac{1}{N} \sum_{i=1}^{N} E_{LAA}[y^j|x, \boldsymbol{\psi}^{(i)}], \tag{2.12}$$

where $E_{LAA}[y^j|x, \boldsymbol{\psi}^{(i)}]$ is the LAA estimation for the jth moment of y given input x and the ith sample $\boldsymbol{\psi}^{(i)}$. The first and the second moments can be expressed as [36]

$$E[y|x, \mathcal{D}] \approx \frac{1}{N} \sum_{i=1}^{N} E_{LAA}[y|x, \boldsymbol{\psi}^{(i)}]$$

$$\text{and} \qquad \text{Var}[y|x, \mathcal{D}] = E[y^2|x, \mathcal{D}] - (E[y|x, \mathcal{D}])^2 \tag{2.13}$$

$$\approx \frac{1}{N} \sum_{i=1}^{N} E_{LAA}[y^2|x, \boldsymbol{\psi}^{(i)}] - \left(\frac{1}{N} \sum_{i=1}^{N} E_{LAA}[y|x, \boldsymbol{\psi}^{(i)}] \right)^2.$$

(e) Laplace asymptotic approximation and robust prediction for Gaussian additive noise

In the case of a Gaussian additive noise ϵ_y with mean zero and standard deviation σ_y, we pick the prior model $p(\boldsymbol{\theta}|\boldsymbol{\psi}, \mathcal{M}_H)$ to be a Gaussian distribution with mean $\boldsymbol{\mu}_\theta$ and covariance matrix Σ_θ to obtain analytical expressions for LAA calculations. To estimate $p(D_i|\boldsymbol{\psi}, \mathcal{M}_H)$ in equation (2.7) and $p(y|x, \boldsymbol{\psi})$ in equation (2.10), the LAA assumes that Gaussian distribution is a good approximation to the posterior PDF of $\boldsymbol{\theta}_i$ given all other parameters ($p(\boldsymbol{\theta}_i|, \boldsymbol{\psi}, \sigma_y)$).

Then, the LAA of the corresponding evidence $p(D_i|\boldsymbol{\psi}, \sigma_y)$ is

$$p(D_i|\boldsymbol{\psi}, \sigma_y) \approx (2\pi)^{n/2} p(D_i|\boldsymbol{\theta}_i^{\mathrm{opt}}, \sigma_y) p(\boldsymbol{\theta}_i^{\mathrm{opt}}|\boldsymbol{\psi}) |\mathcal{H}L(\boldsymbol{\theta}_i^{\mathrm{opt}})|^{-1/2},$$

$$\text{where } L(\boldsymbol{\theta}) = -\ln(p(D_i|\boldsymbol{\theta}, \sigma_y) p(\boldsymbol{\theta}|\boldsymbol{\psi}))$$

$$\text{and } \mathcal{H}L(\boldsymbol{\theta}) = \text{the Hessian of } L(\boldsymbol{\theta}). \tag{2.14}$$

Here, $\boldsymbol{\theta}_i^{\mathrm{opt}}$ denotes the optimal values of $\boldsymbol{\theta}_i$ that minimize $L(\boldsymbol{\theta})$.

We use the Matlab function *fmincon* with an analytical gradient and Hessian function of L

$$\nabla L = \sum_{i=1}^{N} \left(\frac{1}{\sigma_y^2}(f_{j,i} - \hat{y}_{j,i}) \nabla f_{j,i} + \Sigma_\theta^{-1}(\boldsymbol{\theta} - \boldsymbol{\mu}_\theta) \right) \tag{2.15}$$

and

$$\mathcal{H}L = \sum_{j=1}^{N_{D_i}} \left(\frac{1}{\sigma_y^2}(f_{j,i} - \hat{y}_{j,i}) \mathcal{H}f_{j,i} + \frac{1}{\sigma_y^2}(\nabla f_{j,i})^2 + \Sigma_\theta \right), \tag{2.16}$$

where $f_{j,i} = f(\hat{x}_{j,i}, \boldsymbol{\theta})$ and the gradient ∇f and Hessian $\mathcal{H}f$ of the deterministic function f can be found using the Matlab symbolic toolbox functions—*gradient* and *hessian*.

For convenience, we augment the hyperparameter space $\boldsymbol{\psi}$ to include also the parameter σ_y and we draw samples in the augmented parameter space using TMCMC. However, only the likelihood function depends on σ_y, but not the prior model $p(\boldsymbol{\theta}|\boldsymbol{\psi}, \mathcal{M}_H)$. Hence, there is a slight change to the expressions for the robust prediction (cf. equation (2.10)):

$$p(y|x, \mathcal{D}) = \int p(y|x, \boldsymbol{\psi}, \sigma_y) p(\boldsymbol{\psi}, \sigma_y|\mathcal{D}) \, d\boldsymbol{\psi} \, d\sigma_y,$$

$$\text{where } p(y|x, \boldsymbol{\psi}, \sigma_y) = \int p(y|x, \boldsymbol{\theta}, \sigma_y) p(\boldsymbol{\theta}|\boldsymbol{\psi}) \, d\boldsymbol{\theta}. \tag{2.17}$$

Here, $p(y|x, \boldsymbol{\psi}, \sigma_y)$ is estimated using LAA for each TMCMC sample (equation (2.14) with (y, x) substituting D_i).

3. Application to molecular dynamics models

Over the last 50 years, MD has been instrumental to investigations of atomistic phenomena in areas ranging from materials science to medicine. MD simulations integrate Newton's equation of motion for interacting atoms using various empirical force fields that are often calibrated using experimental data [37]. The quality and predictive capabilities of MD simulations hinge on the formulation and choice of these force fields [38,39]. In MD simulations, statistical estimates for quantities of interest are obtained either by sampling large systems or by extended simulations of smaller systems with various thermodynamic control algorithms [40]. In addition to the uncertainties introduced by the force fields [15], algorithmic software and hardware uncertainties can give rise to variations in the values of predicted quantities. If one simply performs analysis by grouping all data together in one large dataset irrespective of the code or the simulation set-up used, the resulting 'cross-experiment' (cross-simulation) uncertainty will mix with the other uncertainties and affect the quality of the posterior PDF used for prediction or model selection. In the formulation of, e.g. Cailliez & Pernot [14], the cross-laboratory uncertainty is essentially given to be estimated as an additional Gaussian component of the prediction error equation. In this study, we show that hierarchical Bayesian modelling provides a rational treatment of the cross-simulation uncertainty. Also, we compare the results of single-level models and hierarchical models to verify the benefits of using the hierarchical one.

Many different transport phenomena (e.g. diffusion, thermal conductivity, electrical conductivity, etc.) of non-zero mass particles can be described by kinetic theory, in which a phase space distribution function is evolved in time. Calculating the correct evolution hinges on having the right collision operator that models how small-distance and small-time interactions between

Table 2. Summary of molecular dynamics simulations.

sim.	kernels	force field	no. atoms	T-coupling	P-coupling	r_c (nm)	ϵ_{LJ}	σ_{LJ}
1	SSE2	Rahman64	6400	v-rescale	Berensden	3.0	120.0	0.340
2	SSE2	Rahman64	6400	v-rescale	Berensden	1.2	120.0	0.340
3	SSE2	BFW	6400	v-rescale	Berensden	3.0	120.0	0.340
4	SSE2	OPLS	6400	v-rescale	Berensden	3.0	120.0	0.340
5	SSE2	OPLS	6400	v-rescale	Parrinello–Rahman	3.0	120.0	0.340
6	AVX128	Rahman64	512	v-rescale	Parrinello–Rahman	2.6	120.0	0.340
7	AVX128	Rahman64	512	v-rescale	Parrinello–Rahman	1.2	120.0	0.340
8	AVX128	Rahman64	512	Andersen-massive	MTTK	2.6	142.1	0.336
9	GPU	Rahman64	512	v-rescale	Berenden	2.6	142.1	0.336
10	GPU	Rahman64	512	Nose–Hoover	Parrinello–Rahman	2.6	140.0	0.335

particles change the distribution. MD simulations provide an ideal sandbox for different classical kinetic theories testing, as all inter-particle interactions are calculated and readily provides access to all necessary collision properties. In this section, we provide an example of assimilating MD data into kinetic theory correlations produced by different hardware platforms, number of atoms and force-fields. All of the simulated systems provide measurements of the diffusion coefficient of argon at eight specific thermodynamic conditions—pressure and temperature (P, T).

(a) Molecular dynamics simulation set-up

We perform MD simulations of argon, using the classical Lennard–Jones (LJ) potential to describe the interaction among its atoms

$$\phi_{LJ}(r_{ij}) = 4\epsilon_{LJ} \left[\left(\frac{\sigma_{LJ}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{LJ}}{r_{ij}} \right)^{6} \right], \tag{3.1}$$

where $r_{ij} = \|\mathbf{r}_{ij}\|$, $\mathbf{r}_{ij} = \mathbf{r}_j - \mathbf{r}_i$ and $\mathbf{r}_i$ and $\mathbf{r}_j$ are the positions of particles i and j, respectively. The parameters ϵ_{LJ}, σ_{LJ} are key components of a chosen force field. In MD simulations, a cut-off radius (r_c) is introduced when computing pairwise interactions, to reduce the computational cost. The respective force shifted LJ potential ($\phi(r_{ij})$) can be expressed as

$$\phi(r_{ij}) = \begin{cases} \phi_{LJ}(r_{ij}) - \phi'_{LJ}(r_c)(r_{ij} - r_c) - \phi_{LJ}(r_c) & r_{ij} < r_c \\ 0 & r_{ij} \geq r_c, \end{cases} \tag{3.2}$$

where $\phi'_{LJ}(r) = d\phi_{LJ}(r)/dr$.

The parameters employed for the MD simulations can be summarized in table 2. We report the self-diffusion coefficient of gaseous argon as measured from these simulations in figure 2.

Each simulation consisted of two phases: the equilibration of the system and the production run used to obtain all the ensemble averages. During the equilibration phase, each simulation was run for 30 ns with a time-step $\Delta t = 1$ fs using a random velocity rescaling algorithm [41] as a thermostat and a Berendsen barostat [42] to obtain the desired (P, T) values.

Various thermostats were used during the production phase. The Nose–Hoover thermostat [43] with a coupling constant of 2.0 ps and the Parrinello–Rahman barostat [44] with a coupling constant of 2.0 ps. The velocity rescaling algorithm [41] was used with a constant of 2.0 ps. The Andersen-massive thermostat [45] was used with a constant of 1.5 ps. The MTTK barostat was
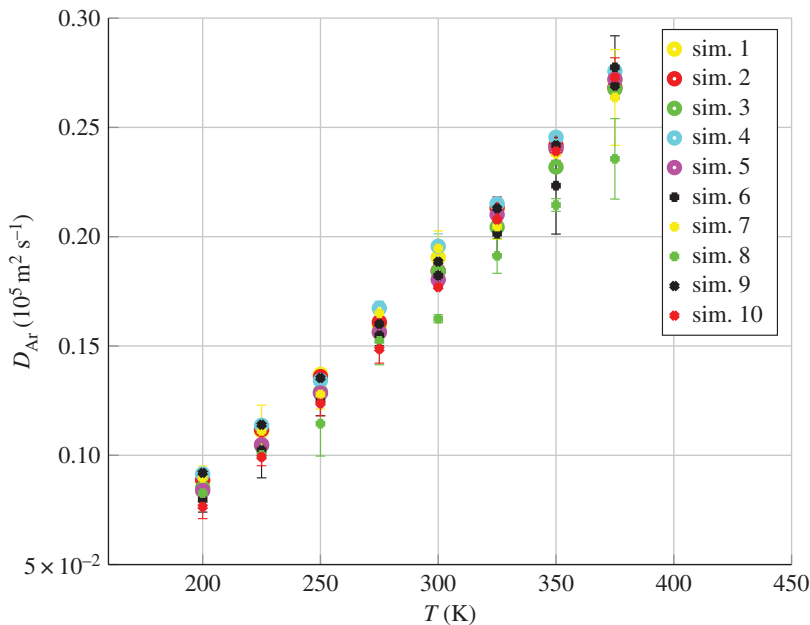
9

rsta.royalsocietypublishing.org Phil. Trans. R. Soc. A **374**: 20150032

Figure 2. Self-diffusion coefficient of gaseous argon D_{Ar} for eight different temperatures using 10 different simulation set-ups. Colour code per simulation as listed in table 2. Error bars depict sampling uncertainty as measured from a linear fit. (Online version in colour.)

used with a constant of 2.0 ps [46]. The production runs were performed for 30 ns, using a time-step of $\Delta t_{\mathrm{prod}} = 1$ fs. Normal periodic boundary conditions were used in all simulations and self-diffusion coefficients were calculated via the mean-squared displacement of the trajectories and the Einstein–Stokes equation.

Equation (3.3) is used to obtain the diffusion coefficient D_{11} given various temperatures T and pressures P under different simulation conditions, in order to infer the model parameters ϵ_{LJ} and σ_{LJ}. In the case of argon, where interatomic interactions are described by simple LJ potentials (equation (3.1)), we can use analytical expressions to calculate the self-diffusion coefficient D_{11} [47]. This semi-empirical correlation is based on Chapman–Enskog gas theory and employs expansions of the collision integrals [48]:

$$
\left.
\begin{aligned}
D_{11} &= \frac{3}{8}\left(\frac{k^3 T^3}{\pi m}\right)^{1/2} \frac{f_D}{P\sigma_{\mathrm{LJ}}^2 \Omega^{(1,1)*}}, \\
f_D &= 1 + \frac{1}{8}(6C^* - 5)^2(2A^* + 5)^{-1}, \\
A^* &= \frac{\Omega^{(2,2)*}}{\Omega^{(1,1)*}} \\
C^* &= 1 + \frac{T^*}{3}\frac{\mathrm{d}\ln\Omega^{(1,1)*}}{\mathrm{d}T^*},
\end{aligned}
\right\}
\tag{3.3}
$$

and

where m denotes the molar mass of argon, and k denotes the Boltzmann factor. The collision integrals Ω are specified as follows, given the non-dimensional temperature $T^* = kT/\epsilon_{\mathrm{LJ}}$:

for $\quad 1.2 \leq T^* \leq 10$

$\Omega^{(1,1)*} = 1.1874(C_6^*/T^*)^{1/3}$

$$\cdot \, [1 + b_1(T^*)^{1/3} + b_2(T^*)^{2/3} + b_3(T^*) + b_4(T^*)^{4/3} + b_5(T^*)^{5/3} + b_6(T^*)^2]$$

$$b_1 = 0, \quad b_2 = 0, \quad b_3 = 10.0161 - 10.5395(C_6^*)^{-1/3}, \quad b_4 = -40.0394 + 46.0048(C_6^*)^{-1/3}$$

$$b_5 = 44.3202 - 53.0817(C_6^*)^{-1/3}, \quad b_6 = -15.2912 + 18.8125(C_6^*)^{-1/3}$$

$$\Omega^{(2,2)*} = \exp[0.46641 - 0.56991(\ln T^*) + 0.19591(\ln T^*)^2 - 0.03879(\ln T^*)^3 + 0.00259(\ln T^*)^4],$$

where (C_6^*) is $(C_6^*) = C_6/\epsilon_{LJ}\sigma_{LJ}^6$ and $C_6 = 0.6041995$.

(b) Comparison of probabilistic models

In this study, we compare four probabilistic models: two single-level and two hierarchical models. We emphasize that the pre-processing (e.g. grouping) of the data are considered a crucial part of the hierarchical Bayesian modelling. Furthermore, we reiterate that the data use for Bayesian inference in this study are obtained from MD simulations using *a priori* specified parameters. Hence we process 80 data points resulting from 10 MD simulation performed at eight different values of (T, P). These data are then grouped into different datasets that are used to infer the model parameters of the analytical expressions given above. A particular aspect of using data from these simulations is that we can then compare the inferred parameters with the respective ones that have been used in the simulations. In summary, we consider the following models:

$\mathcal{M}_{S1}$: A single-level model that groups the 80 data points into one dataset without distinguishing between T, P and other simulation set-ups.

$\mathcal{M}_{S2}$: A single-level model with a reduced dataset. The reduced dataset has eight points, defined by the weighted mean and standard deviation of the values obtained from the 10 different simulations for the same T and P.

$\mathcal{M}_{H1}$: A hierarchical model that groups data obtained from the same simulation set-ups into individual datasets (a total of 10 datasets, each with eight data points for different T and P).

$\mathcal{M}_{H2}$: A hierarchical model that groups data with the same T and P into individual datasets (a total of eight datasets, each with 10 data points from each simulation set-ups).

$\mathcal{M}_{S1}$ represents a typical treatment of a structured dataset where the hierarchical structure is ignored [15] and the uncertainty variable ϵ_y is expected to capture all uncertainties in the inference process. $\mathcal{M}_{S2}$ represents another treatment to a structured dataset where the uncertainty across different datasets is represented by the weighted standard deviation of the data. Hence, the uncertainty variable ϵ_y accounts for all other uncertainties. $\mathcal{M}_{H1}$ represents the appropriate treatment to a structured dataset under the Bayesian approach, where the knowledge about the structure of the datasets is used. Hence, the hyperparameters are expected to account for the uncertainty across different datasets, and ϵ_y will account for all other uncertainties. $\mathcal{M}_{H2}$ represents a similar Bayesian treatment to a structured dataset as in $\mathcal{M}_{H1}$, but the structure of the datasets is incorrectly identified. Hence, we do not expect the same results for ϵ_y and the hyperparameters as in $\mathcal{M}_{H1}$.

(c) Algorithmic details

Hyperparameters: A bivariate Gaussian prior is chosen for the model parameters ϵ_{LJ} and σ_{LJ} in $\mathcal{M}_{H1}$ and $\mathcal{M}_{H2}$ along with five hyperparameters—μ_ϵ, σ_ϵ, μ_σ, σ_σ and ρ, which are the mean and standard deviation of ϵ_{LJ} and σ_{LJ} and the correlation coefficient between ϵ_{LJ} and σ_{LJ}, respectively. Hence, $\boldsymbol{\mu}_\theta$ and Σ_θ in §2e are

$$\boldsymbol{\mu}_\theta = \begin{pmatrix} \mu_\epsilon \\ \mu_\sigma \end{pmatrix} \quad \text{and} \quad \Sigma_\theta = \begin{bmatrix} \sigma_\epsilon^2 & \rho\sigma_\epsilon\sigma_\sigma \\ \rho\sigma_\epsilon\sigma_\sigma & \sigma_\sigma^2 \end{bmatrix}. \tag{3.4}$$

Table 3. Boundaries of the uniform priors for the parameters.

variable	lower bound	upper bound	units
ϵ_{LJ}	60	220	K
σ_{LJ}	0.2	0.5	nm
μ_ϵ	90	200	K
σ_ϵ	0.014	35	K
μ_σ	0.25	0.45	nm
σ_σ	0.00034	0.085	nm
ρ	−0.99	0.99	—
σ_y	0.0001	0.015	10^5 m^2 s^{-1}

Likelihood: This study employs an uncertainty variable ϵ_y with a Gaussian distribution of zero mean and σ_y standard deviation, with σ_y one of the parameters to be inferred. The function f (see equation (2.1)) is computed from equation (3.3), with inputs $x = \{T, P\}$ (temperature and pressure) and model parameters $\theta = \{\epsilon_{LJ}, \sigma_{LJ}\}$. The likelihood function $p(y_i|x_i, \theta, \sigma_y, \mathcal{M}_k) = N(y_i|f(\{T_i, P_i\}, \{\epsilon_{LJ}, \sigma_{LJ}\}), \sigma_y)$ is a Gaussian distribution with mean f and standard deviation σ_y, for all data points $(y_i, x_i) = (y_i, \{T_i, P_i\})$ in single-level and hierarchical models ($k = S$ or H).

Prior: Uniform priors are chosen for the model parameters θ in $\mathcal{M}_{S1}$ and $\mathcal{M}_{S2}$, the hyperparameters ψ in $\mathcal{M}_{H1}$ and $\mathcal{M}_{H2}$ and the parameter σ_y. Table 3 shows the values of the upper and lower bounds of the uniform priors based on the physical constraints of the chosen MD model.

Posterior estimate: For the single-level models $\mathcal{M}_{S1}$ and $\mathcal{M}_{S2}$, we infer three parameters (ϵ_{LJ}, σ_{LJ} and σ_y). We use MCS to estimate the evidence by using samples of the uniform prior obtained from a uniform grid, which leads to an Importance Sampling approximation (with proposal PDF that equals to the prior PDF) of the posterior PDF $p(\epsilon_{LJ}, \sigma_{LJ}, \sigma_y|\mathcal{D}, \mathcal{M}_{S1 \text{ or } S2})$. Five hundred grid points are used for ϵ_{LJ} and σ_{LJ}, and 50 grid points are used for σ_y for the uniform prior range (table 3), with a total of two million samples for the three-dimensional parameter space. The importance sampling approximation of the posterior with N samples from the prior PDF can be written as

$$p(\epsilon_{LJ}, \sigma_{LJ}, \sigma_y|\mathcal{D}, \mathcal{M}_{1a \text{ or } 1b}) \approx \sum_{i=1}^{N} w_i \delta(\epsilon_{LJ} - \epsilon_{LJ}^{(i)})\delta(\sigma_{LJ} - \sigma_{LJ}^{(i)})\delta(\sigma_y - \sigma_y^{(i)}),$$

where $w_i \propto p(\mathcal{D}|\epsilon_{LJ}^{(i)}, \sigma_{LJ}^{(i)}, \sigma_y^{(i)}, \mathcal{M}_{1a \text{ or } 1b})$, $\sum_{i=1}^{N} w_i = 1$

and $\{\epsilon_{LJ}^{(i)}, \sigma_{LJ}^{(i)}, \sigma_y^{(i)}\} \sim p(\epsilon_{LJ}, \sigma_{LJ}, \sigma_y|\mathcal{M}_{1a \text{ or } 1b})$. (3.5)

For the hierarchical models $\mathcal{M}_{H1}$ and $\mathcal{M}_{H2}$, the posterior distributions for all parameters (θ, ψ and σ_y) are computed using the algorithm described in §2c.

TMCMC parameters: TMCMC generates posterior PDF samples from the prior PDF based on multiple stages of MCMC sampling that follows an annealing schedule for the likelihood function [33]. In each stage, MCMC is performed based on a Gaussian local random walk in order to propose a new sample. The number of samples per stage, which is also the final number of posterior samples, is 20 000. TMCMC has two parameters to tune: τ_{CV} controls the annealing speed and β_{COV}^2 controls size of the Gaussian proposal. Suggested by Ching & Chen [33], we choose $\tau_{CV} = 100\%$ to achieve optimal convergence from annealing between the prior and the posterior. We choose $\beta_{COV}^2 = 0.09$, which is slightly higher than the suggested value by Ching & Chen [33], so that the random walk samples explore a larger area in the hyperparameter space.

4. Bayesian inference using data obtained from simulations

We run MD simulations to obtain diffusion coefficients of gaseous argon under different environmental settings and prescribed ϵ_{LJ} and σ_{LJ} values. We compare the results of Bayesian inference using $\mathcal{M}_{S1}$, $\mathcal{M}_{S2}$, $\mathcal{M}_{H1}$ and $\mathcal{M}_{H2}$ with our simulation set-ups to verify the benefits of using a hierarchical model over a single-level model.

(a) Verification of the Laplace asymptotic approximation

We observe that the posterior PDF $p(\boldsymbol{\theta}_i|D_i, \boldsymbol{\psi}, \sigma_y)$ for all datasets D_i is close to a Gaussian distribution in the problems studied herein. Figure 3 shows an instance of the likelihood functions for one dataset. Similar shapes are obtained for all other datasets. Figure 4 shows the prior $p(\boldsymbol{\theta}_i|\boldsymbol{\psi})$ and posterior PDF $p(\boldsymbol{\theta}_i|D_i, \boldsymbol{\psi}, \sigma_y)$ of three random samples of the hyperparameters $\boldsymbol{\psi}$ and σ_y based on the likelihood in figure 3. The resulting posterior PDF can be well approximated by a single bivariate Gaussian distribution. To further verify the quality of LAA in our studies, we compare the estimation of the evidence $p(D_i|\boldsymbol{\psi}, \sigma_y)$ using LAA with the estimate using MCS. Figure 5 illustrates that the MCS and LAA estimates of the evidence are consistent using two random samples of $\boldsymbol{\psi}$ and σ_y.

(b) Verification of Transitional Markov Chain Monte Carlo

TMCMC is a multi-stage sampling scheme where samples from the intermediate stages are used to estimate the evidence, but are not used to approximate the posterior PDF. The scheme achieves a high quality of PDF estimation in the final stage. To verify the performance of TMCMC, the resulting posterior PDF is compared with the posterior estimated by direct MCS with 5 million samples on the six dimension parameter space (five hyperparameters and the likelihood uncertainty parameter)—μ_ϵ, σ_ϵ, μ_σ, σ_σ, ρ and σ_y. Figures 6 and 7 show the resulting samples in the six-dimensional hyperparameter space using TMCMC and MCS, respectively. We observe a significantly better convergence for the TMCMC. The total number of stages equals 8 with 20 000 samples in each stage; thus, a total of 0.16 million samples are generated using TMCMC. Hence, with an order of magnitude less samples, TMCMC achieves a better convergence than MCS in this six-dimensional parameter space.

(c) Model comparison

Figures 8 and 9 show the marginalized posterior PDF of the model parameters $p(\epsilon_{LJ}, \sigma_{LJ}|\mathcal{D}, \mathcal{M}_k)$ for $k = S1, S2, H1, H2$. One can observe that, as expected, the hierarchical model $\mathcal{M}_{H1}$ provides the best posterior that represents the data from the 10 simulations. While the other three models show a strong correlation between ϵ_{LJ} and σ_{LJ}, $\mathcal{M}_{H1}$ gives a posterior PDF distributed as a bivariate Gaussian that accurately covers the parameter values from the 10 datasets (figures 8 and 9). This indicates that a hierarchical model can separate the cross-simulation uncertainty from the other types of uncertainties.

How can we verify whether the hierarchical model can indeed separate the uncertainties correctly? Figure 10 shows the marginalized posterior PDF of σ_y for all four models. Recall that $\mathcal{M}_{S2}$ is designed such that it excludes the cross-simulation uncertainty from σ_y, while $\mathcal{M}_{S1}$ will include all uncertainties to σ_y. One can see that σ_y posterior of $\mathcal{M}_{H1}$ agrees with $\mathcal{M}_{S2}$, while the σ_y posterior of $\mathcal{M}_{H2}$ agrees with that of $\mathcal{M}_{S1}$. This implies that a hierarchical model can, indeed, correctly separately out the cross-simulation uncertainty. On the other hand, a hierarchical model that does not reflect the structure of the data can be misleading.

Figures 11 and 12 show the marginalized posterior PDF of the five hyperparameters. We observe that both $\mathcal{M}_{H1}$ and $\mathcal{M}_{H2}$ give a similar optimal value of the mean for the model parameters, but the uncertainty is significantly larger for $\mathcal{M}_{H2}$. However, both σ_ϵ and σ_σ are approaching zero in $\mathcal{M}_{H2}$, but their values are higher in $\mathcal{M}_{H1}$. These two parameters are
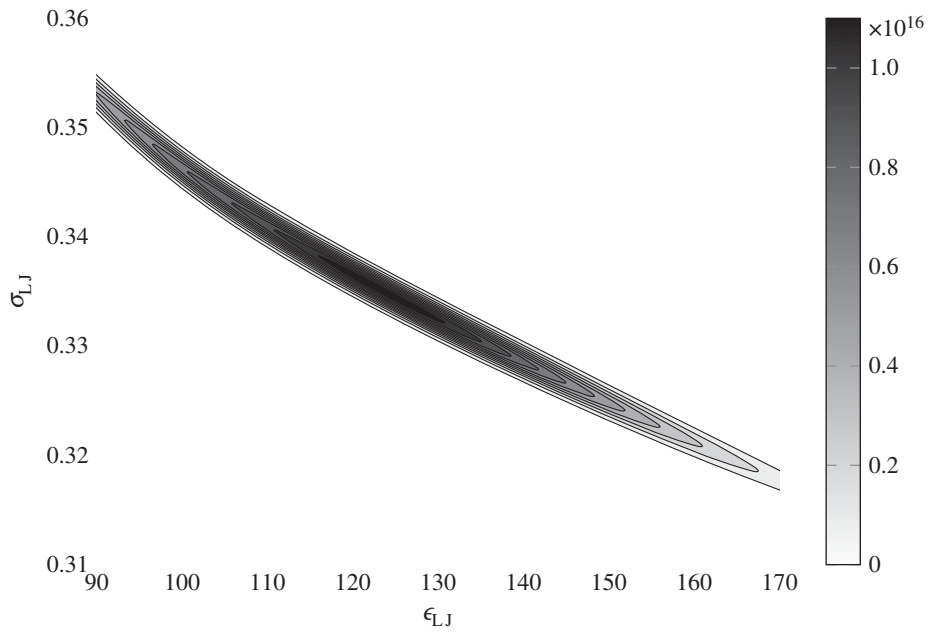
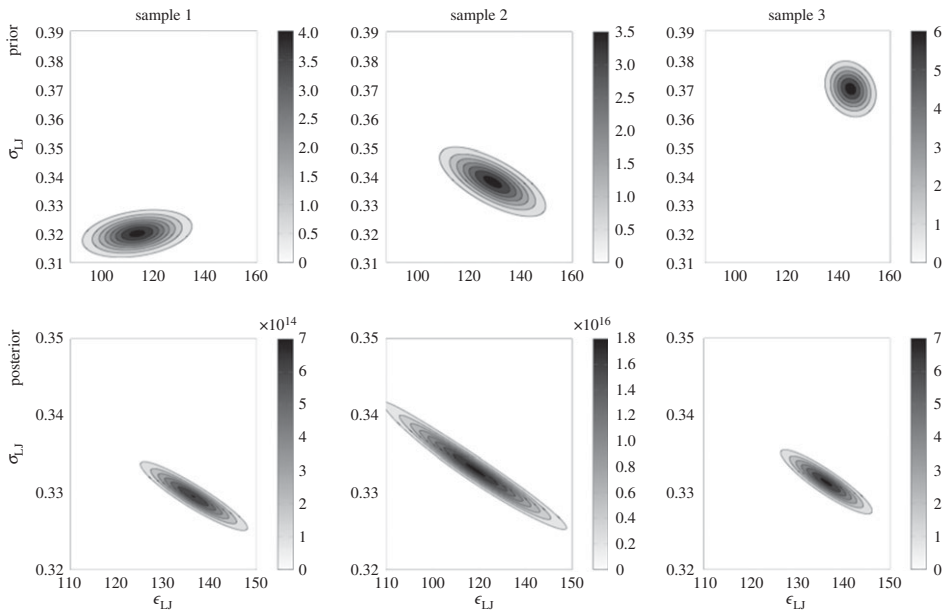**Figure 3.** Likelihood function of dataset from simulation 4.

Figure 4. Posterior PDF $p(\theta_i|D_i, \psi, \sigma_y)$ (second row) of three random samples of ψ and σ_y given the corresponding prior PDF $p(\theta_i|\psi)$ (first row).

representing the effect of cross-simulation uncertainty on the two model parameters. The correct hierarchical model $\mathcal{M}_{H1}$ can identify the cross-simulation uncertainty through σ_ϵ and σ_σ directly, while the incorrect hierarchical model $\mathcal{M}_{H2}$ cannot do so. Instead, $\mathcal{M}_{H2}$ leaves high uncertainty to the mean value of the model parameters. The marginal posterior of ρ is more peak around -1
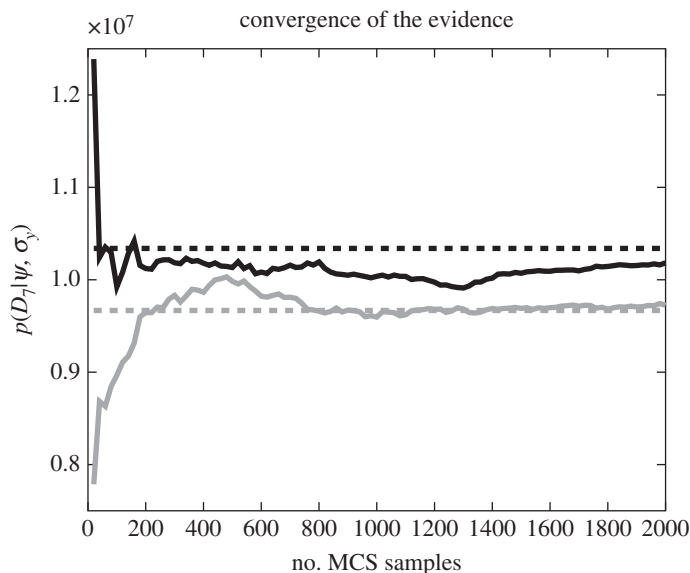
Figure 5. Comparing the estimation of the evidence $p(D_7|\psi, \sigma_y)$ using LAA with the one using MCS estimation. Solid lines are estimates using MCS as a function of the number of samples. Dashed lines are estimates using LAA. Black and grey lines are results based on two random samples of the hyperparameters ψ and σ_y.
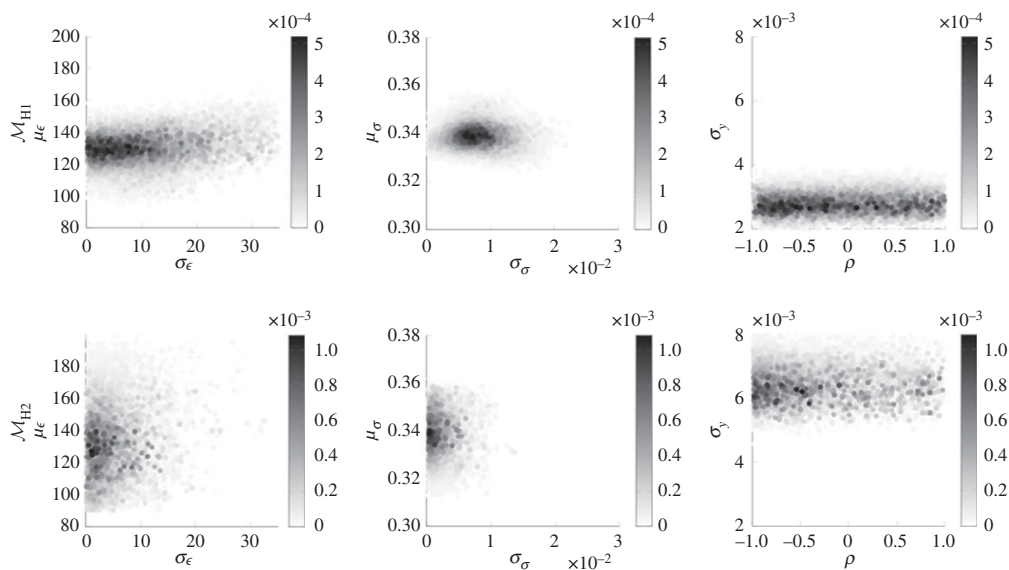
Figure 6. Approximate hyperparameter posteriors using 20 000 TMCMC samples. First row is for $\mathcal{M}_{H1}$ and second row is for $\mathcal{M}_{H2}$.

in $\mathcal{M}_{H2}$ than in $\mathcal{M}_{H1}$ as the model parameters posterior PDF clearly shows a stronger correlation in figure 9.

Table 4 shows the log evidence and the posterior probability of each model. The hierarchical model $\mathcal{M}_{H1}$ has a selection probability ~ 1 over the other models, as one may expect. $\mathcal{M}_{S1}$ and $\mathcal{M}_{H2}$ are ranked next. $\mathcal{M}_{S1}$ is slightly better (in log-scale) than $\mathcal{M}_{H2}$ because the more complex model $\mathcal{M}_{H2}$ did not improve the interpretation of the data much as compared with $\mathcal{M}_{S1}$, and
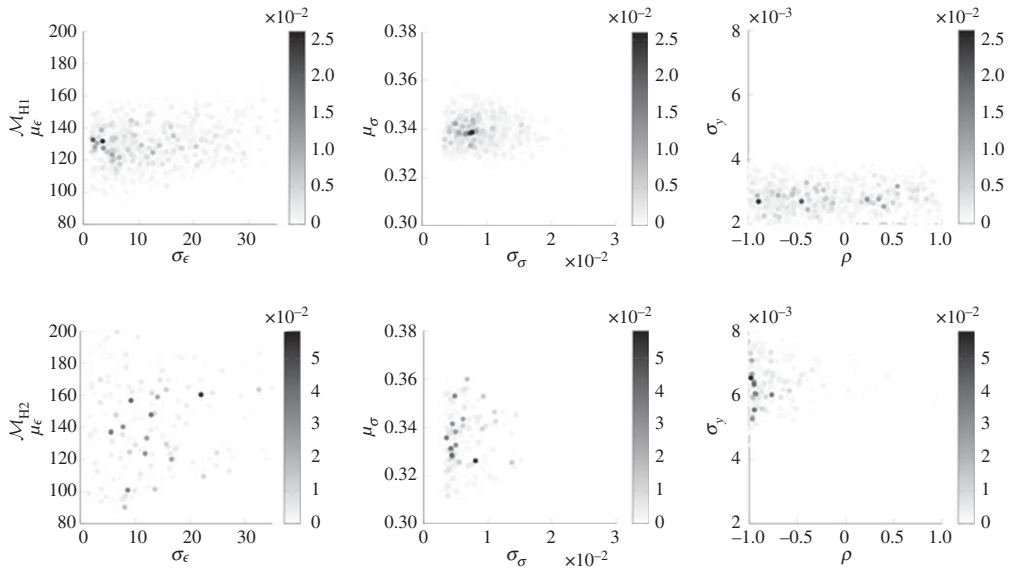
Figure 7. Approximate hyperparameter posteriors using 5 million MCS samples. First row is for $\mathcal{M}_{H1}$ and second row is for $\mathcal{M}_{H2}$.
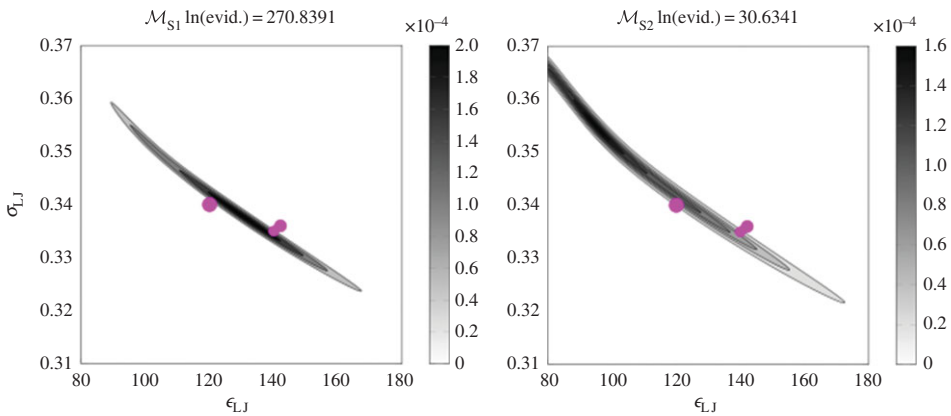
Figure 8. Approximate posterior of model parameters for $\mathcal{M}_{S1}$ and $\mathcal{M}_{S2}$ using 2 million MCS samples. Dots show the actual values from the 10 simulations (table 2). (Online version in colour.)
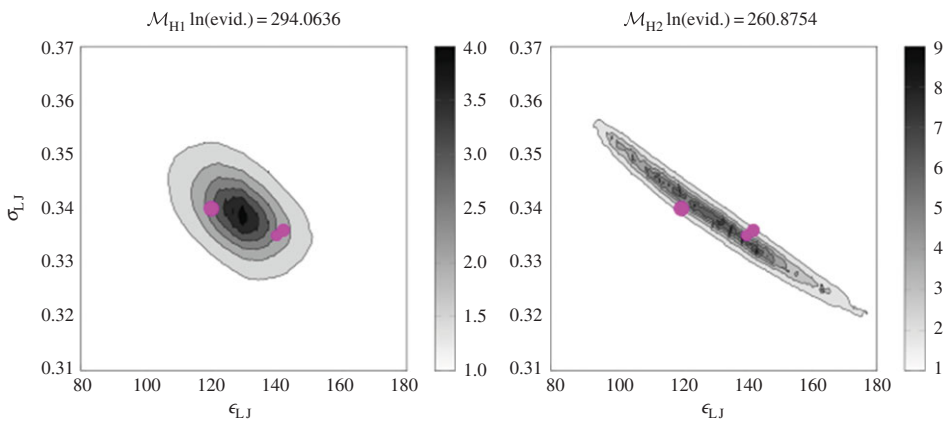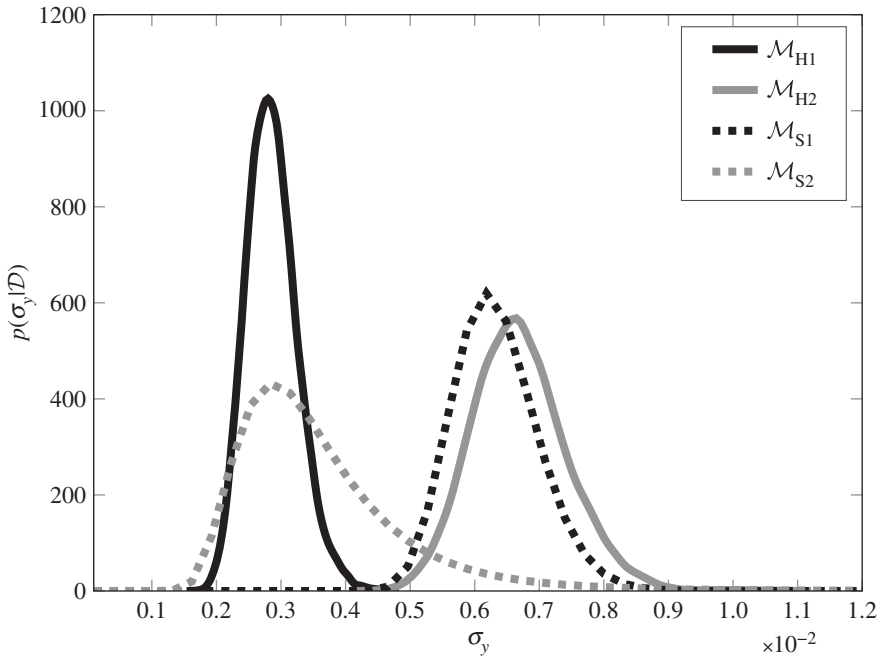

Figure 9. Approximate posterior of model parameters for $\mathcal{M}_{H1}$ and $\mathcal{M}_{H2}$ using 20 000 TMCMC samples. Dots show the actual values from the 10 simulations (table 2). (Online version in colour.)

Figure 10. Approximate posterior of σ_y for $\mathcal{M}_{S1}$ and $\mathcal{M}_{S2}$ using 2 million MCS samples and for $\mathcal{M}_{H1}$ and $\mathcal{M}_{H2}$ using 20 000 TMCMC samples.
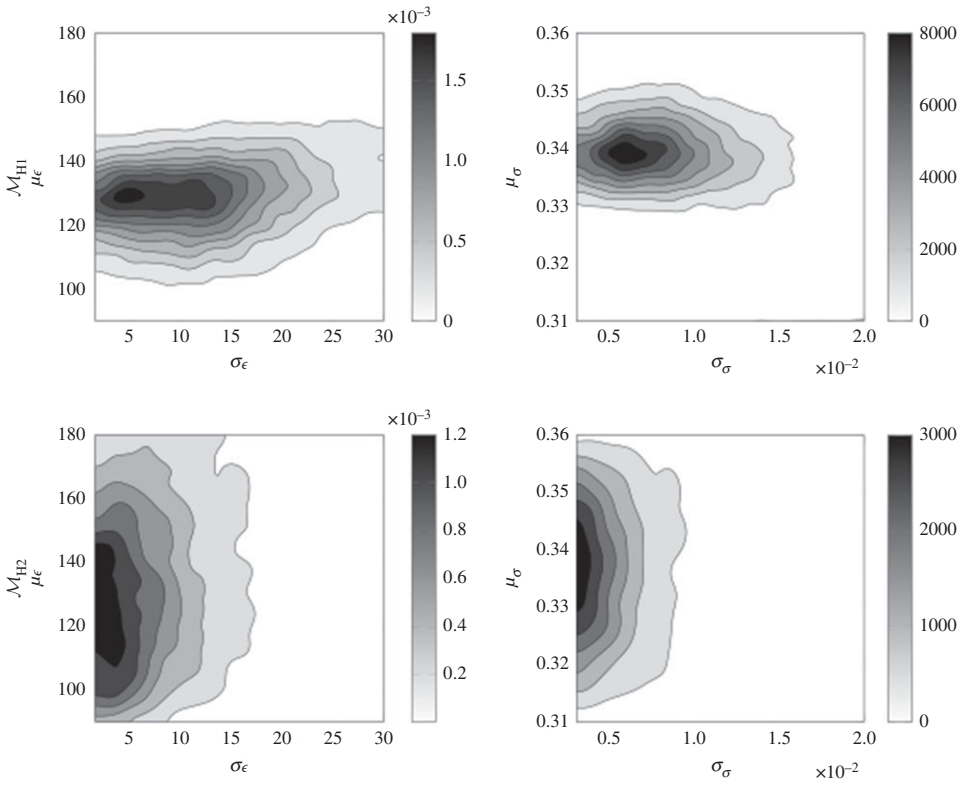
Figure 11. Approximate posterior of μ_ϵ, σ_ϵ, μ_σ and σ_σ using 20 000 TMCMC samples. First row is for $\mathcal{M}_{H1}$ and second row is for $\mathcal{M}_{H2}$.
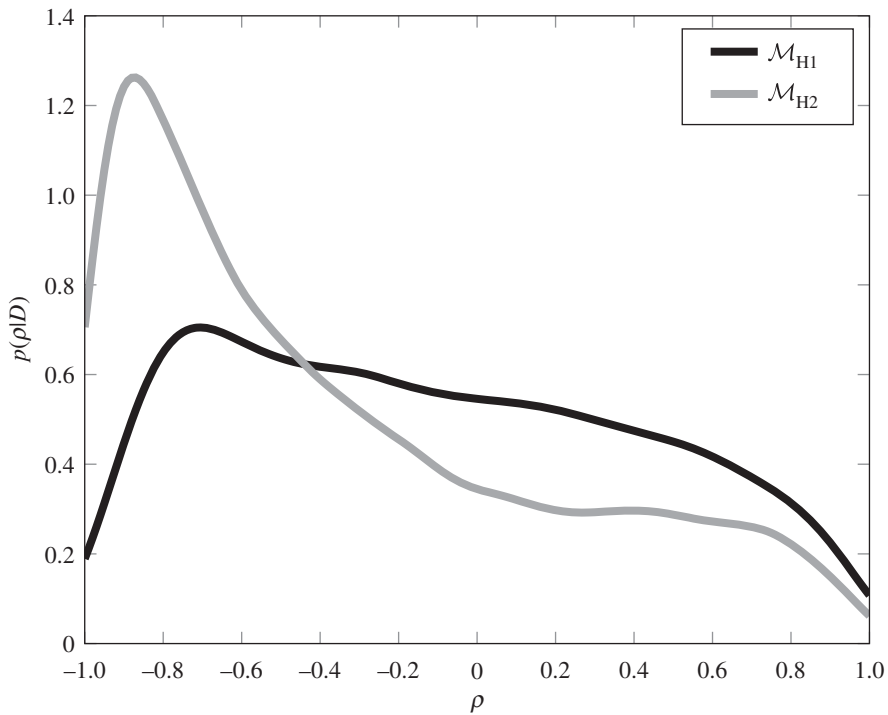
Figure 12. Approximate posterior of ρ for $\mathcal{M}_{H1}$ and $\mathcal{M}_{H2}$ using 20 000 TMCMC samples.
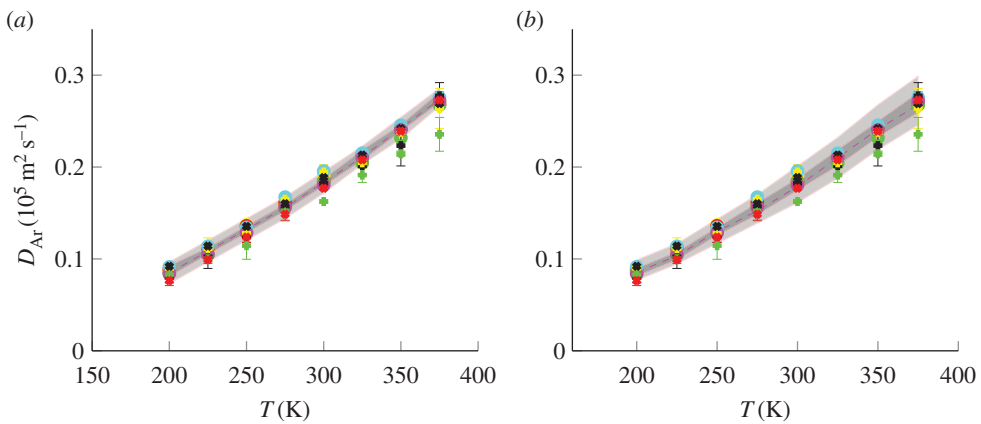
Figure 13. Robust prediction based on a single-level ($\mathcal{M}_{S1}$) and a hierarchical models ($\mathcal{M}_{H1}$). Colour code follows figure 2. (*a*) Single-level model and (*b*) hierarchical model. (Online version in colour.)

Table 4. Posterior probability of the four models given the simulation datasets.

model	ln(evidence)	$P(\mathcal{M}_k|\mathcal{D})$
$\mathcal{M}_{S1}$	270.84	0.00
$\mathcal{M}_{S2}$	30.63	0.00
$\mathcal{M}_{H1}$	294.06	1.00
$\mathcal{M}_{H2}$	260.88	0.00

a simpler model is always preferred in this case due to Ockham's Razor [23,24]. $\mathcal{M}_{S2}$ has a significantly lower probability even in log-scale because the dataset used is significantly smaller than the other models.

Finally, figure 13 shows the effect on posterior PDF by Bayesian robust prediction using the hierarchical ($\mathcal{M}_{H1}$) and single-level models ($\mathcal{M}_{S1}$), respectively. A lower prediction uncertainty is observed from the single-level model because the very narrow and elongated posterior captures the pairs of highly correlated ϵ_{LJ} and σ_{LJ} values that result in similar predictions. On the other hand, the single-mode Gaussian posterior from the hierarchical model leads to a higher prediction uncertainty that scales with the output values. The significantly different posterior prediction from the two models demonstrate the importance of finding the right probabilistic model, even when using a consistent deterministic model. Our results indicate that using the single-level model will lead to an underestimated prediction uncertainty, although the mean values are consistent.

5. Results for experimental data

We now extend our formulation to experimental data. We note that unlike the case of using data obtained from MD simulations, here we cannot *a posteriori* verify the correct distribution of parameters. The same methodology is applied to experimental data of krypton instead of the MD simulation data of argon. Data on viscosity are used instead of diffusion based on Kestin *et al.* [47]. Again, we compare the single-level model (similar to $\mathcal{M}_{S1}$) with the hierarchical model (similar to $\mathcal{M}_{H1}$) using the same datasets. As shown in figure 14, there are nine sets of experimental data coming from different sources covering different temperature ranges, and different experimental set-ups. In the case of viscosity η, the deterministic function f (equation (2.1)) is now expressed as [14]

$$f(T, \{\epsilon_{LJ}, \sigma_{LJ}\}) = \eta = 26.69 \frac{\sqrt{mT}}{\sigma_{LJ}^2 \Omega_v}$$

and

$$\Omega_v = \frac{A}{(T^*)^B} + \frac{C}{\exp(DT^*)} + \frac{E}{\exp(FT^*)}. \tag{5.1}$$

The parameter $T^* = kT/\epsilon_{LJ}$, m denotes the molar mass of krypton, and k denotes the Boltzmann factor. The other coefficients are $A = 1.16145$, $B = 0.14874$, $C = 0.52487$, $D = 0.77320$, $E = 2.16178$ and $F = 2.43787$. The model parameters to be inferred in this case are also ϵ_{LJ} and σ_{LJ}.

Figure 15 shows the likelihood function $p(D_i|\theta, \sigma_y)$ of each experimental dataset D_i, which is proportional to the posterior PDF of the parameters when a uniform prior is assumed. Although the nine datasets have a visually consistent trend, the inferred parameter values in the given model for each dataset are actually very different and highly uncertain. In particular, experiments 1–3 show very inconsistent inferred parameter values when compared to the other six experiments (the plotted ranges of ϵ_{LJ} and σ_{LJ} for experiment 3 is different from the one for the other experiments). This is because the viscosity value is very sensitive to the parameters ϵ_{LJ} and σ_{LJ}. Hence, we use the hierarchical model to infer the uncertainty of the model parameters due to an inaccurate model. If one performs a Bayesian analysis based on a dataset that groups all the nine experimental datasets together, a high model uncertainty value σ_y is expected. In order to balance the error to each data point, the inferred most probable parameter values are expected to be highly affected by the 'outliers' (data from experiments 1 to 3). On the contrary, we expect the hierarchical model to capture the cross-experiment uncertainty in the hyperparameters, thus providing a low model uncertainty value. The uncertainty of ϵ_{LJ} and σ_{LJ} is expected to be high because of the large difference in the inferred parameter values from some of the experiments. However, the most probable value shall be less affected by the 'outliers'.

The single-level model uses the TMCMC method from Angelikopoulos *et al.* [15], and the hierarchical model uses the method introduced in this paper. Figures 16 and 17 show the resulting joint posterior PDF of the model parameters ϵ_{LJ} and σ_{LJ} for both models. In contrast to the single-level model, the hierarchical model results in $p(\sigma_y|\mathcal{D})$ that gives a smaller model uncertainty value
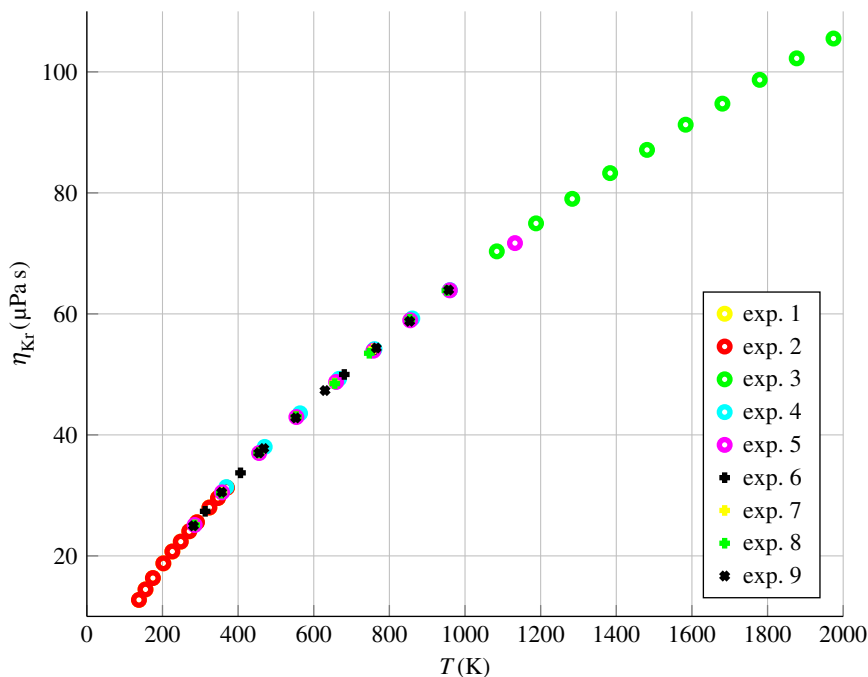
19

rsta.royalsocietypublishing.org Phil. Trans. R. Soc. A **374**: 20150032

Figure 14. Viscosity coefficient data of krypton η_{Kr} from nine different experiments [47]. (Online version in colour.)
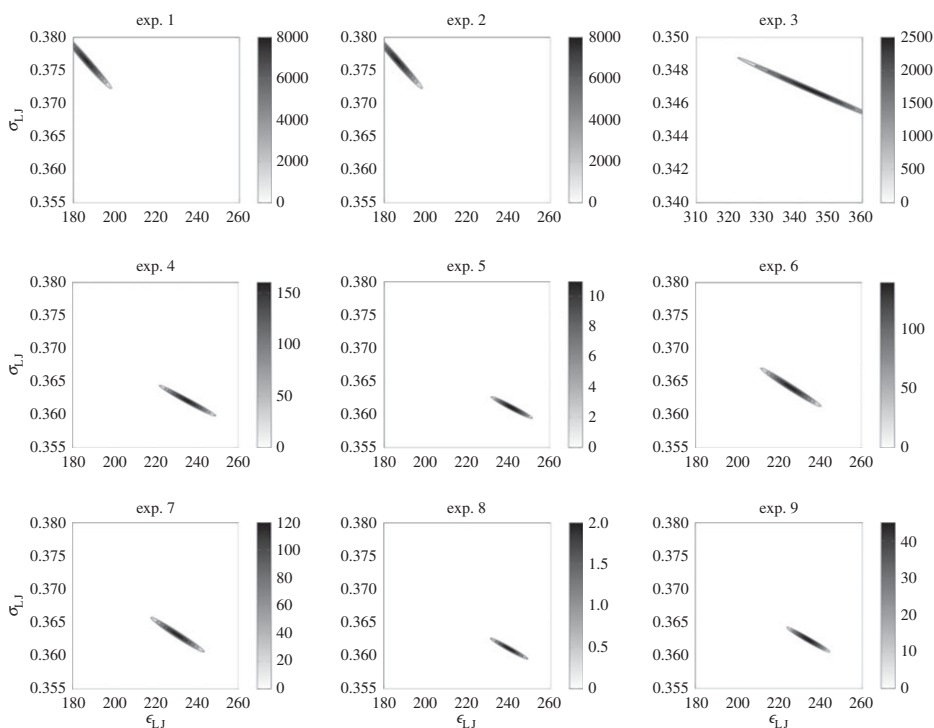

Figure 15. Likelihood value $p(D_i|\boldsymbol{\theta}, \sigma_y)$ of each of the nine experimental datasets (D_1–D_9). All axes are in the same scale except experiment 3.
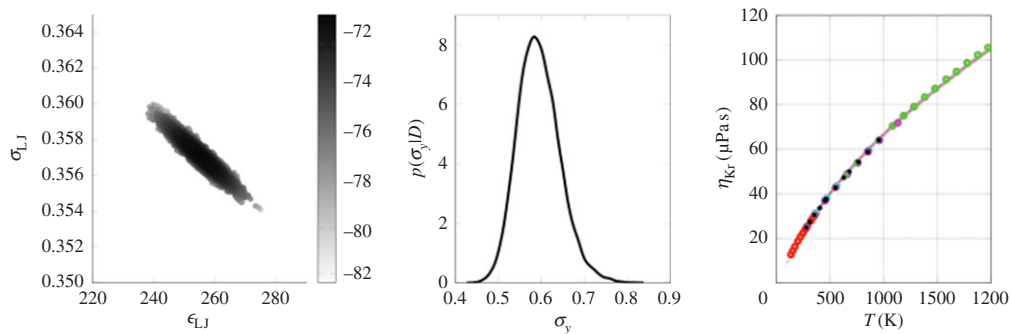
Figure 16. (left to right) Parameter posterior, model uncertainty posterior and robust prediction of the single-level model for viscosity of krypton. TMCMC uses a total of nine stages with 20 000 samples at each stage. (Online version in colour.)
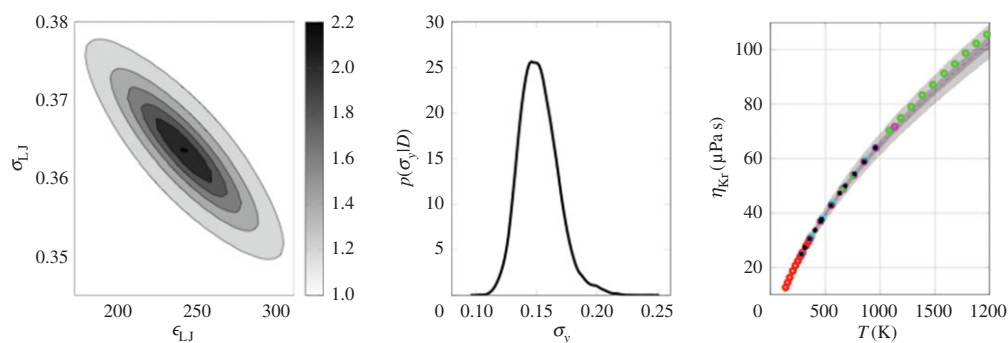

Figure 17. (left to right) Parameter posterior, model uncertainty posterior and robust prediction of the hierarchical model for viscosity of krypton. TMCMC uses a total of 10 stages with 20 000 samples at each stage. (Online version in colour.)

Table 5. Posterior probability of the single-level and hierarchical models given the experimental datasets.

model	ln(evidence)	$P(\mathcal{M}|\mathcal{D})$
single-level	−77.62	0.00
hierarchical	−20.84	1.00

σ_y, whereas a higher parameter uncertainty is observed from $p(\epsilon_{LJ}, \sigma_{LJ}|\mathcal{D})$. We observe that the Bayesian model selection prefers the hierarchical model, as shown in table 5.

6. Conclusion

We present an efficient computational framework to perform for the first time a full Bayesian analysis and model selection between different hierarchical and single-level probabilistic models. We investigate the benefits of a hierarchical Bayesian model over a single-level model using MD simulations. To tackle the computational challenge that typically arises from a full Bayesian analysis on a hierarchical model, we propose a novel efficient approximation scheme that combines LAA and parallelized TMCMC to approximate the posterior PDF and model selection.

Using our results from both simulation and experimental data, we summarize as the benefit of the hierarchical model in three aspects:

(i) Correct separation of the cross-experiment uncertainty from other uncertainties, which helps analysing where the most contribution of the uncertainties comes from.
(ii) Improvement of the posterior PDF of the parameters, which helps improving the accuracy of predictions of future experiments.
(iii) Identification of the most plausible model for a given data from a set of models that can contain multiple hierarchical and single-level models, which helps to discover the underlying data structure if it is not originally known.

In the MD models, the parameter uncertainty has a larger influence on the prediction uncertainty than the model uncertainty σ_y. The hierarchical model is able to identify the cross-experiment uncertainty, and, thus, gives a higher parameter uncertainty, but a lower model uncertainty. This results in a higher prediction uncertainty than the one obtained from the single-level model, which captures the data more accurately.

It is important to note that the knowledge of the data structure affects the implementation of a hierarchical model. Different assumptions on the data structure lead to different implementations, thus, different hierarchical models. If none of the hierarchical models capture the actual data structure, the models may result in misleading conclusions. It is also important to check if the pair of likelihood and prior model can be well approximated by LAA given the dataset. If a multi-modal posterior is expected, the variational Bayesian EM algorithm can be a good alternative to the LAA method.

References

1. Majumder M, Chopra N, Andrews R, Hinds BJ. 2005 Enhanced flow in carbon nanotubes. *Nature* **438**, 44. (doi:10.1038/438044a)
2. Holt JK, Park HG, Wang Y, Stadermann M, Artyukhin AB, Grigoropoulos CP, Noy A, Bakajin O. 2006 Fast mass transport through sub-2-nanometer carbon nanotubes. *Science* **312**, 1034–1037. (doi:10.1126/science.1126298)
3. Falk K, Sedlmeier F, Joly L, Netz RR, Bocquet L. 2010 Molecular origin of fast water transport in carbon nanotube membranes: superlubricity versus curvature dependent friction. *Nano Lett.* **10**, 4067–4073. (doi:10.1021/nl1021046)
4. Suk ME, Aluru NR. 2010 Water transport through ultrathin graphene. *J. Phys. Chem. Lett.* **1**, 1590–1594. (doi:10.1021/jz100240r)
5. Thomas JA, McGaughey AJH. 2008 Reassessing fast water transport through carbon nanotubes. *Nano Lett.* **8**, 2788–2793. (doi:10.1021/nl8013617)
6. Walther JH, Ritos K, Cruz-Chu ER, Megaridis CM, Koumoutsakos P. 2013 Barriers to superfast water transport in carbon nanotube membranes. *Nano Lett.* **13**, 1910–1914. (doi:10.1021/nl304000k)
7. Qin X, Yuan Q, Zhao Y, Xie S, Liu Z. 2011 Measurement of the rate of water translocation through carbon nanotubes. *Nano Lett.* **11**, 2173–2177. (doi:10.1021/nl200843g)

8. Park HG *et al*. 2014 Fabrication of flexible, aligned carbon nanotube/polymer composite membranes by in-situ polymerization. *J. Membrane Sci.* **460**, 91–98. (doi:10.1016/j.memsci.2014.02.016)

9. Park HG, Jung Y. 2014 Carbon nanofluidics of rapid water transport for energy applications. *Chem. Soc. Rev.* **43**, 565–576. (doi:10.1039/C3CS60253B)

10. Park HG. 2007 Mass transport thorugh carbon nanotubes. PhD thesis, University of Berkeley, CA, USA.

11. Ulissi ZW, Shimizu S, Lee CY, Strano MS. 2011 Carbon nanotubes as molecular conduits: advances and challenges for transport through isolated sub-2 nm pores. *J. Phys. Chem. Lett.* **2**, 2892–2896. (doi:10.1021/jz201136c)

12. Mattia D, Leese H, Lee KP. 2015 Carbon nanotube membranes: from flow enhancement to permeability. *J. Membrane Sci.* **475**, 266–272. (doi:10.1016/j.memsci.2014.10.035)

13. Shih C-J, Lin S, Strano MS, Blankschtein D. 2015 Understanding the stabilization of single-walled carbon nanotubes and graphene in ionic surfactant aqueous solutions: large-scale coarse-grained molecular dynamics simulation-assisted DLVO theory. *J. Phys. Chem. C* **119**, 1047–1060. (doi:10.1021/jp5093477)

14. Cailliez F, Pernot P. 2011 Statistical approaches to forcefield calibration and prediction uncertainty in molecular simulation. *J. Chem. Phys.* **134**, 054124. (doi:10.1063/1.3545069)

15. Angelikopoulos P, Papadimitriou C, Koumoutsakos P. 2012 Bayesian uncertainty quantification and propagation in molecular dynamics simulations: a high performance computing framework. *J. Chem. Phys.* **137**, 144103. (doi:10.1063/1.4757266)

16. Cornell WD *et al*. 1995 A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J. Amer. Chem. Soc.* **117**, 5179–5197. (doi:10.1021/ja00124a002)

17. Oostenbrink C, Villa A, Mark AE, Van Gunsteren WF. 2004 A biomolecular force field based on the free enthalpy of hydration and solvation: the gromos force-field parameter sets 53a5 and 53a6. *J. Comput. Chem.* **25**, 1656–1676. (doi:10.1002/jcc.20090)

18. Lee EH, Hsin J, Sotomayor M, Comellas G, Schulten K. 2009 Discovery through the computational microscope. *Structure* **17**, 1295–1306. (doi:10.1016/j.str.2009.09.001)

19. Buongiorno J *et al.* 2009 A benchmark study on the thermal conductivity of nanofluids. *J. Appl. Phys.* **106**, 6269. (doi:10.1063/1.3245330)

20. Lee KP, Leese H, Mattia D. 2012 Water flow enhancement in hydrophilic nanochannels. *Nanoscale* **4**, 2621–2627. (doi:10.1039/c2nr30098b)

21. Congdon PD. 2010 *Applied Bayesian hierarchical methods*. Boca Raton, FL: CRC Press.

22. Gelman A, Hill J. 2006 *Data analysis using regression and multilevel/hierarchical models*. Cambridge, UK: Cambridge University Press.

23. Beck JL, Yuen KV. 2004 Model selection using response measurements: Bayesian probabilistic approach. *J. Eng. Mech.-ASCE* **130**, 192–203. (doi:10.1061/(ASCE)0733-9399(2004)130:2(192))

24. Beck JL. 2010 Bayesian system identification based on probability logic. *Struct. Control Health Monit.* **17**, 825–847. (doi:10.1002/stc.424)

25. Nagel JB, Sudret B. 2014 *A Bayesian multilevel framework for uncertainty characterization and the NASA Langley multidisciplinary UQ challenge*. Zurich, Switzerland: ETH-Zürich, Institute of Structural Engineering, Chair of Risk, Safety & Uncertainty Quantification. (doi:10.3929/ethz-a-010078997)

26. Ballesteros G, Angelikopoulos P, Papadimitriou C, Koumoutsakos P. 2014 Bayesian hierarchical models for uncertainty quantification in structural dynamics. In *Vulnerability, uncertainty, and risk: quantification, mitigation, and management*, vol. 162 (eds M Beer, SK Au, JW Hall), pp. 1615–1624. Reston, VA, USA: American Society of Civil Engineers (ASCE).

27. Behmanesh I, Moaveni B, Lombaert G, Papadimitriou C. 2015 Hierarchical Bayesian model updating for structural identification. *Mech. Syst. Signal Process.* **64–65**, 360–376. (doi:10.1016/j.ymssp.2015.03.026)

28. Nilsson H, Rieskamp J, Wagenmakers E-J. 2011 Hierarchical Bayesian parameter estimation for cumulative prospect theory. *J. Math. Psychol.* **55**, 84–93. (doi:10.1016/j.jmp.2010.08.006)

29. Scheibehenne B, Pachur T. 2014 Using Bayesian hierarchical parameter estimation to assess the generalizability of cognitive models of choice. *Psychon. Bull. Rev.* 22, 391–407. (doi:10.3758/s13423-014-0684-4)

30. Friston KJ, Harrison L, Penny W. 2003 Dynamic causal modelling. *Neuroimage* **19**, 1273–1302. (doi:10.1016/S1053-8119(03)00202-7)

31. Stephan KE, Penny WD, Daunizeau J, Moran RJ, Friston KJ. 2009 Bayesian model selection for

23

rsta.royalsocietypublishing.org *Phil. Trans. R. Soc. A* **374**: 20150032

group studies. *Neuroimage* **46**, 1004–1017. (doi:10.1016/j.neuroimage.2009.03.025)

32. Papadimitriou C, Beck JL, Katafygiotis LS. 1997 Asymptotic expansions for reliability and moments of uncertain systems. *J. Eng. Mech.-ASCE* **123**, 1219–1229. (doi:10.1061/(ASCE)0733-9399(1997)123:12(1219))

33. Ching JY, Chen YC. 2007 Transitional markov chain monte carlo method for bayesian model updating, model class selection, and model averaging. *J. Eng. Mech.-ASCE* **133**, 816–832. (doi:10.1061/(ASCE)0733-9399(2007)133:7(816))

34. Hadjidoukas PE, Angelikopoulos P, Papadimitriou C, Koumoutsakos P. 2015 Π4U: A high performance computing framework for Bayesian uncertainty quantification of complex models. *J. Comput. Phys. J. Eng. Mech.-ASCE* **284**, 1–21. (doi:10.1016/j.jcp.2014.12.006)

35. Beck JL, Taflanidis AA. 2013 Prior and posterior robust stochastic predictions for dynamical systems using probability logic. *Int. J. Uncertain. Quantif.* **3**, 271–288. (doi:10.1615/Int.J. UncertaintyQuantification.2012003641)

36. Tierney L, Kadane JB. 1986 Accurate approximations for posterior moments and marginal densities. *J. Am. Stat. Assoc.* **81**, 82–86. (doi:10.1080/01621459.1986.10478240)

37. Alder BJ, Wainwright TE. 1959 Studies in molecular dynamics. 1. General method. *J. Chem. Phys.* **31**, 459–466. (doi:10.1063/1.1730376)

38. Stillinger FH, Rahman A. 1974 Improved simulation of liquid water by molecular dynamics. *J. Chem. Phys.* **60**, 1545–1557. (doi:10.1063/1.1681229)

39. Berendsen HJC, Grigera JR, Straatsma TP. 1987 The missing term in effective pair potentials. *J. Phys. Chem.* **91**, 6269–6271. (doi:10.1021/j100308a038)

40. Basconi JE, Shirts MR. 2013 Effects of temperature control algorithms on transport properties and kinetics in molecular dynamics simulations. *J. Chem. Theory Comput.* **9**, 2887–2899. (doi:10.1021/ct400109a)

41. Bussi G, Donadio D, Parrinello M. 2007 Canonical sampling through velocity rescaling. *J. Chem. Phys.* **126**, 014 101–014 107. (doi:10.1063/1.2408420)

42. Berendsen HJC, Postma JPM, van Gunsteren WF, DiNola A, Haak JR. 1984 Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* **81**, 3684–3690. (doi:10.1063/1.448118)

43. Nose S. 1984 A unified formulation of the constant temperature molecular dynamics methods. *J. Chem. Phys.* **81**, 511–519. (doi:10.1063/1.447334)

44. Parrinello M, Rahman A. 1980 Crystal structure and pair potentials: a molecular-dynamics study. *Phys. Rev. Lett.* **45**, 1196–1199. (doi:10.1103/PhysRevLett.45.1196)

45. Andersen HC. 1980 Molecular dynamics simulations at constant pressure and-or temperature. *J. Chem. Phys.* **72**, 2384–2393. (doi:10.1063/1.439486)

46. Martyna GJ, Tobias DJ, Klein ML. 1994 Constant pressure molecular dynamics algorithms. *J. Chem. Phys.* **101**, 4177–4189. (doi:10.1063/1.467468)

47. Kestin J, Knierim K, Mason EA, Najafi B, Ro ST, Waldman M. 1984 Equilibrium and transport properties of the noble gases and their mixtures at low density. *J. Phys. Chem. Ref. Data* **13**, 229–303. (doi:10.1063/1.555703)

48. Liboff RL 2003 *Graduate texts in contemporary physics*, ch. 4, pp. 278–328. New York, NY: Springer.