Supplementary Material:

Data driven inference for the repulsive exponent of the Lennard-Jones potential in molecular dynamics simulations

Lina Kulakova\(^1\), Georgios Arampatzis\(^1\), Panagiotis Angelikopoulos\(^{1,2}\), Panagiotis Hadjidoukas\(^1\), Costas Papadimitriou\(^3\), and Petros Koumoutsakos\(^1,\ast\)

\(^1\)Computational Science and Engineering Laboratory, Clausiusstrasse 33, ETH Zrich, CH-8092, Switzerland
\(^2\)Department of Mechanical Engineering, University of Thessaly, Pedion Areos, GR-38334 Volos, Greece
\(^3\)Currently at D.E.Shaw Research LLC, 10016 NY USA
\(\ast\)petros@ethz.ch

S1 Uncertainty quantification

This section provides a detailed description of the UQ theory used in the current work.

S1.1 Sampling the posterior distribution

For the posterior distribution \(p(\vartheta | d, M)\) which is known up to a normalizing constant \(p(d | M)\), available Markov Chain Monte Carlo (MCMC) methods can be used to efficiently generate samples that quantify the uncertainty in \(\vartheta\) \([1, 2, 3, 4, 5]\). In our work we use the Transitional Markov Chain Monte Carlo (TMCMC) algorithm \([6]\) with a slight modification on the MCMC proposal covariance used to overcome low acceptance rates in several runs of TMCMC which we observed. Instead of setting the covariance matrix to the scaled sample covariance of the previous generation, we use the landscape around the chain leaders to construct local covariance matrices. The sampling algorithm automatically tries to increase the radius of the neighborhood starting from 10\% of the domain size until the local covariance matrix is positive definite.

S1.2 Robust posterior prediction

The uncertainty in the model parameters can be further propagated to the uncertainty in the QoI \(y\) produced by \(f(x; \vartheta)\). Under the assumption on the likelihood function

\[ p(d | \vartheta, x) = N(d | f(x; \vartheta), \Sigma), \]  

the probability of the model prediction conditioned on the parameters \(\vartheta\) is given by \(p(y | \vartheta, M) = N(y | f(x; \vartheta), \Sigma)\). The probability of the model prediction conditioned on the observations \(d\) is known as the robust posterior prediction and is given by \([7, 8]\):

\[ p(y | d, M) = \int p(y, \vartheta | d, M) d\vartheta = \int p(y | \vartheta, M) p(\vartheta | d, M) d\vartheta \approx \frac{1}{N_s} \sum_{k=1}^{N_s} p(y | \vartheta^{(k)}, M), \]  

where \(\vartheta^{(k)} \sim p(\vartheta | d, M)\) and \(N_s\) is sufficiently large.

When a new QoI \(z\) produced by \(g(x; \vartheta)\) is considered, \(\Sigma\) is unknown and only the parametric uncertainty is propagated into \(g(x; \vartheta)\). Namely, one should estimate the density of \(\sum_{k=1}^{N_s} g(x; \vartheta^{(k)})\), where \(\vartheta^{(k)} \sim p(\vartheta | d, M)\) and \(N_s\) is sufficiently large.
Figure S1: Bayesian networks: (a) simple network, (b) HB network, (c) network for the i-th dataset.

S1.3 Model selection

The Bayesian framework allows one to select the probabilistic model which best fits the data. The criterion for the model selection comes from the Bayes’ theorem, which computes the probability of a probabilistic model $\mathcal{M}$ as

$$p(\mathcal{M} | d) = \frac{p(d | \mathcal{M}) p(\mathcal{M})}{p(d)}$$  \hspace{1cm} (S3)

where $p(\mathcal{M})$ is the prior PDF of the model $\mathcal{M}_i$, and the evidence $p(d | \mathcal{M})$ of model $\mathcal{M}$ is computed as a by-product of TMCMC.

S1.4 Hierarchical Bayesian models

In our work we follow the methodology developed in [9]. We assume that data comes split in $N$ different datasets: $\vec{d} = \{d_1, \ldots, d_N\}$ and the likelihood in the probabilistic model $\mathcal{M}_i$ is $\mathcal{N}(d_i | f(x_i; \theta_i), \Sigma_i)$. We assume that the probability of $\theta_i$ depends on a hyper-parameter $\psi \in \mathbb{R}^{N_{psi}}$ and is given by a PDF $p(\theta | \psi, \mathcal{M})$, where $\mathcal{M}$ corresponds to the graph describing the relations between $\psi$, $\theta_i$ and $d_i$, see Figure S1. Our goal is to obtain samples from the posterior distribution, $p(\theta_i | \vec{d}, \mathcal{M})$, where $\vec{d} = \{d_1, \ldots, d_N\}$:

$$p(\theta_i | \vec{d}, \mathcal{M}) = \int p(\theta_i | \psi, \vec{d}, \mathcal{M}) p(\psi | \vec{d}, \mathcal{M}) d\psi.$$  \hspace{1cm} (S4)

The dependency assumptions from Fig. S1 allow to simplify: $p(\theta_i | \psi, \vec{d}, \mathcal{M}) = p(\theta_i | \psi, d_i, \mathcal{M})$, and equation (S4) can be rewritten using the Bayes’ theorem:

$$p(\theta_i | \vec{d}, \mathcal{M}) = \int \frac{p(d_i | \theta_i, \psi, \mathcal{M}) p(\theta_i | \psi, \mathcal{M})}{p(d_i | \psi, \mathcal{M})} p(\psi | \vec{d}, \mathcal{M}) d\psi.$$  \hspace{1cm} (S5)

Since $p(d_i | \theta_i, \psi, \mathcal{M}) = p(d_i | \psi, \mathcal{M})$, equation (S5) simplifies to

$$p(\theta_i | \vec{d}, \mathcal{M}) = \int p(d_i | \theta_i, \mathcal{M}) \frac{p(\theta_i | \psi, \mathcal{M})}{p(d_i | \psi, \mathcal{M})} p(\psi | \vec{d}, \mathcal{M}) d\psi.$$  \hspace{1cm} (S6)

Finally, the posterior distribution (S4) can be approximated as

$$p(\theta_i | \vec{d}, \mathcal{M}) \approx \frac{p(d_i | \theta_i, \mathcal{M})}{N_s} \sum_{k=1}^{N_s} \frac{p(\theta_i | \psi^{(k)}, \mathcal{M})}{p(d_i | \psi^{(k)}, \mathcal{M})},$$  \hspace{1cm} (S7)

where, $\psi^{(k)} \sim p(\psi | \vec{d}, \mathcal{M})$ and $N_s$ is sufficiently large. Thus, in order to obtain $\theta_i$ samples, we first have to sample the probability distribution $p(\psi | \vec{d}, \mathcal{M})$, which, according to Bayes’ theorem, is equal to

$$p(\psi | \vec{d}, \mathcal{M}) = \frac{p(\vec{d} | \psi, \mathcal{M}) p(\psi | \mathcal{M})}{p(\vec{d} | \mathcal{M})}.$$  \hspace{1cm} (S8)
where \(p(\psi | \mathcal{M})\) is the prior PDF on \(\psi\) and \(p(\mathbf{d} | \mathcal{M})\) is the normalizing constant. Exploiting the dependency assumption of Fig. S1 we see that

\[
p(\mathbf{d} | \psi, \mathcal{M}) = \prod_{i=1}^{N} p(\mathbf{d}_i | \psi, \mathcal{M}), \tag{S9}
\]

and the likelihood of \(i\)-th dataset can be expressed according to the total probability theorem as

\[
p(\mathbf{d}_i | \psi, \mathcal{M}) = \int p(\mathbf{d}_i | \vartheta_i, \mathcal{M}) p(\vartheta_i | \psi, \mathcal{M}) d\vartheta_i. \tag{S10}
\]

Here we introduce the model \(\mathcal{M}_i\), described in Fig. S1. The posterior distribution of this model will be used as instrumental density for important sampling. Under the modeling assumption \(p(\mathbf{d}_i | \vartheta_i, \mathcal{M}_i) = p(\mathbf{d}_i | \vartheta_i, \mathcal{M}_1)\) (see [9]) and the use of the Bayes’ theorem, equation (S10) is written as

\[
p(\mathbf{d}_i | \psi, \mathcal{M}) = \int \frac{p(\vartheta_i | \mathbf{d}_i, \mathcal{M}_i) p(\mathbf{d}_i | \mathcal{M}_i)}{p(\vartheta_i | \mathcal{M}_i)} p(\vartheta_i | \psi, \mathcal{M}) d\vartheta_i, \tag{S11}
\]

or, equivalently, as

\[
p(\mathbf{d}_i | \psi, \mathcal{M}) = p(\mathbf{d}_i | \mathcal{M}_i) \int \frac{p(\vartheta_i | \psi, \mathcal{M})}{p(\vartheta_i | \mathcal{M}_i)} p(\vartheta_i | \mathbf{d}_i, \mathcal{M}_i) d\vartheta_i. \tag{S12}
\]

Finally, equation (S10) can be approximated as

\[
p(\mathbf{d}_i | \psi, \mathcal{M}) \approx \frac{p(\mathbf{d}_i | \mathcal{M}_i)}{N_s} \sum_{k=1}^{N_s} \frac{p(\vartheta_i^{(k)} | \psi, \mathcal{M})}{p(\vartheta_i^{(k)} | \mathcal{M}_i)}, \tag{S13}
\]

where \(\vartheta_i^{(k)} \sim p(\vartheta_i | \mathbf{d}_i, \mathcal{M}_i)\) and \(N_s\) is sufficiently large. Note that in general \(N_s\) can be different for each data set \(\mathbf{d}_i\). The advantage of this approach is that the likelihoods \(p(\mathbf{d}_i | \vartheta_i, \mathcal{M}_i), i = 1, \ldots, N\), which are the most expensive part of the computations, are not re-evaluated for each \(\psi\).

### S2 Information about hyper-parameter models for hierarchical inference

**Inference for LJ 6-12** We assume that,

\[
p(\vartheta | \psi, \mathcal{M}) = \prod_{j=1}^{3} p(\vartheta_j | \psi, \mathcal{M}), \tag{S14}
\]

and we consider the following two models:

1) uniform: \(p(\vartheta_j | \psi, \mathcal{M}) = \mathcal{U}(\vartheta_j | \psi_{2j-1}, \psi_{2j-1} + \psi_{2j})\), where \(\mathcal{U}(|a, b)\) is the uniform distribution with parameters \(a, b\) and \(\mathcal{M}\) is set to \(\mathcal{U}\).

2) log-normal: \(p(\vartheta_j | \psi, \mathcal{M}) = \mathcal{L}(\vartheta_j | \psi_{2j-1}, \psi_{2j})\), where \(\mathcal{L}(|a, b)\) is the log-normal distribution of \(\xi\) with parameters \(a, b\) and \(\mathcal{M}\) is set to \(\mathcal{L}\).

The prior distribution on the hyper-parameters is modeled as independent uniform,

\[
p(\psi | \mathcal{M}) = \prod_{j=1}^{6} \mathcal{U}(\psi_j | a_j^{\mathcal{M}}, b_j^{\mathcal{M}}), \tag{S15}
\]

where \(\mathcal{M} \in \{\mathcal{U}, \mathcal{L}\}\) and the constants \(a_j^{\mathcal{M}}, b_j^{\mathcal{M}}\) are given in Table S1, along with the values of the log-evidences for the two models. The model \(\mathcal{U}\) is according to the Bayesian model selection criterion, an order of magnitude more plausible and thus will be used for the further inference.

3
Table S1: *HB*$_{12,R}$ inference: lower bound and width for each of the hyper-parameters defined in equation (S15). Log-evidences for each hyper-prior model at the last row.

| $M = \mathcal{U}$          | $M = \mathcal{L}$          | $M = \mathcal{T}$          |
|-----------------------------|-----------------------------|-----------------------------|
| $[a_1^M, b_1^M]$            | $[-1.000, 2.0]$             | $[0.0500, 10.0]$            |
| $[a_2^M, b_2^M]$            | $[0.001, 2.5]$              | $[0.0010, 3.00]$            |
| $[a_3^M, b_3^M]$            | $[-3.000, 0.5]$             | $[3.0000, 4.00]$            |
| $[a_4^M, b_4^M]$            | $[0.001, 4.0]$              | $[0.0010, 0.30]$            |
| $[a_5^M, b_5^M]$            | $[-3.500, 0.5]$             | $[3.0000, 4.00]$            |
| $[a_6^M, b_6^M]$            | $[0.001, 2.5]$              | $[0.0010, 0.30]$            |

Log-ev. -19.1401 -22.0198

Table S2: *HB*$_{p,R}$ inference: lower bound and width for each of the hyper-parameters defined in equation (S17). Log-evidences for each hyper-prior model at the last row.

| $M = \mathcal{U}$          | $M = \mathcal{L}$          | $M = \mathcal{T}$          |
|-----------------------------|-----------------------------|-----------------------------|
| $[a_1^M, b_1^M]$            | $[-1.000, 2.0]$             | $[0.0500, 10.0]$            |
| $[a_2^M, b_2^M]$            | $[0.001, 2.5]$              | $[0.0010, 3.00]$            |
| $[a_3^M, b_3^M]$            | $[-3.000, 0.5]$             | $[3.0000, 4.00]$            |
| $[a_4^M, b_4^M]$            | $[0.001, 4.0]$              | $[0.0010, 0.30]$            |
| $[a_5^M, b_5^M]$            | $[-3.500, 0.5]$             | $[3.0000, 4.00]$            |
| $[a_6^M, b_6^M]$            | $[0.001, 2.5]$              | $[0.0010, 0.30]$            |
| $[a_7^M, b_7^M]$            | $[0.0001, 1.00]$            |
| $[a_8^M, b_8^M]$            | $[0.0001, 1.00]$            |

Log-ev. -24.0202 -27.3889 -25.3927

Inference for LJ 6-\(p\) We assume

\[ p(\theta | \psi, M) = \prod_{j=1}^{4} p(\theta_j | \psi, M), \]  

and consider the following three models:

1) uniform: \( p(\theta_j | \psi, \mathcal{U}) = \mathcal{U}(\theta_j | \psi_{2j-1}, \psi_{2j-1} + \psi_{2j}) \) and \( M \) is set to \( \mathcal{U} \),

2) log-normal: \( p(\theta_j | \psi, \mathcal{L}) = \mathcal{L}(\theta_j | \psi_{2j-1}, \psi_{2j}) \) and \( M \) is set to \( \mathcal{L} \),

3) truncated normal: \( p(\theta_j | \psi, \mathcal{N}) = \mathcal{N}(\theta_j | \psi_{2j-1}, \psi_{2j}) \), where \( \mathcal{N}(\cdot | a, b) \) is the truncated normal distribution with parameters \( a, b \) and \( M \) is set to \( \mathcal{T} \).

The prior distribution on the hyper-parameters is modeled as independent uniform,

\[ p(\psi | \mathcal{M}) = \prod_{j=1}^{8} p(\psi_j | \mathcal{M}) = \prod_{j=1}^{8} \mathcal{U}(\psi_j | a_j^M, b_j^M), \]  

where \( \mathcal{M} \in \{ \mathcal{U}, \mathcal{L}, \mathcal{T} \} \) and the constants \( a_j^M, b_j^M \) are given in Table S2, along with the values of the log-evidences for the three models. As it can be seen, the uniform model is the most plausible one.

S3  Posterior parameter distribution for LJ 6-12 and LJ 6-\(p\)

This section presents the posterior distributions for LJ 6-12 and LJ 6-\(p\) potentials obtained in HB TMCMC runs for each thermodynamic condition, as well as, the distribution for the quantum dimer-based Bayesian inference. Each plot contains all the TMCMC samples of the last stage and it consists of the following parts: histograms of marginal distributions of parameters are shown on the diagonal, projections of the samples to all pair 2-d subspaces in the parameter space colored by the log-likelihood values are given above the diagonal, the corresponding densities constructed via a bivariate kernel estimate are depicted below the diagonal. Green star shows the parameters from Ref. [10], green square indicates the parameters from Ref. [11], and green circle marks the parameters from Ref. [12, 13].
Figure S2: LJ parameters distributions obtained in $HB_{p,R}$ for $L_3$ (top left), $L_4$ (top right), $L_5$ (bottom left) and $V$ (bottom right).
Figure S3: LJ parameters distributions obtained in $HB_{12, R}$ for $L_3$ (top left), $L_4$ (top right), $L_5$ (bottom left) and $V$ (bottom right).
Figure S4: LJ parameters distributions obtained in $HB_{p,R}$ for $L_1$ (left) and $L_2$ (right).

Figure S5: LJ parameters distributions obtained in $B_{p,Q}$.

Figure S6: LJ parameters distributions obtained in $B_{p,R}$ for $L_3$ with the prior for $p$ restricted to $[12, 14]$. 
References

[1] W. K. Hastings. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, 57(1):97–109, 1970.

[2] W. R. Gilks. *Markov chain Monte Carlo*. John Wiley and Sons, Ltd., 2005.

[3] D. Gamerman and H. F. Lopes. *Markov chain Monte Carlo: stochastic simulation for Bayesian inference*. CRC Press, 2006.

[4] C. J. Geyer. Practical Markov chain Monte Carlo. *Statistical Science*, pages 473–483, 1992.

[5] P. H. Peskun. Optimum Monte Carlo sampling using Markov chains. *Biometrika*, 60(3):607–612, 1973.

[6] J. Ching and Y. Chen. Transitional Markov chain Monte Carlo method for Bayesian model updating, model class selection, and model averaging. *J. Eng. Mech.*, 133(7):816–832, 2007.

[7] P. Angelikopoulos, C. Papadimitriou, and P. Koumoutsakos. Bayesian uncertainty quantification and propagation in molecular dynamics simulations: A high performance computing framework. *J. Chem. Phys.*, 137:144103, 2012.

[8] P. E. Hadjidoukas, P. Angelikopoulos, C. Papadimitriou, and P. Koumoutsakos. Pi4U: A high performance computing framework for Bayesian uncertainty quantification of complex models. *J. Comp. Phys.*, 284:1–21, 2015.

[9] S. Wu, P. Angelikopoulos, G. Tauriello, C. Papadimitriou, and P. Koumoutsakos. Fusing heterogeneous data for the calibration of molecular dynamics force fields using hierarchical Bayesian models. *J. Chem. Phys.*, 145:244112, 2016.

[10] J. A. Barker, R. A. Fisher, and R. O. Watts. Liquid argon: Monte Carlo and molecular dynamics calculations. *Mol. Phys.*, 21:657–673, 1971.

[11] John A. White. Lennard-Jones as a model for argon and test of extended renormalization group calculations. *J. Chem. Phys.*, 111:9352–9356, 1999.

[12] A. Rahman. Correlations in the motion of atoms in liquid argon. *Phys. Rev.*, 136:A405–A411, 1964.

[13] L. A. Rowley, D. Nicholson, and N. G. Parsonage. Monte Carlo grand canonical ensemble calculation in a gas-liquid transition region for 12-6 argon. *J. Comp. Phys.*, 17:401–414, 1975.