LA-UR-17-25985
Approved for public release; distribution is unlimited.

Title: Sensitivity Analysis and Uncertainty Quantification for the LAMMPS Molecular Dynamics Simulation Code

Author(s): Picard, Richard Roy
Bhat, Kabekode Ghanasham

Intended for: Report

Issued: 2017-07-18

Sensitivity Analysis and Uncertainty Quantification for the LAMMPS Molecular Dynamics Simulation Code

Rick Picard and K. Sham Bhat

Abstract

We examine sensitivity analysis and uncertainty quantification for molecular dynamics simulation. Extreme (large or small) output values for the LAMMPS code often occur at the boundaries of input regions, and uncertainties in those boundary values are overlooked by common SA methods. Similarly, input values for which code outputs are consistent with calibration data can also occur near boundaries. Upon applying approaches in the literature for imprecise probabilities (IPs), much more realistic results are obtained than for the complacent application of standard SA and code calibration.

1. Introduction

Molecular dynamics simulation codes based on density functional theory, e.g., Tuckerman and Martyna (2000), are very time consuming to run. Surrogate codes such as LAMMPS (Large-scale Atomic/Molecular Massively Parallel Simulator) provide approximate but much faster results, and are often useful. In what follows, we examine the LAMMPS code coupled with the Tersoff parameterization for the potential energy function. LAMMPS takes an initial atomic configuration as an input, which consists of the positions and velocities of elemental atoms, as well as parameters defining the potential energy function, and outputs physical properties of interest. More extensive documentation on LAMMPS is available at the LAMMPS website, http://lammps.sandia.gov.

LAMMPS parameters are calibrated to outputs from aforementioned code runs based on density functional theory (DFT), which take as input the same atomic configuration as LAMMPS but do not use a parameterized potential energy model. These computationally intensive DFT code runs are regarded as more accurate than LAMMPS. Once calibrated, LAMMPS parameters providing good agreement between LAMMPS and DFT results are identified − parameter values that are often not well known in advance.

The primary purpose of this report to show that, under many circumstances, standard sensitivity analysis and/or Bayesian calibration with uniform priors often results in misleading and overconfident estimates of uncertainty in quantities of interest.

2. The Computer Model

2.1 Notation and Equations

Following Juslin et al. (2005), relevant physics equations and parameters for the Tersoff model for potential energy (e.g., Tersoff (1988), Tersoff (1989)) are briefly described below. For more details about the potential, see Juslin et al. (2005) and Tersoff (1988).

The total energy E for the atomic configuration is written

$$E \;=\; \sum_{i>j} f^c(r_{i,j}) \left[V^R\left(r_{i,j}\right) \;+\; \frac{b_{i,j}+b_{j,i}}{2}\, V^A\left(r_{i,j}\right) \right] , \tag{1}$$

where the summation is over pairs of atoms (i, j) in the configuration. Here, the cut-off function is

$$f^c(r_{i,j}) = \begin{cases} 1 & \text{if } r_{i,j} \;\leq\; R \,-\, D, \\ \frac{1}{2} \;-\; sin\left(\frac{\pi}{2}\, \frac{r_{i,j}-R}{D} \right) & \text{if } r_{i,j} \;\in\; (R-D, R+D), \\ 0 & \text{if } r_{i,j} \;\geq\; R \,+\, D\ , \end{cases}$$

where $r_{i,j}$ denotes the distance between atoms i and j in the configuration. Parameters R and D determine the cutoff region and are primarily used to expedite computation.

Atom pairs (i, j) whose atom-to-atom distances $r_{i,j}$ are large contribute minimal quantities to the energy E, but can add greatly to the computation time since the total number of atom pairs is proportional to the square of the number of atoms in the configuration. The cut-off function f^c may considerably reduce computational complexity by eliminating calculations for atom pairs for large distances $r_{i,j} \geq R + D$ that contribute minimally to

the total result. For all of what follows, cut-off function values are fixed at $R = 4$ Å and $D = 0.2$ Å.

The repulsive and attractive pair potentials in (1) are

$$V^R(\,r_{i,j}\,) \;=\; \frac{D_0}{S-1}\; exp\Big(-\beta\,\sqrt{2\,S}\;(r_{i,j}-r_0)\Big)\;,\;\text{and}$$

$$V^A(\,r_{i,j}\,) \;=\; \frac{SD_0}{S-1}\; exp\Big(-\beta\,\sqrt{2\,/\,S}\;(r_{i,j}-r_0)\Big)\;,$$

where D_0 is the dimer bond energy, r_0 is the dimer bond distance, β is a parameter that can be determined from the ground-state oscillation frequency of the dimer, and S is an adjustable parameter.

The two-body bond-order parameter $b_{i,j}$ in (1) has the form

$$b_{i,j} \;=\; 1\,/\,\sqrt{1\,+\,\chi_{i,j}}\;\;,$$

where

$$\chi_{i,j} \;=\; \sum_{k\neq i,j} f^c(r_{i,k})\,g(\,\theta_{i,j,k}\,)\,\omega_{i,j,k}\;exp(\alpha_{i,j,k}\,(r_{i,j}-r_{i,k}))\;. \tag{2}$$

and the three-body angular function

$$g(\,\theta_{i,j,k}\,) \;=\; \gamma\;\omega_{i,j,k}\;\left(\;1\;+\;\frac{c^2}{d^2}-\;\frac{c^2}{d^2+(h+cos(\theta_{i,j,k}))^2}\;\right)\;. \tag{3}$$

Here, $\theta_{i,j,k}$ is the angle between line segment (i,j) connecting atoms i and j and the segment (i,k) connecting atoms i and k, while γ, c, d, and h are two-body model parameters of the atoms i and k. The three-body parameters $\omega_{i,j,k}$ and $\alpha_{i,j,k}$ must obey certain symmetry relations that depend on the types of atoms (in the 29-atom example to follow,

there are only types of atoms: tungsten and hydrogen). The parameters described here are referred to as Juslin parameters, and are listed in Table 1.

For the rest of this report, only two-body parameters with interactions of one tungsten and one hydrogen atom (W-H) are varied; two-body parameters with H-H and W-W interactions are taken to be constant. The summation for $\chi_{i,j}$ in (2) is somewhat complex, using periodic boundary conditions which are standard for MD simulations (Schlick (2010)). And for each (i, j) atom pair, the k-th atom is chosen among all 27 k-th atoms based on a minimum image distance criterion.

2.2 "Non-Physical" Code Runs and Correlated LAMMPS Inputs

The LAMMPS code inputs are described in the LAMMPS online documentation. In particular (see page 4 of http://lammps.sandia.lanl.gov/doc/pair_tersoff.html), LAMMPS relies on the quantities

$$A \;=\; \frac{D_0}{S-1}\; exp\left(\beta\, r_0\, \sqrt{2\,S}\right)\;,$$

$$B \;=\; \frac{S D_0}{S-1}\; exp\left(\beta\, r_0\, \sqrt{2\,/\,S}\right)\;,$$

$$\lambda_1 \;=\; \beta\,\sqrt{2\,S}\;,\;\text{and}$$

$$\lambda_2 \;=\; \beta\,\sqrt{2\,/\,S}\;.$$

Using this reparameterization, for the (i, j) atom pair, the repulsive and attractive pair potentials are written as

$$V^R(\,r_{i,j}\,) \;=\; A\; exp(\,-\lambda_1\, r_{i,j}\,)\;,\;\text{and}$$

$$V^A\,(\,r_{i,j}\,)\;=\;B\;exp\,(\,-\,\lambda_2\,r_{i,j}\,)\;\;.$$

The LAMMPS code *does not verify* whether the parameter values for A and B are feasible, i.e., it does not check whether there exist specific values of the Juslin parameters $\{\,D_0, S, \beta, r_0\,\}$ within their allowable ranges for which the LAMMPS parameter values A and B are simultaneously obtained. Nor does LAMMPS check whether there exist legitimate values of S and β that produce LAMMPS inputs (λ_1, λ_2).

Along similar lines, the user must ensure that LAMMPS code inputs $\{\,\alpha_{i,j,k}\,\}$ and $\{\,\gamma_{i,j,k}\,\}$ obey symmetry relations. For example, if i and j denote tungsten atoms while atom k denotes a hydrogen atom, then the three-body term $\alpha_{i,j,k} = \alpha_{HHW}$ must exactly equal $\alpha_{i,k,j} = \alpha_{HWH}$ for the system to be physically feasible; LAMMPS does not automatically enforce this requirement.

The responsibility lies with the user to ensure that all LAMMPS input values are indeed physically legitimate. When LAMMPS input values do not correspond with physically legitimate combinations of Tersoff parameters, *non-physical* LAMMPS outputs are obtained. The functional dependence of LAMMPS parameters A, B, λ_1, and λ_2 on the same underlying physical parameters induces a strong correlation among these inputs. An arguably extreme example is the ratio $\lambda_1/\lambda_2 \equiv S$, and other examples are given in the next section.

3 Uncertainty Quantification

3.1 Background

As is typically the case with many simulations of physical processes, the "true" LAMMPS input parameter values $(A, B, \lambda_1, \lambda_2, \ldots)$ are not known exactly. Physical measurements used to determine parameter values are not perfect, and there are uncertainties in the Juslin parameters (on which the LAMMPS parameters are based). The "Nominal Prior" columns of Table 1 summarize the corresponding two-body W-H and three-body model parameters (see Juslin et al. (2005)) and the best guesses from local materials science experts of plausible ranges for their values.

Although Juslin's implementation framework allows for varying all two-body parameters, the W-H parameters are considered to be more sensitive than the respective H-H and W-W two-body terms; hence for the duration of this report, only W-H two-body parameters will be allowed to vary.

Consider the first four parameters, D_0, r_0, β, and S, which entirely determine the LAMMPS code inputs $(A, B, \lambda_1, \lambda_2)$. The ranges of Nominal Prior values in Table 1 were obtained through a calibration exercise, running the DFT code with 20 two-atom configurations of tungsten and hydrogen. W-on-H forces were obtained in the x, y, and z directions for each of these configurations. Juslin parameter values producing LAMMPS output forces in good agreement with DFT forces (as quantified by adding in quadrature the force differences in the three dimensions) are provided in the Nominal Prior columns

Table 1. Nominal ranges for other parameters in Table 1 were obtained upon consulting local materials science experts, and are discussed later.

Table 1. Plausible Ranges for Juslin Parameter Values.

Input	Nominal Prior		Low SA Prior		High SA Prior	
Parameter	Lower & Upper Limits		Lower & Upper Limits		Lower & Upper Limits	
D_0 (eV)	2.41	5.11	2.41×1.1	5.11×1.1	2.41×1.1	5.11×1.1
r_0 (Å)	1.27	1.91	1.27	1.91	1.27×1.1	1.91×1.1
β $\left(\text{Å}^{-1}\right)$	0.99	1.85	0.99/1.1	1.85/1.1	0.99/1.1	1.85/1.1
S (λ_1/λ_2)	1.25	2.55	1.25	2.55	1.25/1.1	2.55/1.1
c	0	3	0	2	0	4
d	0	3	0	4	0	2
h	-1	2	-1	2	-1	2
γ	0	0.1	0	0.07	0	0.11
$\alpha_{HHW} = \alpha_{HWH}$	0	10	0	5	0	12
α_{HWW}	0	10	0	5	0	12
α_{WHH}	0	10	0	5	0	12
$\alpha_{WHW} = \alpha_{WWH}$	0	10	0	5	0	12
$\omega_{HHW} = \omega_{HWH}$	0	10	0	5	0	12
ω_{HWW}	0	10	0	5	0	12
ω_{WHH}	0	10	0	5	0	12
$\omega_{WHW} = \omega_{WWH}$	0	10	0	5	0	12

The subject of sensitivity analysis is concerned with the propagation of uncertainties through computer codes such as LAMMPS. That is, as code inputs are varied over their plausible ranges, LAMMPS outputs such as E, $V^R(r_{i,j})$ and $V^A(r_{i,j})$ also vary. Quantifying the extent of this variation, and understanding which specific code inputs most affect LAMMPS code outputs, are the main goals of sensitivity analysis.

A second subject of interest is code calibration, i.e., identifying which LAMMPS input values provide the best match of LAMMPS outputs to more precise DFT results. This exercise is similar to the effort leading to the ranges in Table 1 of the four physical parameters. In general, calibration can be done using a frequentist approach, similar to solving a nonlinear least squares problem, where the sum of squared differences between the LAMMPS and DFT code outputs is minimized as a function of LAMMPS code inputs. The calibration can alternatively be done using a Bayesian approach, where prior information on code input values is combined with the fit of the code outputs to the data. Either way, a "best" (MLE or HPD) set of code inputs can be identified, together with corresponding uncertainties.

3.2 Sensitivity Analysis: Important Inputs and Rare Events

Traditional sensitivity analysis generally begins by assigning uniform distributions to the input parameters of the computer code of interest, and propagating the corresponding uncertainty to the code output(s) of interest. To illustrate, consider the plausible ranges of

nominal prior values in Table 1 and focus initially on the repulsive pair potential $V^R(r_{i,j})$ for a particular atom pair (i, j), here involving a single hydrogen and a single tungsten atom.

Because the output $V^R(r_{i,j})$ can be quickly computed, sensitivity analysis properties can be obtained by brute force, i.e., simulating many thousands of sets of inputs and obtaining corresponding output values. For more complex codes, the brute force approach is not practical, and sensitivity analysis requires more sophisticated experimental designs for code runs, such as the Latin hypercube for common output summaries (McKay et al. (1979)) or importance sampled designs for rare events (Picard and Williams (2013)).

Note that only the inputs D_0, r_0, β, and S affect the outputs $V^R(r_{i,j})$ and $V^A(r_{i,j})$ in (1). Further, only the W-H two-body terms γ, c, d, h, and the three-body terms $\alpha_{i,j,k}$ and $\omega_{i,j,k}$ affect the outputs $b_{i,j}$ and $b_{j,i}$. Hence, the sensitivity analysis can be partitioned to some extent.

Figure 1 shows the resulting scatter of LAMMPS inputs A, B, λ_1, and λ_2 from a 1K uniform sample of the Juslin parameters (D_0, r_0, β, S). All points in the plots lie above the x=y lines, so that $A < B$ and $\lambda_1 < \lambda_2$ in all cases. Inputs A and B have highly skewed induced distributions, while λ_1, and λ_2 are reasonably Gaussian. Finally, there are large regions of the (A, B) and (λ_1, λ_2) spaces that correspond to non-physical code runs.

These plots indicate that applying standard sensitivity analysis methods to LAMMPS inputs directly − methods that too casually assume independence among input parameters

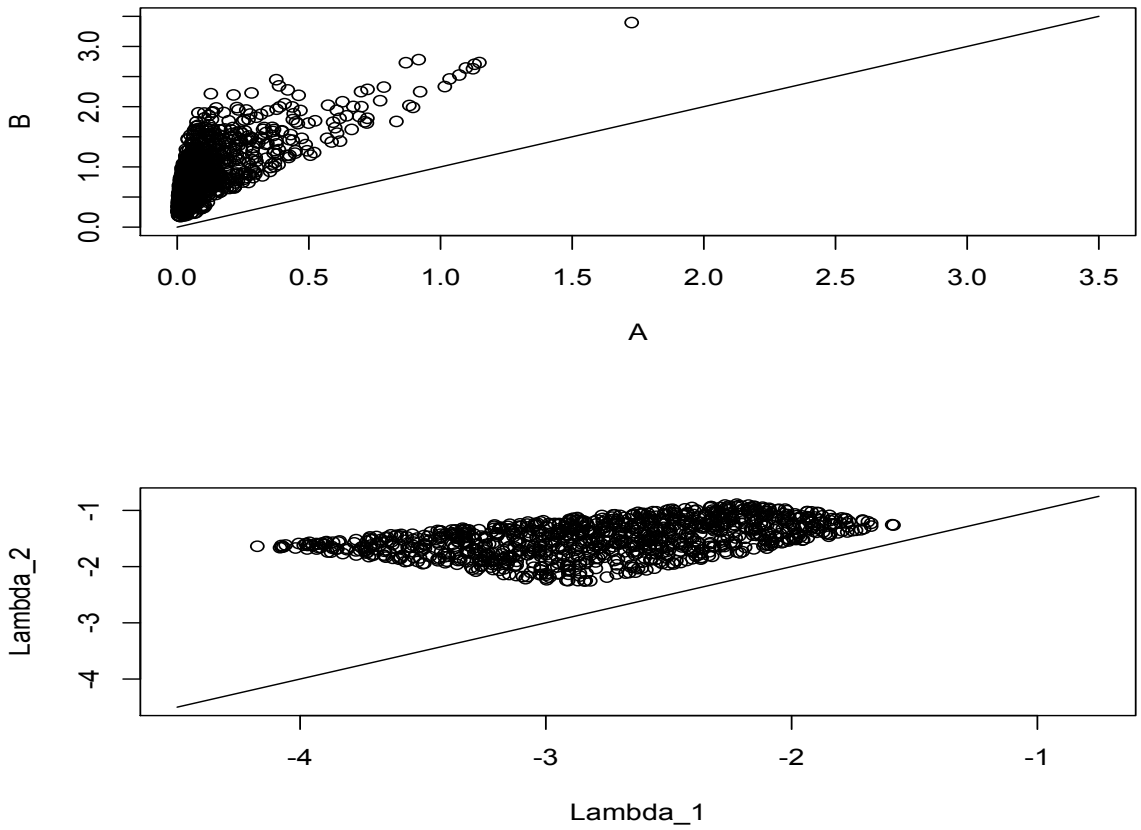

Figure 1: Sample of LAMMPS Input Parameters.

and uniform marginal distributions − often lead to poor conclusions.

Figure 2 suggests the pairwise potentials are strongly correlated. Furthermore, the pairwise potentials span a wide range (from just above zero to almost 40 eV for this atom pair having $r_{i,j} \approx 1.5\,\text{Å}$). In terms of the effects on total energy and on its uncertainty, large values have the greatest impact.

Consequently, understanding extreme V^R and V^A values is important. Consider conducting a larger simulation, using 100K samples in place of the 1K samples to this point.

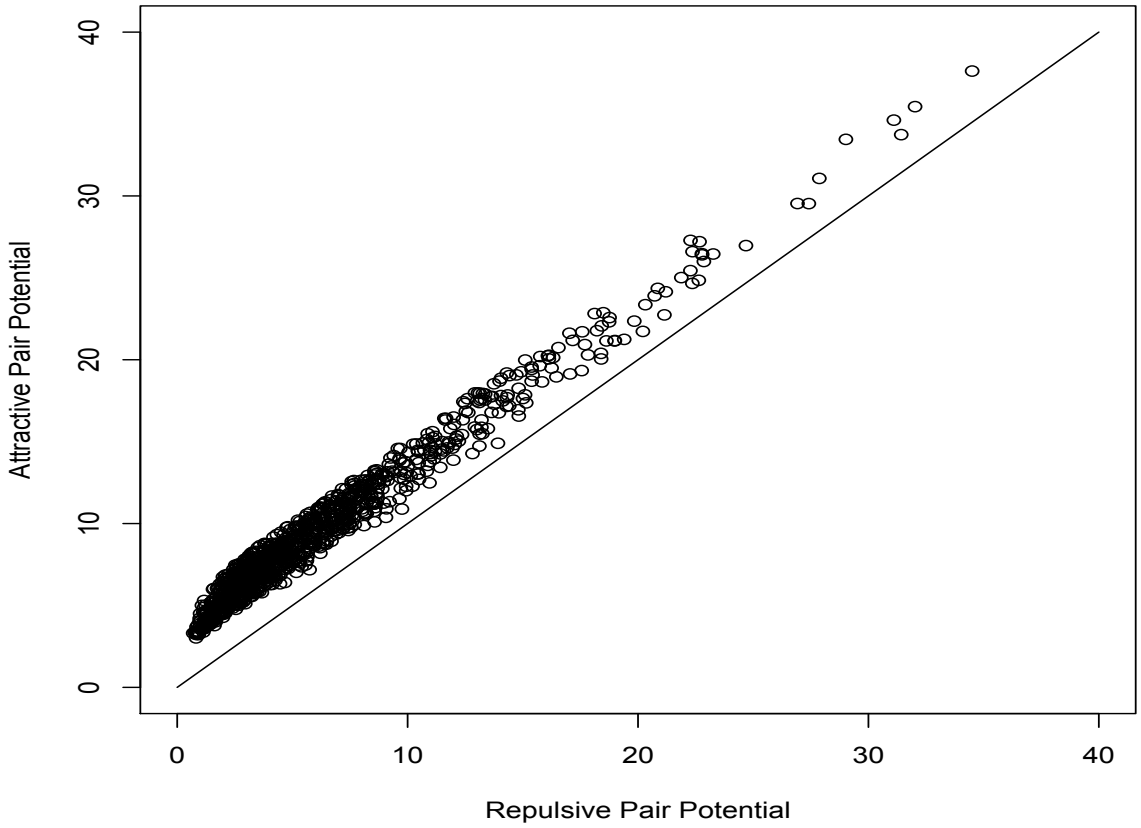

Figure 2: 1K Sample of LAMMPS Input Parameters.

Quantities of interest are then more accurately determined. The average repulsive pair potential here is 6.34, estimated to a standard deviation of slightly more than 0.01; the average attractive pair potential is 9.78, again estimated to a standard deviation of slightly more than 0.01. The standard deviations of the skewed V^R and V^A distributions, as induced by the uncertainty distribution on the inputs, are 5.0 and 5.1, respectively.

In principle, the average and standard deviation of each pair potential distribution can be determined to arbitrary accuracy simply by increasing the number of code runs.

Taken literally, this implies that statistical properties of these outputs could be determined exactly. This counterintuitive property of sensitivity analysis is revisited shortly, and is an artifact of treating input distributions as if they were known perfectly, when in fact they are not.

The 100K data sample also allows for a more complete exploration on the portion of the output space where V^R and V^A are large. For example, only 33 of the 100K runs resulted in a repulsive pair potential exceeding 40 eV. In other words, such large output values are rare ($\approx 0.03\%$) events. Depending on the specific use of the code (e.g., its aid in understanding material science properties for nuclear reactor safety), rare event probabilities could be of great interest. Similar to the output uncertainty of the average value $V^R(r_{i,j})$, quantified by the standard deviation of $V^R(r_{i,j})$ divided by $\sqrt{n}$, where n denotes the number of ordinary code runs, rare event probabilities can also be determined with essentially zero uncertainty by letting n become extremely large.

The study of rare events involves exploring which inputs "cause" these rare events to happen. Figure 3 displays two inputs correlated with rare ($V^R(r_{i,j}) > 40$ eV) events. The plot displays the entire assumed ranges of the input quantities as per the nominal prior in Table 1, and shows that only a small portion of the input space is correlated with this rare event, which exclusively involves small values of S (the ratio of LAMMPS exponential factors λ_1, and λ_2) and large values of the dimer bond distance r_0. Admittedly, this conclusion could have been reached by visual inspection of the equation for $V^R(r_{i,j})$

or from a first-principles understanding of the physics involved. For more complex codes, however, rare event plots such as Figure 3 may be needed.

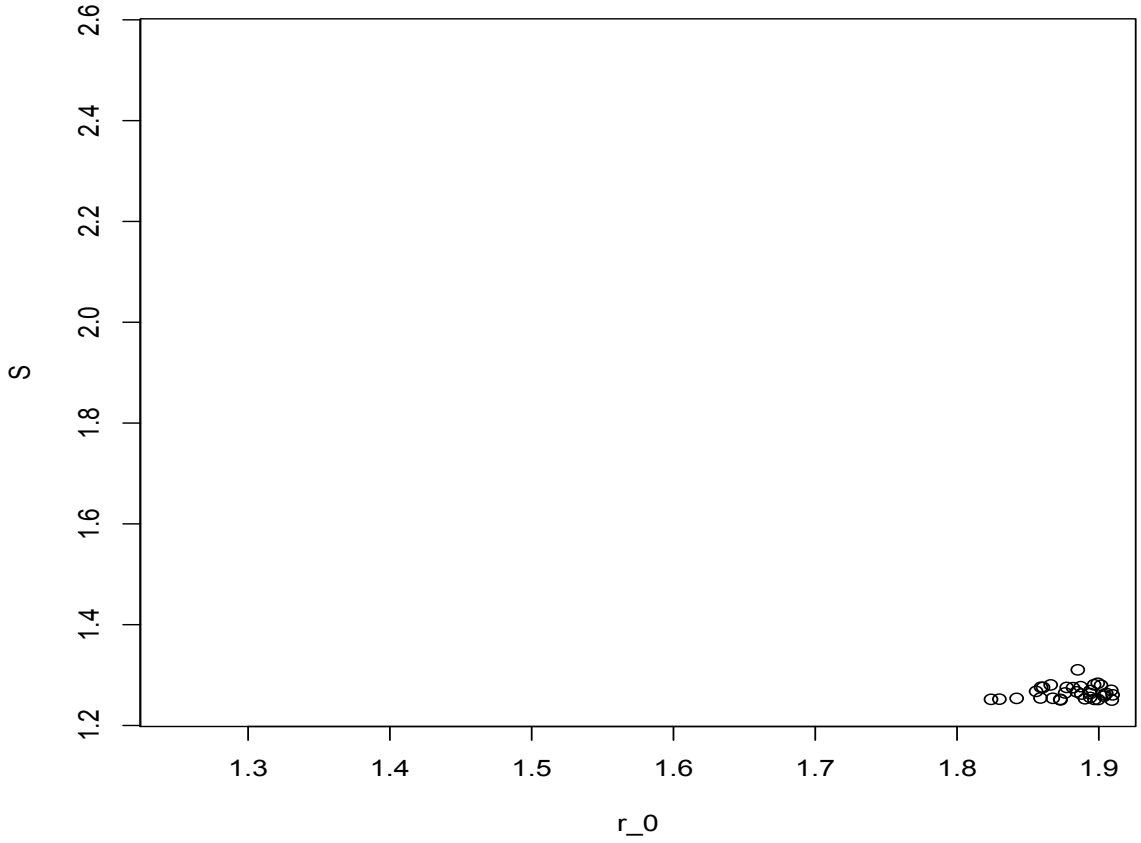

Figure 3: Juslin Input Parameters for Large ($V^R(r_{i,j}) > 40$ eV) Pairwise Potentials.

As should be intuitively obvious from Figure 3, the rare event probability is clearly dependent on the lower limit 1.25 for S and the upper limit 1.91 for r_0. If these limits were truly known exactly, then the rare event probability Pr($V^R(r_{i,j}) > 40$) could then be determined to great accuracy via a large number of code runs.

Importantly, the above limits are not known exactly − they are quantitative estimates

from a calibration exercise. For the other input parameters in Table 1, the "plausible limits" are essentially best guesses from subject matter experts, and they are not precisely known either, primarily because the experts' intuition regarding input parameters is more limited than their intuition about the underlying physics. Consequently, quantities such as rare event probabilities can be highly dependent on uncertainties in presumed ranges of values such as in Table 1.

One common statistical tool to deal with this situation involves hierarchical modeling. Probability distributions are assigned to the unknown limits, such as 1.25 for S and 1.91 for r_0. In straightforward hierarchical sensitivity analysis, input quantities are obtained by first sampling limits from the hierarchical prior, and then sampling input quantities from uniform distributions conditional on those limits. Thus, ordinary sensitivity analysis quantities and rare event probabilities are averaged with respect to the hierarchical prior.

This approach has substantial downsides as follows:

a) these hierarchical prior probability distributions on the limits are *themselves* not known exactly, and thus should not be taken literally, and

b) the end result of the hierarchical analysis is *still* pretentiously precise because there is only a single posterior distribution from which a large number of MCMC samples are simulated, allowing quantities such as a rare event probability to be obtainable with essentially zero uncertainty when sample sizes are very large.

What is needed here is a sensitivity analysis for the sensitivity analysis. Such an ap-

proach goes under the moniker of *imprecise probability* (Walley (1991); Picard and Vander-Wiel (2017)).

Classic sensitivity conclusions can be greatly affected by comparatively modest uncertainties on input bounds, in part due to rare events occurring near the presumed bounds on certain inputs (Figure 3). To illustrate, suppose that the D_0, r_0, β, and S values in Table 1 has a plausible range of $\pm 10\%$. For example, suppose the lower bound for D_0, 2.41 eV, could plausibly be between the values 2.41/1.1 and 2.41×1.1.

The calculations can be run quickly, and so it is possible to consider all $2^8 = 256$ cases, that is, there are two endpoints for each of the four inputs D_0, r_0, β, and S affecting the repulsive pair potential $V^R(r_{i,j})$. For each of these 8 endpoints, lower and upper limits such as 2.41/1.1 and 2.41×1.1 for D_0, are considered. For each of these 256 cases, a 1M sample of input sets is obtained. Recall that the average repulsive pair potential is equal to 6.34±0.01 eV for the nominal prior parameters in Table 1. Across the 256 cases, however, the "6.34 eV" ranges from 2.97 eV to 15.38 eV. The lower plausible value 2.97 eV is achieved when endpoints for D_0 and r_0 are reduced by 10% and endpoints for β and S are increased by 10% from their nominal values in Table 1. The upper plausible value, 15.38 eV, is achieved when endpoints for D_0, r_0, and β are increased by 10% while S, is decreased by 10% from their nominal values in Table 1.

Even though the ±10% nominal uncertainties on limits for the inputs may seem modest, the effects of these uncertainties in input parameter bounds create more than a fivefold

range of the average repulsive pair potential $V^R(r_{i,j})$. The situation with the rare event probability $\Pr(V^R(r_{i,j}) > 40)$, whose nominal value was $\approx 0.03\%$, is even more extreme. Here, modest 10% uncertainties in bounds for inputs lead to rare event probabilities that range from zero to 6%, the latter value indicating that the nominally "rare" event may not even be all that rare. And the nominal standard deviation of $V^R(r_{i,j})$, 5.0 eV, ranges from 2.1 to 14.5 eV over the 256 cases.

The bottom line is that classic sensitivity analysis, in treating the nominal prior endpoints in Table 1 as if known exactly, yields grossly misleadingly and overly precise results. Accounting for uncertainties in the assumed bounds for the inputs sometimes leads to wide ranges of plausible output values, especially when extreme outputs occur near the endpoints of the ranges as displayed in Figure 3. As would be expected from Figure 3, the most important inputs are the dimer bond distance r_0 and the ratio S of LAMMPS parameters λ_1 and λ_2.

This sensitivity analysis is specific to one atom pair (i, j), and is done for expository purposes, to show that in this easily understood case, uncertainty in input bounds can sometimes swamp the nominal sensitivity analysis uncertainties in quantities of interest.

Similar effects can occur in more complex applications. Return to the 29-atom configuration of interest and consider the total energy E in (1). Taking the ranges of inputs in Table 1 as being perfect, and propagating 100K sets of input values through the code produces the spectrum of total energy values displayed in Figure 4. Total energy values

span a range from -246 eV to -75 eV; the negative sign reflects that the attractive pair potentials generally exceed the repulsive pair potentials.

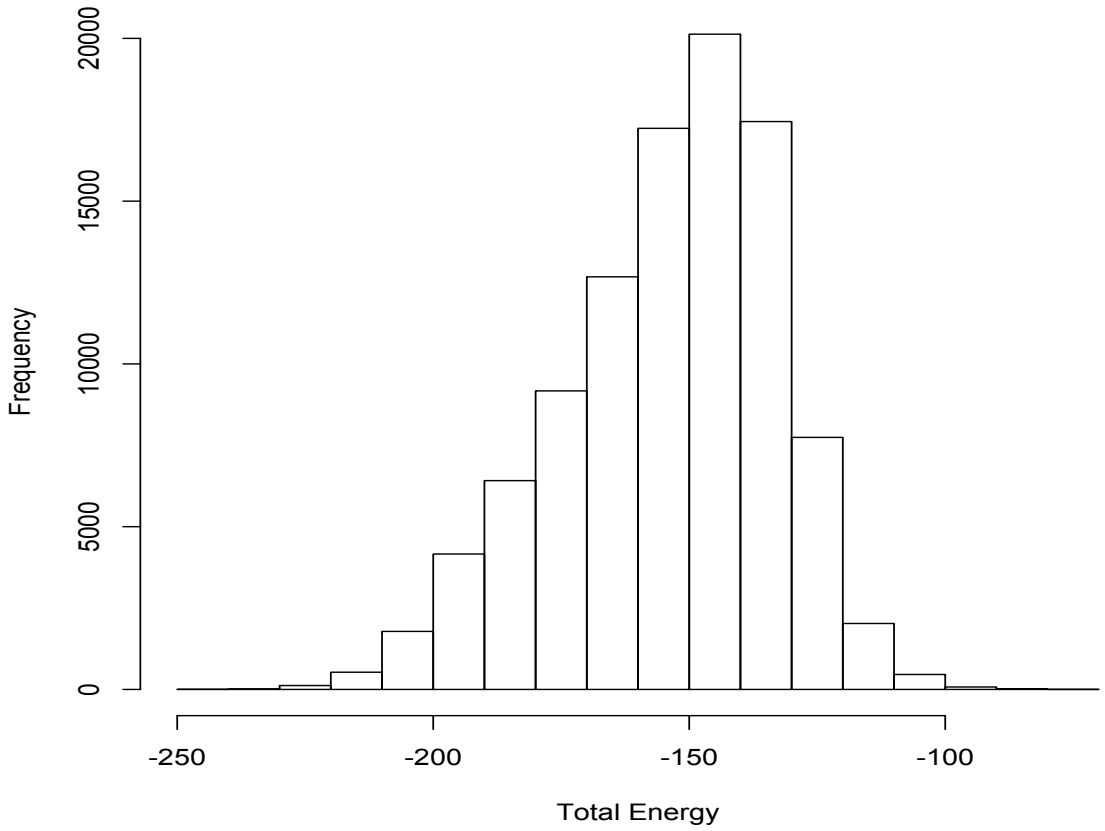

Figure 4: Total Energy Distribution.

The most important inputs for the total energy E are the γ and ω_{HHW} parameters. A rare event plot, Figure 5, displays the smallest 100 (of the 100K) output energies. Note the clustering of inputs for these rare (one-in-1000) large values, with low parameter values corresponding to low energies. Conversely, large values of γ and ω_{HHW} and large values of r_0 and c are also related to extremely high energies, as seen in Figure 6.

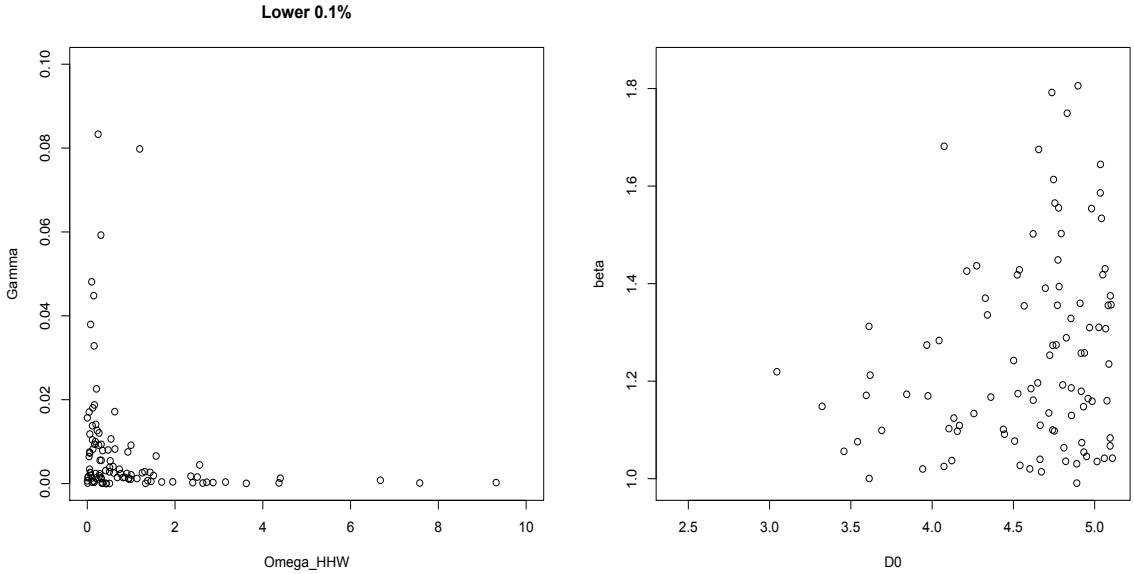

Figure 5: Input Values for Very Low Energies.

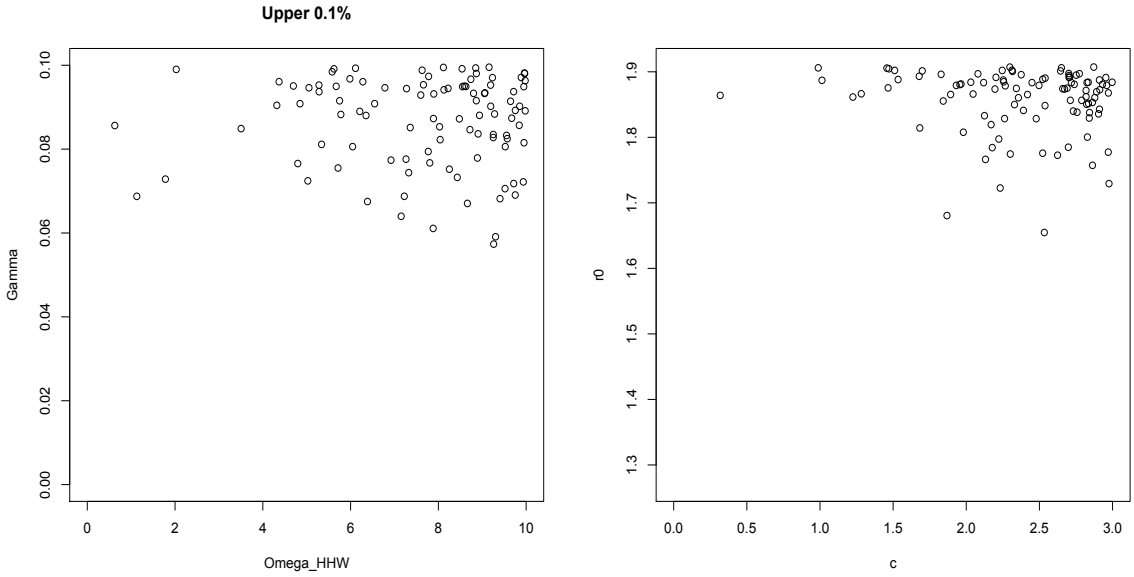

Figure 6: Input Values for Very High Energies.

Assuming the same ±10% uncertainties in input bounds as before for the D_0, r_0, β, and S inputs, which are based on calibration to the DFT code. For the other inputs based on expert opinion, consultation with the subject matter experts revealed that the nominal upper endpoints of the $U(0,3)$ distributions for c and d could be plausibly as low as 2, or as large as 4. Similarly, the upper endpoint for the nominal $U(0,0.1)$ prior for γ could range from 0.07 to 0.11, and the upper endpoints of the nominal $U(0,10)$ priors for the $\omega_{i,j,k}$ and $\alpha_{i,j,k}$ parameters could be as small as 5 or as large as 12 (aside: the "WWW" and "HHH" three-body parameters are not varied herein). Lastly, the lower endpoint of the nominal $U(-1,2)$ prior for h could plausibly be as small as -1.5, and the upper endpoint, 2, could range between 1 and 2.5. These values are expressed in Table 1.

Recall that the nominal prior produces an induced distribution (Figure 4) for total energy E having mean -154 eV, a value that is essentially determined exactly based on 100K brute force samples. Other properties of this distribution, such as rare event probabilities, are also determined with great accuracy. Upon investigating the spectrum of prior distributions having plausible endpoints, histograms of two (chosen to be best and worst case) induced total energy distributions are shown in Figure 7. Endpoint values for these two cases are given in Table 1, where parameters deemed less important are not varied.

Rare event plots such as Figures 5 and 6 are useful in guiding this process when code runs are too time consuming to consider all possible combinations of priors. Means of the displayed distributions for the two cases are -185 eV and -110 eV. Both distributions are

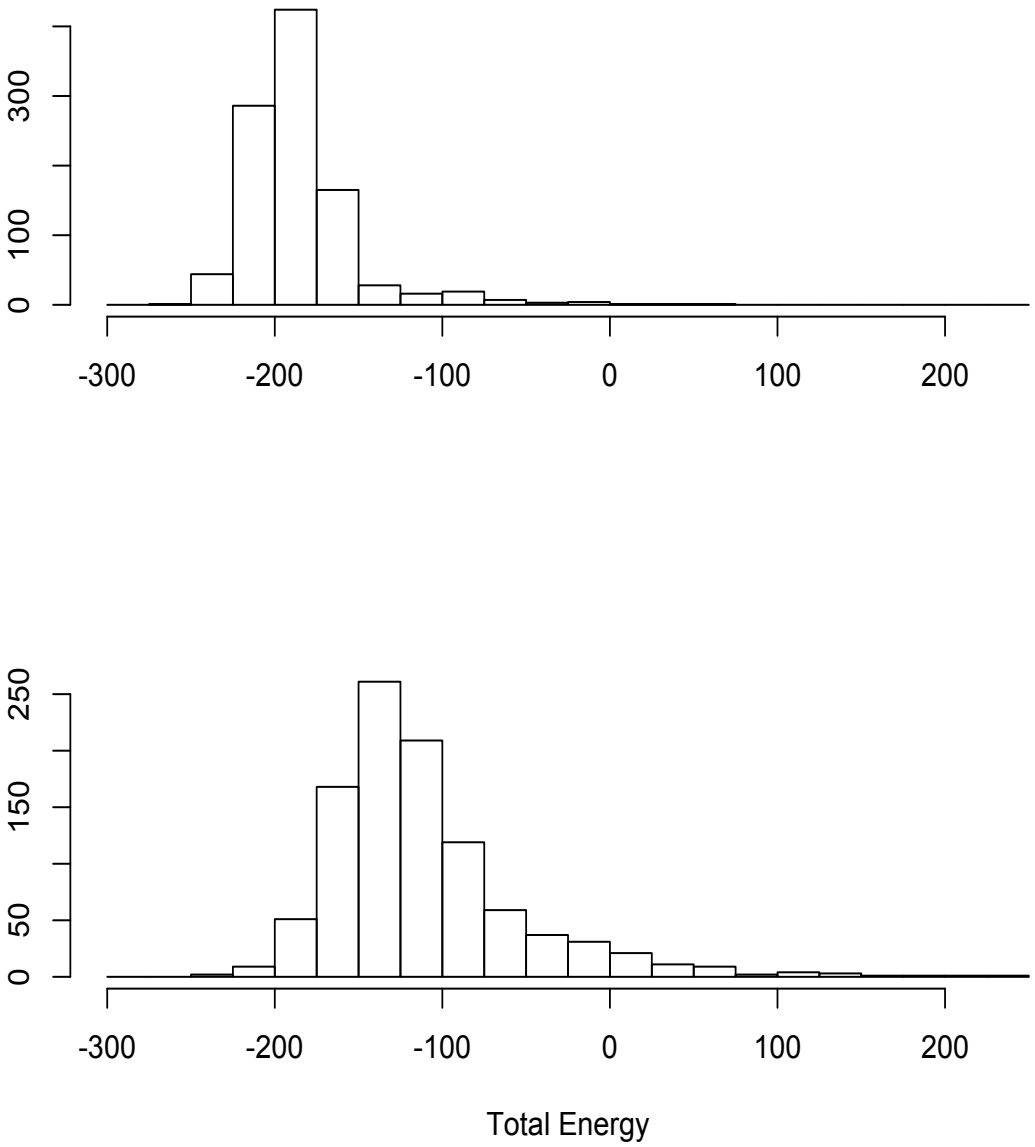

Figure 7: Two Extreme Plausible Priors.

skewed to the right, implying that the usual "plus-or-minus-2σ" confidence intervals do not have the usual Gaussian probabilities. Also greatly affected are rare event probabilities.

As is typical for IP (imprecise probability) analyses, results are quoted for the nominal prior (e.g., the mean of the induced energy distribution is -154 eV) as well as for the two extreme cases (e.g., the values -185 eV and -110 eV). For the nominal prior, the 90^{th} percentile of the output distribution is -129.7 eV (based in 100K simulated samples), i.e., there is only a 10% chance of a larger total energy. For the best and worst case priors, the corresponding IP range of probabilities for the nominal 10% figure is from 5.5% to 56.5%.

For computer models where the output(s) of interest are monotonic in certain input variables, ordinary SA results using only a nominal prior distribution can have apparent precisions that are highly misleading. It can easily (and wrongly) be concluded that output quantities of interest are determined with little uncertainty because of the large number of samples generated from the nominal prior. This misleading state of affairs is clearly bogus, but is a logical consequence of the "perfect-input-in, perfect-output-out" presumption of the Bayesian paradigm. As noted previously, replacing a nominal non-hierarchical prior with a hierarchical prior does nothing to change this "perfect-in-perfect-out" phenomenon.

Sensitivity to assumed prior conditions should be considered when assessing such results. A possible complication here is that subject matter experts can be overconfident in their abilities to determine prior parameters, with the consequence of such overconfidence being an understatement of IP ranges.

3.3 Code Calibration

Using a much more accurate and time consuming code based on density functional theory (DFT) as ground truth (e.g., Tuckerman and Martyna (2000)), code calibration here involves finding values of Juslin input parameters such that the LAMMPS code outputs best match ground truth. The DFT code takes as input only the configuration and does a lengthy calculation for the forces on individual atoms. The results herein are generated using DFT output of eight different configurations. Upon comparing LAMMPS values to those from DFT, calibrated input values can be obtained; as noted in Section 3.1, computed W-on-H forces are used as the basis for comparison. In more general code calibration contexts, code outputs are often compared with corresponding experimental data. The calibration approach pioneered by Kennedy and O'Hagan (2001) is used for the analysis to follow.

Similar to ordinary sensitivity analysis, uncertainty in the nominal prior distribution may (or may not) be large enough to matter. When calibrated inputs are well inside assumed prior bounds, so that the nominal prior is flat over the entire region of parameter values consistent with the data, then there may be little sensitivity to those assumed prior bounds. Otherwise, when the parameter values most consistent with the data are closer to the assumed bounds (or worse, outside of the assumed bounds), an IP approach is needed.

On the next several pages, simulated distributions of Juslin parameters (as induced by the calibrated posterior) are displayed. The first two plots (Figure 8 and 9) are the two-

and three-body Juslin parameters, respectively, for the nominal prior. The following plots are the two- and three-body Juslin parameters for the low and high energy SA priors.

Ideally, the volume of calibration data would be considerable and the best calibrated parameters would be well inside the prior bounds, so that essentially the same posterior would be obtained in all cases. In the case at hand, however, results are somewhat sensitive to the nominal prior.

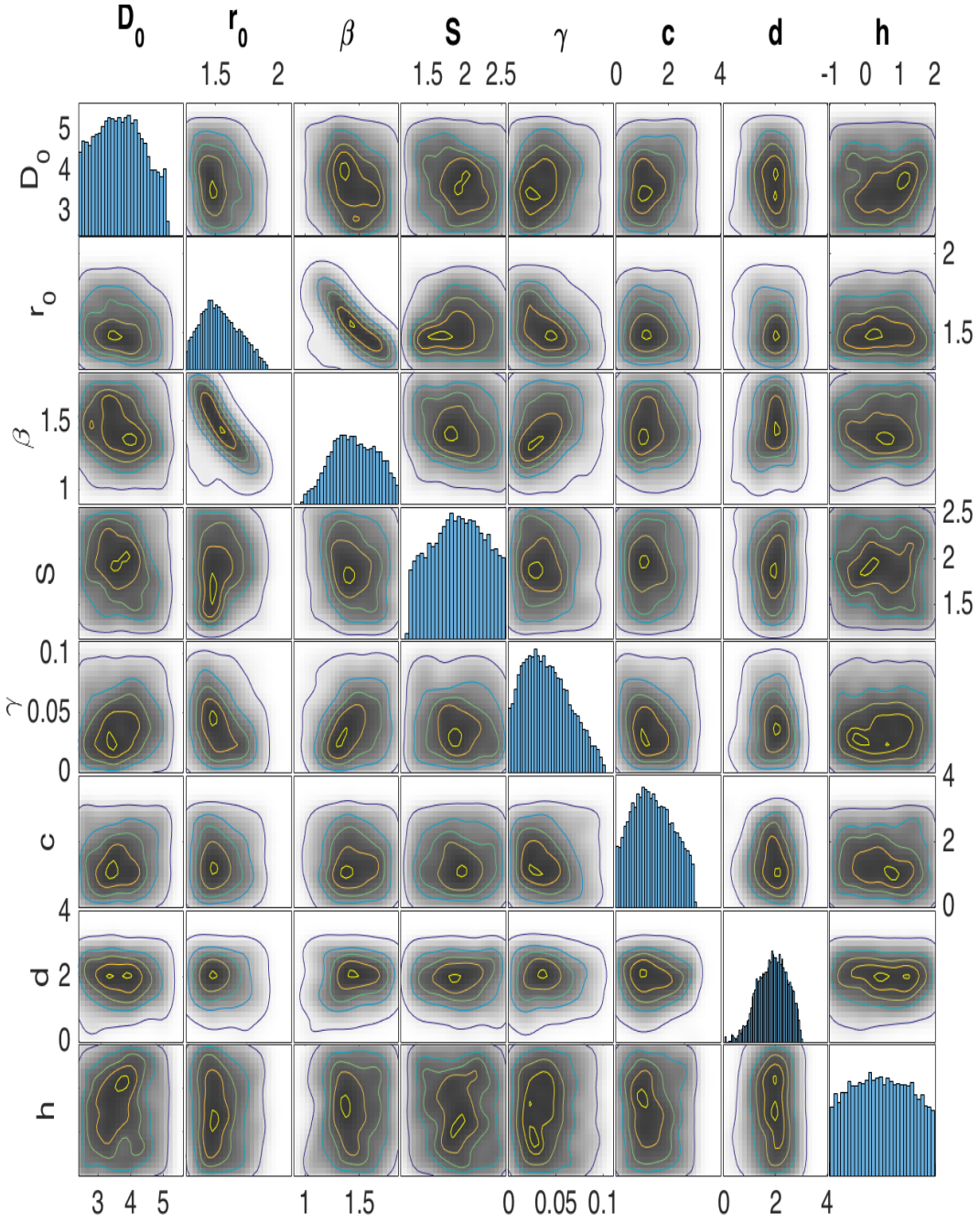

Figure 8: Posterior Densities for Two Body Parameters using the Nominal Prior

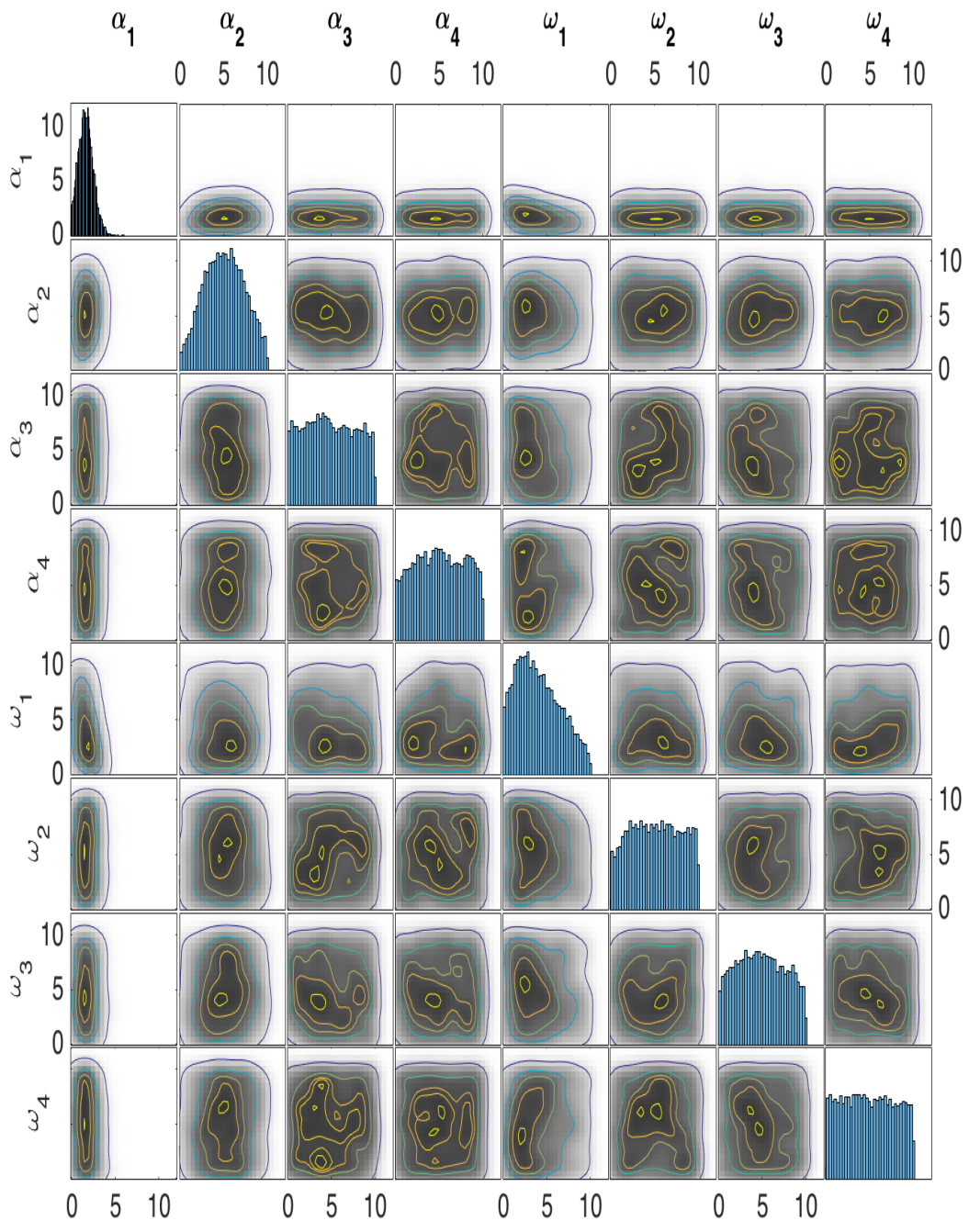

Figure 9: Posterior Densities for Three Body Parameters using the Nominal Prior

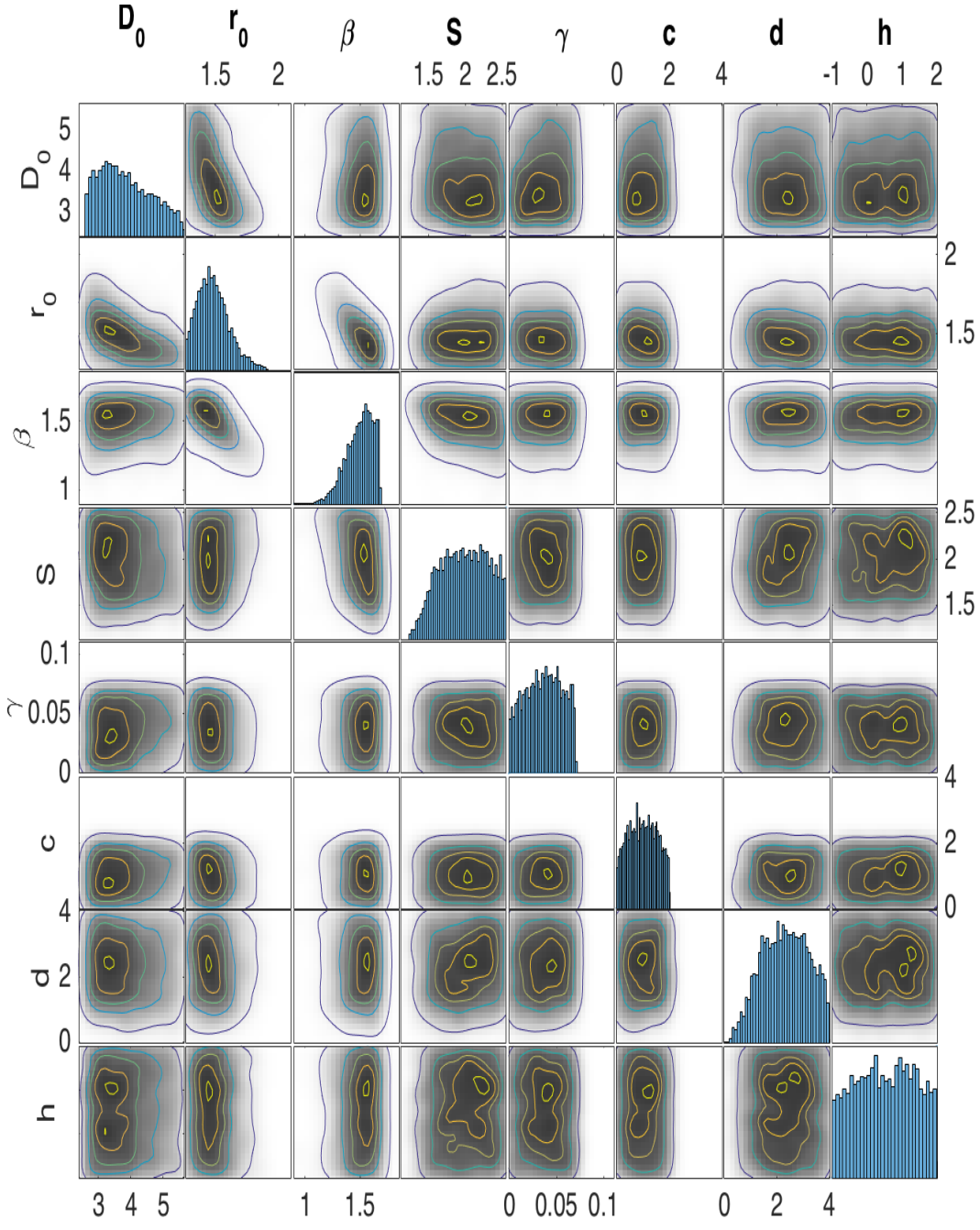

Figure 10: Posterior Densities for Two Body Parameters using the Low Energy SA Prior

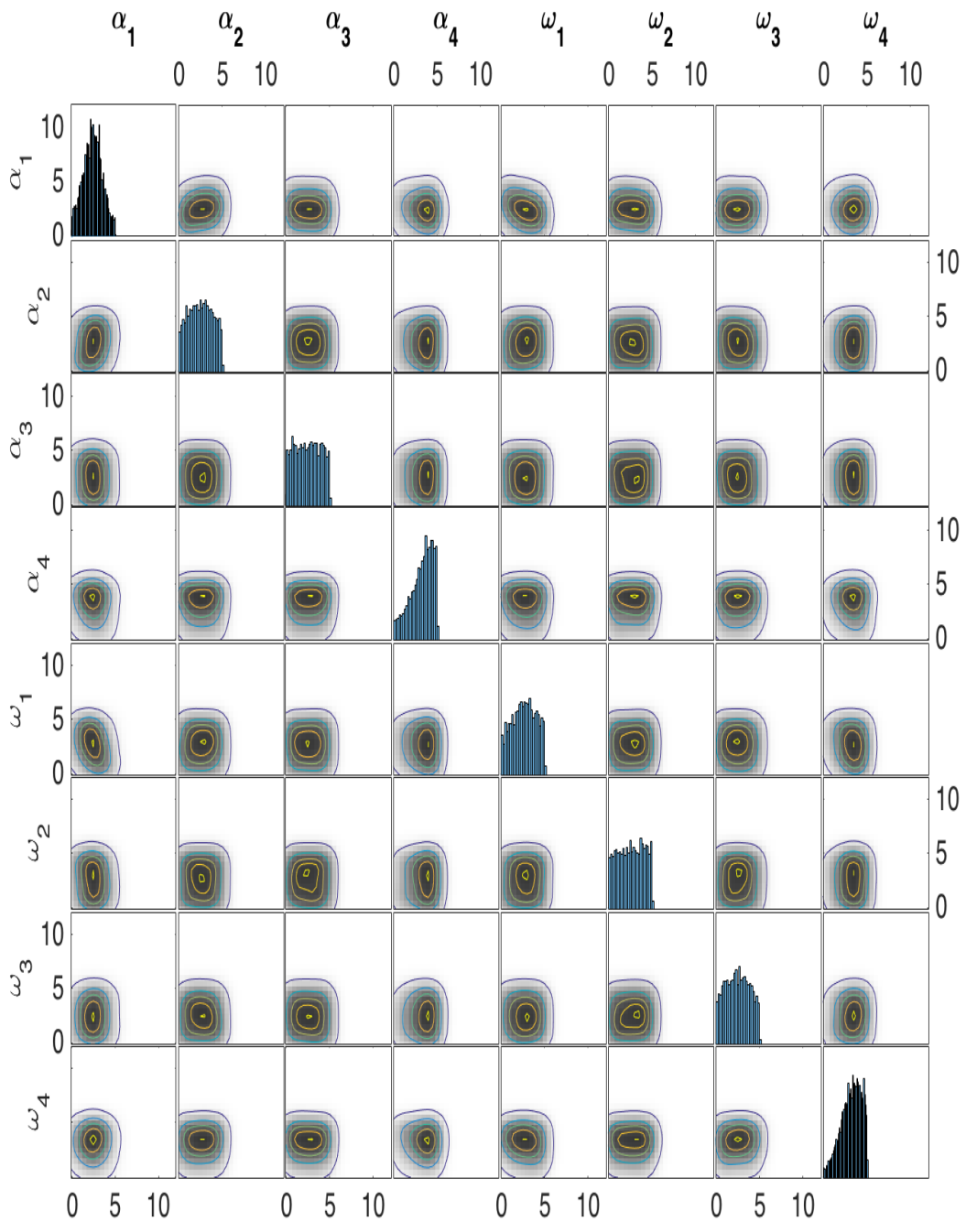

Figure 11: Posterior Densities for Three-Body Parameters using the Low Energy SA Prior

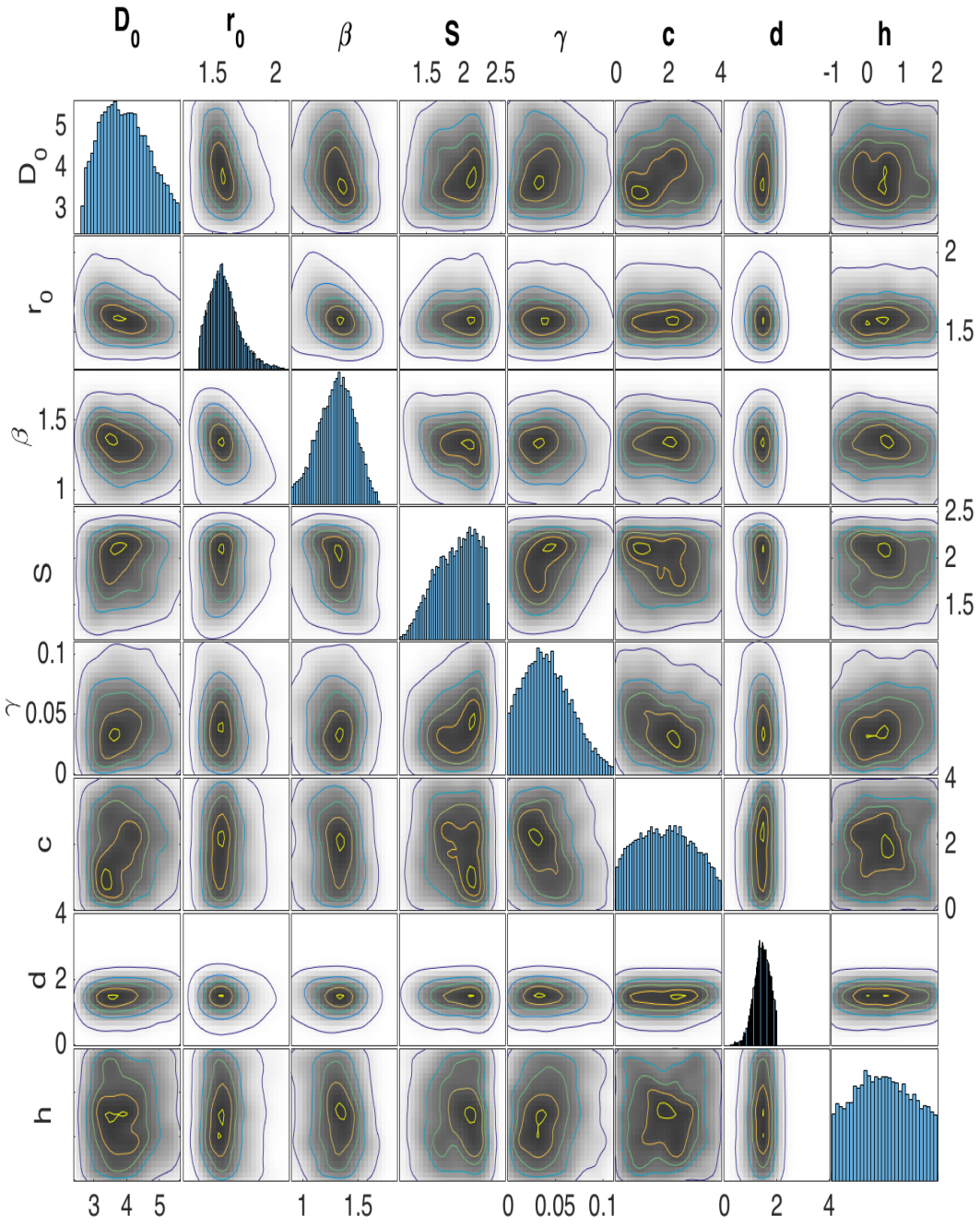

Figure 12: Posterior Densities for Two Body Parameters using the High Energy SA Prior

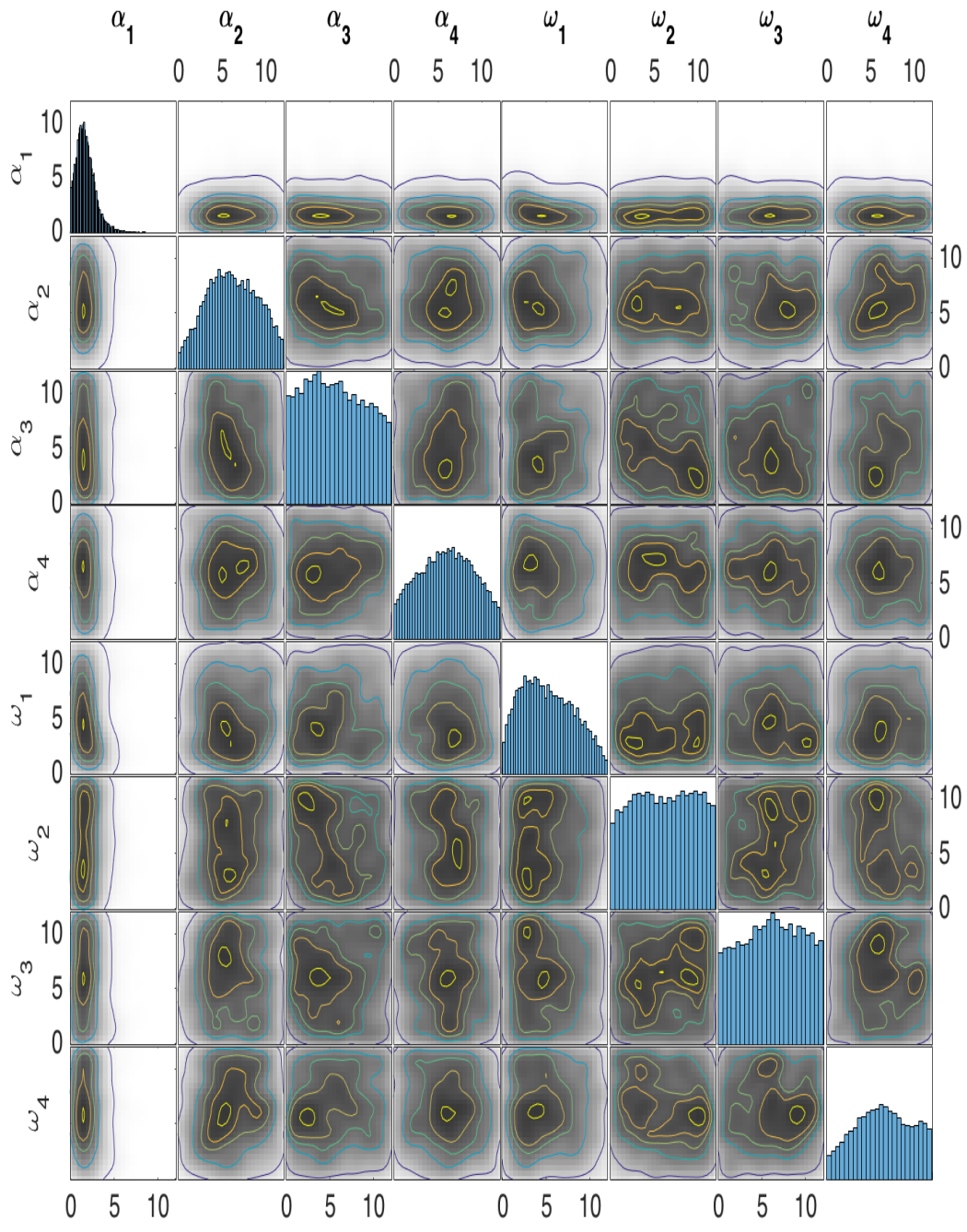

Figure 13: Posterior Densities for Three Body Parameters for the High Energy SA Prior

Figures 8 and 9 show that for the nominal prior, it is not entirely clear that all parameters lie within the initially prescribed bounds. For only r_0, d, and α_{HHW} do the calibrated distributions appear clearly within the assumed bounds, so that values outside these bounds are not implausible. For some parameters, such as h and the α_{WHH}, α_{HWH}, and ω_{WHH}, ω_{HWH} parameters, the posterior distribution strongly resembles the prior. This is an indication that energy is not sensitive, within the stated ranges, to these parameter values (recall that these parameters were not important in the sensitivity analysis, either). In which case, the DFT runs themselves contain little information on the parameters, and the information used in forming the prior is, comparatively speaking, more relevant.

Posterior means and standard deviations for the input parameters are given in Table 2; 5K MCMC samples were obtained for each case. In a true IP analysis, the two extreme priors would be determined upon first choosing a quantity of interest, such as the posterior mean of the energy E, and finding which priors produce the smallest and largest posterior means. Because the calibration runs here are time consuming, the "best" and "worst" priors are taken to be the ones from the sensitivity analysis. As such, true IP ranges of quantities of interest are likely larger than those in Table 2.

Table 2. Calibrated Juslin Parameter Values.

Input	Nominal Posterior		Low SA Posterior		High SA Posterior	
Parameter	Mean	Std.Dev.	Mean	Std.Dev.	Mean	Std.Dev.
D_0 (eV)	3.68	0.72	3.86	0.78	3.96	0.73
r_0 (Å)	1.54	0.15	1.48	0.12	1.59	0.12
β (Å^{-1})	1.45	0.20	1.50	0.12	1.30	0.16
S (λ_1/λ_2)	1.91	0.35	1.99	0.32	1.89	0.27
c	1.41	0.78	1.00	0.54	1.91	1.05
d	1.90	0.56	2.33	0.87	1.44	0.30
h	0.48	0.83	0.53	0.84	0.48	0.82
γ	0.041	0.024	0.036	0.019	0.042	0.025
$\alpha_{HWW} = \alpha_{WWH}$	1.65	0.84	2.42	1.09	1.74	1.09
α_{WHW}	5.10	2.32	2.52	1.36	6.19	2.88
$\alpha_{WHH} = \alpha_{HHW}$	4.89	2.86	2.49	1.43	5.70	3.39
α_{HWH}	5.10	2.81	3.21	1.27	5.98	3.07
$\omega_{HWW} = \omega_{WWH}$	4.11	2.52	2.63	1.35	5.23	2.98
ω_{WHW}	5.14	2.82	2.56	1.45	6.13	3.43
$\omega_{WHH} = \omega_{HHW}$	4.90	2.74	2.47	1.36	6.16	3.38
ω_{HWH}	4.93	2.90	3.17	1.15	6.44	3.14
V^R(r=1.5) (eV)	5.22	2.28	4.24	2.14	6.16	2.31
V^A(r=1.5) (eV)	8.70	2.45	8.02	2.32	9.95	2.44
Total Energy E (eV)	-179	10.9	-192	10.8	-172	11.0

The range of mean energy values in Table 2, from -192 eV to -172 eV, conveys a much better measure of uncertainty in the nominal mean energy -179 eV than does the intrinsic variability in the sample mean from the nominal MCMC simulation − a sample mean that could be computed to arbitrary accuracy by simply increasing the MCMC sample size. Similarly, the range of values for the mean repulsive pair potential, V^R, and the mean attractive pair potential, V^A in Table 2 presents a better measure of uncertainty than the nominal mean V^R and V^A values of 5.22 eV and 8.70 eV.

In terms of input parameters, certain mean values are sensitive to uncertainty in the assumed limits for their ranges; note the factor-of-2.5 range in mean values for ω_{WHH}, for example. For other code inputs, such as the ratio S of LAMMPS parameters λ_1 and λ_2, the relative range of mean values is only 5%. In cases where a particular parameter is of interest, in and of itself, it may be worthwhile to invest additional effort in identifying best and worst case priors than in the LAMMPS example here.

Rare event probabilities, which can be essential in reliability-based work, are often most sensitive to uncertainty in priors. Based on 5K MCMC samples for the three priors in Table 2, the upper 99.8% quantiles of the energy distribution have been obtained. For the nominal prior, the estimated upper quantile is -139 eV; for the Low SA prior, the estimated quantile is -163 eV. In reliability safety applications, where rare event probabilities matter, uncertainties in the prior distribution should almost always be incorporated into bottom-line conclusions.

4. Related Issues

An important problem related to the calibration of inputs is producing predictions of DFT output (recall that DFT is the ground truth here) for new configurations different than those used for calibration. The construction of a fast surrogate for DFT predictions using the calibration procedure and the LAMMPS code could save much time and effort in comparison to running the DFT code. This approach however requires learning a "distance metric" between two general configurations, perhaps using proxies that characterize the configuration. This framework is conceptually similar to the computational methods used in drug discovery (e.g., Sliwoski et al. (2014)), which assign measures of similarity to previous known drugs. A large database of existing drugs against which to compare is usually required to be successful; it is unclear whether a similarly large database (vs. a smaller set) of DFT output from configurations is required.

Calibration approaches using Gaussian processes (e.g. Kennedy and O'Hagan (2001), Higdon et al. (2008)) may provide a good framework to construct this surrogate for DFT predictions by incorporating this distance metric.Another issue to note is the construction of an emulator (i.e., a fast-running surrogate for computer code) for the LAMMPS (or other computer codes) codes when they are too time consuming to use directly for sensitivity analysis or calibration (e.g. Kennedy and O'Hagan (2001), Higdon et al. (2004)). Gaussian processes are often used to build emulators trained on a relatively small set of code runs (e.g. Sacks et al. (1989), Welch et al. (1992)).

An alternative approach for emulator construction and calibration using polynomial chaos expansions (PCEs) is also popular in some quarters (e.g., Ghanem and Spanos (2003), Debusschere et al. (2004)), although the use of PCEs is somewhat controversial in the statistics community (e.g., O'Hagan (2013)).

Because these issues go well beyond the scope of this report, they are not pursued here, although they may arise in the future.

5. Conclusion

The complacent approach to sensitivity analysis and code calibration, based on inputs for a computer code being assigned uniform distributions over prescribed ranges, is not inherently robust. As shown in the examples, uncertainty in the prior distribution itself can propagate to be large or small relative to other uncertainties in a problem. Also, certain quantities of interest, such as rare event probabilities, may be strongly affected by uncertainty in prior distributions, while other quantities in the same problem may be barely affected.

Tools exist to help data analysts detect situations where care is needed. Rare event plots, such as Figures 5 and 6, can reveal whether extreme code output values are achieved near a boundary of the nominal input space. Similarly, posterior contour plots, such as Figures 8-13, can reveal whether parameter values that are consistent with calibration data are near input boundaries. When these circumstances occur, special attention should be

given to the uncertainties in assumed limits for nominal input parameters.

For population-based quantities, such as the mean value of a posterior distribution for an input parameter, an imprecise probability (IP) approach to inference is essential. Population-based quantities of interest can, in principle, be determined with essentially no error by running an MCMC simulation until eternity. The apparent accuracy of resulting quantities is fully dependent on the presumed perfection of the prior distribution in representing prior information. Only when the prior achieves such perfection (which, frankly, never happens in reality) is the minuscule output precision believable. Rectifying this quandary is *not* achieved by the standard Bayesian approach of using hierarchical priors.

In quoting results relative to sensitivity analysis or code calibration, the standard IP approach is to find "extreme" prior distributions (representing best and worst case prior scenarios that are consistent with available prior information). Once such priors are identified, quantities of interest are computed for the two extreme cases, and are presented in conjunction with corresponding quantities based on the nominal prior. In this way, the nominal uncertainties in baseline quantities, which can be greatly understated for large MCMC sample sizes, are not subject to misinterpretation.

References

Debusschere, B. J., Najm, H. N., Pébay, P. P., Knio, O. M., Ghanem, R. G., and Le Maıtre, O. P. (2004). Numerical challenges in the use of polynomial chaos representations for stochastic processes. *SIAM Journal on Scientific Computing*, 26(2):698–719.

Ghanem, R. G. and Spanos, P. D. (2003). *Stochastic finite elements: a spectral approach*. Courier Corporation.

Higdon, D., Gattiker, J., Williams, B., and Rightley, M. (2008). Computer model calibration using high-dimensional output. *Journal of the American Statistical Association*, 103(482):570–583.

Higdon, D., Kennedy, M., Cavendish, J. C., Cafeo, J. A., and Ryne, R. D. (2004). Combining field data and computer simulations for calibration and prediction. *SIAM Journal on Scientific Computing*, 26(2):448–466.

Juslin, N., Erhart, P., Traskelin, P., J, N., Henriksson, K. O., Nordlund, K., Salonen, E., and Albe, K. (2005). Analytical interatomic potential for modeling nonequilibrium processes in the w-c-h system. *Journal of Applied Physics*, 98:123520.

Kennedy, M. and O'Hagan, A. (2001). Bayesian calibration of computer models. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, 63(3):425–464.

McKay, M. D., Conover, W. J., and Beckman, R. J. (1979). A comparison of three methods

of selecting values of input variables in the analysis of output from a computer code. *Technometrics*, 21:239–245.

O'Hagan, A. (2013). Polynomial chaos: A tutorial and critique from a statistician's perspective. *Journal of Uncertainty Quantification*, 20:1–20.

Picard, R. and VanderWiel, S. (2017). Accommodating uncertainty in prior distributions. *Los Alamos National Laboratory Technical Report LA-UR-17-20370*.

Picard, R. and Williams, B. (2013). Rare event estimation for computer models. *American Statistician*, 67:22–32.

Sacks, J., Welch, W. J., Mitchell, T. J., and Wynn, H. P. (1989). Design and analysis of computer experiments (C/R: P423-435). *Statistical Science*, 4:409–423.

Schlick, T. (2010). *Molecular modeling and simulation: an interdisciplinary guide: an interdisciplinary guide*, volume 21. Springer Science & Business Media.

Sliwoski, G., Kothiwale, S., Meiler, J., and Lowe, E. W. (2014). Computational methods in drug discovery. *Pharmacological Reviews*, 66:334–395.

Tersoff, J. (1988). New empirical approach for the structure and energy of covalent systems. *Physical Review B*, 37(12):6991.

Tersoff, J. (1989). Modeling solid-state chemistry: Interatomic potentials for multicomponent systems. *Physical Review B*, 39:5566–5568.

Tuckerman, M. E. and Martyna, G. J. (2000). Understanding modern molecular dynamics: Techniques and applications. *Journal of Physical Chemistry B*, 104:159–178.

Walley, P. (1991). *Statistical Reasoning with Imprecise Probabilities*. Chapman Hall, London.

Welch, W. J., Buck, R. J., Sacks, J., Wynn, H. P., Mitchell, T. J., and Morris, M. D. (1992). Screening, predicting, and computer experiments. *Technometrics*, 34(1):15–25.