



**HAL**  
open science

# Global ranking of the sensitivity of interaction potential contributions within classical molecular dynamics force fields

Wouter Edeling, Maxime Vassaux, Yiming Yang, Shunzhou Wan, Serge Guillas, Peter V Coveney

## ► To cite this version:

Wouter Edeling, Maxime Vassaux, Yiming Yang, Shunzhou Wan, Serge Guillas, et al.. Global ranking of the sensitivity of interaction potential contributions within classical molecular dynamics force fields. npj Computational Materials, 2024, 10 (1), pp.87. 10.1038/s41524-024-01272-z . hal-04598826

**HAL Id: hal-04598826**

**<https://hal.science/hal-04598826>**

Submitted on 3 Jun 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

<https://doi.org/10.1038/s41524-024-01272-z>

# Global ranking of the sensitivity of interaction potential contributions within classical molecular dynamics force fields

Check for updates

Wouter Edeling<sup>1</sup>, Maxime Vassaux<sup>2</sup>, Yiming Yang<sup>3</sup>, Shunzhou Wan<sup>4</sup>, Serge Guillas<sup>3</sup> & Peter V. Coveney<sup>4,5,6</sup>

Uncertainty quantification (UQ) is rapidly becoming a sine qua non for all forms of computational science out of which actionable outcomes are anticipated. Much of the microscopic world of atoms and molecules has remained immune to these developments but due to the fundamental problems of reproducibility and reliability, it is essential that practitioners pay attention to the issues concerned. Here a UQ study is undertaken of classical molecular dynamics with a particular focus on uncertainties in the high-dimensional force-field parameters, which affect key quantities of interest, including material properties and binding free energy predictions in drug discovery and personalized medicine. Using scalable UQ methods based on active subspaces that invoke machine learning and Gaussian processes, the sensitivity of the input parameters is ranked. Our analyses reveal that the prediction uncertainty is dominated by a small number of the hundreds of interaction potential parameters within the force fields employed. This ranking highlights what forms of interaction control the prediction uncertainty and enables systematic improvements to be made in future optimizations of such parameters.

Classical molecular dynamics simulation, originally created in the late 1950s, has become one of the most common computer-based methods used to investigate the behavior of molecular and condensed matter systems whether in the context of physics, chemistry, materials, life or medical research<sup>1,2</sup>. It is routinely used in an attempt to understand the physico-chemical mechanisms underlying not only microscopic properties of such systems but also to explore their macroscopic properties, such as thermodynamic and mechanical behavior and, in combination with quantum mechanical MD, can account for up to 50% of the cycles consumed on large scale supercomputers<sup>3,4</sup>.

Remarkably enough, molecular dynamics is used routinely in many important areas of science and technology without much attention being paid to its reliability and reproducibility. Practically speaking, along with many other microscopic modeling and simulation methods, it is rarely encountered at the sharp end of decision making, which requires actionable outcomes from simulations executed in a timely manner. However, in recent times the field of validation, verification and uncertainty quantification (VVUQ) has come to the fore as a means of assessing the reliability and

reproducibility of computer simulation techniques, with particular focus on areas in which modeling and simulation are applied to assist decision making in safety-critical situations such as those arising in tsunami modeling<sup>5,6</sup>, the design of environmentally friendly energy generation plants<sup>7,8</sup>, engineering and construction projects<sup>9,10</sup> as well as disaster evasion<sup>11,12</sup>.

Such (macroscopic) studies generally involve the application of VVUQ methods to systems of (nonlinear) partial differential equations, or (graph-based) discrete structures. Many microscopic descriptions of matter have a different algorithmic structure. Relatively speaking, these algorithms have been exposed to less uncertainty quantification. They are particulate in nature so the treatment of uncertainty and reliability is somewhat different. Nonetheless, the general principles of VVUQ remain the same. In this article, we shall be focusing on uncertainty quantification (UQ), assessing the source and quantifying the size of errors originating from the equations used to describe the behavior of matter. These equations are, of course, very well known - they are Newton's equations of motion, applied to the detailed molecular motion of assemblies of atoms within molecules and, in condensed phases, large assemblies of such molecules in the solid or liquid state.

<sup>1</sup>Scientific Computing Group, Centrum Wiskunde & Informatica, Amsterdam, The Netherlands. <sup>2</sup>Université de Rennes, CNRS, Institut de Physique de Rennes, UMR 6251, Rennes 35000, France. <sup>3</sup>Department of Statistical Science, University College London, London, United Kingdom. <sup>4</sup>Centre for Computational Science, University College London, London, United Kingdom. <sup>5</sup>Advanced Research Computing Centre, University College London, London, United Kingdom. <sup>6</sup>Informatics Institute, University of Amsterdam, Amsterdam, The Netherlands. ✉e-mail: [p.v.coveney@ucl.ac.uk](mailto:p.v.coveney@ucl.ac.uk)

In UQ one often starts by first dividing the sources of uncertainty into aleatoric and epistemic components. In forward parametric UQ, the latter are concerned with how uncertainties in the parameters used to set up, control and run a simulation impact the quantities of interest (QoIs) being computed. These parameters are epistemic, as with more knowledge and/or data it is possible to obtain better estimates of their values. In a classical molecular dynamics simulation, the dynamical behavior is determined by the forces which each atom exerts on the others, as these forces directly enter Newton's equations of motion, the solution of which provides a description of the trajectories of all the particles, or, more technically, of a single point moving in the  $(6N+1)$ -dimensional phase space comprising all the positions and velocities of the  $N$  particles in the simulation and the time  $t$ . The sum of all the forces is known as the force field, and it is obtained from the interaction potentials that describe these atomistic interactions. As is well known, there exists a range of force fields (or interaction potential parameterizations) which are known to do a generally reasonable job of describing many different systems, from condensed matter physics, chemistry and materials to biological and biomedical cases of concern. While we have studied the effect of 14 uncertain simulation parameters (epistemic parameters not related to the force field, e.g. thermodynamic inputs associated with the temperature or pressure) in a previous study<sup>13</sup>, there are orders of magnitude more force-field parameters to contend with in most all-atom MD simulations. Some earlier attempts have been made to investigate parametric uncertainty and its propagation in MD simulations for limited force field parameters in simple molecular systems, such as for three force-field parameters in TIP4P water<sup>14</sup> and for two Lennard-Jones energy parameters in an aqueous NaCl solution<sup>15</sup>. In more complicated molecular systems, there are hundreds of force field parameters as a consequence of the wide range of different atoms present in real-world situations (see Section "Selection of models and interaction potential parameters"). It is the study of uncertainty in these parameters which is the focal point of this paper.

In addition, we also deal with sources of aleatoric uncertainty. By this, we mean the inherent uncertainty that cannot be reduced with increased knowledge, which in our case arises from the presence of random number generators or 'seeds' in the MD codes. Molecular dynamics provides an example *par excellence* of chaos, by which we mean that the trajectories manifest extreme sensitivity to the initial conditions. Remarkably, this is a facet of the method which is seldom discussed in any MD publication, and yet the sheer fact of its existence immediately points to a problem in terms of its reproducibility - it is essentially impossible to reproduce the result of any molecular dynamics simulation. That the dynamics is chaotic is also deeply connected to the existence of thermodynamic equilibrium states: in the hierarchy of ergodic theory, a necessary and sufficient condition for a dynamical system to exhibit an approach to equilibrium is that the dynamics is mixing. Mixing in turn implies chaos. Moreover, mixing is a stronger property than ergodicity itself: mixing implies ergodicity but is not implied by it<sup>16</sup>.

In order to get control of such aleatoric uncertainty, one must perform ensembles of simulations, by which we mean that a large number of 'replicas' (model simulations at the same epistemic parameters and a different random seed) are executed concurrently on a large enough computer, and statistical averages from all the simulations are calculated. In this way we obtain results which are statistically robust, reproducible and can be treated as scientifically meaningful. For those who regard a single MD simulation as computationally expensive, the requirement to perform ensembles of such simulations comes as a bitter pill. Although it is now dawning on many that they must run more than one simulation, they prefer to think that they can run three and that will suffice. This is based on the experimentalists' oft-cited protocol which is sufficient to produce a mean and a standard deviation to report on a measurement. By now, it is evident that this minimal size of ensemble, often referred to as a set of 'repeats', will not cut the mustard. Even to assess the true nature of a normal distribution, one often needs many more such measurements; but - and much to the surprise of many - the distribution of quantities of interest is frequently non-normal, possessing

many more than only two moments and leading to unexpected behavior of the QoIs<sup>13,17</sup>.

Existing forms of UQ often suffer from the curse of dimensionality (e.g. refs. 18,19), which is to say that in order to investigate how the uncertainty in the input parameters affects the output QoIs one must perform a number of simulations which increases exponentially with  $D$ , the number of uncertain parameters under investigation. That said, such methods as Polynomial Chaos (PC) expansions can show very rapid convergence at low parameter counts and have been applied in the context of MD<sup>14</sup>. PC expansions can also subsequently be used to accelerate Bayesian calibration of (low-dimensional) force-field parameters, see e.g. ref. 20. For any non-trivial computational model, such as MD with  $D > 10$ , this quickly becomes impractical, and the expense is rendered even greater as in many cases one must perform ensemble averaging over random seeds before one can assess the uncertainty in the epistemic parameters. New UQ methods capable of handling high-dimensional parameter spaces are required to address such situations.

These high-dimensional UQ methods bank on the existence of a low-dimensional 'effective dimension'<sup>21</sup>, where a (hopefully small) subset of the input parameters is responsible for most of the variance in the model output. The effective dimensionality of a model is closely related to the concept of 'sloppiness'; see ref. 22 for a recent review. Sloppy models are characterized by an insensitivity to changes in parameter values in certain directions of the parameter space. This poses challenges for (inverse) UQ problems, as sloppy combinations of parameters are not informed by data, yet they do increase the dimension of the UQ problem, thereby slowing down convergence. Interaction potentials have also been shown to be sloppy, see e.g. the recent work of Kurniawan et al.<sup>23</sup>. To obtain a local measure of sloppiness, an eigenvalue decomposition of the Fisher Information Matrix (FIM) evaluated at the best-fit parameters is performed<sup>23,24</sup>. Sloppiness is said to occur if these eigenvalues span orders of magnitude, and have only a few large and many small values. The corresponding eigenvectors identify the stiff/sloppy directions in the parameter space respectively, where stiff directions are those along which the model can be informed by data.

A downside of the aforementioned approach is that the FIM is a local quadratic approximation of the cost surface (the metric that quantifies how well the parameters fit the available data, i.e. a loss function), valid in the neighborhood of the best-fit parameters. Instead, we will use global, ensemble-based methods to identify the effective dimensionality, which require a sampling algorithm to draw new parameter values. One could employ an adaptive strategy where the most important parameters are found via an iterative sampling algorithm<sup>13,25,26</sup>, refining only the dimensions of influential inputs. Instead, the authors of refs. 27,28 introduced a coordinate transformation (applied in the context of (Gaussian) PC expansions), that looked at linear combinations of all parameters instead of individual inputs. Projection operators were developed that project to a very low-dimensional structure around which the probabilistic content of the (scalar) model output is concentrated. Similarly, active-subspace methods<sup>29</sup> also look for low-dimensional linear combinations of all parameters, along which the model varies the most on average. These directions of the strongest variability, i.e. the active subspace, form a new low-dimensional coordinate system which is rotated from the original coordinate system of the individual inputs. Note that the active-subspace directions are the equivalent of the stiff directions in the sloppy model literature, although the active subspace is determined from an eigenvalue decomposition of a global matrix constructed from the gradient of the model with respect to its input parameters. By looking at linear combinations of inputs rather than individual input parameters, active-subspace methods can lead to a much more significant reduction in the input dimension compared to adaptive sampling techniques. In practice, it is not uncommon to find a one-dimensional active subspace; see for instance references<sup>30-35</sup>. Once such an active subspace has been identified, a low-dimensional surrogate model (an approximation which can be evaluated rapidly) of the high-dimensional simulation code is constructed in this subspace.

The original version of the active subspace method requires evaluations of gradients of the model; here we would need to differentiate the MD code

**Table 1 | Data set summary per application**

Application	input types	$D$	$n_{data}$	$n_{replica}$	output
Epoxy	FF only	103	500	20	Young's modulus, Poisson ratio
ESMACS	FF only	153	428	25	binding free energy
	FF + SIM	167	308	25	binding free energy
TIES	FF only	157	273	5	relative binding free energy
	FF + SIM	169	215	5	relative binding free energy

Here, the input type (force-field (FF) only and force-field & simulation (SIM) parameters), number of inputs ( $D$ ), number of data points, number of replica simulations and output type are described.

predictions with respect to simulation and force-field parameters, which is practically infeasible. For these circumstances several derivative-free active-subspace methods have been developed. One such method has been applied to a simple MD model<sup>34</sup>, using a local linear interpolant to estimate the derivatives. Unlike the study we report here, however, the parameter space considered was low-dimensional (7) and the simulation setup was much simpler, involving a single atom type and crystalline order of the atoms, avoiding the need for ensemble averaging. The study was also limited to a single macroscopic quantity of interest.

Instead, we will use a recently developed neural network based approach which is scalable in  $D$ , called the Deep Active Subspace (DAS) method<sup>36,37</sup>, which allows us to probe the uncertainty of at least hundreds of parameters. In addition, we compare our findings with a derivative-free kernel-based method which uses a Gaussian Process (GP) as a surrogate model<sup>38</sup>. The latter approach is denoted as the Kernel-based Active Subspace GP (KAS-GP) method in the following discussion. We describe our methods here and apply them to perform uncertainty quantification and sensitivity analysis for several different real-world molecular dynamics systems comprising:

- (i) Epoxy-resin thermosetting polymer materials, predicting mechanical properties, namely  $E$ , the Young's modulus, an indicator of the stiffness of the material, and the Poisson ratio, which is the ratio of lateral to axial deformation when straining the material in a given direction.
- (ii) Protein-ligand biomolecular systems, where binding free energies are computed. Two types of binding free energies are investigated: the absolute binding free energy (shortened as 'binding free energy' hereafter) which is a quantitative measure of the strength of protein-ligand binding, and the relative binding free energy which is the difference of binding free energies between two ligands. Reliable prediction of such properties plays an important role in drug discovery and personalized medicine.

Within this study, we have used two different molecular dynamics engines, LAMMPS<sup>39</sup> for epoxy-resin polymer materials and NAMD<sup>40</sup> for protein-ligand biomolecular systems. The absolute and relative binding free energies are calculated using the ESMACS (enhanced sampling of molecular dynamics with approximation of continuum solvent)<sup>41</sup> and TIES (thermodynamic integration with enhanced sampling)<sup>42</sup> protocols, respectively.

The objective of the present paper is thus to perform a high-dimensional active-subspace based uncertainty quantification analysis to assess the influence of aleatoric, simulation and particularly force-field parameter uncertainty on classical molecular dynamics predictions. Although in each case we vary more than 100 simulation and force-field parameters, we find evidence of a one-dimensional active subspace in all of these MD applications. Beside identifying active subspaces, we also extract the sensitivity of the QoIs to individual input parameters. Here the existence of a low effective dimension is not only manifest but striking, as our MD predictions typically show a sensitivity to less than 10 individual inputs. This shows for instance that for the protein-ligand biomolecular systems, the uncertainties in non-bonded van der Waals parameters control the uncertainty in the binding free energy predictions.

## Results

### Selection of models and interaction potential parameters

In the context of classical MD simulation, a force field is used to describe the interactions between atoms, which is a collection of equations and associated constants designed to reproduce selected properties of a given molecular system<sup>1</sup>. The Amber<sup>43</sup> and OPLS<sup>44</sup> force fields we are using, as with many other contemporary force fields in widespread use, consist of two parts: the bonded terms and the non-bonded terms. The former describes the interactions of bonds, angles, torsions and improper torsions, while the latter comprises electrostatic interactions and van der Waals (vdW) interactions. Atom types are used to assign the force-field parameters for both the bonded and nonbonded types. For the electrostatic interactions, the partial charges for the atoms are also required, which are typically calculated using quantum or (semi-)empirical approaches<sup>45</sup>. Although electrostatic interactions are crucial in MD simulations, we do not investigate their sensitivity here because the partial charges cannot be varied independently (the total net charge needs to be consistent and integer-valued in units of the electronic charge). There are still 836 force-field parameters in the ESMACS model and a few more in the TIES model. We further reduce the number of parameters by only including parameters which describe the interactions within the selected ligand or ligand pair and its vdW interactions with its environment. This includes all parameters for bonds and angles, force constants for dihedrals within the ligand and the vdW parameters for all atom types. This leads to a total number of (force-field) parameters under study here between 100 and 200; see Table 1 for the exact number per application. The main reason for this a priori reduction is to investigate if a low-dimensional active subspace exists when replacing 100–200 inputs with random variables.

### Selection of input distributions and generation of training data

For all input parameters  $x_i$  we prescribe independent uniform input distributions such that the joint probability density function (pdf) is  $p(\mathbf{x}) = \prod_{i=1}^D p(x_i)$ , with  $x_i \sim \mathcal{U}(0.85\bar{x}_i, 1.15\bar{x}_i)$ , where  $\bar{x}_i$  are the default values. The only exception is the temperature simulation parameter (`setTemperature`), for which  $-15\%$  from a default value of 300K would yield a freezing temperature. Instead,  $\pm 7.5\%$  is used. Random samples from  $p(\mathbf{x})$  are drawn via simple Monte Carlo sampling; see also Section "Statistics". To find all parameter names and their default values we refer the reader to Supplementary Tables 1–6.

Note that the choice of the pdf will surely affect the observed uncertainty in the MD simulations. The uniform distributions represent our lack of prior knowledge on the most-likely parameter values. Furthermore, we do not argue that the  $\pm 15\%$  range is optimal, although the generated output uncertainty (see Section "Epistemic vs aleatoric uncertainty") is large enough to expect it to envelop the true value of the outputs. Overall, the results we present should thus be interpreted as a sensitivity study conditional on our choice of  $p(\mathbf{x})$ . Obtaining more realistic, data-informed, posterior distributions is the domain of (Bayesian) inference; see e.g. refs. 23,46 for examples in MD.

For the epoxy application we have one training data set with  $n_{data} = 500$  randomly drawn input-output samples, all inputs being force-field parameters. In the case of the ESMACS and TIES applications we generated two separate data sets, namely one in which only the force-field parameters are



varied with the simulation parameters being fixed at their default value and one in which both are varied, allowing us to evaluate the relative importance of each parameter class. In fact, our sensitivity analysis (Section “Global first-order sensitivity analysis”) shows that the force-field parameters completely dominate the simulation parameters. Unless otherwise stated, we therefore combine both data sets in the ESMACS and TIES application to increase the training data size.

Finally, to assess the aleatoric uncertainty of the random seed involved in the initial condition,  $n_{\text{replica}}$  replica simulations were executed for each random input value, similar to ref. 13. All data sets are summarized in Table 1.

### Epistemic vs aleatoric uncertainty

Since we have MD output data over both the epistemic (force-field & simulation) parameters and the random seed of the replicas, we can estimate marginal distributions in order to analyze the respective contribution of epistemic and aleatoric uncertainty. The results for the epoxy-resin, the ESMACS and the TIES applications are given in Fig. 1. The marginal distributions are created by averaging over either the replicas or the parameters with bootstrapping. For all applications, given the input distributions described in Section “Selection of input distributions and generation of training data”, the epistemic uncertainty dominates over the aleatoric uncertainty introduced by the random seeds.

We also display the mean, standard deviation, skewness and kurtosis computed over all replica and parameter samples in Table 2. This analysis reveals notable characteristics in terms of skewness and kurtosis. Skewness, a measure of the asymmetry of a distribution, indicates a departure from perfect symmetry in the calculated data. Excess kurtosis indicates the presence of heavier tails or outliers compared to a normal distribution. Higher kurtosis values suggest the presence of more extreme values or outliers in the data, indicating a greater concentration of observations away from the mean. A common rule of thumb for the presence of significant skewness are (absolute) values larger than 1, and values larger than 3 for the excess kurtosis. With the exception of the ESMACS case, we find significant skewness and/or kurtosis values, see Table 2. Both the skewness and kurtosis estimations fall within the reported bootstrap 95% confidence intervals, which demonstrates their robustness against random sampling error and confirms the non-normality of the relevant QoIs. Ensemble methods enable a deeper understanding of the system dynamics and uncover insights that are not apparent through traditional one-off simulations<sup>47</sup>, enhancing the reliability and validity of subsequent analyses and interpretations, thus leading to more informed decision-making.

### Dimension reduction

In the original active subspace method<sup>29</sup>, the eigenvalue spectrum of an (uncentered) covariance matrix of the gradient of the MD output  $f(\mathbf{x})$  is examined, i.e.  $C = \int (\partial f / \partial \mathbf{x})(\partial f / \partial \mathbf{x})^T p(\mathbf{x}) d\mathbf{x}$ . A large eigengap between consecutive (ordered) eigenvalues (defined as  $\lambda_d - \lambda_{d+1} \geq 0$ ) is evidence of the existence of a low-dimensional active subspace, provided that  $d \ll D$ . Denote  $U_1 \in \mathbb{R}^{D \times d}$  to be the  $d$  dominant eigenvectors of  $C$ , which span the identified rotated active-subspace coordinate system along which  $f(\mathbf{x})$  varies most on average. The high-dimensional input vector  $\mathbf{x}$  is then linearly projected onto the  $d$ -dimensional active subspace as  $\mathbf{y} = U_1^T \mathbf{x} \in \mathbb{R}^d$ .

The active-subspace inspired methods we employ, that is DAS and KAS-GP, each use different means to approximate  $U_1$  with standard supervised input-output data  $(\mathbf{x}_k, f(\mathbf{x}_k))$ ,  $\mathbf{x} \sim p(\mathbf{x})$ ,  $k = 1, \dots, n_{\text{data}}$ , i.e. without requiring the MD gradient  $\partial f / \partial \mathbf{x}$ . More information is presented in the Supplementary Methods. Once  $U_1$  is approximated, the DAS method creates a surrogate of the MD output as a function of  $\mathbf{y}$  by means of a feed-forward neural network, while the KAS-GP method utilizes a GP for this purpose.

In Fig. 2 we show the largest 4 eigenvalues  $\lambda_1 \geq \lambda_2 \geq \lambda_3 \geq \lambda_4$  of the  $C$  matrix of both DAS and KAS-GP  $C$  matrix, for all three applications.

Especially in the DAS results, we observe an eigengap  $\lambda_1 - \lambda_2$  of one order of magnitude, which is evidence of the existence of a 1D active subspace, although not as strong as in some other (non-MD) models, see e.g. ref. 31.

Hence we set the dimension of the active subspace to  $d = 1$  for all applications. All subsequent results are from a DAS or KAS-GP surrogate model trained with data averaged over the random seeds of the MD initial condition, unless otherwise specified. Besides  $d$ , there are other hyper-parameters that must be suitably chosen, e.g. the number of neurons per hidden layer of the DAS network, and the kernel length scale in the KAS method. We have performed a hyper-parameter grid search for both approaches to select optimal values, the results of which are shown in the Supplementary Methods. This includes error analysis on unseen test data, which shows TIES to be the most challenging application.

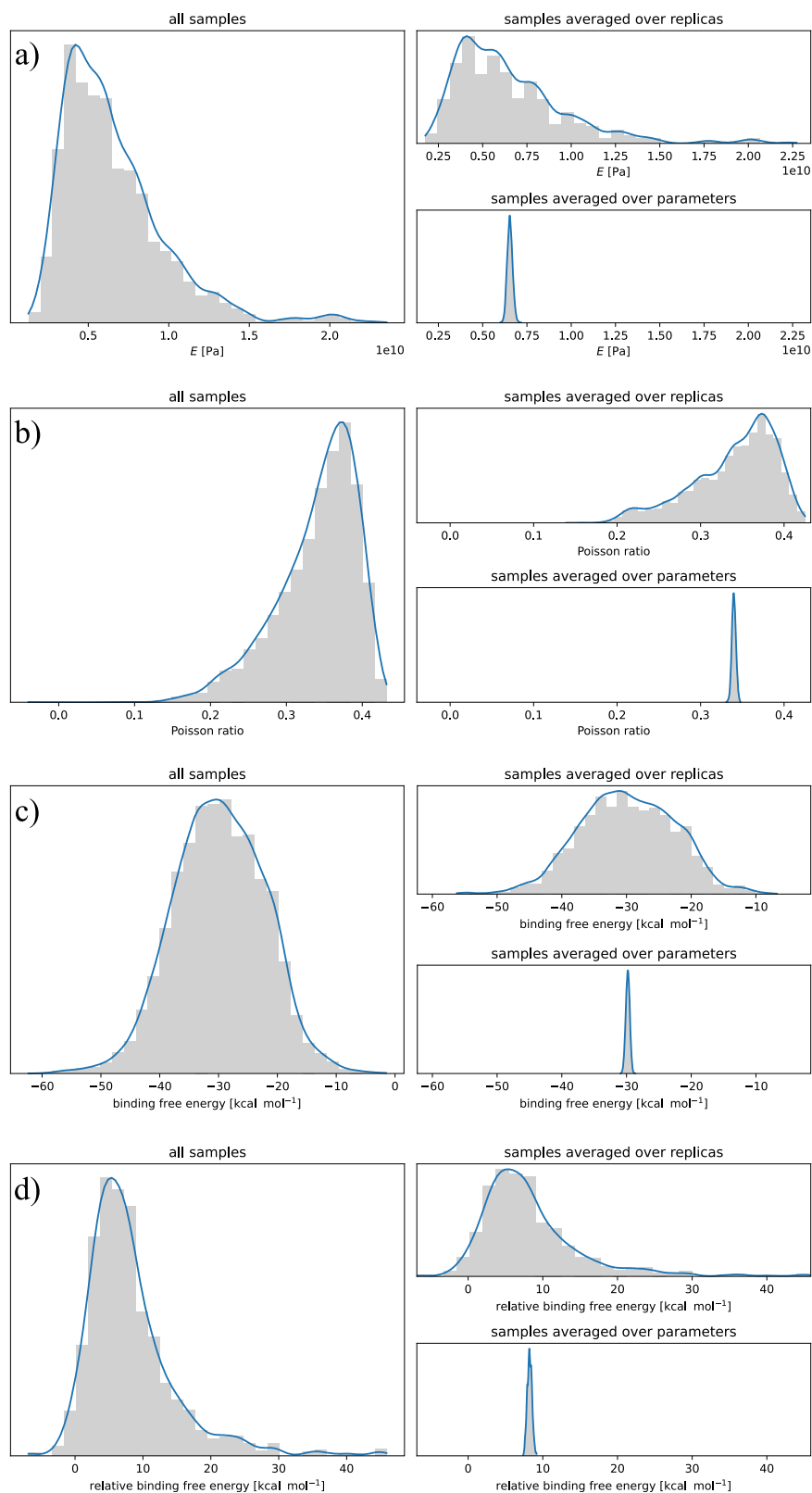
### Active subspace approximation

The epoxy DAS and KAS-GP surrogates for the Young’s modulus  $E$  and the Poisson ratio are plotted vs the  $d = 1$  dimensional active subspace  $y_1$  in Fig. 3a, b. Note that the 1D function captures the overall trend of the data well, for both QoIs. Importantly, the variation of the MD data  $f(\mathbf{x})$  at a given location in the active subspace, i.e.  $\mathbb{V}\text{ar}[f(\mathbf{x})|y_1]$ , is heavily concentrated around the prediction, especially for the DAS surrogate. This holds for both the training data and the test data (10% of the data set), which was not used in constructing the surrogates. Hence, while the original MD model is a function of a 103-dimensional input space (see Table 1), it is well approximated by a 1D surrogate. The 1D ESMACS binding-energy surrogates are shown in Fig. 3c. A 1D active subspace is clearly visible, although the variance of the training and test data around the surrogate is larger than for the epoxy surrogates. Note that here the trend of the DAS surrogate is reversed from the trend of its KAS-GP counterpart. This is due to use of stochastic gradient descent in the DAS training procedure. Since eigenvectors are defined up to multiplication with  $-1$ , it is possible that the learned active subspace flips upon retraining. This is not a serious issue: the main criterion for success is that the eigenvector spans the active subspace correctly. The TIES 1D relative binding free energy surrogates are shown in Fig. 3d. The 1D active subspace is again visible, although (like the ESMACS case) it is not of the same quality as for the epoxy surrogates.

Thus far all surrogates were trained with data that was averaged over the replica MD simulations. We can contrast this with surrogates that were trained on data without replicas, i.e. of a single random seed. In many cases we obtained similar performance in the sense that  $\mathbb{V}\text{ar}[f(\mathbf{x})|y_1]$  remained similar or decreased only slightly when averaging out the random seed. An exception is shown in Fig. 4, which compares the DAS Poisson ratio surrogate with and without replica-averaged training data. Here  $\mathbb{V}\text{ar}[f(\mathbf{x})|y_1]$  is visibly smaller when  $f(\mathbf{x})$  is averaged over the replicas. Hence, even though marginal distributions of Fig. 1 show that the random seed is relatively unimportant overall, this does not mean the aleatoric uncertainty can be safely ignored. While the seed accounts only for a marginal fraction of the total output variance  $\mathbb{V}\text{ar}[f(\mathbf{x})]$ , its influence on the conditional variance in the active subspace  $\mathbb{V}\text{ar}[f(\mathbf{x})|y_1]$  can be more pronounced, as shown in Fig. 4. This is important as it directly influences the quality of the surrogate that can be constructed in the (1D) active subspace. Moreover, the relative importance of the seed is of course wholly dependent upon the type of epistemic parameters involved, as well as the distributions imposed upon them. For instance, in ref. 13 we employed a dimension-adaptive UQ technique, varying the seed and the simulation parameters while keeping the force-field fixed. Under such circumstances the aleatoric uncertainty significantly impacts the overall variance  $\mathbb{V}\text{ar}[f(\mathbf{x})]$ ; see also refs. 16,48 for other examples where the aleatoric uncertainty is comparatively significant.

When comparing the DAS surrogates with the KAS-GP surrogates in Fig. 3, note that the variance  $\mathbb{V}\text{ar}[f(\mathbf{x})|y_1]$  of the training and test data around the DAS prediction is smaller than the corresponding variance around the GP mean. While overall the DAS and KAS-GP results are similar, this does show that the DAS method found a more pronounced active subspace, which is consistent with the eigenvalue results of Section

**Fig. 1 | MD output distributions.** Left column: the probability density functions of all output samples (from both epistemic parameters and the aleatoric random seed replicas). Right column: the marginal epistemic parameter distribution averaged over the replicas, and the marginal aleatoric replica distribution averaged over the epistemic parameters. Results are shown for **a** epoxy resin application,  $E$  output, **b** epoxy resin application, Poisson ratio output, **c** ESMACS binding free energy, **d** TIES relative binding free energy.



“Dimension reduction”, where the DAS method was able to find a larger eigengap. The underperformance of KAS-GP can be attributed to the selected kernel functions. The chosen kernels (see Supplementary Methods) should match the regularity of the active subspace which is challenging in the absence of prior knowledge. Conversely, in contrast to DAS, the KAS-GP intrinsically provides an uncertainty for the surrogate output which is

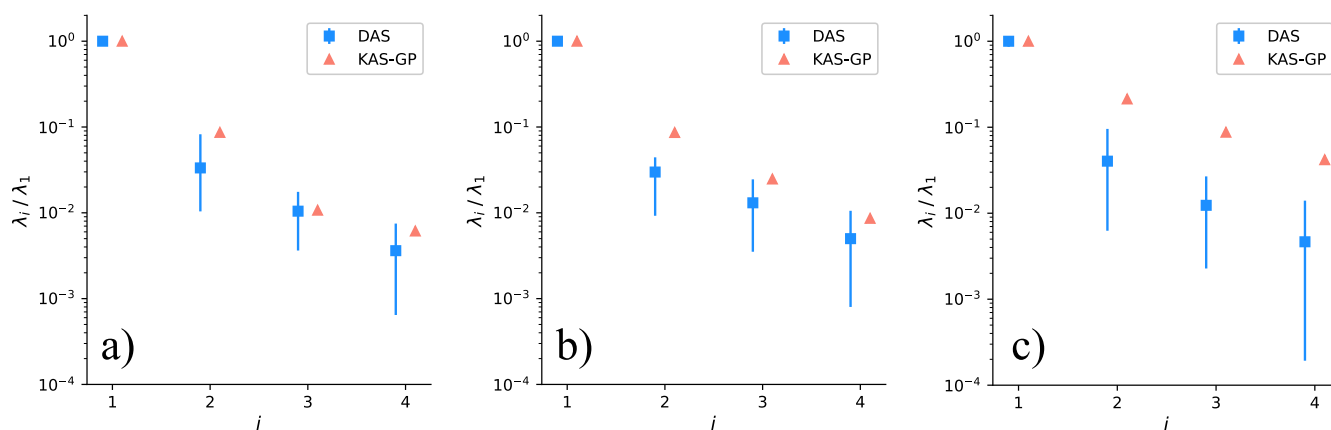
the standard deviation or confidence interval of the model response, a direct consequence of the GP surrogate model.

Finally, note that so far all our QoI haven been scalar. In fact, the original active subspace method is limited to scalar outputs due to the construction of  $C = \int (\partial f / \partial \mathbf{x})(\partial f / \partial \mathbf{x})^T p(\mathbf{x}) d\mathbf{x} \in \mathbb{R}^{D \times D}$  with  $f \in \mathbb{R}$ , although it can be applied pointwise on vector-valued outputs. Adaptations

**Table 2 | The first four sample moments**

	Output	Mean	Std dev	Skew	Kurt
Epoxy	$E$	6.54 (6.49/6.59) $\times 1e9$	3.22 (3.16/3.29) $\times 1e9$	1.61 (1.541/1.666)	3.76 (3.46/4.05)
	Poisson ratio	0.34 (0.339/0.341)	0.053 (0.0520/0.0535)	-1.01 (-1.076/-0.956)	0.97 (0.63/1.42)
ESMACS	Bind fr energy	-29.85 (-29.95/-29.76)	7.68 (7.61/7.74)	-0.08 (-0.117/-0.050)	0.07 (0.003/0.14)
TIES	Rel bind fr energy	8.23 (8.00/8.45)	6.70 (6.38/7.01)	1.95 (1.77/2.11)	5.94 (4.91/6.87)

Includes 95% bootstrap confidence intervals in brackets, for all applications and outputs. The statistics were computed using all parameter and replica samples, giving a sample size of 10,000/18,400/2240 for the epoxy/ESMACS/TIES application respectively. We use the excess kurtosis definition, which yields zero for a normal distribution.



**Fig. 2 | The first 4 eigenvalues  $\lambda_i/\lambda_1$  of  $\bar{C}_{DAS}$  and  $\bar{C}_{KAS}$ .** These are the DAS and KAS-GP equivalents of the gradient  $C$  matrix Eq. (1), Section “Methods”. They are plotted off-axis from the  $i$  indices to avoid overlap of squares and triangles. In the case of the DAS method, we generate 95% confidence intervals on the eigenvalues by retraining the neural network 20 times, each time recomputing the eigenvalues. The variation is

caused by the use of stochastic gradient descent in training the DAS neural network. The squares indicate the sample means over the replica DAS networks. Results are shown for **a** epoxy resin application,  $E$  output, **b** ESMACS binding free energy, **c** TIES relative binding free energy.

to the active-subspace method that can handle vector-valued functions have appeared only relatively recently, see e.g. ref. 49. The DAS method can also handle vector-valued outputs by just increasing the number of output neurons<sup>36</sup>. Whether the network converges to a meaningful active subspace or not depends on the degree to which the different outputs share the same active subspace, which could be unrealistic. To investigate, we have performed an additional study for the materials case, where we build a single DAS surrogate for the Young’s modulus, Poisson ratio, bulk and shear modulus simultaneously. The results (included in Supplementary Section 2) show that in this case we can approximate all 4 QoIs in the same 1D active subspace.

### Global first-order sensitivity analysis

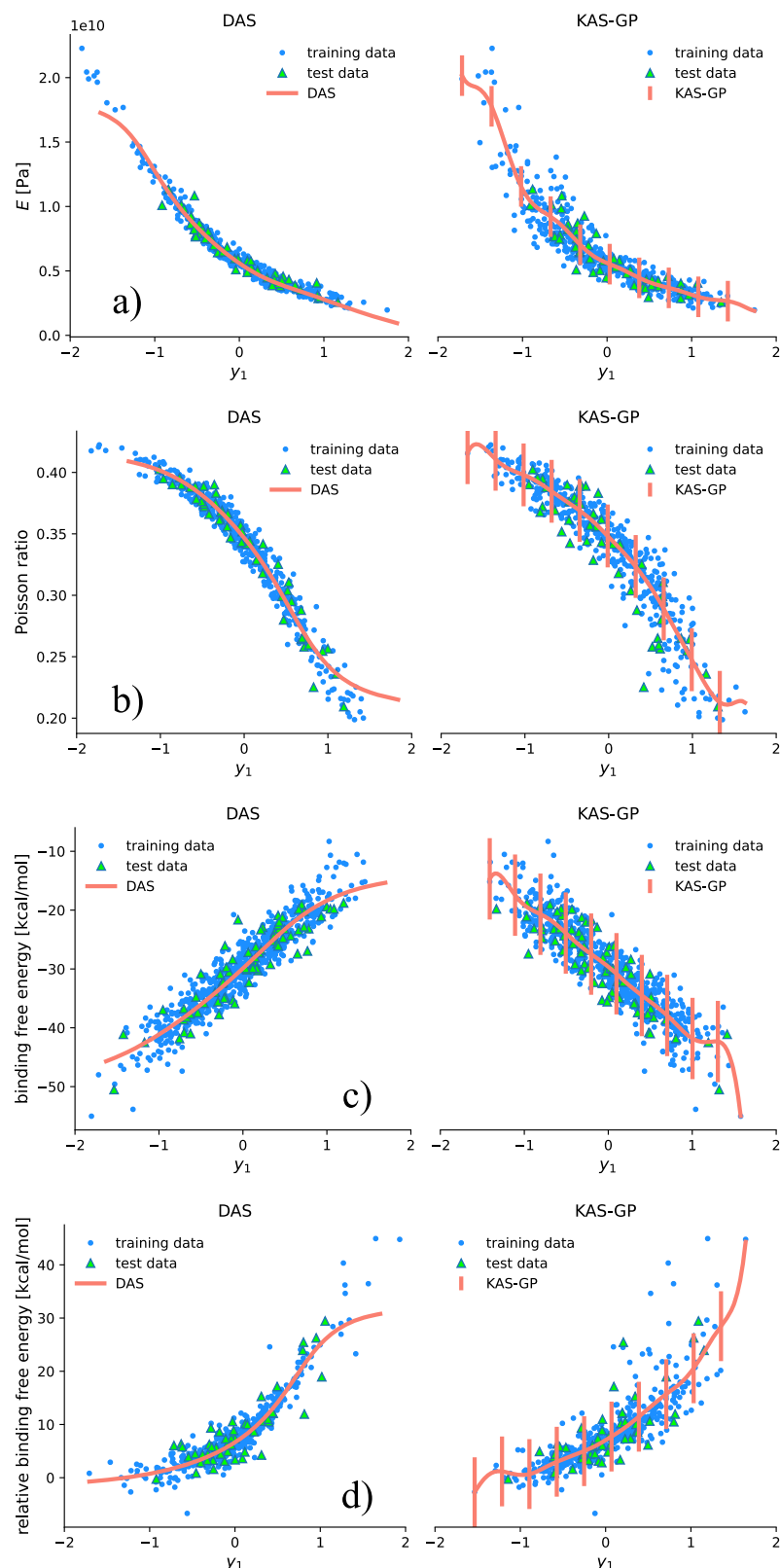
We use global derivative-based sensitivity indices  $v_i$  to measure the impact of individual input parameters,  $i = 1, \dots, D$ , see (5) of Section “Global derivative-based sensitivity analysis”. Figure 5a, b displays the sorted 25 largest sensitivity indices  $v_i$  of the epoxy  $E$  and Poisson ratio surrogates. Roughly the same subset of important parameters emerge for both QoIs, as 8 out of the  $E$  top 10 also appear in the top 10 of the Poisson ratio. The order within the top 10 most sensitive parameters differs between the two outputs. Figure 5c, d shows the ESMACS and TIES sensitivity indices for the (relative) binding free energies. To interpret the parameter naming convention we refer to the Supplementary Methods. Overall, while there are small differences in the ranking put forward by the DAS and KAS-GP methods, both methods flag the same input parameters as important for all applications.

For the materials application, bond (b) and pairwise (p) interactions (see Supplementary Table 1) for a limited number of atom types and bond types dominate. Bond interaction b12 is ideal to control the Poisson

coefficient without modifying the Young’s modulus. The parameters b12 and b42 refer to the equilibrium distance of C-H and C-C atom types. Conversely, b42 offers independent control of the Young’s modulus. Pairwise interactions p12, p42 and p62 do affect both QoIs significantly. The parameters p12 and p42 control the minimum potential energy radii of two different hydrogen atom types, while p62 controls this same potential energy for some carbon atoms. Comparatively to bond and pairwise interactions, the angle (a) and dihedral (d) parameters have much less influence on the elastic mechanical properties of the pristine epoxy resin. Interestingly, at these low strain amplitudes, the Young’s modulus of the material is essentially controlled by non-bonded hydrogen atom interactions, while the Poisson ratio results from a more complex combination of bonded and non-bonded interactions involving carbon and hydrogen atoms.

To identify the sources of the dominant ESMACS contributions (Fig. 5c), we demonstrate the positions of the atoms to which the most sensitive parameters are assigned. Most of the important ESMACS parameters in Fig. 5c are pNNr<sub>m</sub> and pNNev, representing the pairwise equilibrium internuclear distance and well depth of vdW interactions for atoms with index NN (see Supplementary Table 4). It is reasonable to observe that the non-covalent binding free energies are sensitive to changes in the non-bonded parameters. The first parameter (p03r<sub>m</sub>) is the equilibrium internuclear distance of vdW interactions for most sp<sup>3</sup> hybridized carbon atoms (which have one 2s-orbital and three 2p-orbitals to create four hybrid orbitals), while the second and fourth (p08r<sub>m</sub> and p08ep) are the equilibrium internuclear distance associated with the well depth of the vdW interactions for hydrogen atoms attached to these sp<sup>3</sup> carbon atoms. These parameters are the most sensitive ones as (i) many atoms close to the ligand are carbon and hydrogen atoms with these parameters (Fig. 6), and (ii) most

**Fig. 3 | The DAS and KAS-GP surrogates of all QoI plotted along the first active variable  $y_1$ .** Note that the KAS-GP surrogates includes 95% confidence intervals. Results are shown for **a** epoxy resin application,  $E$  output, **b** epoxy resin application, Poisson ratio output, **c** ESMACS binding free energy, **d** TIES relative binding free energy.



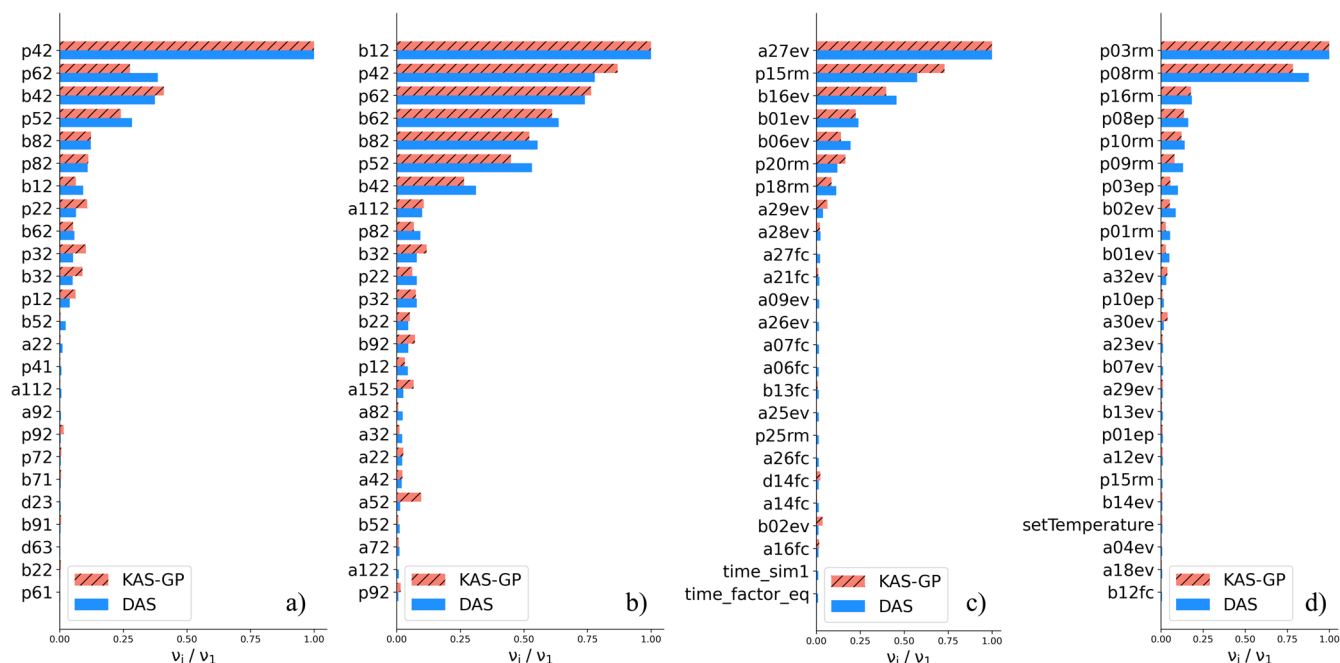
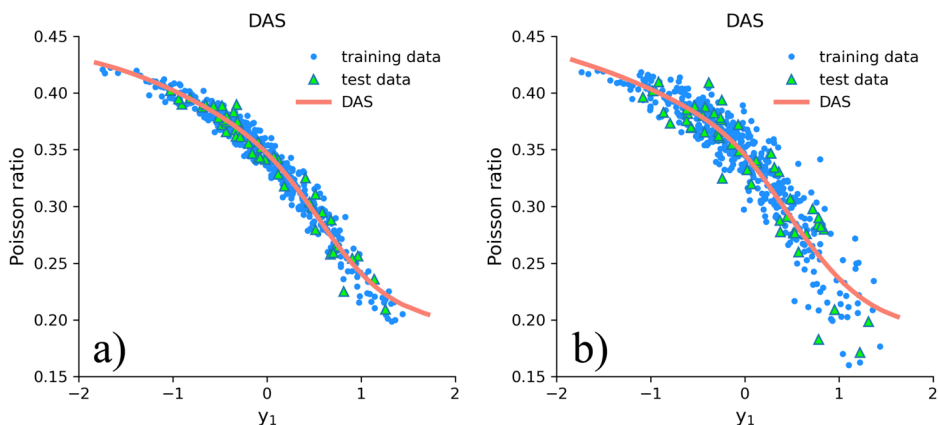
rotatable bonds involve  $sp^3$  carbon atoms, which are mainly affected by these parameters. The third parameter ( $p16rm$ ) is the equilibrium internuclear distance of vdW interactions for hydrogen atoms on the ligand, which are the most common atom type on the surface of the ligand. The fifth parameter is for oxygen on tyrosine residues; two tyrosine residues are in close proximity with the ligand (Fig. 6). As the ligand binds to the protein non-

covalently, it is not surprising that the distance-related parameters for non-bonded interactions can be found among the highest sensitivity indices.

For the TIES case (Fig. 5d), in addition to the vdW parameters as in the ESMACS study, several bonded interactions, including force constants ( $[a, b]$  NNfc, see Supplementary Table 6) and equilibrium values ( $[a, b]$  NNev) for angle and bond terms respectively, also contribute substantially



**Fig. 4 | The largest effect of aleatoric uncertainty on DAS surrogates.** Compare a the DAS Poisson ratio surrogate when using training data averaged over the replica simulations, and b the same surrogate when training data of a single random seed is used.



**Fig. 5 | The 25 largest global derivative-based sensitivity indices.** Displayed for both the DAS and KAS-GP method, ordered as  $v_1 \geq v_2 \geq \dots \geq v_{25}$ , see Section “Global derivative-based sensitivity analysis” for more details. The indices are normalized by  $v_1$  and ordered according to the DAS ranking, although the KAS-GP ranking is

similar. The DAS indices are averaged over a set of 20 replica networks. Results are shown for **a** epoxy resin application,  $E$  output, **b** epoxy resin application, Poisson ratio output, **c** ESMACS binding free energy, **d** TIES relative binding free energy.

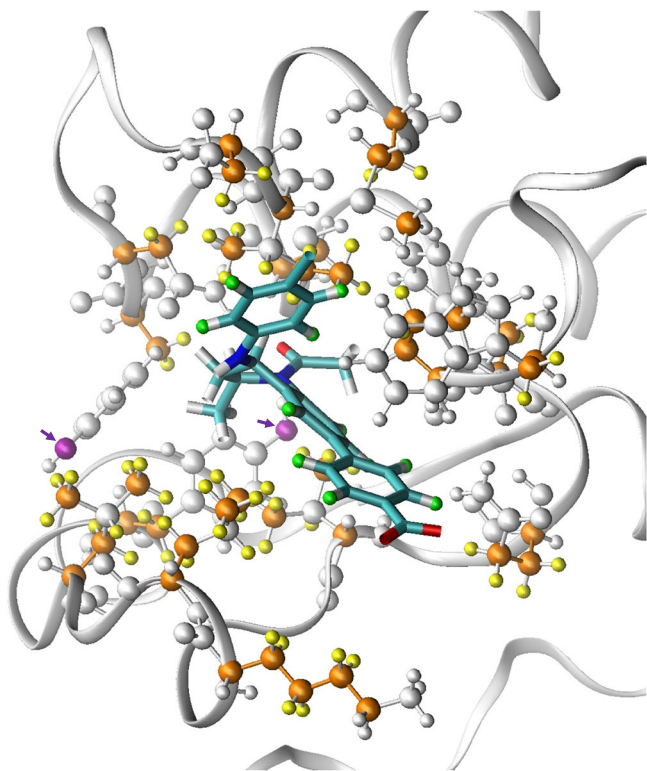
to the sensitivity of the predictions. In the TIES simulations, we need to perform separate alchemical simulations in both protein and solvent environments, in which the uncertain parameters are varied independently. The bonded parameters (bonds and angles) therefore also contribute significantly to the uncertainty of the final predicted free energy differences.

This study thus allows us to identify the most sensitive parameters upon which the QoIs depend and thus the ones whose values are most important to pin down as accurately as possible. In all cases only force-field inputs are influential, see Fig. 5. The simulation inputs (i.e. the non force-field inputs), which have been given the same uniformly-distributed input distribution (Section “Epistemic vs aleatoric uncertainty”), are comparatively insignificant. This is clear from Fig. 5, as simulation parameters such as `setTemperature` or `time_sim1` rank very low if they appear at all in the top 25 most sensitive parameters. Note however that we varied only force-field parameters for the epoxy case, see Table 1. Moreover, in all cases the number of important parameters is much smaller than the total input

dimension  $D$ . Hence, such analysis can be utilized to focus and enhance the optimization of existing and future force fields. By exploring the uncertainties associated with different parameters, we gain a deeper understanding of force field behavior and the terms that dominate a system’s behavior. By incorporating UQ techniques, informed decisions can be made on the continuous refinement and optimization of the force-field parameters, ultimately leading to more accurate and reliable simulations. It should be noted that our method provides a first-order sensitivity analysis; higher-order interaction effects between input parameters are not considered.

## Discussion

We have shown that it is possible to perform a comprehensive uncertainty quantification analysis of all-atom chemically specific classical molecular dynamics simulations that embraces the interaction potential parameterizations present in force fields used in real-world research. We use



**Fig. 6 | Positions of the atoms in the binding site of the ligand-protein complex, of which the force-field parameters are most sensitive in the ESMACS study.** The ligand is represented as bond, and the protein is shown as ribbon in white. The residues at the binding site are shown as ball and stick. The sp<sup>3</sup> carbon atoms (colored orange) and attached hydrogen atoms (yellow) from protein have the most sensitive parameters, along with hydrogen atoms (green) from ligand. The parameters for the oxygen atoms (purple with an arrow) from two tyrosine residues are also important to the sensitivity.

deep active subspace and kernel-based Gaussian process methods to handle these problems in a scalable manner suitable for use on supercomputers. Both work well and yield results in close agreement with one another. We illustrate the approach with current research applications taken from advanced materials and drug discovery.

In all cases studied, we are able to identify a one-dimensional active subspace. Especially in the epoxy-resin polymer materials application, we could reduce the input dimension from 103 to 1, with little loss of accuracy. This shows the potential of active-subspace methods for dimension reduction in MD which is significant as, today, commonly used force fields come with high-dimensional parameter sets which admit no notion of uncertainty - they simply provide fixed numbers, often to two or three decimal places.

While our parametric spaces are high-dimensional, we have imposed an a priori reduction in the number of parameters for reasons discussed in Section “Selection of models and interaction potential parameters”. We consider  $\approx 1000$  parameters as a case study for future research, and expect our DAS/KAS-GP methods to scale to this dimension, although this will again require substantial computational resources to sample the MD code (with replicas). An interesting question is if a low-dimensional active subspace will again exist, and how much data is required to approximate it. Regarding scalability, random sampling is done by Monte Carlo, which does not suffer from the curse of dimensionality. Secondly, increasing the input dimension will not cause significant scaling issues. Neural networks are regularly trained with more than 200 input neurons. The KAS-GP method, using the kernel functions to approximate the covariance matrix of the model response gradient, has the same convergence rate as Monte Carlo methods<sup>50,51</sup>, thus avoiding the curse of dimensionality. Furthermore, the

kernel functions can be defined for arbitrary types of data, including non-Euclidean data such as graphs and images. By incorporating the intrinsic interactions (see Fig. 6) among input parameters within the kernel function<sup>52</sup>, a more effective and interpretable active subspace could be identified with less data.

Some light can be shed on the existence of the active subspaces. By combining classical dimensional analysis with active subspaces, the authors of refs. 53–55 show that a physical law for a QoI  $f(\mathbf{x}) = f(x_1 \cdots, x_D) \in \mathbb{R}$  has links with the active subspace of that QoI. While this analysis does link the active subspace to the physical law that generated a (non-dimensional) QoI, it does not explain why the active subspace is often very low-dimensional in practice,  $d = 1$  in our case. However, if this also proves to be the case in other high-dimensional MD applications, active-subspace methods are well suited for use in the future design of MD surrogate models. Moreover, a low-dimensional active subspace might also be exploited in an inverse UQ context, e.g. for the data-driven calibration of force-field inputs via Bayesian inference. This typically requires running a Markov chain to draw samples from the data-driven posterior distribution<sup>56</sup>, which does not scale well to high-dimensional input spaces. The problem can be made computationally feasible by only running the Markov chain on the active variables, while sampling the inactive variables through the prior distribution<sup>57</sup>. Some recent work on statistical inference of force-field parameters can be found in reference<sup>23</sup>, which takes place in the context of sloppy models. As such, to determine the effective dimension, the eigenvalue decay of the Fisher Information matrix is examined, which is a local quantity. Their effective dimension is the number of so-called ‘non-evaporated’ parameters, essentially those parameters which can be informed by data via Bayesian (or Frequentist) inference procedures. The closest analog in this article would be the number of parameters for which the sensitivity indices  $v_i$  are larger than a certain cutoff. Depending on the cutoff value, we would obtain an effective dimension between 5 and 10 (see Fig. 5), similar to the results from ref. 23. Since active-subspace methods are not tied to the original coordinates axes of the parameters they are able to reduce the dimension further, if we consider a linear combination of parameters as a single input.

As mentioned, within the active subspace it is also possible to rank the influence of the individual interaction potential parameters on the quantities of interest, which include materials properties and free energies. This ranking tells us which physicochemical parameters dominate the uncertainty and therefore require most attention when determining what improvements to make to such all-atom force fields. This physics-based approach provides a level of physicochemical insight and understanding that wholly surpasses what is forthcoming from machine-learning (ML) derived potentials. There are orders of magnitude more parameters (the connection weights in the neural networks) in any ML based interatomic potentials<sup>58</sup>. In a recent study of machine learning interatomic potentials (MLIPs) for coarse-grained (CG) proteins<sup>59</sup>, the ensuing MLIP contains 294,565 parameters in it, whereas the CG models had far fewer parameters than the all-atom model. While physics-based force fields are comprised of interpretable and understandable parameters, none of these 294,565 parameters in the MLIP model have any meaning – they are simply and purely fitting parameters. Uncertainty quantification has only been attempted for MLIPs in the case of very simple molecular systems<sup>60,61</sup>, even in such cases it is not possible to obtain any insight into the parameters as they have no physicochemical bases.

## Methods

Here we briefly describe the molecular models, the simulation protocols, and our statistical active-subspace and global derivative-based sensitivity methods. The Supplementary Methods contains more details on the derivative-free DAS and KAS-GP methods.

## Models and simulation protocols

The materials application relies on epoxy resin molecular models, and more precisely tetraglycidyl methylene dianiline (TGMDA) cured with

polyetheramine (PEA) in a 1:1 ratio<sup>62</sup>. We use an in-house code to pack the epoxy and cross-link the monomers<sup>63</sup>. By this method, we produce an ensemble of 20 replicas for which  $91.8\% \pm 0.5\%$  of the functional groups have reacted. We already performed an extensive aleatoric uncertainty quantification analysis of these epoxy resins<sup>48</sup>. All materials properties simulations were run with LAMMPS<sup>39</sup> on SuperMUC-NG at the Leibniz Supercomputing Center, Germany, following the *ELASTIC* procedure from the LAMMPS examples database.

The protein target of the ESMACS simulation is the bromodomain-containing protein 4 (BRD4)<sup>64</sup>, while that of the TIES simulation is myeloid cell leukemia sequence 1 (MCL1)<sup>65</sup>. Both proteins are promising anticancer drug targets currently under extensive research in academia and the pharmaceutical industry. They have recently become something of a benchmark system for free energy calculations, which we have investigated extensively using our binding affinity calculator for diverse ligand data sets<sup>42,65,66</sup>. Here we use one of the ligands<sup>64</sup> and one ligand pair<sup>65</sup> studied previously, and investigate the sources of uncertainty along with the quality of binding free energy predictions.

The standard ESMACS and TIES protocols<sup>41,42</sup> have been applied to investigate the binding free energy for the ligand to BRD4 and the binding free energy difference for the ligand pair to MCL1. The ESMACS protocol employs an ensemble of 25 replicas<sup>41,67</sup>, while the TIES protocol employs an ensemble of 5 replicas for each of the 13 intermediate alchemical states (represented by a coupling parameter  $\lambda$  which is introduced to connect the thermodynamic end states)<sup>42,67</sup>. A 10 ns and 4 ns production run is used for all replicas in ESMACS and TIES respectively. Our extensive studies over several years demonstrate good convergence and reproducibility from these protocols<sup>48,67</sup>. All ESMACS and TIES simulations were run with NAMD<sup>40</sup> on SuperMUC-NG and on ARCHER2 at Edinburgh Parallel Computing Center, UK.

## Statistics

We define  $\mathbf{x} \in \mathbb{R}^D$  as our high-dimensional vector of uncertain (simulation and force field) input parameters, that are distributed according to a given probability density function;  $\mathbf{x} \sim p(\mathbf{x})$ . Let a (scalar) output of the MD code be generally denoted by  $f(\mathbf{x})$ , which is ensemble-averaged over the replicas unless otherwise noted. The active subspace method can achieve significant dimension reduction in the input space by noting that  $f(\mathbf{x})$  will likely not show the greatest variation in a direction that is exactly aligned with some coordinate axes of  $\mathbf{x}$ . Instead, a rotated coordinate system is sought that is aligned with the directions along which  $f$  varies the most on average. A low-dimensional approximation of  $f$  (a surrogate model) can then be created if  $d$  rotated directions of greatest variability exist, where  $d \ll D$ . To search for these directions, the following gradient matrix is constructed:

$$C = \mathbb{E} \left[ (\nabla f(\mathbf{x})) (\nabla f(\mathbf{x}))^T \right] = \int (\nabla f(\mathbf{x})) (\nabla f(\mathbf{x}))^T p(\mathbf{x}) d\mathbf{x}. \quad (1)$$

$C$  is symmetric positive semi-definite and therefore has the following spectral decomposition

$$C = U \Lambda U^T = [U_1 \ U_2] \begin{bmatrix} \Lambda_1 & 0 \\ 0 & \Lambda_2 \end{bmatrix} [U_1 \ U_2]^T, \quad (2)$$

with orthonormal eigenvectors contained in the columns of  $U_1$ ,  $U_2$ , and real eigenvalues  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_D \geq 0$  along the diagonals of  $\Lambda_1 := \text{diag}(\lambda_1, \dots, \lambda_d)$  and  $\Lambda_2 := \text{diag}(\lambda_{d+1}, \dots, \lambda_D)$ . Note that the  $d$  largest eigenvalues are grouped into  $\Lambda_1$  such that the column vectors of  $U_1$  point in the direction of largest (on-average) variability. If a clear separation in magnitude between  $\Lambda_1$  and  $\Lambda_2$  exists, most variability of  $f$  is retained along directions obtained by linearly projecting the input  $\mathbf{x} \in \mathbb{R}^D$  to a low dimensional active subspace  $\mathbf{y} \in \mathbb{R}^d$  via the matrix  $U_1 \in \mathbb{R}^{D \times d}$  of

orthonormal basis vectors, i.e.;

$$\mathbf{y} = U_1^T \mathbf{x}. \quad (3)$$

In a similar vein,  $\mathbf{z} = U_2^T \mathbf{x}$  are the so-called inactive variables along which  $f$  varies relatively little. If a suitable active subspace  $\mathbf{y}$  exist, one can approximate  $f(\mathbf{x}) = f(UU^T \mathbf{x}) = f(U_1 U_1^T \mathbf{x} + U_2 U_2^T \mathbf{x}) = f(U_1 \mathbf{y} + U_2 \mathbf{z})$  via the following conditional expectation,

$$f(\mathbf{x}) \approx G(\mathbf{y}) = \mathbb{E}_{\mathbf{z}} [f|\mathbf{y}] = \int f(U_1 \mathbf{y} + U_2 \mathbf{z}) p(\mathbf{z}|\mathbf{y}) d\mathbf{z} \approx \frac{1}{N} \sum_{i=1}^N f(U_1 \mathbf{y} + U_2 \mathbf{z}_i). \quad (4)$$

While the integration in (4) over the inactive variables is typically a high-dimensional problem,  $f$  will show little variation over these variables, again provided that most variability takes place over  $\mathbf{y}$ . If this is the case, the Monte Carlo approximation shown on the right will not require a large number of samples. If the active subspace is especially pronounced, one can even ignore the inactive variables altogether<sup>29</sup>.

Finally, note that in order to use the active subspace method as described here, one needs access to  $\nabla f(\mathbf{x})$ . Once again, the DAS and KAS-GP methods are derivative-free adaptations as discussed in the Supplementary Methods. The DAS method does not incorporate the effect of the inactive variables, while the KAS method does work with the conditional expectation (4).

We trained the DAS and KAS-GP surrogates using Monte Carlo sampling from  $p(\mathbf{x})$ . The parametric configurations for the different  $\mathbf{x}$  samples of each application have been generated using the Python package EasyVVUQ<sup>68</sup>. We generated 500 Monte Carlo samples for each application, each containing an ensemble of replicas, see Table 1. However, for the ESMACS and TIES computation, not all samples converged properly; see the computational setup of the Supplementary Methods. We used direct job submission on the different HPC machines employed. Training the DAS surrogate model based on the simulated Monte Carlo data was performed using EasySurrogate<sup>69</sup>; see the Data & Code availability statement to retrieve the scripts associated with this paper. Both EasyVVUQ and EasySurrogate are distributed as part of the open-source SEAVEA-toolkit<sup>70</sup>. The KAS-GP approach consists of two parts: 1) estimating the active subspace by Kernel Dimension Reduction, and 2) Fitting the Gaussian Process surrogate model over active subspace to model response. We used the Multi-Output Gaussian Process Emulator (MOGP)<sup>71</sup> for the implementations of both KAS and GP.

## Global derivative-based sensitivity analysis

To assess which inputs are most influential, commonly-used options are global variance-based sensitivity methods (e.g. ref. 72). In our case it is convenient to use global derivative-based methods, e.g.

$$\nu_i := \int \left( \frac{\partial f}{\partial x_i} \right)^2 p(\mathbf{x}) d\mathbf{x}. \quad (5)$$

These indices measure the (average) sensitivity of  $f$  to small perturbations in the inputs  $\mathbf{x}$ , and are especially suited for identifying non-influential parameters<sup>73</sup>. To connect (5) to the active subspace method, note that the  $\nu_i$  are the diagonal elements of the  $C$  matrix<sup>74</sup>:

$$[\nu_1, \dots, \nu_D]^T = \text{diag}(C). \quad (6)$$

While we do not have access to  $C$ , we compute the exact analytic derivatives  $\partial f / \partial x_j$ ,  $j = 1, \dots, D$ , either through the DAS neural network or the KAS Gaussian Process. Here  $\tilde{f}$  denotes the surrogate model approximation of  $f$  as modeled by the neural network or GP. An approximation of  $\nu_i$  or  $C$ , replacing  $\partial f / \partial x_i$  by  $\partial \tilde{f} / \partial x_i$ , is therefore available. The integrals involved can be computed using Monte Carlo sampling from  $p(\mathbf{x})$ , since  $\partial \tilde{f} / \partial x_i$  is evaluated very rapidly.



## Data availability

The data required to reproduce the results presented above can be found in ref. 75. Details on our NAMD and LAMMPS simulations can also be found there.

## Code availability

The source code (Jupyter notebooks) can be found in ref. 75.

Received: 10 October 2023; Accepted: 11 April 2024;

Published online: 03 May 2024

## References

- Frenkel, D. & Smit, B. *Understanding molecular simulation: from algorithms to applications*, 3rd edn. (Academic Press, San Diego, 2023).
- Hollingsworth, S. & Dror, R. Molecular dynamics simulation for all. *Neuron* **99**, 1129–1143 (2018).
- Software usage data on ARCHER2. <https://www.archer2.ac.uk/news/2022/02/07/software-usage-data.html> (2022).
- Service Status. <https://www.archer2.ac.uk/support-access/status.html#usage-statistics> (2024).
- Giles, D., Gopinathan, D., Guillas, S. & Dias, F. Faster than real time tsunami warning with associated hazard uncertainties. *Front. Earth Sci.* **8**, 597865 (2021).
- Gopinathan, D., Heidarzadeh, M. & Guillas, S. Probabilistic quantification of tsunami current hazard using statistical emulation. *Proc. R. Soc. A* **477**, 20210180 (2021).
- Coster, D. et al. Building a turbulence-transport workflow incorporating uncertainty quantification for predicting core profiles in a tokamak plasma. *Nucl. Fusion* **61**, 126068 (2021).
- Eidi, A., Ghiassi, R., Yang, X. & Abkar, M. Model-form uncertainty quantification in RANS simulations of wakes and power losses in wind farms. *Renew. Ener.* **179**, 2212–2223 (2021).
- Zarghami, S. & Dumrak, J. Aleatory uncertainty quantification of project resources and its application to project scheduling. *Reliab. Eng. Sys. Safe.* **211**, 107637 (2021).
- Giannetti, C. & Ransing, R. Risk based uncertainty quantification to improve robustness of manufacturing operations. *Comput. Indus. Eng.* **101**, 70–80 (2016).
- de Baar, J. & Roberts, S. Multifidelity sparse-grid-based uncertainty quantification for the Hokkaido Nansei-oki tsunami. *Pure Appl. Geophys.* **174**, 3107–3121 (2017).
- Gernay, T., Van Coile, R., Khorasani, N. & Hopkin, D. Efficient uncertainty quantification method applied to structural fire engineering computations. *Eng. Struc.* **183**, 1–17 (2019).
- Vassaux, M., Wan, S., Edeling, W. & Coveney, P. V. Ensembles are required to handle aleatoric and parametric uncertainty in molecular dynamics simulation. *J. Chem. Theory Comput.* **17**, 5187–5197 (2021).
- Rizzi, F. et al. Uncertainty quantification in md simulations. part i: Forward propagation. *Multiscale Model. Simul.* **10**, 1428–1459 (2012).
- Rizzi, F., Jones, R. E., Debusschere, B. J. & Knio, O. M. Uncertainty quantification in MD simulations of concentration driven ionic flow through a silica nanopore. II. Uncertain potential parameters. *J. Chem. Phys.* **138**, 194105 (2013).
- Coveney, P. V. & Wan, S. On the calculation of equilibrium thermodynamic properties from molecular dynamics. *Phys. Chem. Chem. Phys.* **18**, 30236–30240 (2016).
- Bhati, A. P., Hoti, A., Potterton, A., Bieniek, M. K. & Coveney, P. V. Long time scale ensemble methods in molecular dynamics: Ligand-protein interactions and allosterism in SARS-CoV-2 targets. *J. Chem. Theory Comput.* **19**, 3359–3378 (2023).
- Xiu, D. & Karniadakis, G. The Wiener–Askey polynomial chaos for stochastic differential equations. *SIAM J. Sci. Comput.* **24**, 619–644 (2002).
- Babuška, I., Nobile, F. & Tempone, R. A stochastic collocation method for elliptic partial differential equations with random input data. *SIAM J. Num. Anal.* **45**, 1005–1034 (2007).
- Rizzi, F. et al. Uncertainty quantification in md simulations. part ii: Bayesian inference of force-field parameters. *Multiscale Model Simul.* **10**, 1460–1492 (2012).
- Rabitz, H. & Aliş, Ö. General foundations of high-dimensional model representations. *J. Math. Chem.* **25**, 197–233 (1999).
- Quinn, K. N., Abbott, M., Transtrum, M., Machta, B. & Sethna, J. Information geometry for multiparameter models: new perspectives on the origin of simplicity. *Rep. Prog. Phys. Phys. Soc. (Gt. Br.)* **86**, 035901 (2022).
- Kurniawan, Y. et al. Bayesian, frequentist, and information geometric approaches to parametric uncertainty quantification of classical empirical interatomic potentials. *J. Chem. Phys.* **156**, 214103 (2022).
- Brown, K. & Sethna, J. Statistical mechanical approaches to models with many poorly known parameters. *Phys. Rev. E* **68**, 021904 (2003).
- Gerstner, T. & Griebel, M. Dimension-adaptive tensor-product quadrature. *Comput* **71**, 65–87 (2003).
- Edeling, W. et al. The impact of uncertainty on predictions of the CovidSim epidemiological code. *Nat. Comput. Sci.* **1**, 128–135 (2021).
- Tipireddy, R. & Ghanem, R. Basis adaptation in homogeneous chaos spaces. *J. Comput. Phys.* **259**, 304–317 (2014).
- Tsilifis, P. & Ghanem, R. Reduced wiener chaos representation of random fields via basis adaptation and projection. *J. Comput. Phys.* **341**, 102–120 (2017).
- Constantine, P., Dow, E. & Wang, Q. Active subspace methods in theory and practice: applications to kriging surfaces. *SIAM J. Sci. Comput.* **36**, A1500–A1524 (2014).
- Coster, D. Quantification of the uncertainty arising from atomic physics in edge plasmas. *Nucl. Mat. Ener.* **33**, 101282 (2022).
- Loudon, T. & Pankavich, S. Mathematical analysis and dynamic active subspaces for a long term model of HIV. *Math. Bio. Eng.* **14**, 709–733 (2017).
- Constantine, P., Emory, M., Larsson, J. & Iaccarino, G. Exploiting active subspaces to quantify uncertainty in the numerical simulation of the HyShot II scramjet. *J. Comput. Phys.* **302**, 1–20 (2015).
- Jefferson, J., Gilbert, J., Constantine, P. & Maxwell, R. Active subspaces for sensitivity analysis and dimension reduction of an integrated hydrologic model. *Comput. Geosci.* **83**, 127–138 (2015).
- Vohra, M. & Mahadevan, S. Discovering the active subspace for efficient UQ of molecular dynamics simulations of phonon transport in silicon. *Int. J. Heat. Mass Trans.* **132**, 577–586 (2019).
- Lin, K., Zhou, Z., Wang, Y., Law, C. & Yang, B. Using active subspace-based similarity analysis for design of combustion experiments. *Proc. Combust. Inst.* **39**, 5177–5186 (2023).
- Edeling, W. On the deep active-subspace method. *SIAM/ASA J. Uncer. Quant.* **11**, 62–90 (2023).
- Tripathy, R. & Billionis, I. Deep active subspaces: A scalable method for high-dimensional uncertainty propagation. In *ASME 2019 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference* (American Society of Mechanical Engineers Digital Collection, 2019).
- Liu, X. & Guillas, S. Dimension reduction for Gaussian process emulation: An application to the influence of bathymetry on tsunami heights. *SIAM/ASA J. Uncer. Quant.* **5**, 787–812 (2017).
- Thompson, A. et al. LAMMPS - a flexible simulation tool for particle-based materials modeling at the atomic, meso, and continuum scales. *Comput. Phys. Comm.* **271**, 108171 (2022).
- Phillips, J. et al. Scalable molecular dynamics on CPU and GPU architectures with NAMD. *J. Chem. Phys.* **153**, 044130 (2020).
- Wan, S., Knapp, B., Wright, D., Deane, C. & Coveney, P. V. Rapid, precise, and reproducible prediction of peptide-MHC binding affinities from molecular dynamics that correlate well with experiment. *J. Chem. Theory Comput.* **11**, 3346–3356 (2015).

42. Bhati, A., Wan, S., Wright, D. & Coveney, P. V. Rapid, accurate, precise, and reliable relative free energy prediction using ensemble based thermodynamic integration. *J. Chem. Theory Comput.* **13**, 210–222 (2017).
43. Case, D. et al. The Amber biomolecular simulation programs. *J. Comput. Chem.* **26**, 1668–1688 (2005).
44. Jorgensen, W., Maxwell, D. & Tirado-Rives, J. Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. *J. Am. Chem. Soc.* **118**, 11225–11236 (1996).
45. Mobley, D., Dumont, E., Chodera, J. & Dill, K. Comparison of charge models for fixed-charge force fields: Small-molecule hydration free energies in explicit solvent. *J. Phys. Chem. B* **111**, 2242–2254 (2007).
46. Frederiksen, S., Jacobsen, K., Brown, K. & Sethna, J. Bayesian ensemble approach to error estimation of interatomic potentials. *Phys. Rev. Lett.* **93**, 165501 (2004).
47. Soares, T. et al. Guidelines for reporting molecular dynamics simulations in JCI publications. *J. Chem. Inf. Model.* **63**, 3227–3229 (2023).
48. Wan, S., Sinclair, R. & Coveney, P. V. Uncertainty quantification in classical molecular dynamics. *Philos. Trans. Roy. Soc. A* **379**, 20200082 (2021).
49. Zahm, O., Constantine, P., Prieur, C. & Marzouk, Y. Gradient-based dimension reduction of multivariate vector-valued functions. *SIAM J. Sci. Comput.* **42**, A534–A558 (2020).
50. Fukumizu, K., Bach, F. & Jordan, M. Dimensionality reduction for supervised learning with reproducing kernel hilbert spaces. *J. Mach. Learn. Res.* **5**, 73–99 (2004).
51. Fukumizu, K. & Leng, C. Gradient-based kernel dimension reduction for regression. *J. Am. Stat. Assoc.* **109**, 359–370 (2014).
52. Ferré, G., Haut, T. & Barros, K. Learning molecular energies using localized graph kernels. *J. Chem. Phys.* **146**, 114107 (2017).
53. Constantine, P., del Rosario, Z. & Iaccarino, G. Many physical laws are ridge functions. Preprint at <https://arxiv.org/abs/1605.07974> (2016).
54. del Rosario, Z., Constantine, P. & Iaccarino, G. Developing design insight through active subspaces. In *19th AIAA Non-Deterministic Approaches Conference*, 1090 (AIAA, 2017).
55. del Rosario, Z., Lee, M. & Iaccarino, G. Lurking variable detection via dimensional analysis. *SIAM/ASA J. Uncer. Quant.* **7**, 232–259 (2019).
56. Gelman, A., Carlin, J., Stern, H. & Rubin, D. *Bayesian data analysis* (Chapman and Hall/CRC, 1995).
57. Constantine, P., Kent, C. & Bui-Thanh, T. Accelerating Markov Chain Monte Carlo with active subspaces. *SIAM J. Sci. Comput.* **38**, A2779–A2805 (2016).
58. Pouchard, L., Reyes, K., Alexander, F. & Yoon, B. A rigorous uncertainty-aware quantification framework is essential for reproducible and replicable machine learning workflows. *Digital Discov.* **2**, 1251–1258 (2023).
59. Majewski, M. et al. Machine learning coarse-grained potentials of protein thermodynamics. *Nat. Commun.* **14**, 5739 (2023).
60. Bartók, A. & Kermode, J. Improved uncertainty quantification for Gaussian process regression based interatomic potentials. Preprint at <https://arxiv.org/abs/2206.08744> (2022).
61. Thaler, S., Doehner, G. & Zavadlav, J. Scalable Bayesian Uncertainty Quantification for Neural Network Potentials: Promise and Pitfalls. *J. Chem. Theory Comput.* **19**, 4520–4532 (2023).
62. Vassaux, M., Sinclair, R., Richardson, R., Suter, J. & Coveney, P. V. The role of graphene in enhancing the material properties of thermosetting polymers. *Adv. Theory Simul.* **2**, 1800168 (2019).
63. Epoxy builder. [https://github.com/velocirobbie/epoxy\\_builder](https://github.com/velocirobbie/epoxy_builder). GitHub repository (2016).
64. Wan, S. et al. Rapid and reliable binding affinity prediction of bromodomain inhibitors: A computational study. *J. Chem. Theory Comput.* **13**, 784–795 (2017).
65. Bhati, A. & Coveney, P. V. Large scale study of ligand-protein relative binding free energy calculations: Actionable predictions from statistically robust protocols. *J. Chem. Theory Comput.* **18**, 2687–2702 (2022).
66. Wright, D. et al. Application of ESMACS binding free energy protocols to diverse datasets: Bromodomain-containing protein 4. *Sci. Rep.* **9**, 6017 (2019).
67. Wan, S., Bhati, A., Zasada, S. & Coveney, P. V. Rapid, accurate, precise and reproducible ligand-protein binding free energy prediction. *Interface Focus* **10**, 20200007 (2020).
68. Richardson, R. et al. EasyVVUQ: A library for verification, validation and uncertainty quantification in high performance computing. *J. Open Res. Soft.* **8**, 11 (2020).
69. Edeling, W. EasySurrogate (GitHub repository). <https://github.com/wedeling/EasySurrogate> (2021).
70. Groen, D. et al. VECMAtk: a scalable verification, validation and uncertainty quantification toolkit for scientific simulations. *Philos. Trans. R. Soc. A* **379**, 20200221 (2021).
71. Daub, E. Multi-output GP emulator documentation. <https://mogp-emulator.readthedocs.io/en/latest/> (2022).
72. Sobol, I. Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates. *Math. Comput. Sim.* **55**, 271–280 (2001).
73. Sobol, I. & Kucherenko, S. Derivative based global sensitivity measures and their link with global sensitivity indices. *Math. Comput. Sim.* **79**, 3009–3017 (2009).
74. Constantine, P. & Diaz, P. Global sensitivity metrics from active subspaces. *Reliab. Eng. Sys. Safe.* **162**, 1–13 (2017).
75. Edeling, W. wedeling/md-active-subspace: final release. *Zenodo*, <https://doi.org/10.5281/zenodo.10817481> (2024).

## Acknowledgements

The authors acknowledge funding support from (i) the UK EPSRC for the UK High-End Computing Consortium (EP/R029598/1), the Software Environment for Actionable & VVUQ-evaluated Exascale Applications (SEAVEA) grant (EP/W007762/1), the UK Consortium on Mesoscale Engineering Sciences (UKCOMES grant no. EP/L00030X/1), and the Computational Biomedicine at the Exascale (CompBioMedX) grant (EP/X019276/1); (ii) the UK MRC Medical Bioinformatics project (grant no. MR/L016311/1); (iii) the European Commission for EU H2020 CompBioMed2 Center of Excellence (grant no. 823712) and EU H2020 EXDCI-2 project (grant no. 800957). We made use of SuperMUC-NG at Leibniz Supercomputing Center under project COVID-19-SNG1, and the ARCHER2 UK National Supercomputing Service under the SEAVEA grant (EP/W007762/1).

## Author contributions

The study was designed by P.V.C., S.W., M.V. and W.E. The DAS results were generated by W.E. and the GP-KAS results by Y.Y. and S.G. The MD simulations were conducted by M.V. The paper was written by W.E., P.V.C., Y.Y., M.V. with input from the other authors.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41524-024-01272-z>.

**Correspondence** and requests for materials should be addressed to Peter V. Coveney.

**Reprints and permissions information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2024