Model Specific to Model General Uncertainty for Physical Properties

Ni Zhan and John R. Kitchin

Department of Chemical Engineering, Carnegie Mellon University, 5000 Forbes, Ave, Pittsburgh, PA 15213

E-mail:

Abstract

Many physical properties are derived from models, and we would like to be able to report the property with its uncertainty in an interpretable way. Different frameworks for interpreting uncertainty have been proposed. For example, aleatoric refers to randomness in the experiment or observations, while epistemic refers to ignorance about the best model. In addition, there are many ways to calculate uncertainty, including frequentist confidence interval and Bayesian probability distribution. In this work, we improve the understanding of which sources of uncertainty are captured by different methods. We use an example of obtaining physical properties based on function derivatives from an equation of state for Pd and Au. We obtain uncertainties using three methods: the delta method and nonlinear regression, Bayesian nonlinear regression, and Gaussian process regression. We develop a Gaussian process with joint covariance over a function and its first and second derivatives. The delta method and Bayesian regression uncertainties are consistent with each other and capture epistemic uncertainty, specifically model parameter uncertainty. The Gaussian process extends this to uncertainty arising from model selection.

1 Introduction

Many physical properties are derived from models. For example, reaction rates and enthalpy in a chemical reaction network can be determined from a kinetics model.[1–4] The thermal diffusivity of a material can be derived from a model that includes the differential equations governing heat transfer.[5] We calculate properties such as diffusion, viscosity, density, and the elastic modulus from molecular dynamics simulations and further modeling such as Arrhenius equation, adjustments for finite size effects, extrapolating to alternate conditions.[6–8]

When reporting the physical property it is important to also report its sensitivity to design choices in model and data selection. We can classify uncertainty into its sources, and different frameworks have been proposed.[9] For deriving physical properties, the sources of uncertainty may have a clear interpretation, such as different sources of experimental error, model selection, and parameter uncertainty.[10] Another common framework describes uncertainty as aleatoric or epistemic. Aleatoric refers to randomness in the experiment, or observations, while epistemic refers to ignorance about the best model.[11,12] Based on this description, uncertainty from process variation and noise in observations can be considered aleatoric, while uncertainty in parameters, model specification, and dataset shift can be classified as epistemic uncertainty. Sometimes the different uncertainties are not clear cut[12] or not feasible to separate.[9]

The uncertainty could be in the form of a confidence interval or even a full probability distribution. The frequentist approach derives confidence intervals, and Bayesian approach is concerned with the full probability distribution, with additional differences between them.[13,14] In previous work, we described some methods of obtaining uncertainty including ensembles, delta method, and Gaussian process, while focusing on the delta method.[15] Apart from Gaussian processes, the methods were specific to standard nonlinear regression. A point estimate of the parameters was obtained by minimizing mean squared error, and the parameter and model prediction standard errors were obtained from the separate methods described. Determining a point estimate of the parameters and the confidence intervals falls

under the frequentist approach. In an alternate setting, one could update a distribution over parameters as data is observed, and obtain uncertainties using the distribution. This would be a Bayesian approach, and examples of methods under this setting include probabilistic graphical models, Bayesian regression, and Gaussian processes.[16]

Given the various possible sources of uncertainty and methods of calculation, this work aims to answer "what is the uncertainty really telling us". We use an example of obtaining physical properties from an equation of state for a solid. The physical properties are the equilibrium volume, energy, and bulk modulus. We obtain uncertainties using three methods: the delta method, Bayesian nonlinear regression, and a Gaussian process. Through comparison between uncertainties obtained from the methods, we show that the delta method and Bayesian nonlinear regression give model specific uncertainty while Gaussian process gives model general uncertainty, although they can all be considered epistemic uncertainty. In Section 1.1, we provide background and motivation for the equation of state problem. In Section 1.2, we provide brief description of probabilistic modeling and further applications in engineering.

1.1 Equation of state

The equation of state (EOS) of a solid relates pressure-volume or energy-volume. We use it here as a prototypical nonlinear model that describes material properties via model evaluations and model derivatives. The EOS is important in fields such as materials, condensed matter, and geophysics, and can be used to extract the equilibrium volume and bulk modulus.[17] The pressure P, bulk modulus B can be expressed as:

$$P = -\frac{dE}{dV} \tag{1}$$

$$B = -V\frac{dP}{dV} = V\frac{d^2E}{dV^2} \tag{2}$$

where E is energy and V is volume. From Eq. 2, in an energy-volume EOS, the bulk modulus depends on the second derivative, thereby increasing its variance. Many physical properties of interest rely on derivatives, and have higher variance, making it more important to quantify their uncertainty.

Researchers have developed many EOS of different analytical forms, such as Vinet, Poirier-Tarantola, and they aim to describe a range of solids accurately.[18–22] Typically the analytical function has some parameters which are fitted to experimental or computational data using nonlinear regression. It is not necessarily clear which analytical function is most accurate for obtaining physical properties of a particular solid. Therefore uncertainty exists with regard to model selection and a specific model's parameters.

Many of the analytical functions have the equilibrium energy, volume, and bulk modulus as parameters themselves, which makes their standard error available from the inverse Fisher information matrix.[15] For other functions, it is possible to calculate the uncertainty of the physical property under nonlinear regression and the delta method when the derivative of the property w.r.t. function parameters exists.

1.2 Probabilistic models in engineering applications

A probabilistic graphical model (PGM) uses a graph to represent a joint probability over variables. The graph encodes a set of conditional independence assumptions between its nodes, which represent the random variables. A probability distribution that satisfies the independencies associated with the graph can be factorized according to the graph, which is a theorem.[23] This is helpful because it is often more intuitive to first specify a graph over variables and then derive a factorized joint distribution. Some examples of PGMs are Hidden Markov models, Naive Bayes, and Bayesian regression.

Fig. 1 shows a graphical representation of Bayesian regression. An equivalent interpretation is the joint probability factorized according to the chain rule and conditional probabilities of the graph, shown in Eq. 3. In regression, we have data with input x and target y

which is fit to a model f with parameters $\boldsymbol{\theta}$, and ϵ is residual error. In our EOS example for Bayesian regression, f is one of the analytical EOS, such as a third order polynomial with four coefficients as $\boldsymbol{\theta}$ or the analytical Vinet EOS with its parameters as $\boldsymbol{\theta}$.

In Eq. 3, $p(\boldsymbol{\theta})$ and $p(\epsilon)$ are prior distributions, or initial belief about the values of $\boldsymbol{\theta}$ and ϵ, and $p(y|x, \boldsymbol{\theta}, \epsilon)$ is the likelihood, or conditional probability of data given the parameters. The joint, prior, likelihood, along with the posterior, defined as conditional probability of parameters given data, are critically important probability distributions in the Bayesian view. Applying Bayes' theorem, the posterior is proportional to prior times likelihood.

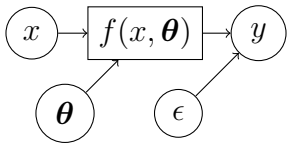


Figure 1: Graphical representation of Bayesian regression. x is input data, y is target data, f is a model with parameters $\boldsymbol{\theta}$, and ϵ is residual error.

$$p(x, y, \boldsymbol{\theta}, \epsilon) = p(x)p(\boldsymbol{\theta})p(\epsilon)p(y|x, \boldsymbol{\theta}, \epsilon) \tag{3}$$

A PGM can be advantageous to use when some random variables have a clear natural distribution, such as Poisson, Bernoulli, or when variables have clear dependence or independence relationships. Domain knowledge can identify high relevance variables and reasonable independence relations to build a PGM that closely represents reality.

Many engineering and science applications can benefit from PGMs, and we review some here. One benefit of probabilistic models is further customization of error structure among observations, as shown by Miki et al. to deduce ionization rate of atomic nitrogen.[1] Napp et al. used PGM to model chemical reaction networks.[2] Savara et al. calculated adsorption enthalpy and desorption rate using Bayesian parameter estimation and their developed CheKiPEUQ software, thereby including uncertainty from experimental data and model parameters.[4,24] Other physical properties such as modulus of elasticity[25] and crack propagation in materials[26] were found using probabilistic models. Researchers have integrated chemical

process knowledge with probabilistic modeling. Bayesian regression was used to model pressure for safe and reliable operation in a natural gas regulating and metering station.[27] Lu et al. used time-series PGM for regression of variables in a hydrocracking process,[28] and Chen et al. used probabilistic principal components analysis and mixture model for trend and fault identification of industrial propylene polymerization.[29] There are clearly many applications of PGMs and Bayesian methods, also in the pharmaceutical industry[30] and astronomy field.[31]

Another useful probabilistic model is the Gaussian process. A Gaussian process (GP) is a collection of random variables such that any finite number of them have a joint multivariate normal distribution. The distribution of the GP is a distribution of functions with a continuous domain. For the EOS example, in order to easily find a GP distribution over the minimum volume and bulk modulus, we specifically develop a GP with joint distribution over function observations, and its first and second derivatives. We describe our implementation in Section 2.2. Including derivatives in the collection of GP random variables can be useful to train with function observations and derivatives, such as energy and forces. Indeed, the Gaussian approximation potential uses GP to fit energies, forces, and virial stress (derivative of energy w.r.t. lattice deformation) of atomic configurations.[32–39] GP models have also been used as surrogate models to speed up nudged elastic band calculation or geometry optimization, sometimes using their uncertainty to guide training dataset selection.[40–42] A GP model was used to predict yield and lower heating value of fluidized bed gasifier.[43]

This work is also related to uncertainty in density functional theory (DFT) functionals such as Bayesian error estimation functional (BEEF).[44] BEEF ensemble comes from an inverse Hessian of a regression cost function used as ensemble covariance, which is similar to the inverse Fisher of the delta method. In other related work, Kim et al. develop hIPPYlib-MUQ software for Bayesian inference problems governed by partial differential equations.[45] Hirschfeld et al. compare quantitative uncertainty of various ensembling methods, distance-based uncertainty, and pipelined neural networks with GP and random forests for predicting

small organic molecule properties.[46]

2 Methods

2.1 Approximate inference for PGMs

Given a PGM, we typically want to answer queries about its probability distribution. This is called inference. One of the most common queries is computing the posterior distribution, or conditional probability of random variables in the PGM given data. In the Bayesian regression in Section 1.2, the posterior distribution is $p(\boldsymbol{\theta}, \epsilon|x, y)$. To compute the distribution, we use approximate inference methods, meaning they approximate the target distribution. Specifically we use variational inference and Markov chain Monte Carlo (MCMC). We focus on these methods because they are widely used for many Bayesian models. Variational inference is usually faster than MCMC, while MCMC is guaranteed to asymptotically produce samples from the true target distribution.[47,48]

2.1.1 Variational inference

In variational inference, we approximate the posterior using a surrogate distribution called the variational distribution q. Continuing the Bayesian regression example, our variational distribution is $q(\boldsymbol{\theta}, \epsilon)$. We assume a form for q, and some common choices are multivariate normal or mean field, which is a product of independent normals for each parameter. To find q, we minimize the Kullback-Leibler divergence between q and the true posterior, $KL(q(\boldsymbol{\theta}, \epsilon)||p(\boldsymbol{\theta}, \epsilon|x, y))$. The KL has an intractable term $\log p(x, y)$, so we remove it from our optimization, and actually minimize $E_q[\log q(\boldsymbol{\theta}, \epsilon) - \log p(\boldsymbol{\theta}, x, y, \epsilon)]$, which is the negative ELBO (evidence lower bound).[47] Minimizing KL is equivalent to maximizing ELBO.

Based on Eq. 3, we can factorize the joint probability, and our minimization problem is

$$q^*(\boldsymbol{\theta}, \epsilon) = \underset{q \in \mathcal{Q}}{\operatorname{argmin}} E_q[\log q(\boldsymbol{\theta}, \epsilon) - \log p(\boldsymbol{\theta}) - \log p(\epsilon) - \log p(y|\boldsymbol{\theta}, \epsilon, x) - \log p(x)] \quad (4)$$

We note the term $\log p(x)$ does not affect the optimization solution for q. Also, the terms on the right appear in maximum likelihood estimation (MLE) and maximum a posteriori (MAP) estimation. Specifically one could solve

$$\hat{\boldsymbol{\theta}}_{\mathrm{MLE}}, \hat{\epsilon}_{\mathrm{MLE}} = \underset{\boldsymbol{\theta},\epsilon}{\operatorname{argmax}} \log p(y|\boldsymbol{\theta}, \epsilon, x) \tag{5}$$

$$\hat{\boldsymbol{\theta}}_{\mathrm{MAP}}, \hat{\epsilon}_{\mathrm{MAP}} = \underset{\boldsymbol{\theta},\epsilon}{\operatorname{argmax}} \log p(\boldsymbol{\theta}) + \log p(\epsilon) + \log p(y|\boldsymbol{\theta}, \epsilon, x). \tag{6}$$

The difference here is that MLE and MAP return point estimates while the optimization of Eq. (4) returns a distribution q.

In our model, we must define prior probability distributions for $\boldsymbol{\theta}$ and ϵ. We also select a form for q, and we used either multivariate normal or mean field. Now we estimate the parameters of q, in our case mean and covariance matrix, by gradient descent and estimates of the ELBO gradient. This method is stochastic variational inference (SVI),[49–51] and software such as `pyro` and `tensorflow-probability` have implemented SVI and automatic calculation of ELBO and its gradients.

We used `pyro` and defined normal distribution priors for parameters in the EOS models. The prior for ϵ was normal with mean zero and standard deviation with prior Uniform$(0, 1)$. The negative ELBO was minimized with the Adam optimizer[52] until it stopped decreasing.

2.1.2 Markov chain Monte Carlo

Markov chain Monte Carlo methods generate samples from a probability distribution. A proposed sample is accepted or rejected according to some criterion related to the target distribution. The next sample is generated by random walk or another algorithm. Hamiltonian Monte Carlo (HMC) is an efficient and well-studied MCMC method, and uses Hamiltonian dynamics from physics to propose the next sample.[53] We used the No-U-Turn Sampler method[54] to generate HMC samples of posterior, implemented in `pyro`, and inspected the sample trajectory to retain samples after convergence.

2.2 Gaussian process joint including derivatives

A function f is a Gaussian process with mean $m(x)$ and covariance kernel $k(x_i, x_j)$ if

$$[f(x_1), ..., f(x_n)] \sim \mathcal{N}(\mu, K) \tag{7}$$
$$\mu_i = m(x_i)$$
$$K_{ij} = k(x_i, x_j).$$

To perform inference with a GP (predict on test data), we note that the joint distribution between noisy training data $\mathbf{y} = f(X) + \epsilon$ and test data $f(X_*) = \mathbf{f}_*$, where $\epsilon \sim N(0, \sigma^2)$, is

$$\begin{bmatrix} y \\ f_* \end{bmatrix} \sim \mathcal{N}\left(0, \begin{bmatrix} K(X,X) + \sigma^2 I & K(X, X_*) \\ K(X_*, X) & K(X_*, X_*) \end{bmatrix}\right). \tag{8}$$

GP is assumed to have zero mean for convenience. To deal with nonzero means, we can subtract $m(X)$ to create a zero mean GP and add back $m(X_*)$ after inference.

We use the conditional rule for multivariate normals, therefore $p(\mathbf{f}_*|X, \mathbf{y}, X_*)$ is multivariate normal with mean $\bar{\mathbf{f}}_*$ and $\text{cov}(\mathbf{f}_*)$, where

$$\bar{f_*} = K(X_*, X)[K(X,X) + \sigma^2 I]^{-1} y \tag{9}$$
$$\text{cov}(f_*) = K(X_*, X_*) - K(X_*, X)[K(X,X) + \sigma^2 I]^{-1} K(X, X_*).$$

Now suppose $\mathbf{y}$ and $\mathbf{f}_*$ may include arbitrary number and location of derivative and second derivative observations, for example

$$X = \begin{bmatrix} X_0 \\ X_1 \\ X_2 \end{bmatrix}, \qquad \mathbf{y} = \begin{bmatrix} f(X_0) \\ f'(X_1) \\ f''(X_2) \end{bmatrix}. \tag{10}$$

Then the covariance matrices $K(X,X), K(X,X_*), K(X_*,X), K(X_*,X_*)$ must include covariances between functions and derivatives, functions and second derivatives, derivatives and derivatives, etc. The covariances are

$$\text{cov}\left(f(x_i), \frac{\partial f(x_j)}{\partial x_j}\right) = \frac{\partial k(x_i,x_j)}{\partial x_j}, \quad \text{cov}\left(\frac{\partial f(x_i)}{\partial x_i}, \frac{\partial f(x_j)}{\partial x_j}\right) = \frac{\partial^2 k(x_i,x_j)}{\partial x_i \partial x_j} \tag{11}$$

etc.[55] Suppose $X_0 \in R^{n\times 1}, X_1 \in R^{m\times 1}$, we define an "element-wise" derivative

$$\frac{\partial K(X_0,X_1)}{\partial_\odot X_1^T} = \left[\frac{\partial K(X_0,x_{1,1})}{\partial x_{1,1}}, ..., \frac{\partial K(X_0,x_{1,m})}{\partial x_{1,m}}\right] \in R^{n\times m}. \tag{12}$$

Therefore $K(X,X), K(X,X_*)$, etc. have blockwise covariances,[56] as an example

$$K(X,X) = \begin{bmatrix} K(X_0,X_0) & \frac{\partial K(X_0,X_1)}{\partial_\odot X_1^T} & \frac{\partial^2 K(X_0,X_2)}{(\partial_\odot X_2^T)^2} \\ \frac{\partial K(X_1,X_0)}{\partial_\odot X_1} & \frac{\partial^2 K(X_1,X_1)}{\partial_\odot X_1 \partial_\odot X_1^T} & \frac{\partial^3 K(X_1,X_2)}{\partial_\odot X_1 (\partial_\odot X_2^T)^2} \\ \frac{\partial^2 K(X_2,X_0)}{(\partial_\odot X_2)^2} & \frac{\partial^3 K(X_2,X_1)}{(\partial_\odot X_2)^2 \partial_\odot X_1^T} & \frac{\partial^4 K(X_2,X_2)}{(\partial_\odot X_2)^2 (\partial_\odot X_2^T)^2} \end{bmatrix}, \tag{13}$$

and we apply Eq. 9 as before.

In practice, the kernel function k is selected, and its parameters are optimized. Some kernel functions are linear, periodic, and radial basis function (RBF), and selecting a kernel is described in Duvenaud 2014.[57] We used the RBF kernel, Eq. 14, because it works well for many functions.

$$k(x_i,x_j) = \lambda \exp\left(-\frac{(x_i - x_j)^2}{2l^2}\right) \tag{14}$$

The RBF kernel has two parameters, scale λ and lengthscale l. We optimized these

parameters using `gpytorch` and by maximizing the log marginal likelihood $p(\mathbf{y}|X)$.[55] We used vectorized-map and automatic differentiation in `jax` to efficiently calculate derivatives as those required in Eq. 13.

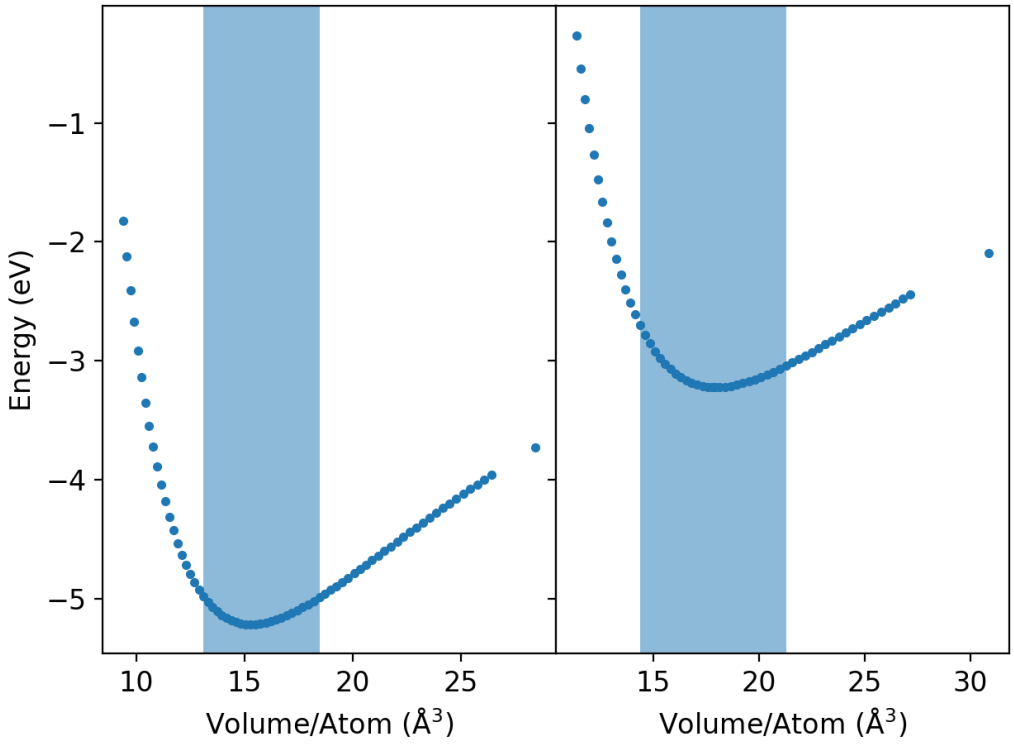
To find the posterior over minimum volume, energy, and bulk modulus, we took 1000 joint samples of the function, first and second derivatives at 1000 linspaced points across the volume range and used `scipy.interpolate.interp1d` for other points. To find the minimum volume, we used the root-finder `brentq` in `scipy` over the first derivative.[58] The corresponding sampled energies at the volumes were the distribution of minimum energies, and corresponding second derivatives multiplied with minimum volumes were the distribution of bulk moduli.

3 Results and Discussion

We discuss the data for the EOS and show results for uncertainties from nonlinear regression and the delta method, Bayesian regression, and the Gaussian process.

3.1 Data

Fig. 2 shows the data for face-centered cubic Pd and Au using DFT calculation. The data are from Boes 2016 and 2017.[59,60] The EOS calculated in the papers used all of the data points shown in Fig. 2. We used the data closer to the minimum, the shaded regions: 24 and 29 points for Pd and Au, respectively. We chose the narrower range to maintain homoscedastic residuals across the different EOS models. The range of data for training makes a significant difference in the physical properties, and the spread across models. In addition, there is not a clear way to determine which range of data should be used. This is further motivation for reporting uncertainties, and the uncertainties reported depend on the data selected.



(a) Pd (b) Au

Figure 2: Data for EOS. Shaded region represents data used in training.

3.2 Nonlinear regression

We perform nonlinear regression for common equations of state, and Table 1 shows the results. The error is low for all of the models. Most minimum volumes fall within 0.01 and 0.03 Å^3 for Pd and Au, respectively. The minimum energies are almost the same for all models, and the bulk moduli fall within 15 and 20 GPa for Pd and Au, respectively. If we used the entire dataset from Fig. 2, the ranges for minimum energies, volumes, and bulk moduli would be much wider, and the point estimates would be different. For example, using stabilized jellium (SJ) and the entire datasets, the bulk moduli for Pd and Au are both 6 GPa larger than our results.[59,60]

Table 1: Equilibrium volume, equilibrium energy, bulk modulus from different equations of state

	Equation of State	V ($Å^3$)	E (eV)	B (GPa)	RMSE (eV)	MAE (eV)
Pd	Stabilized Jellium[17]	15.304	-5.215	169	1e-4	8.5e-5
	Anton-Schmidt[61]	15.303	-5.214	168	7.4e-5	6.6e-5
	Polynomial3/Taylor	15.331	-5.216	180	1.9e-3	1.7e-3
	Murnaghan[62]	15.302	-5.214	165	4.6e-4	4.0e-4
	Birch/Birch-Murnaghan[63]	15.303	-5.214	168	6.8e-5	6.1e-5
	Poirier-Tarantola[20]	15.307	-5.215	171	3.9e-4	3.3e-4
	Vinet[63]	15.303	-5.215	168	4.1e-5	3.3e-5
Au	Stabilized Jellium	17.960	-3.222	141	2.1e-4	1.7e-4
	Anton-Schmidt	17.968	-3.221	139	3.2e-4	2.6e-4
	Polynomial3/Taylor	17.927	-3.225	160	3.9e-3	3.4e-3
	Murnaghan	17.984	-3.221	137	9.5e-4	8.2e-4
	Birch/Birch-Murnaghan	17.967	-3.221	140	2.8e-4	2.2e-4
	Poirier-Tarantola	17.950	-3.222	144	7.4e-4	6.3e-4
	Vinet	17.963	-3.222	140	1.8e-4	9.6e-5

Next we calculated standard errors for the models, shown in Table 2. The delta method follows previous work.[15] There is a difference between the uncertainty calculation for stabilized jellium (SJ), Anton-Schmidt (AS), and polynomial3 (P3) compared with the other models. For these models, the physical properties are derived quantities or predicted outputs, but for the other models, they are direct parameters. For derived quantities or model outputs, we use the derivative of the property w.r.t. function parameters to apply the delta method. Our results show that the delta method finds reasonable uncertainty estimates for both model parameters and outputs. The standard errors are larger for models with higher RMSE because the inverse Fisher information is scaled with the model error. The standard errors confidence for minimum volume and energy are quantitatively reasonable given the spreads in Table 1. For bulk modulus, if we consider 95% confidence interval ($\pm \sim 2 \cdot$ s.e.), the selected models' intervals do not overlap with each other. Hence the uncertainties from the delta method are inherent to the model used, which we describe as model-specific uncertainty.

Table 2: Standard error confidence of physical properties from delta method

		Standard Error Confidence		
	Equation of State	V (Å^3)	E (eV)	B (GPa)
Pd	Stabilized Jellium	6.8e-4	8e-5	0.069
	Anton-Schmidt	1.1e-3	8e-5	0.13
	Polynomial3/Taylor	2.5e-2	2.1e-3	4.3
	Murnaghan	7.6e-3	4.9e-4	0.75
	Birch/Birch-Murnaghan	1.1e-3	7e-5	0.13
	Poirier-Tarantola	6.1e-3	4.2e-4	0.82
	Vinet	6.7e-4	4e-5	0.075
Au	Stabilized Jellium	1.5e-3	1.7e-4	0.13
	Anton-Schmidt	4.8e-3	3.4e-4	0.44
	Polynomial3/Taylor	5.9e-2	4.0e-3	4.92
	Murnaghan	1.4e-2	1.0e-3	1.40
	Birch/Birch-Murnaghan	4.5e-3	3.0e-4	0.37
	Poirier-Tarantola	1.2e-2	7.7e-4	0.88
	Vinet	2.8e-3	1.9e-4	0.24

3.3 Bayesian regression

We performed Bayesian regression on the same set of models. We used stochastic variational inference (SVI) with multivariate normal variational distribution and HMC to find the posterior distribution over parameters, as described in Sections 2.1.1 and 2.1.2. For SVI, we used priors over parameters that were close to the nonlinear regression solution. This is analogous to choosing a good initial guess for nonlinear regression, and greatly helped the optimization. For certain problems that are ill-conditioned or non-convex, the optimization is hard to solve[64] and using a prior close to the optimal solution is recommended. The Supporting Information has examples of initialization and their optimization result. For SVI, the ELBO estimates were noisy, and optimization for one model required around 5 minutes on a laptop. HMC was more robust to the initial guess, however, it required around 20 minutes for one model. Fig. 3 compares the posteriors from SVI, HMC, and the 95% confidence interval from the delta method for Pd Poirier-Tarantola. Figures for additional models are in Supporting Information. Again the SJ, AS, and P3 models require a slightly different

method; specifically we directly calculate samples of the derived quantities from posterior parameter samples.

In the Bayesian setting, the credible interval is analogous to the prediction interval, and $1-\alpha$ credible interval for θ is defined as $\int_a^b p(\theta|x,y)d\theta = 1-\alpha$. The statistical interpretation is technically different than a confidence or prediction interval, but researchers have shown that they are quantitatively similar.[65–67] We note that Fig. 3 only shows the marginal posteriors over parameters, although we did find the joint posteriors. The HMC posterior means are close to the parameters found using nonlinear regression. Between SVI and HMC, HMC predicts a tighter posterior with a mean closer to the nonlinear regression result. This is expected because HMC can predict the posterior more accurately than variational inference. For this model, the delta method 95% confidence intervals are wider than the SVI or HMC credible intervals. However, they are still quantitatively close, which was also observed for the other models. This indicates that uncertainty from a Bayesian regression in this setting is also model-specific.

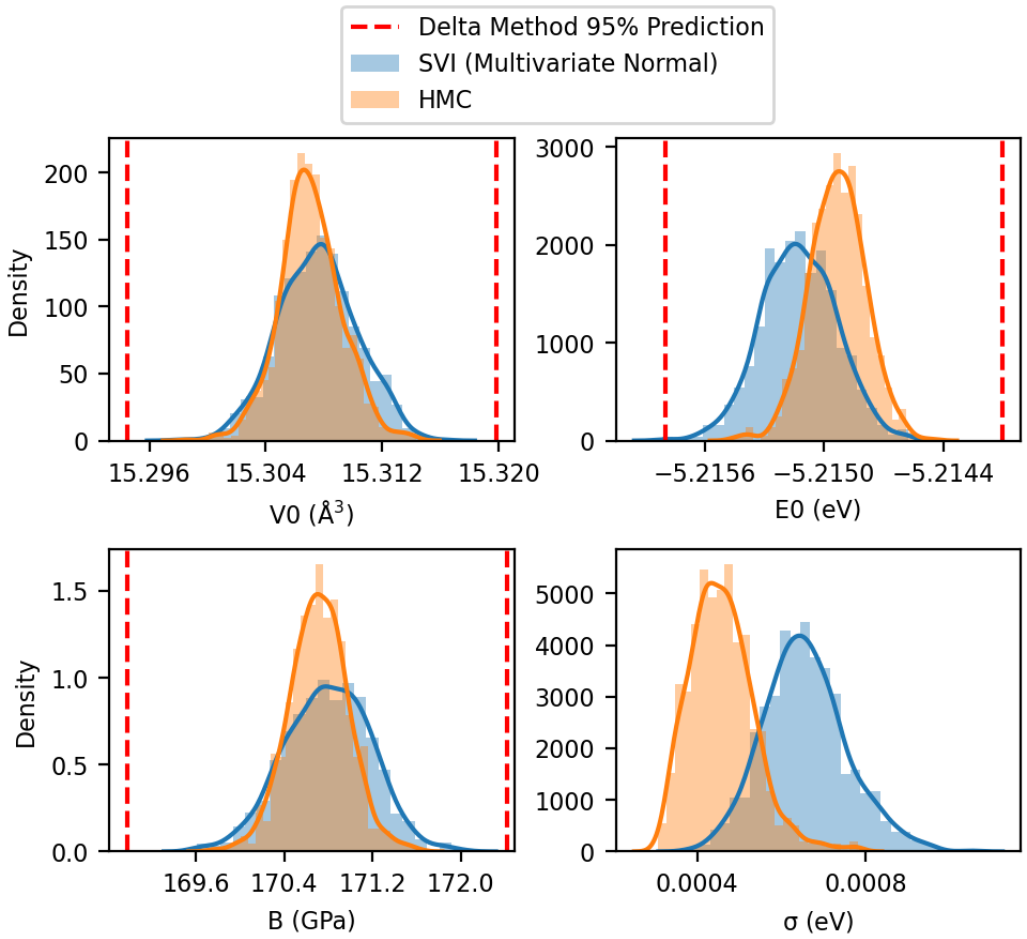


Figure 3: Comparison of HMC, variational inference posteriors and delta method confidence interval for Pd properties using the Poirier-Tarantola EOS.

3.4 Gaussian process

In this section, we show uncertainties from the Gaussian process for Pd. Results for Au are in the Supporting Information. Fig. 4 shows the Gaussian process posterior. Each sample is from the joint distribution over the function, first and second derivatives. This means that for one sample, the function, first and second derivatives correspond with each other. The variance, especially around the lowest and highest volumes, grows larger as the derivative order increases. Clearly we would not expect the GP to extrapolate well outside of the training data range. This is not surprising, the RBF kernel does not contain any physics of the EOS.

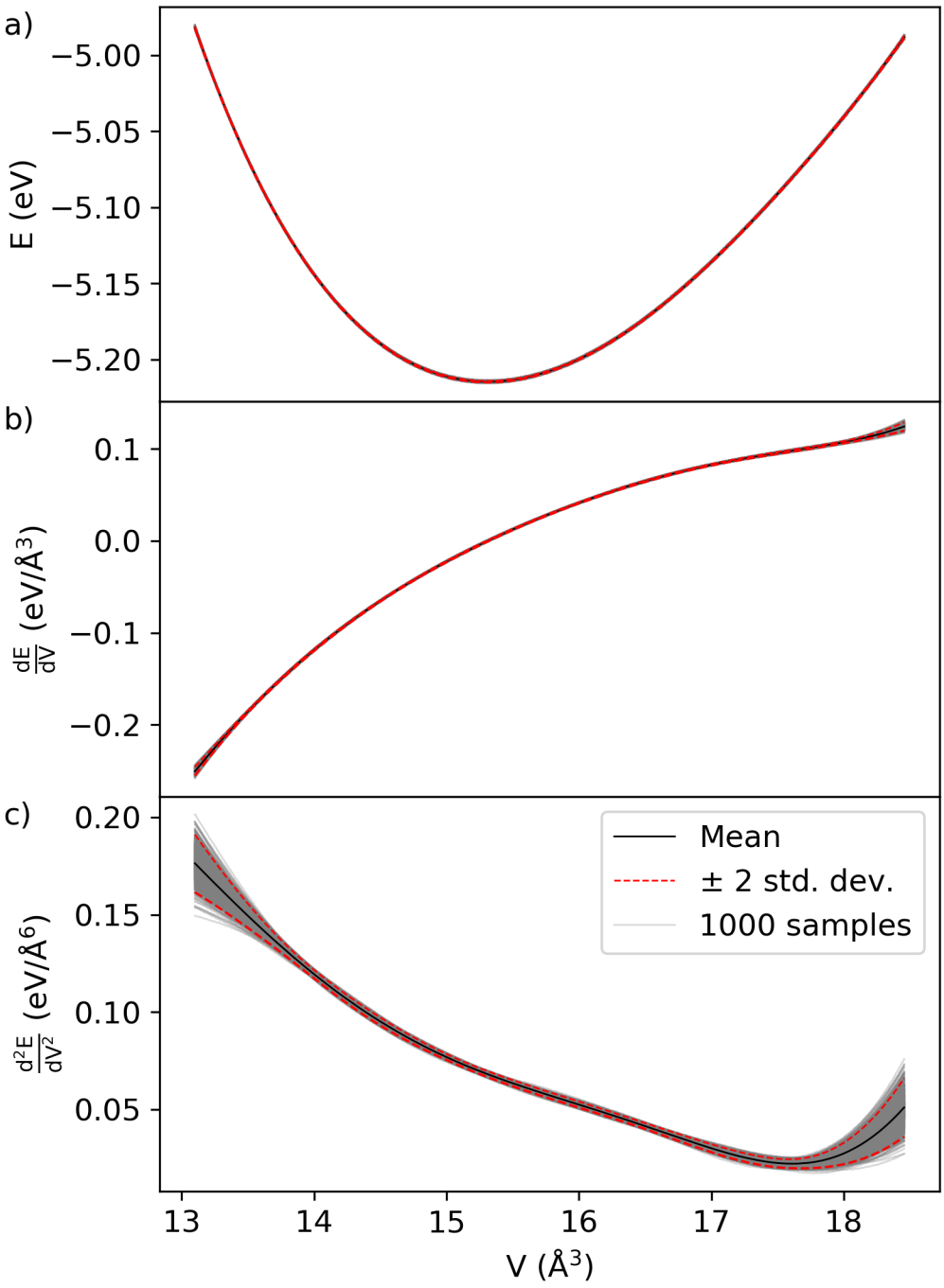


Figure 4: Gaussian process posterior for Pd EOS. a): function, b): first derivative, c): second derivative.

Figs. 5, 6, and 7 show the Gaussian process distribution compared with previous methods for minimum volume, energy, and bulk modulus, respectively. For all of the physical proper-

ties, the GP has the widest distribution compared with other methods. The GP distribution covers the predictions for almost all EOS models, while Bayesian regression posteriors do not always overlap across models, especially for minimum energy and bulk modulus, Figs. 6 and 7. The delta method intervals do not overlap across models for bulk modulus. We define the GP uncertainties to be model-general because their posterior distributions should include almost any model that could be the true EOS.

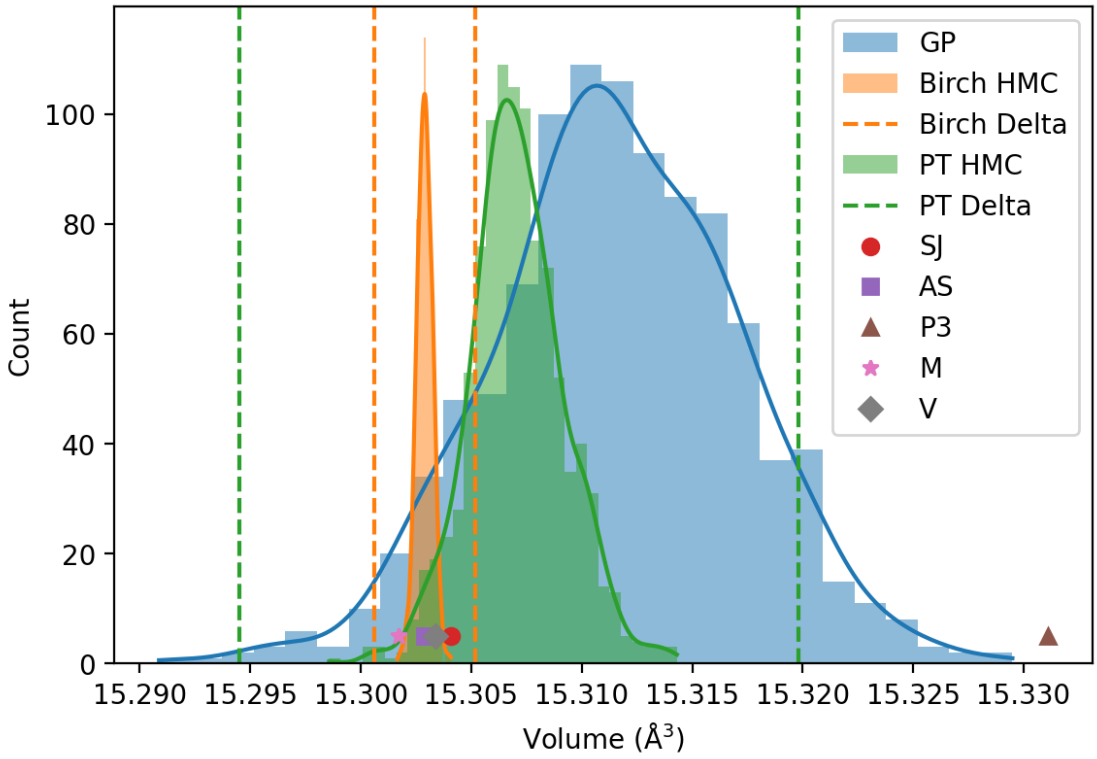


Figure 5: Comparison of GP, Bayesian regression, and nonlinear regression uncertainties with different model predictions for minimum volume.

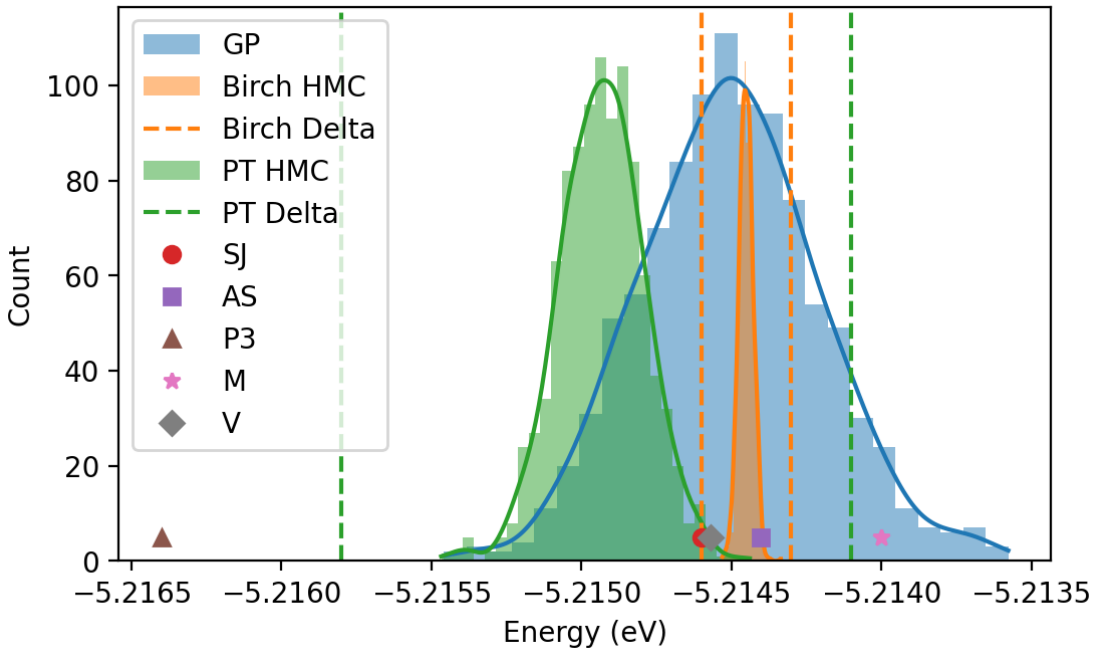


Figure 6: Comparison of GP, Bayesian regression, and nonlinear regression uncertainties with different model predictions for minimum energy.

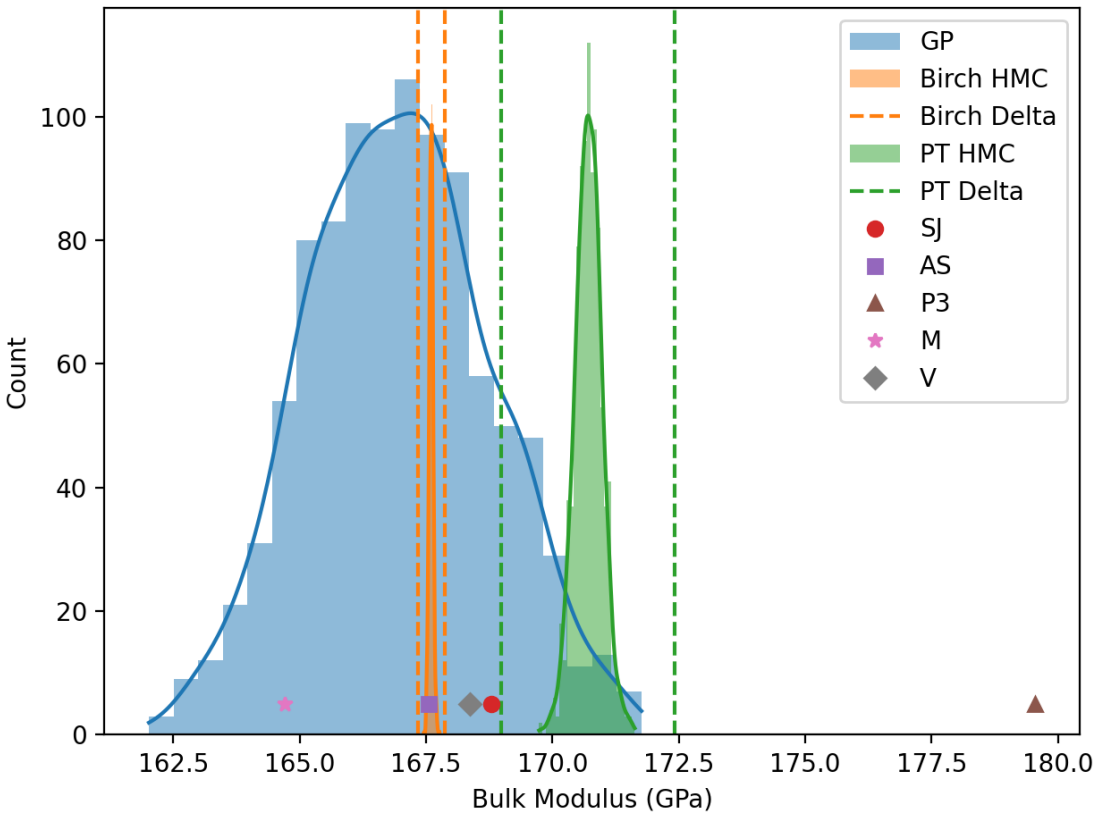


Figure 7: Comparison of GP, Bayesian regression, and nonlinear regression uncertainties with different model predictions for bulk modulus.

Fig. 8 includes the confidence intervals and posterior distributions for all analytical EOS models. Visually, Fig. 8 shows the model-specific property of the delta method and Bayesian regression. Namely, the uncertainty from the delta method or Bayesian regression for any arbitrary EOS model does not cover the physical property prediction of another model. As previously discussed, we also observe that the GP posterior covers the physical properties' predictions of all analytical EOS models (model-general uncertainty), with the exception of P3. This is because P3 has one to two orders of magnitude higher regression error compared with the other models. We can think of the GP posterior as a combined average over all analytical EOSs' distributions. The GP combines the individual model distributions into a smooth distribution, weighting each by their probability, which is related to their model's residual error. If we simply added the individual distributions, it would be difficult

to determine the weighting of each distribution, and the resulting distribution would be multimodal, which may be less desirable.

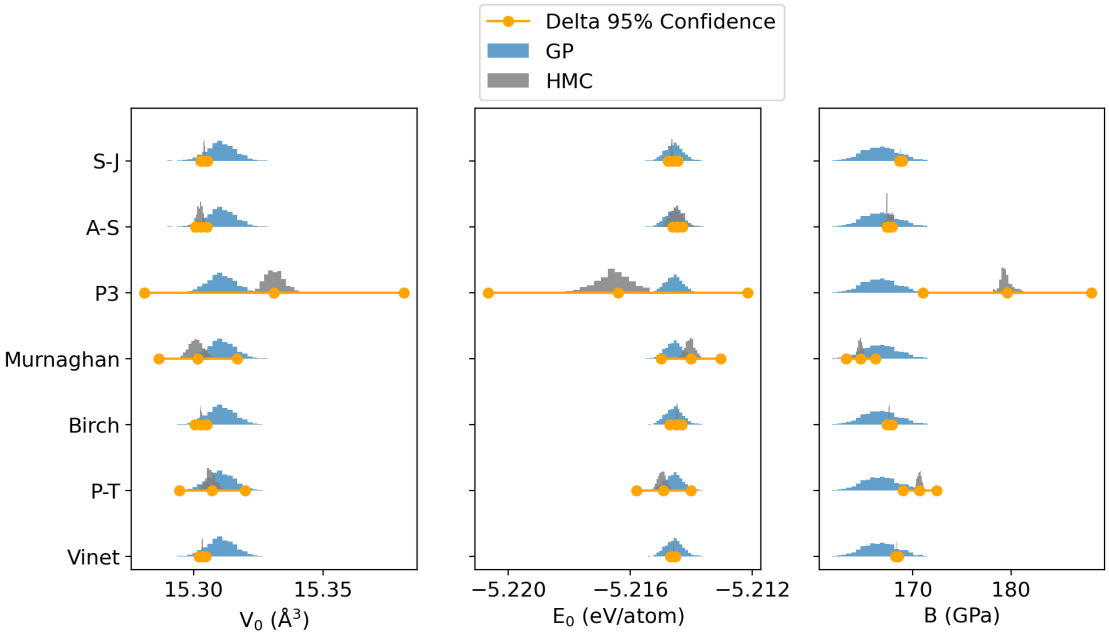


Figure 8: GP and confidence intervals and posterior distributions for all analytical EOS.

3.5 Overall comparison of methods

Here we compare the methods, including their ease of use. Between nonlinear and Bayesian regression, nonlinear regression is easier to set up, runs much faster, and is more commonly used. We showed that the delta method uncertainties are quantitatively close to those from Bayesian regression. Therefore if model parameters are expected to be normally distributed, nonlinear regression is recommended. Delta method scales as $O(d^3)$ with d as number of model parameters and $O(n)$ with n as number of data points, with research to improve scaling.[15]

Bayesian regression is harder to set up. In our experience, `pyro` is generally easier to use than `tensorflow-probability`. For each problem, the probabilistic model and variational distribution must be defined. The software improve ease of use with automatically calculated ELBO and integrated optimizers. In these problems, Bayesian regression was

slower than other methods, requiring around 5 and 20 minutes for SVI and HMC, respectively. Bayesian regression may outperform nonlinear regression if the parameters have a non-normal distribution, such as skewed, uniform, log-normal, discrete distribution, etc.

SVI scales as $O(< n)$ because of mini-batching,[68] and scaling with model dimension depends on the factorization of the variational distribution with best case $O(d)$ using mean field.[69,70] SVI is naturally parallelizable across data because of the mini-batching.[71] HMC is $O(n)$ with data,[72] and $O(d^{5/4})$ with input variables, assuming variables are independent.[53] The minimum (or most optimal) computation cost occurs with an accept rate of 65%,[53] and empirically HMC can require tens of thousands of epochs to produce a sample from the posterior.[48] HMC and MCMC are naturally sequential methods, but some parallel algorithms have been developed.[72,73]

We showed both nonlinear and Bayesian regression give model-specific uncertainty. If uncertainty across models is required, the implementations would need to be modified and rerun for separate models. To avoid this, we can use a Gaussian process. GPs are relatively easy to use with many available software tools such as `scikit-learn` and `gpytorch`. Optimizing the hyperparameters for our models required a few minutes, and calculating the covariance and distribution samples was even faster. However GPs naively scale with $O(n^3)$ with n as number of data points. There are ongoing research to improve GP scaling.[74–79] We showed the GP uncertainty is larger than other methods and includes uncertainty from model selection. Using GP, we can report uncertainty across a larger model space without repeatedly fitting many different model forms.

4 Conclusions

Uncertainty arises from different sources, and there are several, complementary methods of calculating uncertainty. We reviewed the background and literature applications in science and engineering of probabilistic graphical models and Gaussian process. There may be many

analytical forms of a model that are used to calculate physical properties, for example an equation of state. This further motivates an analysis of what sources of uncertainty are included in a particular uncertainty calculation. We compared nonlinear regression, Bayesian regression, and Gaussian process for their uncertainty quantification. Nonlinear regression gives a point estimate of the physical property, dependent on the specific model. The delta method can find a standard error and confidence or prediction interval when its assumption conditions hold. The delta method interval is wider for models with higher error, and is close to the posterior found from Bayesian regression. We used HMC and SVI for Bayesian inference, and HMC gave more consistent posteriors than SVI, at an expense of higher compute. Both the delta method and Bayesian regression give model-specific uncertainty, which is not expected to include uncertainty arising from model selection. We developed the joint GP over a function, its first and second derivatives to conveniently find distributions of physical properties requiring derivatives. The GP uncertainties cover almost all of the alternative EOS models' physical property predictions. Hence we conclude that GP gives model-general uncertainty which includes uncertainty from model selection. In situations when the true model form is unknown, such as Bayesian optimization or sparse data problems, model-general uncertainty is critical, and a method with model-general uncertainty should be used. Potential future applications include expanding model-general uncertainty to DFT functionals, uncertainty quantification for properties derived from molecular dynamics simulations, and discovery of new methods which provide model-general uncertainty.

5 List of Acronyms

EOS Equation of state

PGM probabilistic graphical model

GP Gaussian process

DFT density functional theory

MCMC Markov chain Monte Carlo

KL Kullback-Leibler

ELBO evidence lower bound

SVI stochastic variational inference

HMC Hamiltonian Monte Carlo

RBF radial basis function

SJ stabilized jellium

AS Anton-Schmidt

PT Poirier-Tarantola

P3 third order polynomial

M Murnaghan

V Vinet

Supporting Information Available

Code for accessing the data and all of the analysis in the manuscript are available in Supporting Information.

References

(1) Miki, K.; Panesi, M.; Prudencio, E.; Prudhomme, S. Probabilistic Models and Uncertainty Quantification for the Ionization Reaction Rate of Atomic Nitrogen. *Journal of Computational Physics* **2012**, *231*, 3871–3886.

(2) Napp, N. E.; Adams, R. P. Message passing inference with chemical reaction networks. *Advances in neural information processing systems* **2013**, *26*, 2247–2255.

(3) Zimányi, L.; Sipos, Á.; Sarlós, F.; Nagypál, R.; Groma, G. I. Machine-Learning Model Selection and Parameter Estimation From Kinetic Data of Complex First-Order Reaction Systems. *PLOS ONE* **2021**, *16*, e0255675.

(4) Savara, A.; Walker, E. A. Chekipeuq Intro 1: Bayesian Parameter Estimation Considering Uncertainty Or Error From Both Experiments and Theory**. *ChemCatChem* **2020**, *12*, 5385–5400.

(5) Allard, A.; Fischer, N.; Ebrard, G.; Hay, B.; Harris, P.; Wright, L.; Rochais, D.; Mattout, J. A Multi-Thermogram-Based Bayesian Model for the Determination of the Thermal Diffusivity of a Material. *Metrologia* **2015**, *53*, S1–S9.

(6) Griebel, M.; Hamaekers, J. Molecular Dynamics Simulations of the Elastic Moduli of Polymer-Carbon Nanotube Composites. *Computer Methods in Applied Mechanics and Engineering* **2004**, *193*, 1773–1788.

(7) Kang, Y.; Zhou, D.; Wu, Q.; Duan, F.; Yao, R.; Cai, K. Fully Atomistic Molecular Dynamics Computation of Physico-Mechanical Properties of PB, PS, and SBS. *Nanomaterials* **2019**, *9*, 1088.

(8) Dzulfikar, M. A. E.; Hikmawati, D.; Supardi, A. Molecular dynamics simulation to determine elastic constant and bulk modulus from Mg_xZn. The 2nd International Conference On Physical Instrumentation And Advanced Materials 2019. 2020.

(9) Kennedy, M. C.; O'Hagan, A. Bayesian Calibration of Computer Models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **2001**, *63*, 425–464.

(10) Lukić, B.; Saletti, D.; Forquin, P. Use of Simulated Experiments for Material Characterization of Brittle Materials Subjected To High Strain Rate Dynamic Tension. *Philo-*

sophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences **2017**, *375*, 20160168.

(11) Kendall, A.; Gal, Y. What Uncertainties Do We Need in Bayesian Deep Learning for Computer Vision? *arXiv* **2017**,

(12) Hüllermeier, E.; Waegeman, W. Aleatoric and Epistemic Uncertainty in Machine Learning: An Introduction To Concepts and Methods. *Machine Learning* **2021**, *110*, 457–506.

(13) Guyon, I.; Saffari, A.; Dror, G.; Cawley, G. Model Selection: Beyond the Bayesian/Frequentist Divide. *Journal of Machine Learning Research* **2010**, *11*, 61–87.

(14) Bayarri, M. J.; Berger, J. O. The Interplay of Bayesian and Frequentist Analysis. *Statistical Science* **2004**, *19*.

(15) Zhan, N.; Kitchin, J. R. Uncertainty Quantification in Machine Learning and Nonlinear Least Squares Regression Models. *AIChE Journal* **2021**,

(16) Ghahramani, Z. Probabilistic Machine Learning and Artificial Intelligence. *Nature* **2015**, *521*, 452–459.

(17) Alchagirov, A. B.; Perdew, J. P.; Boettger, J. C.; Albers, R. C.; Fiolhais, C. Energy and Pressure Versus Volume: Equations of State Motivated By the Stabilized Jellium Model. *Physical Review B* **2001**, *63*, 224115.

(18) Rose, J. H.; Smith, J. R.; Guinea, F.; Ferrante, J. Universal Features of the Equation of State of Metals. *Physical Review B* **1984**, *29*, 2963–2969.

(19) Vinet, P.; Ferrante, J.; Smith, J. R.; Rose, J. H. A Universal Equation of State for Solids. *Journal of Physics C: Solid State Physics* **1986**, *19*, L467–L473.

(20) Poirier, J.-P.; Tarantola, A. A Logarithmic Equation of State. *Physics of the Earth and Planetary Interiors* **1998**, *109*, 1–8.

(21) Jacobs, M. H. G.; Oonk, H. A. J. A New Equation of State Based on Grover, Getting and Kennedy's Empirical Relation Between Volume and Bulk Modulus. The High-Pressure Thermodynamics of MgO. *Physical Chemistry Chemical Physics* **2000**, *2*, 2641–2646.

(22) Grady, D. Equation of state for solids. *AIP Conference Proceedings* **2012**, *1426*, 800–803.

(23) Koller, D.; Friedman, N. *Probabilistic graphical models: principles and techniques*; MIT press, 2009.

(24) Walker, E. A.; Ravisankar, K.; Savara, A. Chekipeuq Intro 2: Harnessing Uncertainties From Data Sets, Bayesian Design of Experiments in Chemical Kinetics**. *ChemCatChem* **2020**, *12*, 5401–5410.

(25) Gardoni, P.; Trejo, D.; Vannucci, M.; Bhattacharjee, C. Probabilistic Models for Modulus of Elasticity of Self-Consolidated Concrete: Bayesian Approach. *Journal of Engineering Mechanics* **2009**, *135*, 295–306.

(26) Wu, W.; Ni, C. Probabilistic Models of Fatigue Crack Propagation and Their Experimental Verification. *Probabilistic Engineering Mechanics* **2004**, *19*, 247–257.

(27) BahooToroody, A.; Carlo, F. D.; Paltrinieri, N.; Tucci, M.; Gelder, P. V. Bayesian Regression Based Condition Monitoring Approach for Effective Reliability Prediction of Random Processes in Autonomous Energy Supply Operation. *Reliability Engineering & System Safety* **2020**, *201*, 106966.

(28) Lu, Y.; Peng, X.; Yang, D.; Jiang, C.; Zhong, W. The Probabilistic Discriminative Time-Series Model With Latent Variables and Its Application To Industrial Chemical Process Modeling. *Chemical Engineering Journal* **2021**, *423*, 130298.

(29) Chen, T.; Sun, Y. Probabilistic Contribution Analysis for Statistical Process Monitoring: A Missing Variable Approach. *Control Engineering Practice* **2009**, *17*, 469–477.

(30) Tabora, J. E.; Albrecht, J.; Mack, B. *Chemical Engineering in the Pharmaceutical Industry*; Chemical Engineering in the Pharmaceutical Industry; John Wiley & Sons, Inc., 2019; pp 919–935.

(31) Andreon, S.; Weaver, B. *Bayesian Methods for the Physical Sciences*; Springer Series in Astrostatistics; Springer International Publishing, 2015.

(32) Bartók, A. P.; Csányi, G. Gaussian Approximation Potentials: A Brief Tutorial Introduction. *International Journal of Quantum Chemistry* **2015**, *115*, 1051–1057.

(33) Bartók, A. P.; Kermode, J.; Bernstein, N.; Csányi, G. Machine Learning a General-Purpose Interatomic Potential for Silicon. *Physical Review X* **2018**, *8*, 041048.

(34) Rowe, P.; Csányi, G.; Alfè, D.; Michaelides, A. Development of a Machine Learning Potential for Graphene. *Physical Review B* **2018**, *97*, 054303.

(35) Sivaraman, G.; Krishnamoorthy, A. N.; Baur, M.; Holm, C.; Stan, M.; Csányi, G.; Benmore, C.; Vázquez-Mayagoitia, Á. Machine-Learned Interatomic Potentials By Active Learning: Amorphous and Liquid Hafnium Dioxide. *npj Computational Materials* **2020**, *6*, 104.

(36) Deringer, V. L.; Caro, M. A.; Csányi, G. A General-Purpose Machine-Learning Force Field for Bulk and Nanostructured Phosphorus. *Nature Communications* **2020**, *11*, 5461.

(37) Deringer, V. L.; Bartók, A. P.; Bernstein, N.; Wilkins, D. M.; Ceriotti, M.; Csányi, G. Gaussian Process Regression for Materials and Molecules. *Chemical Reviews* **2021**, *121*, 10073–10141.

(38) Unruh, D.; Meidanshahi, R. V.; Goodnick, S. M.; Zimányi, G. T. Training a Machine-Learning Driven Gaussian Approximation Potential for Si-H Interactions. *arXiv:2106.02946* **2021**,

(39) Rosenbrock, C. W.; Gubaev, K.; Shapeev, A. V.; Pártay, L. B.; Bernstein, N.; Csányi, G.; Hart, G. L. W. Machine-Learned Interatomic Potentials for Alloys and Alloy Phase Diagrams. *npj Computational Materials* **2021**, *7*, 24.

(40) Koistinen, O.-P.; Ásgeirsson, V.; Vehtari, A.; Jónsson, H. Nudged Elastic Band Calculations Accelerated With Gaussian Process Regression Based on Inverse Interatomic Distances. *Journal of Chemical Theory and Computation* **2019**, *15*, 6738–6751.

(41) Torres, J. A. G.; Jennings, P. C.; Hansen, M. H.; Boes, J. R.; Bligaard, T. Low-Scaling Algorithm for Nudged Elastic Band Calculations Using a Surrogate Machine Learning Model. *Physical Review Letters* **2019**, *122*, 156001.

(42) Bisbo, M. K.; Hammer, B. Efficient Global Structure Optimization With a Machine-Learned Surrogate Model. *Physical Review Letters* **2020**, *124*, 086102.

(43) Pan, I.; Pandey, D. S. Incorporating Uncertainty in Data Driven Regression Models of Fluidized Bed Gasification: A Bayesian Approach. *Fuel Processing Technology* **2016**, *142*, 305–314.

(44) Wellendorff, J.; Lundgaard, K. T.; Møgelhøj, A.; Petzold, V.; Landis, D. D.; Nørskov, J. K.; Bligaard, T.; Jacobsen, K. W. Density Functionals for Surface Science: Exchange-Correlation Model Development With Bayesian Error Estimation. *Physical Review B* **2012**, *85*, 235149.

(45) Kim, K.-T.; Villa, U.; Parno, M.; Marzouk, Y.; Ghattas, O.; Petra, N. Hippylib-Muq: a Bayesian Inference Software Framework for Integration of Data With Complex Predictive Models Under Uncertainty. *CoRR* **2021**,

(46) Hirschfeld, L.; Swanson, K.; Yang, K.; Barzilay, R.; Coley, C. W. Uncertainty Quantification Using Neural Networks for Molecular Property Prediction. *Journal of Chemical Information and Modeling* **2020**, *60*, 3770–3780.

(47) Blei, D. M.; Kucukelbir, A.; McAuliffe, J. D. Variational Inference: A Review for Statisticians. *Journal of the American Statistical Association* **2017**, *112*, 859–877.

(48) Izmailov, P.; Vikram, S.; Hoffman, M. D.; Wilson, A. G. What Are Bayesian Neural Network Posteriors Really Like? *arXiv:2104.14421* **2021**,

(49) Hoffman, M. D.; Blei, D. M.; Wang, C.; Paisley, J. Stochastic variational inference. *Journal of Machine Learning Research* **2013**, *14*.

(50) Kingma, D. P.; Welling, M. Auto-Encoding Variational Bayes. *arXiv:1312.6114* **2013**,

(51) Ranganath, R.; Gerrish, S.; Blei, D. Black box variational inference. Artificial intelligence and statistics. 2014; pp 814–822.

(52) Kingma, D. P.; Ba, J. Adam: A Method for Stochastic Optimization. *arXiv:1412.6980v9* **2014**,

(53) Neal, R. M. *Handbook of Markov Chain Monte Carlo*; CRC Press, 2011; Chapter 5.

(54) Hoffman, M. D.; Gelman, A. The No-U-Turn sampler: Adaptively setting path lengths in Hamiltonian Monte Carlo. *J. Mach. Learn. Res.* **2014**, *15*, 1593–1623.

(55) Rasmussen, C.; Williams, C. *Gaussian Processes for Machine Learning*; MIT Press, 2006.

(56) Girard, A. *Approximate methods for propagation of uncertainty with Gaussian process models*; University of Glasgow (United Kingdom), 2004.

(57) Duvenaud, D. Automatic model construction with Gaussian processes. Ph.D. thesis, University of Cambridge, 2014.

(58) Brent, R. P. *Algorithms for minimization without derivatives*; Courier Corporation, 2013.

(59) Boes, J. R.; Groenenboom, M. C.; Keith, J. A.; Kitchin, J. R. Neural Network and ReaxFF Comparison for Au Properties. *International Journal of Quantum Chemistry* **2016**, *116*, 979–987.

(60) Boes, J. R.; Kitchin, J. R. Neural Network Predictions of Oxygen Interactions on a Dynamic Pd Surface. *Molecular Simulation* **2017**, *43*, 346–354.

(61) Mayer, B.; Anton, H.; Bott, E.; Methfessel, M.; Sticht, J.; Harris, J.; Schmidt, P. Ab-Initio Calculation of the Elastic Constants and Thermal Expansion Coefficients of Laves Phases. *Intermetallics* **2003**, *11*, 23–32.

(62) Fu, C. L.; Ho, K. M. First-Principles Calculation of the Equilibrium Ground-State Properties of Transition Metals: Applications To Nb and Mo. *Physical Review B* **1983**, *28*, 5480–5486.

(63) Hebbache, M.; Zemzemi, M. Ab Initio study of High-Pressure Behavior of a Low Compressibility Metal and a Hard Material: Osmium and Diamond. *Physical Review B* **2004**, *70*, 224107.

(64) Alger, N. V. Data-scalable Hessian preconditioning for distributed parameter PDE-constrained inverse problems. Ph.D. thesis, 2019.

(65) Lu, D.; Ye, M.; Hill, M. C. Analysis of Regression Confidence Intervals and Bayesian Credible Intervals for Uncertainty Quantification. *Water Resources Research* **2012**, *48*.

(66) Elster, C.; Wübbeler, G. Bayesian Regression Versus Application of Least Squares-An Example. *Metrologia* **2015**, *53*, S10–S16.

(67) Kok, G. J. P.; van der Veen, A. M. H.; Harris, P. M.; Smith, I. M.; Elster, C. Bayesian Analysis of a Flow Meter Calibration Problem. *Metrologia* **2015**, *52*, 392–399.

(68) Zhang, C.; Butepage, J.; Kjellstrom, H.; Mandt, S. Advances in Variational Inference. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **2019**, *41*, 2008–2026.

(69) Tran, D.; Blei, D. M.; Airoldi, E. M. Copula Variational Inference. *CoRR* **2015**,

(70) Jiang, X.; Hoy, M.; Yu, H.; Dauwels, J. Linear-complexity stochastic variational Bayes inference for SLAM. 2017 IEEE 20th International Conference on Intelligent Transportation Systems (ITSC). 2017.

(71) Zhang, J.; Raman, P.; Ji, S.; Yu, H.-F.; Vishwanathan, S.; Dhillon, I. Extreme Stochastic Variational Inference: Distributed Inference for Large Scale Mixture Models. Proceedings of the Twenty-Second International Conference on Artificial Intelligence and Statistics. 2019; pp 935–943.

(72) Neiswanger, W.; Wang, C.; Xing, E. Asymptotically Exact, Embarrassingly Parallel Mcmc. *CoRR* **2013**,

(73) Chowdhury, A.; Jermaine, C. Parallel and Distributed MCMC via Shepherding Distributions. Proceedings of the Twenty-First International Conference on Artificial Intelligence and Statistics. 2018; pp 1819–1827.

(74) Banerjee, A.; Dunson, D. B.; Tokdar, S. T. Efficient Gaussian Process Regression for Large Datasets. *Biometrika* **2012**, *100*, 75–89.

(75) Hensman, J.; Fusi, N.; Lawrence, N. D. Gaussian Processes for Big Data. *arXiv:1309.6835* **2013**,

(76) Ambikasaran, S.; Foreman-Mackey, D.; Greengard, L.; Hogg, D. W.; O'Neil, M. Fast Direct Methods for Gaussian Processes. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **2016**, *38*, 252–265.

(77) Wang, K. A.; Pleiss, G.; Gardner, J. R.; Tyree, S.; Weinberger, K. Q.; Wilson, A. G. Exact Gaussian Processes on a Million Data Points. *arXiv:1903.08114* **2019**,

(78) Liu, H.; Ong, Y.-S.; Shen, X.; Cai, J. When Gaussian Process Meets Big Data: A

Review of Scalable GPs. *IEEE Transactions on Neural Networks and Learning Systems* **2020**, *31*, 4405–4423.

(79) Xie, Y.; Vandermause, J.; Sun, L.; Cepellotti, A.; Kozinsky, B. Bayesian Force Fields From Active Learning for Simulation of Inter-Dimensional Transformation of Stanene. *npj Computational Materials* **2021**, *7*, 40.