Dear author,

Please note that changes made in the online proofing system will be added to the article before publication but are not reflected in this PDF.

We also ask that this file not be used for submitting corrections.

International Journal of Engineering Science xxx (xxxx) xxx

Contents lists available at ScienceDirect

International Journal of Engineering Science

journal homepage: www.elsevier.com/locate/ijengsci

# Bayesian strategies for uncertainty quantification of the thermodynamic properties of materials

Q1 Noah H. Paulson[a,*], Elise Jennings[b], Marius Stan[a]

[a] Applied Materials Division, Argonne National Laboratory, Lemont, IL 60439, United States
[b] Leadership Computing Facility, Argonne National Laboratory, Lemont, IL 60439, United States

ARTICLE INFO

*Article history:*
Received 2 July 2018
Revised 6 May 2019
Accepted 26 May 2019
Available online xxx

*Keywords:*
Thermodynamic property models
Bayesian statistics
CALPHAD
Uncertainty quantification
Hafnium

ABSTRACT

Reliable models of the thermodynamic properties of materials are critical for industrially relevant applications that require a good understanding of equilibrium phase diagrams, thermal and chemical transport, and microstructure evolution. The goal of thermodynamic models is to capture data from both experimental and computational studies and then make reliable predictions when extrapolating to new regions of parameter space. These predictions will be impacted by artifacts present in real data sets such as outliers, systematic errors and unreliable or missing uncertainty bounds. Such issues increase the probability of the thermodynamic model producing erroneous predictions. We present a Bayesian framework for the selection, calibration and quantification of uncertainty of thermodynamic property models. The modular framework addresses numerous concerns regarding thermodynamic models including thermodynamic consistency, robustness to outliers and systematic errors by the use of hyperparameter weightings and robust Likelihood and Prior distribution choices. Furthermore, the framework's inherent transparency (e.g. our choice of probability functions and associated parameters) enables insights into the complex process of thermodynamic assessment. We introduce these concepts through examples where the true property model is known. In addition, we demonstrate the utility of the framework through the creation of a property model from a large set of experimental specific heat and enthalpy measurements of Hafnium metal from 0 to 4900K.



## 1. Introduction

High quality thermodynamic property models are the key to calculating numerous other properties and predicting complex material behaviors. Researchers construct these models through a process called assessment (Kattner, 1997) whereby experimental and computational information regarding material properties is used to evaluate the optimal mathematical forms that relate properties to parameters such as temperature, pressure and composition. Thermodynamic assessments present numerous challenges: the presence of outliers, missing or underestimated reported uncertainties in datasets and systematic errors are commonplace. Furthermore, it is critical to select model forms that match the trends in the data and are based on the physics and chemistry of the material. These complex judgments are the sole responsibility of the practitioner and are rarely quantitatively expressed in the final models. Bayesian statistical methods provide an opportunity to address each of these challenges in a robust and comprehensive manner.

* Corresponding author.
*E-mail address:* npaulson@anl.gov (N.H. Paulson).

https://doi.org/10.1016/j.ijengsci.2019.05.011
0020-7225/

Over the past several decades, researchers have developed approaches for the quantification of uncertainty in thermodynamic property models. The challenges involved in the calculation of phase diagrams (CALPHAD) have driven many of these efforts. CALPHAD requires the selection and calibration of models for the Gibbs energies of relevant phases from disparate experimental and computational sources (Kattner, 1997). Through these procedures, researchers fit empirical model forms to limited and often conflicting datasets. Both frequentist (Jansson, 1984; Malakhov, 1997) and Bayesian (Chatterjee, Krüger, Haller, & Olbricht, 1998; Chatterjee, Miller, & Olbricht, 1994; Duong et al., 2016; Königsberger, 1991; Olbricht, Chatterjee, & Miller, 1994; Otis & Liu, 2017; Stan & Reardon, 2003) approaches have been employed to construct models that best represent the data, ensure thermodynamic consistency and provide estimates of parameter and model uncertainties. Furthermore, authors have employed such approaches to build thermodynamic property models for single-component systems (Olbricht et al., 1994; Roslyakova, Sundman, Dette, Zhang, & Steinbach, 2016).

While the previously described efforts made significant progress in evaluating uncertainty of thermodynamic properties, a general framework for model selection, calibration and uncertainty quantification is lacking. Bayesian statistical methods have the potential to address each of these elements in an elegant and self-consistent manner. Researchers in a number of fields have employed comprehensive Bayesian modeling approaches to great effect, with particular success in the cosmology community (Ma & Berndsen, 2014; Verde, 2010). In this work, we present a Bayesian framework for the construction of thermodynamic property models, with applications in the greater materials science community. This framework is enabled by recent advances in numerical sampling algorithms in Statistics which are available in several open source packages (Farr & Farr, 2015; Feroz, Hobson, Cameron, & Pettitt, 2013; Foreman-Mackey, Hogg, Lang, & Goodman, 2013).

This Bayesian framework naturally accommodates ancillary methods to address issues commonly seen in real computational and experimental data, including the presence of outliers, systematic errors and inaccurately reported uncertainty intervals. Our novel approach optimally leverages specific heat and enthalpy data (as well as entropies and Gibbs free energies if available) to construct self-consistent thermodynamic models. One major advantage of this approach versus previous attempts is that it does not incur the errors associated with the conversion of all data to the same property basis (e.g. fitting and then differentiating enthalpy data to convert to specific heat). Furthermore, the framework can naturally incorporate the relative weighting of various datasets and provides significant insight into the relative importance of each dataset to the model.

We demonstrate the strengths of the framework through the construction of a thermodynamic property model for Hafnium metal from a diverse collection of experimental measurements of heat capacity and enthalpy for the alpha, beta and liquid phases. Hafnium metal is a good candidate for this study as the available measurements exhibit the previously mentioned issues due to the difficulty of obtaining Hafnium samples without Zirconium content, the potential for oxygen contamination, and the extreme temperatures at which the beta and liquid phases are stable (Hafnium melts at $\sim$2500K and vaporizes at 4900K) (Arblaster, 2014). In developing the thermodynamic property models we investigate model forms that are effective across the entire temperature range. In the low temperature regime we utilize the Einstein and Debye models of specific heat to account for quantum effects. At higher temperatures, both polynomial models and the segmented regression model developed by Roslyakova et al. (2016) are investigated. The resulting property models are compared with assessments performed using the HSC Chemistry software package (Roine, 2002) and the recent assessment by Arblaster (2014). The comparison with these assessments provide an opportunity to compare the strategy of the Bayesian framework with that of a commercial software package and a human expert.

## 2. The Bayesian framework

In Section 2.1 we introduce Bayes' Theorem and statistics. In Section 2.2 we describe some modern statistical sampling techniques which will be used in this paper. In Section 2.3 we discuss the model's robustness to outliers. In Section 2.4 we describe the weighting applied to different datasets and in Section 2.5 we outline how thermodynamic consistency is enforced in the model. We provide example problems to highlight the application of each method. In Appendix A, we present a simple example of fitting a linear model to noisy data.

### 2.1. Bayes' theorem

Bayes' theorem states that given some data, $D$, and a model, $M$, the probability density function of the model parameters, $\boldsymbol{\Theta}$, called the Posterior distribution, is given by,

$$P(\boldsymbol{\Theta}|\boldsymbol{D}, M) = \frac{P(\boldsymbol{D}|\boldsymbol{\Theta}, M)P(\boldsymbol{\Theta}|M)}{P(\boldsymbol{D}|M)}, \tag{1}$$

where $P(\boldsymbol{D}|\boldsymbol{\Theta}, M)$ is the *Likelihood* which describes the conditional probability of the data given the model and associated parameters. $P(\boldsymbol{\Theta}|M)$ is the Prior and describes our previous beliefs or knowledge about the model parameters (e.g. ranges from previous experiments or physical restrictions on the parameter values). $P(\boldsymbol{D}|M)$ is the marginal Likelihood or Evidence and describes the probability of the data being generated by the model. A characteristic of Bayesian statistics that differentiates it from the frequentist approach is that both the parameters and the data are assumed to be random variables and we can make probabilistic statements of certainty given a single realization of the data. This has important implications in the interpretation of Bayesian probabilities.

Please cite this article as: N.H. Paulson, E. Jennings and M. Stan, Bayesian strategies for uncertainty quantification of the thermodynamic properties of materials, International Journal of Engineering Science, https://doi.org/10.1016/j.ijengsci.2019.05.011

The quantities which we want to evaluate are moments of the probability function (e.g. means, variances and higher order moments) which require integrating over the Posterior probability. For many problems this is difficult to evaluate and so the solutions are obtained using numerical methods such as Markov Chain Monte Carlo (MCMC) (Gelman et al., 2013) which seeks an approximate solution to the exact Posterior; or variational inference methods which find an exact solution to an approximation of the Posterior (Blei, Kucukelbir, & McAuliffe, 2017).

An accurate determination of the Posterior probability enables us to measure their moments as well as uncertainty intervals (or credible intervals in the Bayesian vocabulary) for the model response. One advantage of the Bayesian framework is that uncertainties have intuitive explanations; for example, a one standard-deviation interval for a parameter of interest contains $\sim$68% of its probability distribution (assuming Gaussian distributed variables). In the remainder of this work, the 2.5th and 97.5th percentile bounds form the 95th percentile uncertainty intervals, and the best fit parameters and model predictions are given by the 50th percentile levels of the quantities of interest.

Bayesian statistics also provides a strategy for model selection via the marginal Likelihood. The marginal Likelihood is the denominator of Eq. (1) and represents the probability of the data given the model. The marginal Likelihood has the desirable qualities of rewarding models that match the data well and penalizing models that are overly complex (i.e. have too many degrees of freedom or parameters). The marginal Likelihood is given by,

$$P(\boldsymbol{D}|M) = \int_{\Omega_{\boldsymbol{\Theta}}} P(\boldsymbol{D}|\boldsymbol{\Theta}, M) P(\boldsymbol{\Theta}|M) d\boldsymbol{\Theta}, \tag{2}$$

where $\Omega_{\boldsymbol{\Theta}}$ represents the complete parameter space. Note that by integrating over the parameters in the model we are able to compare Bayesian Evidences for two distinct models in a meaningful way, even if they contain different numbers of parameters. This ratio of the Bayesian Evidence for two models is called the Bayes Factor and is given by

$$R = \frac{P(M_A|\boldsymbol{D})}{P(M_B|\boldsymbol{D})} = \frac{P(\boldsymbol{D}|M_A)P(M_A)}{P(\boldsymbol{D}|M_B)P(M_B)}, \tag{3}$$

where $M_A$ and $M_B$ are the two models under consideration (Feroz et al., 2013). Kass and Raftery (1995) provide a commonly used guide to interpret the Bayes factor, in which a value in the range 3–20 indicates a positive strength of evidence for the preference of one model versus another and a factor in the range 20–150 represents strong evidence to prefer one model over another. Unfortunately because of the integral over parameter space Eq. (2) is notoriously difficult to evaluate. Only with recent advances in sampling techniques has the model Evidence become an appealing option for model selection. We briefly discuss some selected sampling approaches in the following section.

### 2.2. Numerical sampling

Analytical methods to evaluate the Posterior probability are only available for a small subset of problems and in practice numerical sampling approaches are employed for this task. The most common and historically important sampler is the Metropolis-Hastings (M-H) algorithm (Chib & Greenberg, 1995), which makes successive jumps in parameter space (based on a proposal distribution), that are accepted when the ratio of the current to the previous local Posterior probability exceeds a uniformly distributed random number between zero and one. While the M-H algorithm is guaranteed to obtain the Posterior distribution after a sufficient number of steps, it is inefficient when applied to non-trivial distributions. For example, it is a significant challenge to determine a proposal distribution that does not lead to a precipitous drop in sampling efficiency for distributions with degeneracy (the Rosenbrock function is a classic example). Furthermore, the M-H algorithm struggles to sample multimodal distributions, as the hops must traverse regions of low Posterior probability (and therefore low acceptance probability).

In this work, we discuss two state-of-the-art samplers that improve on the M-H algorithm: *kombine* (Farr & Farr, 2015) and *MultiNest* (Feroz et al., 2013). kombine is based on Goodman and Weare's affine-invariant ensemble sampler (Goodman & Weare, 2010). Goodman and Weare's algorithm leverages the positions of a set of 'walkers' (i.e. independently initialized samplers) to propose a jump with a high probability of acceptance (even for degenerate distributions). kombine builds on this by defining local estimates of the Posterior probability function defined by $K$ nearest neighbor clusters of walkers which update a local Kernel Density Estimate (KDE) of the distribution (Silverman, 2018). This estimate enables both frequent jumps between modes and the computation of the marginal Likelihood. It is important to note that the researcher must monitor MCMC-based methods such as kombine to ensure that the walkers effectively traverse the Posterior distribution and that subsequent jumps are not overly correlated (Gelman et al., 2013). Many approaches exist to monitor convergence. The Gelman–Rubin statistic compares within-sequence and between-sequence variance to assess how well the walkers explore the Posterior (mixing) and whether each walker has achieved convergence (stationarity) (Gelman et al., 2013). The Gelman–Rubin statistic should be close to one, with 1.1 being a conservative limit. The Gelman–Rubin statistic is discussed further in Appendix A. Additionally, the effective number of independent simulation draws may be employed to ensure that a statistically significant number of independent samples have been drawn from the posterior Gelman et al. (2013). Lastly, the researcher must visually inspect the sequences to assess mixing and stationarity.

In contrast to other Monte Carlo based approaches, nested sampling (Skilling, 2004) does not utilize the concept of the Markov chain. The essential idea of this approach is to transform the multivariate integral for the marginal Likelihood (Eq. (2)) into a one dimensional integral over regions of the Prior distribution bounded by isolikelihood contours. In practice, a set of live points are sampled from the Prior and ranked by their respective Likelihoods. Each iteration, the lowest

Please cite this article as: N.H. Paulson, E. Jennings and M. Stan, Bayesian strategies for uncertainty quantification of the thermodynamic properties of materials, International Journal of Engineering Science, https://doi.org/10.1016/j.ijengsci.2019.05.011

Likelihood point is removed from the set of live points, a new point is sampled from the Prior and this point is added to the set if it has a greater Likelihood than that of the point just removed. With each iteration, the live points occupy a smaller volume of the Prior until the isolikelihood contour encloses an acceptably small volume with an associated error in the marginal Likelihood. This process efficiently computes the marginal Likelihood, and the resulting set of all current and previous live points are samples from the Posterior. MultiNest improves upon basic nested sampling by preferentially sampling regions of the Prior more likely to fall within the current isolikelihood contour. In contrast to the M-H and kombine algorithms, MultiNest requires little supervision, proceeding until the estimated marginal Likelihood reaches a maximum within some pre-specified tolerance (Feroz et al., 2013). In this work, we employ the open source python implementation of the MultiNest algorithm, *pymultinest* (Buchner et al., 2014).

### 2.3. Robustness to outliers

It is common practice in the development of thermodynamic property models to identify outliers and remove them from the analysis. This may be a reasonable approach in many cases, especially when the questionable data points or sets have known sources of uncertainty or capture different physical phenomena. In other cases, however, we may not know the reasons for the discrepancy, and we risk ignoring relevant physics or uncertainty in the property of interest by removing the data. In the following section, we present an approach to incorporate datasets with potential systematic errors.

In the previous example, we assume a Normal distribution for the Likelihood. While this is a reasonable assumption when each datum represents the mean of multiple data points, it does not necessarily hold in all cases. The use of a distribution with higher probability density in the tails, such as Student's-t distribution have a better chance of robustly accommodating outliers (Gelman et al., 2013; Kruschke, 2013). The probability density function of Student's-t distribution is given by

$$f(t) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi}\,\Gamma\left(\frac{\nu}{2}\right)}\left(1+\frac{t^2}{\nu}\right)^{-\frac{\nu+1}{2}}, \tag{4}$$

where $\nu$ is the degrees of freedom and $\Gamma$ is the gamma function. Gelman et al. suggest that it is appropriate to either select a reasonable value for the degrees of freedom, $\nu$, or to include it as a parameter in the Bayesian analysis if finding the specific form of the distribution of the data around the model is important (Gelman et al., 2013). They also caution against employing $\nu <= 2$, as this $\nu$ results in a distribution with infinite variance that is not realistic in the far tails. One significant advantage of the Bayesian framework is that modifying the choice for the Likelihood is, in this case, trivial to implement and does not add to the computational overhead when evaluating the Posterior using numerical samplers. We show an example of this in Appendix B.

### 2.4. Data quality

In this section we introduce a technique in Bayesian inference when multiple datasets of varying quality are available. Several problems often arise in this situation. Firstly, journal articles may report uncertainties for individual datasets, but these are not always reliable and must be independently evaluated. Secondly, some datasets may exhibit systematic errors where their mean behavior significantly deviates from the true trend. In a method introduced by Lahav, Bridle, Hobson, Lasenby, and Sodré (2000), and expanded by Ma and Berndsen (2014), the reported uncertainties for each dataset are weighted using hyperparameters in the Bayesian analysis. In the thermodynamics literature, data variances are often available, though covariances are rarely reported. Consequently, we do not present methods for handling data covariances in this work and we assume independence of the data points.

The hyperparameters for each dataset are incorporated into the Likelihood function as follows. Consider we have N datasets, we write the Likelihood as

$$P(\boldsymbol{D}|\boldsymbol{\Theta},\boldsymbol{\alpha},M) = \prod_{i}^{N}\prod_{j}^{t_i}\mathcal{N}\left(y_j^i|M\left(x_j^i,\boldsymbol{\Theta}\right),\varepsilon_j^i/\alpha^i\right), \tag{5}$$

where $M$ is the model of interest, $i$ is the data-set index, $j$ is the data-point index, $t_i$ is the number of points in data-set $i$, $\varepsilon_j^i$ is the reported uncertainty (expressed as 1 standard deviation) of data point $\left(x_j^i, y_j^i\right)$ from dataset $D_i$, and $\alpha_i$ is the hyperparameter for dataset $D_i$. $\mathcal{N}(x|a,b)$ Denotes the probability density of $x$ for a Normal distribution with mean $a$ and variance $b$. Note that while the reported uncertainties for points within a data set may vary, we assign a single hyperparameter for each dataset. A small value for the hyperparameter increases the effective uncertainties of the dataset and implies that the reported uncertainties underestimate the true uncertainties, while the inverse is true when the hyperparameter is large. Furthermore, a value of unity for the hyperparameter implies that the reported uncertainties are accurate and that no rescaling is required. An alternate interpretation of the hyperparameter is as a measure of the relative importance of each dataset, with a low value indicating that the associated dataset has been de-emphasized in the analysis as compared to its original weighting. We define an Exponential Prior for each hyperparameter. The principal reason for this selection is that the Exponential distribution, with a rate of unity, has a mean of one, which is consistent with the initial expectation

Please cite this article as: N.H. Paulson, E. Jennings and M. Stan, Bayesian strategies for uncertainty quantification of the thermodynamic properties of materials, International Journal of Engineering Science, https://doi.org/10.1016/j.ijengsci.2019.05.011

that the uncertainties for each dataset have been properly reported. Note that in regions of input space where the model is inadequate, this approach will interpret the deviations of the data from the model as systematic errors and will increase the uncertainty budget of such datasets. We do not explicitly address model inadequacy here, but in future work it would be advantageous to combine our approach with the model fusion methodology described by Honarmandi et al. so that data weights are physically meaningful across the entire input space (Honarmandi, Duong, Ghoreishi, Allaire, & Arroyave, 2019). We present an example problem to demonstrate the advantages of this approach in Appendix C.

### 2.5. Thermodynamic consistency

The basic thermodynamic properties of materials, including specific heats, enthalpies, entropies and Gibbs free energies are of great utility to the scientific community. These quantities are connected by thermodynamic relationships that must be considered in the construction of thermodynamic property models. For example, the specific heat of a material is the derivative of the enthalpy with temperature. Material property databases generally report thermodynamic property models in the form of polynomial relationships for room temperature and above, in mathematical forms that include a common set of parameters (Dinsdale, 1991). These parameters are typically obtained through a regression of experimental or computational information for a specific quantity. Sometimes all available data sets are leveraged by converting them to a common quantity. Conversions between quantities may be performed by direct differentiation and integration of the raw quantities or by first fitting models to the data followed by differentiation or integration. Unfortunately, such conversions are subject to unavoidable errors.

In this work, we propose a novel method to simultaneously utilize all available thermodynamic information (without conversion) to construct a common thermodynamic property model family for all quantities. Specifically, we express the overall Likelihood as the product of Likelihoods for each original quantity (and associated property model) and define priors for the parameters. By specifying a single prior distribution for each parameter common to multiple property models, we ensure that the final model family will be thermodynamically consistent and optimally leverage diverse thermodynamic information. We illustrate these concepts in the remainder of this section using synthetic datasets of specific heat and enthalpy.

Firstly, we generate 100 synthetic data points each for specific heat and enthalpy. The specific heats and enthalpies are equally spaced in the 1K–75K and 300K–1800K temperature ranges, respectively. We select these non-overlapping regions of measurement for specific heat and enthalpy to demonstrate that the Bayesian framework can capture robust representations of both quantities when data does not cover the entire temperature range for each quantity. We generate synthetic data for the enthalpy, $H$ and specific heat, $C_p$, via Eq. (6a) and (6b), respectively,

$$H(T) - H(298.15K) = \frac{3R\theta}{e^{\theta/T} - 1} + a\frac{T^2}{2} + b\frac{T^3}{2} \tag{6a}$$

$$C_p(T) = \frac{d}{dT}H(T) = 3R\left(\theta/T\right)^2 \frac{e^{\theta/T}}{e^{\theta/T} - 1^2} + aT + bT^2, \tag{6b}$$

where R is the gas constant and the true model parameters are defined as follows: $\theta = 150$K, $a = 6 \times 10^{-3}$ J/ mol K, $b = 7 \times 10^{-7}$ J/ mol K$^2$ and $c = -5 \times 10^3$ J/ mol We add Gaussian noise with a standard uncertainty of 5000 J/ mol to the enthalpy values. We add Gaussian noise to the specific heat values that scales linearly with the magnitude to a maximum of 3 J/ mol K at 75K. Note that the model form of the specific heat in Eq. (6b) was proposed at the Ringberg workshop in 1995 to be effective across the entire temperature range (Chase et al., 1995). In this case, the first term employs the Einstein model to account for low-temperature behavior, though the Debye or other models may be substituted (Grimvall, 1999).

As mentioned previously, we define our Likelihood as the product of the individual Gaussian Likelihoods for the enthalpy and specific heat, i.e.,

$$P(\boldsymbol{D}; \boldsymbol{\Theta}, M) = \prod_{i=1}^{100} \mathcal{N}(D_{H,i}|H(T_i, \boldsymbol{\Theta}), \varepsilon_i) \prod_{i=1}^{100} \mathcal{N}\left(D_{C_p,i}|C_p(T_i, \boldsymbol{\Theta}), \varepsilon_i\right), \tag{7}$$

where $\boldsymbol{D}$ represents the data, $\boldsymbol{\Theta}$ is the set of parameters, $\{\theta, a, b\}$, $D_{H,i}$ are enthalpy data and $D_{C_p,i}$ are specific heat data. At first, we initialize priors conservatively to cover the entire parameter ranges. We then iteratively narrow the priors to include the bulk and tails of the posterior distribution. The final priors for the four parameters are given as follows,

$$\theta \sim \mathcal{U}(\theta|145, 155) \tag{8a}$$

$$a \sim \mathcal{U}(a|0.003, 0.009) \tag{8b}$$

$$b \sim \mathcal{U}\left(b| - 5 \times 10^{-6}, 5 \times 10^{-6}\right). \tag{8c}$$

where $x \sim \mathcal{U}(x|a, b)$ is a uniformly distributed random variable between [a,b] .

Please cite this article as: N.H. Paulson, E. Jennings and M. Stan, Bayesian strategies for uncertainty quantification of the thermodynamic properties of materials, International Journal of Engineering Science, https://doi.org/10.1016/j.ijengsci.2019.05.011

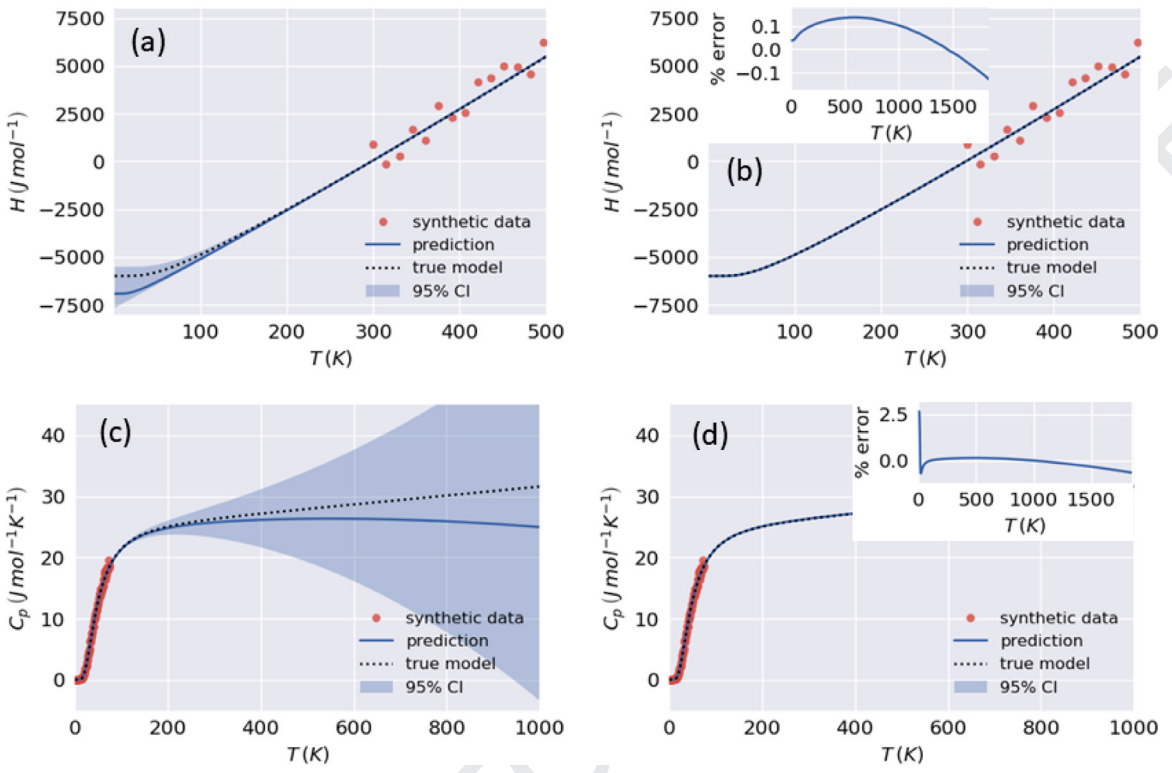


**Fig. 1.** Specific heat data, predictions and uncertainty intervals for enthalpy (a) using only enthalpy data, and (b) through a combined analysis; and for specific heat (c) using only specific heat data, and (d) through a combined analysis. Note that in panels b (d) the inset figure shows the ratio of the analyses using only enthalpy (specific heat) data to the combined analysis as a % difference.

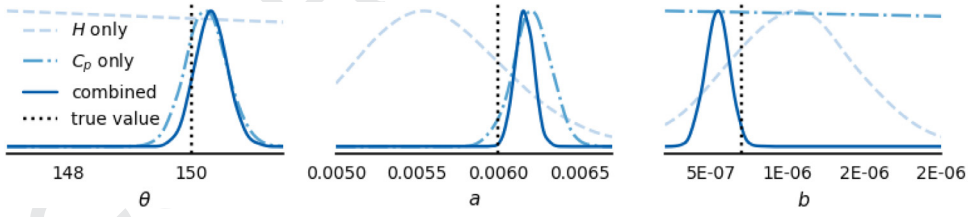


**Fig. 2.** 1D Marginal Posterior distributions from both combined and separate $H$ and $C_p$ analyses. The distributions are vertically rescaled for ease of comparison. True parameter values are indicated by the vertical dotted black lines.

The Posterior probability is then evaluated using the MultiNest sampler with 800 live points. We carry out two different analyses. Firstly, we perform separated analyses where the models for specific heat and enthalpy do not share parameter values. Both Posterior distributions are simultaneously evaluated using MultiNest to obtain the combined marginal Likelihood for enthalpy and specific heat. The Likelihood expression for this analysis is identical to Eq. (7), save that the parameters are not shared between the models for enthalpy and specific heat (e.g., $\theta_{C_p}$ and $\theta_H$ are distinct parameters). Secondly, we perform a combined analysis using the Likelihood given in Eq. (7) where the model parameters are shared between the enthalpy and specific heat models. In other words, the analysis only employs a single parameter each for $\theta$, $a$ and $b$. Note that we select priors for the separate analysis in the same manner as in the combined analysis.

Fig. 1 presents the model predictions and uncertainty intervals for both the separate and combined analyses. For both quantities, the prediction from the combined analysis nearly exactly reproduces the true model and exhibits tight uncertainty intervals. Additionally, the predictive accuracy extends well beyond the temperature ranges where data for the original quantity are available. As expected, the predictions from the separate analysis exhibit large uncertainty intervals and poor model predictions due to the lack of data across the temperature range.

We find it informative to visually inspect the posterior distributions for both analyses. Fig. 2 displays univariate Posterior distributions alongside the true parameter values for each parameter. The best fit parameter values for the analysis where only enthalpy data are employed are poor for $\theta$ and $a$, and the reverse is true for the specific heat only analysis. In comparison, the combined analysis results in reasonable best fit values for all parameters and lower variance than for the separate

Please cite this article as: N.H. Paulson, E. Jennings and M. Stan, Bayesian strategies for uncertainty quantification of the thermodynamic properties of materials, International Journal of Engineering Science, https://doi.org/10.1016/j.ijengsci.2019.05.011

analysis. Lastly, the Bayes Factor between the marginal Likelihoods for the two analyses was 29, corresponding to a strong preference for the combined approach presented here. In part, this preference likely stems from the reduced number of parameters required.

## 3. Case study

In this section we employ the Bayesian framework and methods introduced in the previous sections for the development of a thermodynamic property model for the alpha, beta and liquid phases of Hafnium. In Section 3.1 we give an overview of Hafnium. In Section 3.2 we discuss how the data was collected and any corrections which were applied. In Section 3.3 we present our analysis methods and in Section 3.4 the model selection criteria are presented. Our main results are given in Section 3.5.

### 3.1. Hafnium transition metal

Hafnium (Hf) is a tetravalent transition metal, commonly found in zirconium minerals. It is a good absorber of neutrons, which makes it a good material for control rods in some nuclear energy applications. For a review of the development of hafnium and comparison with other pressurized water-reactor control rod materials, see Keller, Shallenberger, Hollein, and Hott (1982). Hf also has a very high melting point, leading to applications in plasma welding torches (Wang et al., 2012). Superalloys of Hafnium and niobium, tungsten and titanium have excellent mechanical properties (Hou, Guo, Wu, Zhou, & Ye, 2010). Hafnium zirconate is a good gate dielectric for materials in electronics (Hegde, Triyoso, Samavedam, & White, 2007). Hafnium oxides, nitrides and carbides have additional applications that are outside the scope of this paper. The construction of a thermodynamic property model for Hafnium is interesting due to its high melting ($\sim 2500K$) and vaporization temperatures ($\sim 4900K$). This means that many more low temperature investigations are available than high temperature ones. The remainder of this work explores the thermodynamic properties of Hafnium in detail, yielding a comprehensive assessment and, to the best of our knowledge, the only known assessment that includes both parameter uncertainties and model Evidences. Furthermore, through the application of the methods described in Section 2, we explore the relative weighting of various datasets using hyperparameters in a novel way.

### 3.2. Dataset collection and correction

We have collected all available specific heat and enthalpy data of Hafnium metal for the alpha, beta and liquid phases from the published texts, correcting for Zirconium content and temperature scale when appropriate.[1] We only exclude the measurements of Cristescu and Simon (1934) as the work makes no mention of the purity of the Hafnium sample. We prioritize original measurements over the published fits of the measurements. When available, we also collect the reported measurement uncertainties. Experimental measurement uncertainties are most frequently reported as $\pm x\%$ absolute uncertainties. This does not describe the distribution of these uncertainties, so we must make some assumptions to property characterize the uncertainty. In this work, we assume that reported percent absolute uncertainties correspond to the minimum and maximum extents of a Uniform distribution (as opposed to the bounds of a triangular distribution or a 1-sigma standard deviation for a Normal distribution). If uncertainties were not reported, we assume $\pm 5\%$ absolute uncertainty bounds. This is a conservative, but not unreasonable estimate. We then use Eq. (9) as provided in the GUM standard to calculate the standard uncertainty, $\varepsilon$ (reported as a $\pm 1$ standard deviation) corresponding to the Uniform distribution describing the measurement uncertainty (BIPM, IFCC, IUPAC, & ISO, 2008),

$$\varepsilon = \sqrt{\frac{(b-a)^2}{12}} \tag{9}$$

where $a$ and $b$ are the lower and upper bounds of the Uniform distribution, respectively.

We perform corrections to account for differences in temperature scale standards (Goldberg & Weir, 1992) and Zirconium contamination (Arblaster, 2014). Over the past century, several successive standards have been employed for the calibration of temperature measurement devices. As a consequence, temperature measurements vary according to the temperature scale in effect at the time of measurement. As papers rarely report the temperature scale in use at the time of publication, we assume that the standard temperature scale during the period of publication was utilized. Consequently, we convert measurements from the original scales (i.e. IPTS-48 and IPTS-68) to the ITS-90 standard. The maximum temperature difference due to a change in scale is 0.1%, so we do not expect this correction to have a significant impact on the final analysis. Due to the chemical similarities between Zirconium and Hafnium, it is infamously difficult to obtain Hafnium specimens of high purity. Many thermodynamic measurements were performed on specimens with up to 3.12% Zirconium content. If the original report does not contain corrections for Zirconium, we correct the thermodynamic properties for Zirconium content using the Kopp–Neumann Rule (Arblaster, 2014) and the thermodynamic properties of Zirconium assessed by Arblaster (2013). The maximum change in specific heat or enthalpy due to Zirconium content is 1.7% for any dataset.

[1] Note that the measured values were extracted from tables or from figures using the free online WebPlotDigitizer utility (Rohatgi, 2011)

Please cite this article as: N.H. Paulson, E. Jennings and M. Stan, Bayesian strategies for uncertainty quantification of the thermodynamic properties of materials, International Journal of Engineering Science, https://doi.org/10.1016/j.ijengsci.2019.05.011

We provide both uncorrected and corrected datasets, along with original uncertainty types, measurement uncertainties, Zr impurity levels and temperature scales []. Additionally, we provide access to the code employed to perform these corrections []. We hope that these resources will reduce the significant effort required to collect thermodynamic information for future assessments of Hafnium.

### 3.3. Analysis methodology

We construct thermodynamic property models for the three phases of Hafnium metal via the framework introduced in Section 2. As demonstrated in Section 2.5, we can use a Bayesian framework to analyze the specific heat and enthalpy measurements in concert to ensure thermodynamic consistency and to optimally leverage the raw experimental measurements. We employ a Student's-t distribution with $2 + 1 \times 10^{-6}$ degrees of freedom for the Likelihood to correctly account for outliers, and utilize the approach of Section 2.4 to optimally rescale the reported uncertainties. We select Uniform prior distributions for the model parameters (except for the hyperparameters of Section 2.4 for which Exponential prior distributions are used) to simplify the process of restricting parameter values to physically realistic domains. We use the MultiNest algorithm to sample the Posterior distributions because it outperforms the kombine sampler on time to solution when the number of parameters included in the Bayesian analysis becomes large (23 parameters for the most complex model). We use 800 live points in the MultiNest analysis. We have verified that no significant changes occur in the marginal Likelihood or Posterior distribution mean and uncertainty intervals when the number of live points is increased to 1600.[2]

The specific Prior distributions chosen can affect the final Posterior distribution and the marginal Likelihood. Consequently, we take measures to consistently refine the Prior distributions for each model to ensure reliable estimates of these quantities For each model of interest, we first select wide prior distributions which are guaranteed to encompass the expected parameter ranges. We perform the Bayesian analysis to obtain the initial Posterior distribution (note that due to the wide Posteriors, the MultiNest algorithm tends to require additional time to reach a solution). Following this we refine the prior boundaries to encompass the Posterior mean and 5 standard deviations on either side (assuming a Normal distribution). When the value of a parameter is subject to some physical or mathematical constraint, the constraint is prioritized (for example, the Debye temperature must be greater than zero). In Section 3.5, we investigate the effect of this approach on the marginal Likelihood and Posterior distribution of the final model in more detail.

### 3.4. Model selection

A critical aspect of the assessment is the identification of a potential set of model formulations. Model identification for the alpha phase requires the most care as the models have to be reasonable at cryogenic temperatures where quantum effects dominate. In this work, low temperature physics is captured by the Einstein and Debye formulations,

$$C_p^{Ein}(T, \theta_E) = \frac{3R\theta_E}{e^{\theta_E/T} - 1} \tag{10a}$$

$$C_p^{Deb}(T, \theta_D) = 9R\left(\frac{T}{\theta_D}\right)^3 \int_0^{\theta_D/T} \frac{x^4 e^x}{(e^x - 1)^2} dx, \tag{10b}$$

where $\theta_E$ and $\theta_D$ are the Einstein and Debye temperatures, respectively. Unfortunately, Eq. (10b) cannot be evaluated analytically, so in this work we employ Simpsons integration with 100 equally-spaced points to evaluate the integral. Polynomial expressions (not including the constant term) are added to the low-temperature models to capture high-temperature properties (Chase et al., 1995). Additionally, we employ a segmented regression model developed by Roslyakova et al. (2016). In this four parameter model, two linear segments are smoothly connected by a quadratic bend. The bend is parameterized by a location and bend half-width. In principle, this form might both model electronic effects at low temperatures and high temperature behavior. We employ the family of increasing order polynomials given by Eq. (16) to model the higher temperature beta and liquid phases where quantum effects are not important. The development of thermodynamic property models that are effective across the entire temperature range is still an active area of research, so by no means is our model set exhaustive. Also, note that no constant term is employed in the enthalpy relationship, as we force the model to give zero enthalpy at 298.15 K (as is standard in the thermodynamic literature for the enthalpy of elements). The complete set of models evaluated in this work is given in Table 1.

Given a set of appropriate thermodynamic property models for specific heat and enthalpy for each phase, we apply the Bayesian framework to compute the entropy and Gibbs free energy for all phases. Due to the difficulty of analytically integrating the Debye relationship, the change in entropy in each phase relative to the lower transformation temperature, $T_{trans}$, was numerically computed according to the following relationship:

$$S(T) = \int_{T_{trans}}^{T} \frac{C_p(\tilde{T})}{\tilde{T}} d\tilde{T}. \tag{11}$$

[2] Note that we use a threaded version of MultiNest which utilizes all 36 cores on a single E5-2695v4 processor on the Bebop cluster at Argonne National Laboratory. The majority of the analyses run to completion within 3 hours on the compute cluster.

Please cite this article as: N.H. Paulson, E. Jennings and M. Stan, Bayesian strategies for uncertainty quantification of the thermodynamic properties of materials, International Journal of Engineering Science, https://doi.org/10.1016/j.ijengsci.2019.05.011

**Table 1**
Potential models, their log Marginal-Likelihoods and Bayes' Factor with respect to the selected model (marked with an asterisk) are presented for each phase.

| Model | Number of model parameters | Log marginal likelihood | Bayes' factor |
|---|---|---|---|
| $\alpha$-phase | | | |
| Einstein | 1 | −1744.2 | ∼0 |
| Debye | 1 | −1262.9 | ∼0 |
| Debye + Linear | 2 | −1072.6 | ∼0 |
| Debye + Quadratic | 3 | −813.2 | ∼0 |
| Debye + Cubic | 4 | −640.2 | ∼0 |
| Debye + Quartic* | 5 | −623.1 | 1 |
| Debye + Quintic | 6 | −627.4 | $1.4 \times 10^{-2}$ |
| Debye + SR | 5 | −629.7 | $1.4 \times 10^{-3}$ |
| **$\beta$-phase** | | | |
| Constant | 2 | −534.2 | ∼0 |
| Linear | 3 | −511.1 | $3.0 \times 10^{-3}$ |
| Quadratic* | 4 | −505.3 | 1 |
| Cubic | 5 | −518.5 | $1.9 \times 10^{-6}$ |
| **Liquid-phase** | | | |
| Constant | 2 | −491.4 | ∼0 |
| Linear* | 3 | −471.0 | 1 |
| Quadratic | 4 | −476.0 | $6.7 \times 10^{-3}$ |

The entropy associated with each phase transformation is simply given by

$$S_{trans} = \frac{H_{trans}}{T_{trans}}, \tag{12}$$

where $H_{trans}$ is the enthalpy of transformation. The entropy at any temperature is given by the entropy relative to the lower phase transformation temperature (e.g. $T_{\alpha \to \beta}$ for the beta-phase) plus the lower temperature cumulative relative entropies and entropies of transformation. The Gibbs free energy is computed according to the following relationship:

$$G(T) = H(T) - T \cdot S(T). \tag{13}$$

### 3.5. Results

Table 1 summarizes the potential models for the alpha, beta and liquid phases. Model names refer to the form of the specific heat relationship. For example, Debye + Quadratic refers to a specific heat model where the Debye relationship accounts for low temperature effects and a quadratic polynomial accounts for high temperatures. As previously mentioned, the constant term was excluded for the alpha phase models because the specific heat must be 0 at 0 K. Alongside the model names, the table lists total number of parameters, marginal Likelihoods and the Bayes Factors. Note that for each phase the Bayes Factor for that model is the ratio of the marginal Likelihood for the model and the model marked with an asterisk. We only evaluate polynomial models up to the order at which the marginal Likelihood stops increasing. We select the Debye + Quartic, Quadratic and Linear models for the alpha, beta and liquid phases, respectively, as each exhibits at least a strong evidence of preference versus the next best model according to the Bayes Factors. Equations (14a) through (14f) present the selected expressions for the specific heat and enthalpy for all phases.

$$C_p^{\alpha}(T) = C_p^{Deb}(T, \theta_D) + a_2 T + a_3 T^2 + a_4 T^3 + a_5 T^4 \tag{14a}$$

$$H^{\alpha}(T) - H^{\alpha}(298.15K) = \int_0^T C_p^{Deb}\left(\tilde{T}, \theta_D\right) d\tilde{T} + a_2 \frac{T^2}{2} + a_3 \frac{T^3}{3} + a_4 \frac{T^4}{4} + a_5 \frac{T^5}{5} \tag{14b}$$

$$C_p^{\beta}(T) = b_1 + b_2 T + b_3 T^2 \tag{14c}$$

$$H^{\beta}(T) = b_0 + b_1 T + b_2 \frac{T^2}{2} + b_3 \frac{T^3}{3} \tag{14d}$$

$$C_p^{liq}(T) = c_1 + c_2 T \tag{14e}$$

$$H^{liq}(T) = c_0 + c_1 T + c_2 \frac{T^2}{2} \tag{14f}$$

Please cite this article as: N.H. Paulson, E. Jennings and M. Stan, Bayesian strategies for uncertainty quantification of the thermodynamic properties of materials, International Journal of Engineering Science, https://doi.org/10.1016/j.ijengsci.2019.05.011

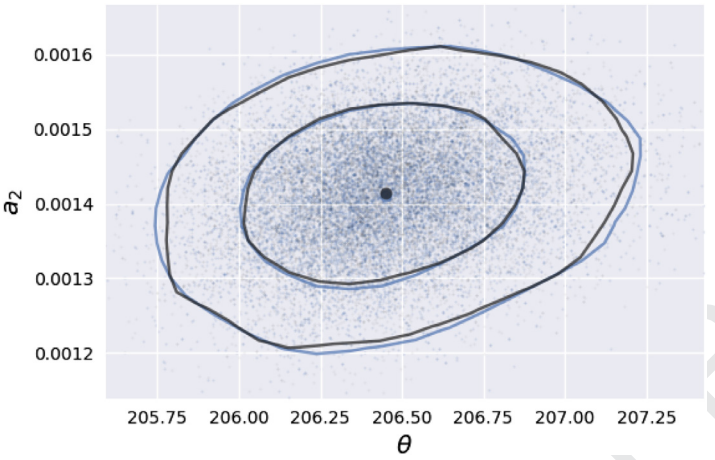


**Fig. 3.** The 68% (95%) uncertainty intervals are shown as the inner (outer) contours. Results using unrestricted priors are shown as blue lines and restricted priors as black lines for both contours. The dots represent samples from the Posterior distribution and the circular markers represent the means of the Posterior distribution for both Prior choices (Note that these lie on top of one another in this figure). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

As mentioned previously, we carry out two stages of analysis; firstly using broad unrestrictive priors to identify high probability regions of parameter space (stage A), followed by a subsequent analysis with tighter priors (stage B). We demonstrate that this does not impact model selection, as we find that the rank-ordering of the models is identical before and after tightening the priors. As an example, we examine the effect of narrowing the priors for the selected alpha-phase model. Changing the priors increases the log marginal Likelihood from −641.1 to −623.1. In contrast, the Posterior distributions change little between the stages A and B, with the maximum relative deviation in mean and standard deviation for any parameter in the Posterior distribution limited by 1% and 5%, respectively. Fig. 3 illustrates the effect of changing the Prior on the Posterior distribution of the $\theta$ and $a_2$ parameters. The original samples, means and uncertainty intervals (68% and 95%) are plotted for both stage A (blue lines) and stage B (black lines). The uncertainty intervals are produced by computing the kernel density estimated probability density function for the posterior samples and identifying an isoprobability-density contour containing the desired volume fraction (Silverman, 2018). We find that the Posterior distribution does not significantly change after changing the Prior distributions.

Table 2 contains the means, standard deviations and uncertainty intervals for the parameters in Eq. (14a) through Eq. (14f) as well as for the hyperparameters that rescale the uncertainties for each experimental dataset. In Fig. 4, we plot the model predictions for each phase versus temperature for the results of two previous assessments and the datasets used. The first assessment is from Arblaster (2014), hereafter Arb2014, whose results leverage their knowledge and expertise to judge the quality of various experimental measurements and come up with a final model. The second assessment was performed using thermodynamic information and HSC Chemistry, a software package which automatically combines existing thermodynamic assessments to find an optimal model, leveraging multiple experts whenever possible. The software uses data from literature, especially review articles. Whenever possible, we trace back the original article and use the original experimental measurement information. At the time of this article, HSC Chemistry does not provide uncertainty intervals for the model parameters or property model predictions. To acquire insights into the effect of the method presented in Section 2.4 on the actual experimental datasets, we provide uncertainty intervals for a single dataset in each subfigure. Note that these uncertainty intervals have been rescaled from the reported values by the associated best-fit hyperparameter found by our model. Percent deviations of the thermodynamic quantities between our models and the two other assessments are shown in Fig. 7.

Our alpha-phase model prediction for specific heat shows reasonable agreement with both the HSC Chemistry and Arb2014 assessments up until 1000K, after which Arb2014 trends consistently lower until the phase transition. The discrepancies are far smaller for enthalpy. From Fig. 4(a) and (b), we can see that our model closely tracks the experimental measurements of Mil2006S2, Haw1963 and Cez1974. In Arb2014, the alpha-phase model is built (for 298.15K and above) solely from the measurements of Haw1963 and Kat1985. The uncertainty intervals for our alpha-phase model prediction are narrow which is likely due to the large number of experimental data points available in the alpha phase. Also note that the prediction of our model is not skewed by a number of discrepant measurements at high temperatures including Aru1972 and Cag2008.

Fig. 4(c) and (d) make it apparent that our beta-phase model prioritizes the experimental measurements of Par2003, Cez1974 and Ros2001. This observation is corroborated by the high values of the hyperparameters for these data sets in Table 2. The lower set of points of Par2003 are from measurements in the liquid phase; the temperature ranges of the different phases overlap as measurements were taken in the metastable regime. In contrast, Arb2014 selected a linear model calibrated solely on the enthalpy measurements of Ros2001 (note that the enthalpy difference in the model of Arb2014 may

Please cite this article as: N.H. Paulson, E. Jennings and M. Stan, Bayesian strategies for uncertainty quantification of the thermodynamic properties of materials, International Journal of Engineering Science, https://doi.org/10.1016/j.ijengsci.2019.05.011

**Table 2**
The mean, standard deviation, 2.5th percentile bound and 97.5th percentile bound are presented for each parameter for the best model for each phase.

| Parameter | Mean | Std. dev. | 2.5% CI | 97.5% CI |
|---|---|---|---|---|
| **$\alpha$-phase** | | | | |
| $\theta_D$ | 206.45 | 0.28 | 205.93 | 207.02 |
| $a_2$ | $1.41 \times 10^{-3}$ | $7.58 \times 10^{-5}$ | $1.26 \times 10^{-3}$ | $1.56 \times 10^{-3}$ |
| $a_3$ | $9.71 \times 10^{-6}$ | $2.29 \times 10^{-7}$ | $9.26 \times 10^{-6}$ | $1.02 \times 10^{-5}$ |
| $a_4$ | $-5.61 \times 10^{-9}$ | $2.44 \times 10^{-10}$ | $-6.07 \times 10^{-9}$ | $-5.12 \times 10^{-9}$ |
| $a_5$ | $1.10 \times 10^{-12}$ | $7.63 \times 10^{-14}$ | $9.41 \times 10^{-13}$ | $1.24 \times 10^{-12}$ |
| $\alpha_{Ade1952}$ Adenstedt (1952) | 1.111 | 0.532 | 0.348 | 2.402 |
| $\alpha_{Aru1972}$ Arutyunov, Banchila, and Filippov (1972) | 0.569 | 0.138 | 0.340 | 0.876 |
| $\alpha_{Bur1958}$ Burk, Estermann, and Friedberg (1958) | 0.188 | 0.023 | 0.147 | 0.237 |
| $\alpha_{Cag2008}$ Cagran, Hüpf, Wilthan, and Pottlacher (2008) | 0.533 | 0.217 | 0.210 | 1.046 |
| $\alpha_{Cez1974}$ Cezairliyan and McClure (1975) | 4.137 | 0.959 | 2.613 | 6.343 |
| $\alpha_{Col1971}$ Collings and Ho (1971) | 0.209 | 0.040 | 0.140 | 0.295 |
| $\alpha_{Fie1961}$ Fieldhouse and Lang (1961) | 0.444 | 0.118 | 0.250 | 0.709 |
| $\alpha_{Fil1971}$ Filippov and Yurchak (1971) | 0.640 | 0.163 | 0.367 | 1.002 |
| $\alpha_{Gol1970}$ Golutvin and Maslennikova (1970) | 0.703 | 0.152 | 0.449 | 1.043 |
| $\alpha_{Haw1963}$ Hawkins, Onillon, and Orr (1963) | 1.512 | 0.234 | 1.101 | 2.027 |
| $\alpha_{Kats1985}$ Kats (1985) | 0.317 | 0.080 | 0.183 | 0.497 |
| $\alpha_{Kne1963}$ Kneip Jr et al. (1963) | 0.004 | 0.001 | 0.003 | 0.005 |
| $\alpha_{Mil2006S1}$ Milošević and Maglić (2006) | 2.409 | 0.427 | 1.670 | 3.351 |
| $\alpha_{Mil2006S2}$ Milošević and Maglić (2006) | 11.272 | 2.225 | 7.361 | 16.081 |
| $\alpha_{Pel1971}$ Peletskii and Druzhinin (1971) | 1.060 | 0.183 | 0.744 | 1.457 |
| $\alpha_{McC1964}$ McClaine (1964) | 0.053 | 0.008 | 0.038 | 0.071 |
| $\alpha_{Wol1957}$ Wolcott (1957) | 0.138 | 0.018 | 0.106 | 0.175 |
| **$\beta$-phase** | | | | |
| $b_0$ | $-3.36 \times 10^4$ | $4.27 \times 10^3$ | $-4.17 \times 10^4$ | $-2.44 \times 10^4$ |
| $b_1$ | 77.43 | 5.92 | 64.69 | 88.71 |
| $b_2$ | $-4.59 \times 10^{-2}$ | $5.48 \times 10^{-3}$ | $-5.63 \times 10^{-2}$ | $-3.42 \times 10^{-2}$ |
| $b_3$ | $1.20 \times 10^{-5}$ | $1.00 \times 10^{-6}$ | $9.00 \times 10^{-6}$ | $1.50 \times 10^{-5}$ |
| $\alpha_{Cag2008}$ Cagran et al. (2008) | 0.717 | 0.260 | 0.315 | 1.324 |
| $\alpha_{Cez1974}$ Cezairliyan and McClure (1975) | 4.144 | 1.017 | 2.507 | 6.475 |
| $\alpha_{Fil1971}$ Filippov and Yurchak (1971) | 0.616 | 0.472 | 0.075 | 1.834 |
| $\alpha_{Kat1985}$ Kats (1985) | 0.123 | 0.040 | 0.059 | 0.215 |
| $\alpha_{Mil2006S1}$ Milošević and Maglić (2006) | 0.792 | 0.285 | 0.350 | 1.451 |
| $\alpha_{Mil2006S2}$ Milošević and Maglić (2006) | 1.090 | 0.532 | 0.349 | 2.368 |
| $\alpha_{Par2003}$ Paradis, Ishikawa, and Yoda (2003) | 6.364 | 2.249 | 2.816 | 11.411 |
| $\alpha_{Pel1971}$ Peletskii and Druzhinin (1971) | 1.025 | 0.343 | 0.501 | 1.846 |
| $\alpha_{Ros2001}$ Rösner-Kuhn, Drewes, Franz, and Frohberg (2001) | 2.368 | 0.376 | 1.718 | 3.171 |
| **liquid-phase** | | | | |
| $c_0$ | $5.86 \times 10^3$ | $1.76 \times 10^3$ | $2.46 \times 10^3$ | $9.30 \times 10^3$ |
| $c_1$ | 29.20 | 1.02 | 27.18 | 31.15 |
| $c_2$ | $5.28 \times 10^{-3}$ | $2.97 \times 10^{-4}$ | $4.73 \times 10^{-}3$ | $5.88 \times 10^{-}3$ |
| $\alpha_{Kor2005}$ Korobenko, Polyakova, and Savvatimskii (2005) | 1.880 | 0.627 | 0.946 | 3.397 |
| $\alpha_{Par2003}$ Paradis et al. (2003) | 0.120 | 0.043 | 0.055 | 0.222 |
| $\alpha_{Cag2008}$ Cagran et al. (2008) | 5.183 | 1.450 | 2.839 | 8.403 |
| $\alpha_{Ros2001}$ Rösner-Kuhn et al. (2001) | 0.713 | 0.126 | 0.496 | 0.991 |

result from differences in the Zr content correction). This explains why the two beta-phase models follow the same trend in enthalpy but are highly discrepant in specific heat. Furthermore, the linear model of Arb2014 for enthalpy forces the specific heat to be constant and to take a value which does not pass through the majority of the experimental specific heat measurements. The difference between our model and the HSC chemistry assessment is small, as the HSC Chemistry software appears to have incorporated specific heat measurements. Uncertainty intervals are broader for the beta-phase model due to the fewer number of data points and their significant spread.

Based on Fig. 4(e) and (f), the Bayesian framework prioritizes the measurements of Kor2005 and Cag2008 for the construction of the liquid-phase model and downweights the measurements of Par2003. As expected, these datasets have the largest and smallest hyperparameter values, respectively, in Table 2. As for the beta-phase, Arb2014 selected a linear model for the liquid-phase based on the measurements from Ros2001, resulting in a constant model for specific heat which is highly discrepant from our model. HSC Chemistry also selects a different model, albeit with a slope that is more consistent with the measurements of Kor2005 for specific heat. The origins of the discrepant enthalpy shift in the HSC Chemistry assessment is unknown, but we estimate that this is due to choices made when correcting for Zr content. The uncertainty intervals of our model are larger than those of the alpha- or beta-phase models due to the relative scarcity of experimental measurements.

Fig. 5 shows the predicted alpha phase model at very low temperatures. The alpha-phase model does a reasonable job fitting the low temperature specific heat measurements. That said, the Debye model that dominates in this temperature

Please cite this article as: N.H. Paulson, E. Jennings and M. Stan, Bayesian strategies for uncertainty quantification of the thermodynamic properties of materials, International Journal of Engineering Science, https://doi.org/10.1016/j.ijengsci.2019.05.011

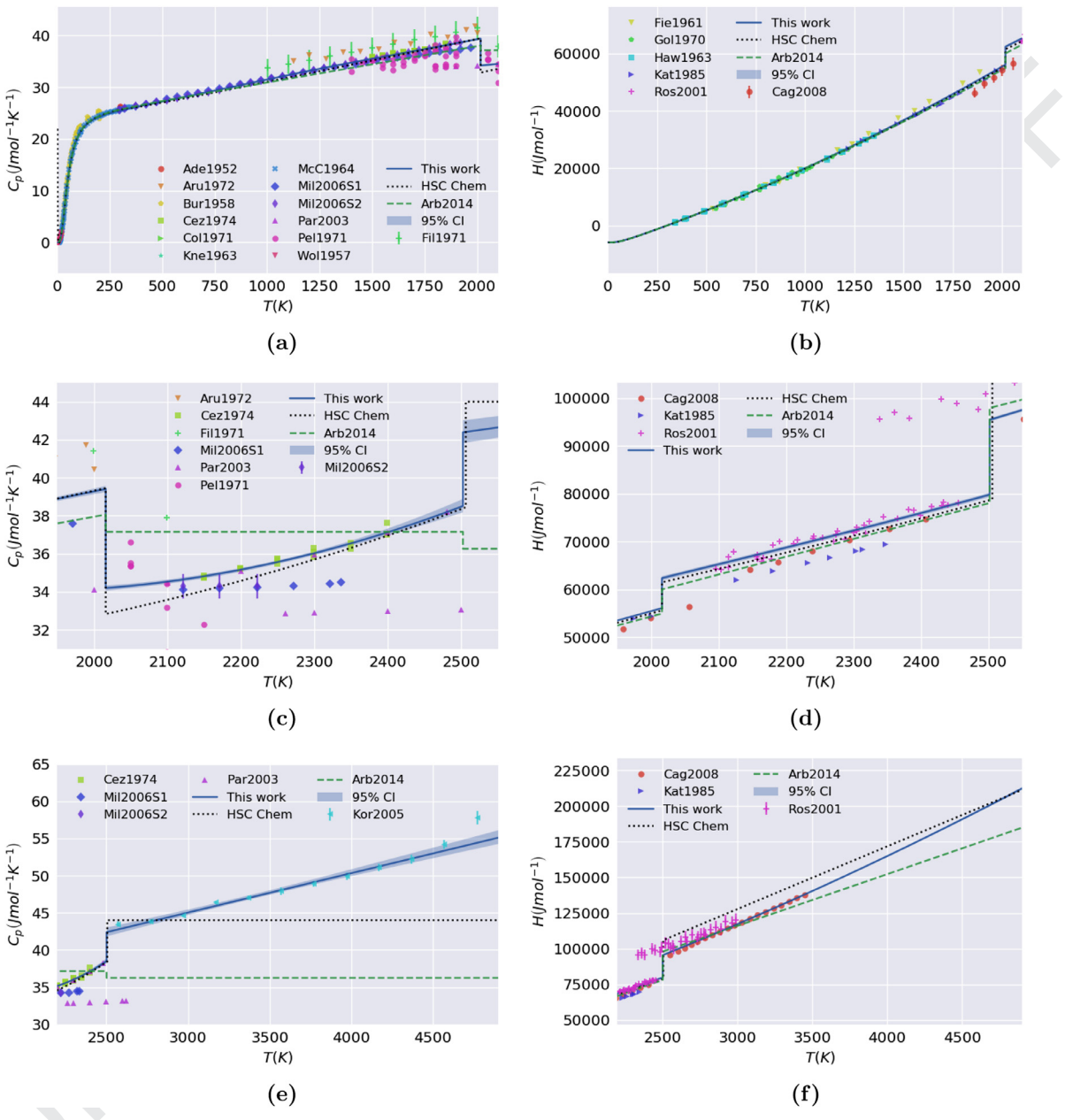


**Fig. 4.** The Bayesian model predictions for each of the phases versus temperature for specific heat (a, c, e) and enthalpy (b, d, f) are plotted as a solid blue line. The results of previous Hf assessments from Arb2014 and HSC Chem are plotted as dashed green lines and dotted black lines respectively. The experimental data points are plotted as markers and labeled in the legend. The uncertainty intervals have been rescaled by the associated best fit hyperparameter. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

regime does not have the flexibility to exactly fit the experimental data. An artifact of this discrepancy is that the hyperparameters for the low temperature datasets are close to zero. For example, the mean value of the marginalized hyperparameter for Kne1963 is 0.004, implying that uncertainties were under-reported. Since these hyperparameter values result from model inadequacy, they do not necessarily represent uncertainties in the experimental dataset with respect to the true thermodynamic properties. Our optimized value of the Debye temperature parameter, $\theta_D = 206.45 \pm 0.48$ K, deviates significantly from what is reported by other authors (250 K) Arblaster (2014). This discrepancy results from employing the full Debye model across the entire temperature range for the alpha phase, whereas previous researchers determined the Debye temperature from a linearization of the specific heat versus temperature below 20 K based on the low temperature limit of the Debye relationship (Kneip Jr, Betterton Jr, & Scarbrough, 1963). To confirm that this is the origin of the discrepancy rather than some implicit error in the Bayesian approach, we have implemented the traditional analysis technique within

Please cite this article as: N.H. Paulson, E. Jennings and M. Stan, Bayesian strategies for uncertainty quantification of the thermodynamic properties of materials, International Journal of Engineering Science, https://doi.org/10.1016/j.ijengsci.2019.05.011

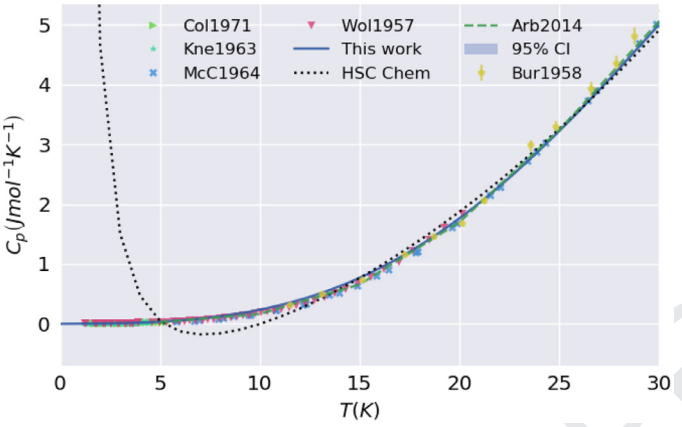


**Fig. 5.** The Bayesian model prediction (solid blue line), the results of previous Hf assessments (dashed lines) and the experimental data points (given by symbols in the legend) versus temperature. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

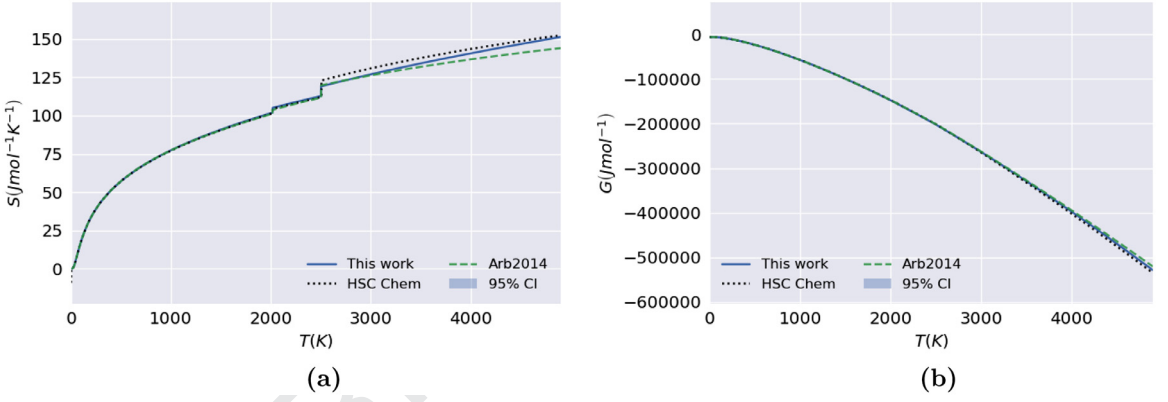


**Fig. 6.** The model predictions for all phases and the previous Hf assessments are plotted versus temperature in Kelvin for (a) entropy and (b) Gibbs free energy.

our Bayesian framework and find a best fit value of $\theta_D = 253.07 \pm 0.26$ which agrees with previous results Arblaster (2014). We find that the Debye temperature of the selected alpha-phase model is not dependent on the temperature range considered (i.e. our Debye temperature parameter does not significantly change when only measurements at 100K and below are included). In future work, we plan to employ more sophisticated relationships in the low temperature region. Our low temperature model compares reasonably well with Arb2014 but shows significant discrepancies with the HSC Chemistry calculations. We attribute the differences to the negative specific heats predicted by HSC Chemistry at around 7 K.

Fig. 6(a) and (b) compare our results for entropy and Gibbs free energy for all phases against the assessments of HSC Chemistry and Arb2014. The results of the three assessments appear to be reasonably close in the alpha and beta regimes, with differences in slope and vertical placement becoming more apparent at high temperatures.

To further investigate the differences between the three assessments, we present the percentage differences between the HSC Chemistry and Arb2014 results and our model in Fig. 7. Each phase is plotted separately due to the significant differences in scale on the x- and y- axes. In general, deviations decrease in severity for quantities in the order of specific heat, enthalpy, entropy and Gibbs free energy. This is not surprising due to the mathematical relationships between the thermodynamic quantities. Also, in most cases, the Arb2014 result is more discrepant than that of HSC Chemistry. For most of the alpha-phase temperature range, deviations are less than 4%, however, at very low temperatures the deviation between our model and Arb2014 approaches 30% due to the inadequacy of the Debye model as previously discussed. We do not evaluate the discrepancy for the HSC Chemistry result below 40 K due to its highly anomalous behavior. Deviations between our model and the Arb2014 assessment in Cp reach 8% for the beta-phase due to the constant behavior of the Arb2014 model. Discrepancies are even larger in the liquid-phase for both HSC Chemistry and Arb2014.

Finally, we evaluate the effect of removing individual datasets on the Bayesian analysis. For each phase, we evaluate an equal number of models as the number of relevant datasets. Each iteration, we remove a single dataset and perform the Bayesian analysis. We directly employ the model forms of Eq. (14) in this analysis, so it is possible that different model forms would be more appropriate depending on the specific dataset dropped in each analysis. Fig. 8 highlights the results

Please cite this article as: N.H. Paulson, E. Jennings and M. Stan, Bayesian strategies for uncertainty quantification of the thermodynamic properties of materials, International Journal of Engineering Science, https://doi.org/10.1016/j.ijengsci.2019.05.011

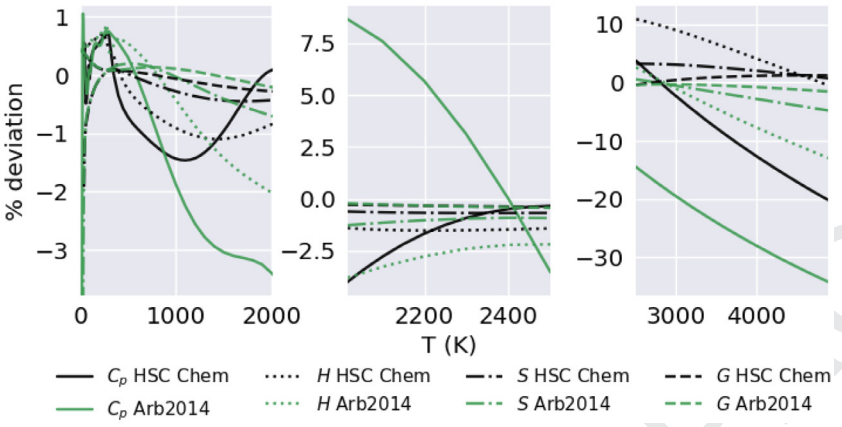


**Fig. 7.** The percentage differences between the assessments of HSC Chemistry and Arb2014 and the Bayesian model predictions given in this work are presented for each phase and thermodynamic quantity versus temperature.

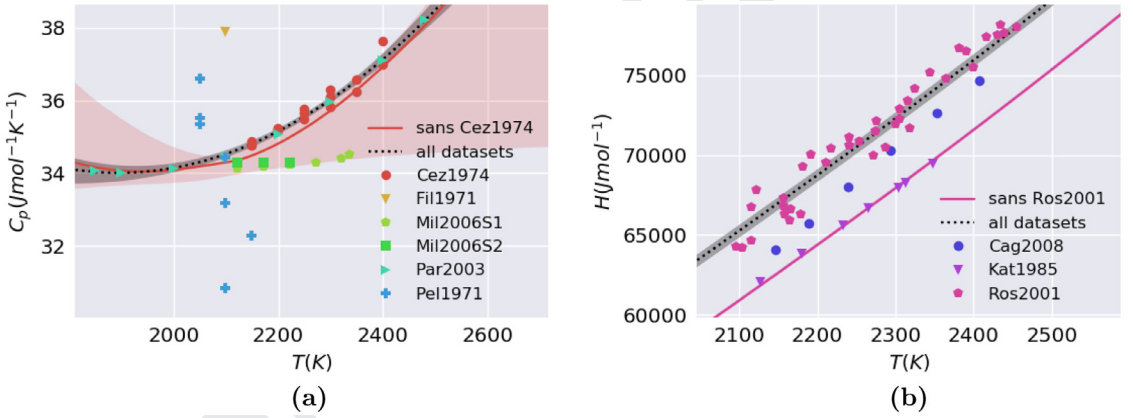


**Fig. 8.** The beta-phase models and 95% uncertainty intervals resulting from Bayesian analyses are plotted versus temperature for (a) specific heat disregarding Cez1974, and (b) enthalpy disregarding Ros2001. The overall beta-phase prediction and 95% uncertainty intervals are also plotted alongside the experimental measurements versus temperature.

of this analysis for the beta-phase. Fig. 8(a) presents the effects of removing the measurements of Cez1974 (highly favored by the framework in the original Bayesian analysis) for specific heat. When the dataset is removed, the 50th percentile model behavior changes moderately and the 95th percentile uncertainty interval signals a drastically increased uncertainty in the model parameters. Interestingly, the Mil2006S1 and Par2003 hyperparameter distributions for this model (not presented) are bimodal, indicating that the Bayesian analysis results in comparable probabilities of these two highly-discrepant datasets having low uncertainty (i.e. being correct). Finally, Fig. 8(b) presents the effect of removing the measurements of Ros2001 on the beta-phase model for enthalpy. While the slope does not appreciably change, the Bayesian framework prioritizes the measurements of Kats1985, resulting in a significant change in the enthalpy of transformation. We observe similar changes in the alpha and liquid phase models when dropping datasets which the framework originally prioritizes. These changes highlight the importance of considering all available measurements in the assessment, and motivate additional computational analyses to help establish expected trends in material properties.

The thermodynamic assessment of Hafnium metal shows that our single framework directly accepts raw, experimental measurements of specific heat and enthalpy, and uses them to simultaneously calibrate models, produce evaluation metrics and provide uncertainty intervals for the parameters and model predictions. Furthermore, the framework can naturally account for tasks that are typically restricted to experts, such as managing outliers, rescaling inaccurate reported uncertainties, weighting datasets and finding consistency among multiple datasets expressed via different quantities. Through examination of the Posterior distributions, we can provide both model parameter uncertainties as well as model comparison using Bayes Factors.

Please cite this article as: N.H. Paulson, E. Jennings and M. Stan, Bayesian strategies for uncertainty quantification of the thermodynamic properties of materials, International Journal of Engineering Science, https://doi.org/10.1016/j.ijengsci.2019.05.011

## 4. Summary and conclusions

In this work we introduce a Bayesian framework for the selection, calibration and quantification of uncertainty of thermodynamic property models. We present the Bayesian framework in detail, and describe modifications to address common data issues in thermodynamic assessment including the presence of outliers, systematic errors, inaccurately reported uncertainties and inconsistencies between thermodynamic quantities. Enabled by cutting-edge numerical sampling algorithms, the framework simultaneously performs these tasks in a single analysis. We accompany the framework and modifications with example problems that clearly demonstrate each method and its benefits. Perhaps most critically, the Bayesian framework provides the researcher with a clear and self-consistent method to perform assessment. Data sets, irrespective of the thermodynamic quantity measured, can be incorporated into the framework without modification, which then determines the best weights and parameters given the model. Our confidence in the best fitting parameters are described by the Posterior distributions.

We culminate the study with an assessment of the thermodynamic properties of Hafnium metal from 0 to 4900K through three distinct phases. The resulting models are broadly effective over a wide range of temperatures and come with parameter and prediction uncertainty intervals. We find that the Debye model plus a quartic polynomial performs best to model the specific heat versus temperature for the low temperature alpha phase. The model fits the data well, except at very low temperatures where the Debye model has deficiencies in modeling the thermodynamic properties. This model compares well with those of HSC Chemistry and Arb2014 through the majority of the temperature range, showing the largest percent deviations with Arb2014 at higher temperatures. We find that a quadratic model for specific heat for the beta phase, captures the trends of the Cez1974, Par2003 and Ross2001 measurements. While this model is comparable to the HSC Chemistry prediction, larger deviations are seen versus the Arb2014 model as that work employed a linear model for enthalpy versus temperature and therefore a constant model for specific heat. Our results indicate that a linear model performs well at describing the temperature dependence of specific heat in the liquid phase, aligning most closely with the Kor2004 and Cag2008 data sets. Our liquid phase model shows significant discrepancies with HSC Chemistry and Arb2014 for specific heat as both employ linear models to describe the dependence of enthalpy versus temperature, and therefore have constant specific heat versus temperature. These discrepancies highlight the Bayesian framework's ability to select and calibrate models, optimally leveraging both specific heat and enthalpy measurements.

We hope that the community will take advantage of these publicly available tools to enhance the assessment of a variety material properties. The approaches presented in this work have direct application to future efforts to improve the assessment of binary, ternary and multi-component systems, lessening the burden on the expert in managing unwieldy collections of data.

## Data availability

Q3 **This section will be filled out prior to publication.**

## Acknowledgments

**This section is left blank to ensure anonymity.**

## Appendix A

In this section, we demonstrate Bayesian model selection, calibration and uncertainty quantification through a simple example problem. We employ synthetic data because it provides a convenient ground truth for comparison. In this example, we generate 20 pairs of x and y values according to the following relationships,

$$x \sim \mathcal{U}([0, 1]), \tag{15a}$$

$$\varepsilon \sim \mathcal{N}(0, 0.05), \tag{15b}$$

$$y = 1 + x + \varepsilon, \tag{15c}$$

where $\mathcal{U}$ is the uniform distribution between [0, 1) and $\mathcal{N}$ is the Normal (Gaussian) distribution with mean 0 and variance 0.05.

In the first step of the Bayesian analysis we select a set of potential models. One of the strengths of this framework is that models of any form with different numbers of parameters may be compared. In this example, we select polynomials of order $N$ represented as follows,

$$M(x, \boldsymbol{\Theta}) = \sum_{i=0}^{N} \Theta_i x^i, \tag{16}$$

Please cite this article as: N.H. Paulson, E. Jennings and M. Stan, Bayesian strategies for uncertainty quantification of the thermodynamic properties of materials, International Journal of Engineering Science, https://doi.org/10.1016/j.ijengsci.2019.05.011

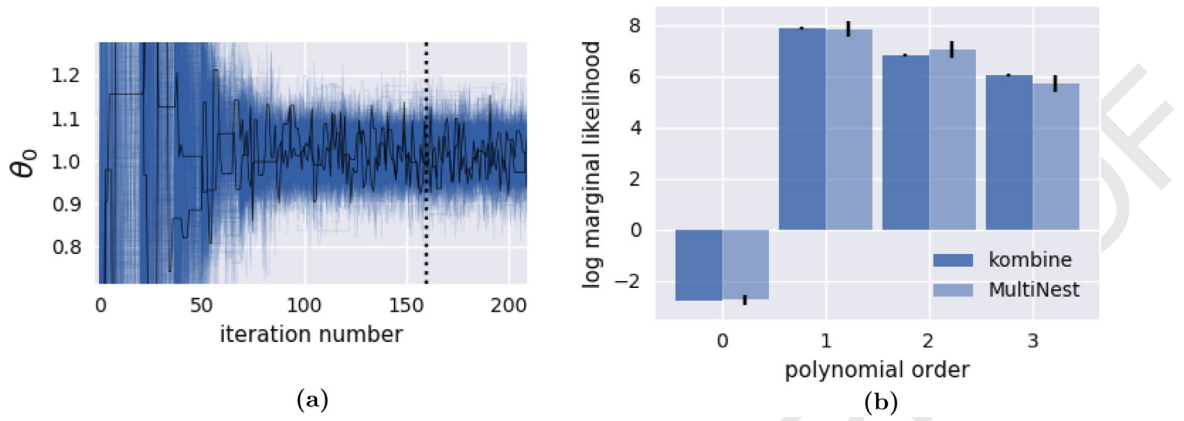


**Fig. 9.** (a) Sequences of 400 walkers from the kombine sampler over 150 iterations. Two randomly selected sequences are highlighted in black. All samples to the left of the black dashed line are discarded. (b) The marginal Likelihoods and 2-sigma standard deviations are plotted for both samplers and all models.

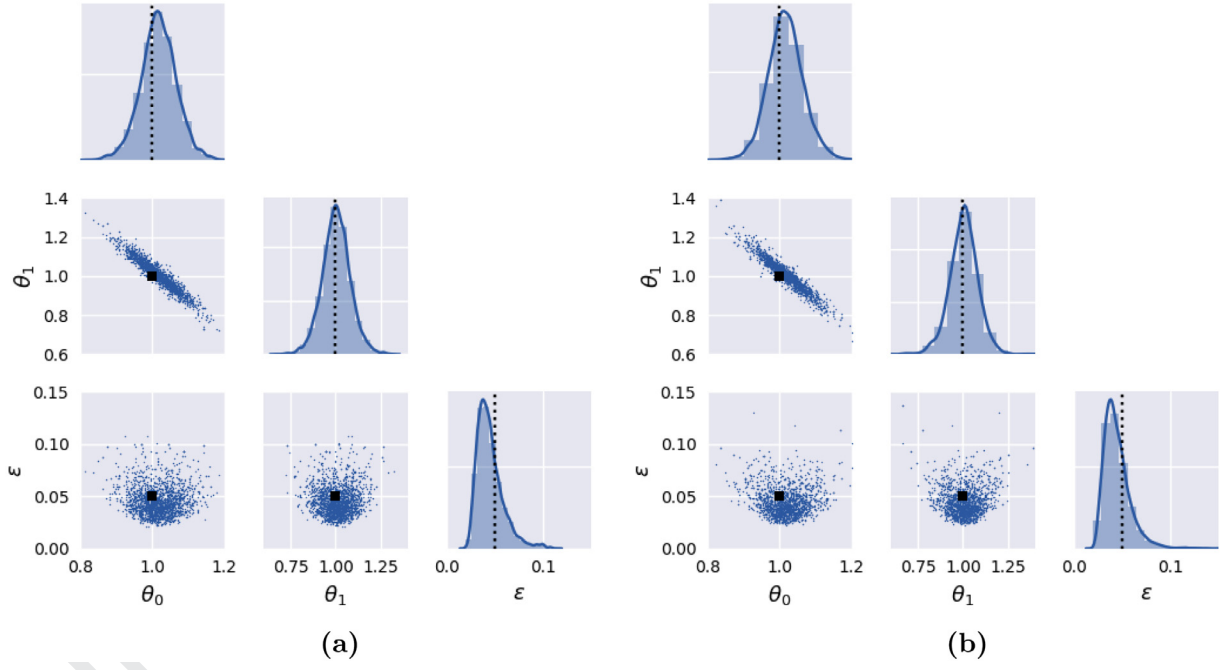


**Fig. 10.** Histograms, kernel density estimated distributions and scatter plots of the posterior distributions obtained via (a) kombine, and (b) MultiNest, respectively. The true parameter values are marked as a black dashed line.

where $i$ is the exponent of $x$ and indexes the model parameter vector. Next, we define the Likelihood and priors on the model parameters. We assume independence of the data points and write the Likelihood as the product of the individual Gaussian Likelihoods for each data point,

$$P(\boldsymbol{D}|\boldsymbol{\Theta}, \varepsilon, M) = \prod_j \mathcal{N}\big(y_j|M\big(x_j, \boldsymbol{\Theta}\big), \varepsilon\big), \tag{17}$$

where $j$ indexes each data point and $\varepsilon$ is the standard uncertainty parameter, which is included in the Bayesian analysis and captures the standard deviation of the data generation process. Note that, in practice, the product is replaced by a summation of log-Likelihoods; this alleviates the issues of arithmetic underflow and overflow that are common when multiplying large sets of probability densities. Priors are chosen based on the best available knowledge about the parameters and the relationships between parameters. In this simple example we select uniform priors as follows,

$$P(\Theta_i|M) = \mathcal{U}(\Theta_i| -2, 2), \tag{18a}$$

Please cite this article as: N.H. Paulson, E. Jennings and M. Stan, Bayesian strategies for uncertainty quantification of the thermodynamic properties of materials, International Journal of Engineering Science, https://doi.org/10.1016/j.ijengsci.2019.05.011

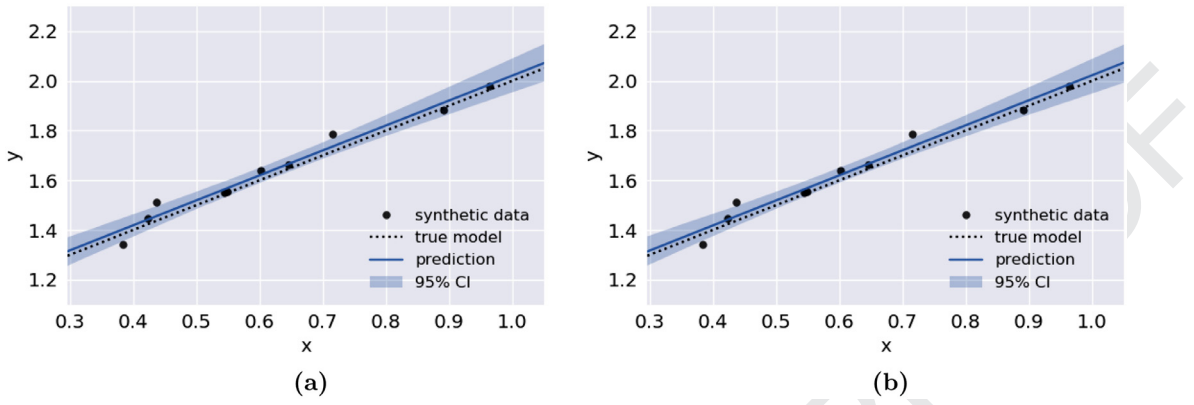


**Fig. 11.** The model prediction, uncertainty intervals, true model and synthetic data points are plotted versus x for (a) kombine and (b) MultiNest, respectively.

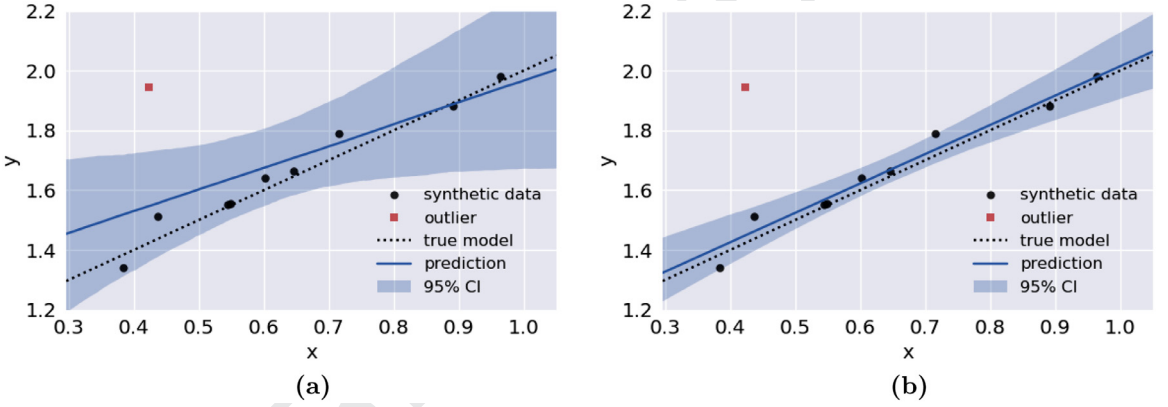


**Fig. 12.** The model prediction, uncertainty interval, true model and synthetic data points are plotted with (a) Normal Likelihood (b) Student-t Likelihood with $2 + 10^{-6}$ degrees of freedom.

$$P(\varepsilon|M) = \mathcal{U}(\varepsilon|0, 1). \tag{18b}$$

By defining the Likelihood and priors we have specified the Bayesian model. Even though there are different numerical techniques available to evaluate the Posterior, the result should not depend on the sampler. As a practical demonstration, we employ both samplers to evaluate the posterior distributions and marginal Likelihoods for polynomial models of order 0 through 3. First, we initialize 400 points (denoted walkers and active points for kombine and MultiNest, respectively) from the prior distributions. We monitor the kombine sequences for convergence based on the GR metric introduced Section 2.2. Fig. 9(a) shows the paths of the walkers in kombine for the first parameter in the 1st order polynomial model. Note that the sampler automatically discards all samples to the left of the black dashed line to allow the walkers to begin randomly exploring the Posterior area (i.e. burn-in). Fig. 9(b) plots the marginal Likelihoods and $2\sigma$ standard deviation bounds for each sampler and polynomial model. Based on the computed Bayes Factors, the Evidence for the linear model is weak versus the quadratic, positive versus the cubic, and very strong versus the constant model. Consequently, it would be reasonable to select the linear model as it exhibits the highest marginal Likelihood and has fewer parameters than the next best (the quadratic) model. Furthermore, the marginal Likelihoods agree within the uncertainties for both samplers.

Finally, we examine the Posterior distribution and uncertainty intervals. Fig. 10 summarizes the posterior distributions obtained via the kombine (left) and MultiNest (right) samplers. We plot histograms and KDE distributions for each univariate parameter distribution. We provide multivariate summaries for each pair of coefficients via scatter plots of 2000 randomly selected samples. True parameter values are indicated by black markers. We first note that the samplers produce similar posterior parameter distributions. Furthermore, regions of high density in the Posterior closely align with the true parameter values. Fig. 11 displays the model prediction and 95th percentile uncertainty intervals obtained via the kombine (left) and MultiNest (right) samplers versus the synthetic data points. The uncertainty intervals are obtained by randomly sampling the Posterior distributions of the model parameters and computing the 2.5th and 97.5th percentile levels of the resulting model

Please cite this article as: N.H. Paulson, E. Jennings and M. Stan, Bayesian strategies for uncertainty quantification of the thermodynamic properties of materials, International Journal of Engineering Science, https://doi.org/10.1016/j.ijengsci.2019.05.011

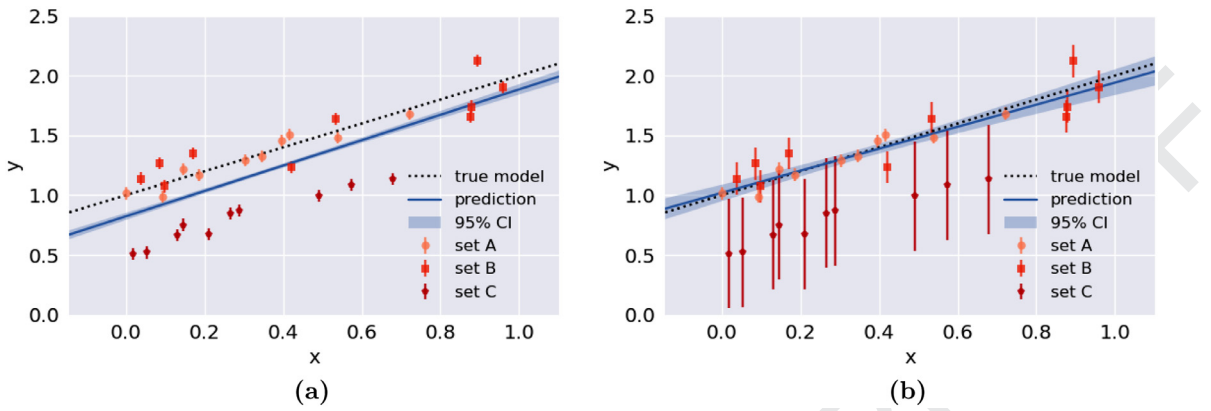


**Fig. 13.** The model prediction, uncertainty intervals and synthetic data points are plotted versus x for the posterior distributions (a) employing the reported uncertainties, and (b) weighing the uncertainties using hyperparameters. Note that for visual effect we have rescaled the uncertainty intervals in (b) by the best fit hyperparameters.

predictions. The uncertainty intervals show the expected spread of model predictions from the distribution of $\theta_0$ and $\theta_1$. The models fit the synthetic data well, and both the model predictions and uncertainty intervals are nearly indistinguishable between the two samplers. The results of Figs. 9(b)–11 give us confidence that different samplers obtain the same posterior parameter distributions given a single set of Likelihood and prior definitions.

## Appendix B

In this section we demonstrate the ability of the Bayesian framework to accommodate outliers through the analysis of a data set obtained by perturbing a single datum from Appendix A, shown as a red square in Fig. 12. Fig. 12(a) shows the results of this analysis using the Likelihood given in Eq. (17). Clearly, the model prediction is biased by the outlier, and the uncertainty intervals show significant spreads in the slope and intercept. In this example, we replace the Normal distribution with a Students-t distribution and $2 + 10^{-6}$ for the degrees of freedom. Clearly, by employing Students-t distribution for the Likelihood definition the robustness to the outlier is greatly improved. The model prediction is comparable to that from Fig. 11 of the previous section.

## Appendix C

In this Section, we demonstrate the approach described in Section 2.4 to rescale the uncertainties of datasets and accommodate systematic errors. We use three synthetic data sets of 10 points each. We generate Set A via Eq.(15) and assign uncertainty intervals equal to the standard deviation of the $\varepsilon$ standard uncertainty parameter (equal to 0.05). In this example we will assume that the reported uncertainties for Set A are accurate. The points in Set B follow the same trend, but with a 0.2 standard uncertainty (4 times larger than for Set A). The reported uncertainties in Set B underestimate the standard uncertainty by a factor of 4. We generate Set C with an intercept 0.5 less than the true value and a 0.05 standard uncertainty (also reported as 0.05). We then perform the analysis with and without the use of hyperparameters. In the method without hyperparameters, we directly use the reported standard uncertainties in the Likelihood definition (we do not need to employ a parameter for the data variance as in Eq. (17)). In the hyperparameter method, we initialize one hyperparameter for each of the three data sets.

Fig. 13(a) and (b) display the results of these analyses with and without hyperparameters. Without hyperparameters, the best fit model lies between data sets A and B, which follow the same trend, and set C, which has a different intercept. Furthermore, the model displays misleadingly narrow uncertainty intervals, especially considering that it does not represent any of the data sets well. In contrast, the best fit model using hyperparameters, shown in Fig. 13(b), has larger uncertainty intervals which represent the data better, and more importantly, the best fit model now agrees with the true model used to generate the data in this example. In this figure we have scaled the uncertainty intervals on each dataset by the best fit hyperparameter values of 0.9, 0.4 and 0.1 for hyperparameters $\alpha_A$, $\alpha_B$ and $\alpha_C$, respectively. Fig. 14 displays the histograms and KDE distributions for each parameter and scatter plots for each pair of parameters. Each true parameter value (marked by the black dashed line) falls within the high probability regions of the associated posterior distributions. Considering the log marginal Likelihood for both models we find a values $-2.6$ for the model with hyperparameters and $-323.2$ without. In this case the Bayes Factor favors the model with hyperparameters.



Please cite this article as: N.H. Paulson, E. Jennings and M. Stan, Bayesian strategies for uncertainty quantification of the thermodynamic properties of materials, International Journal of Engineering Science, https://doi.org/10.1016/j.ijengsci.2019.05.011

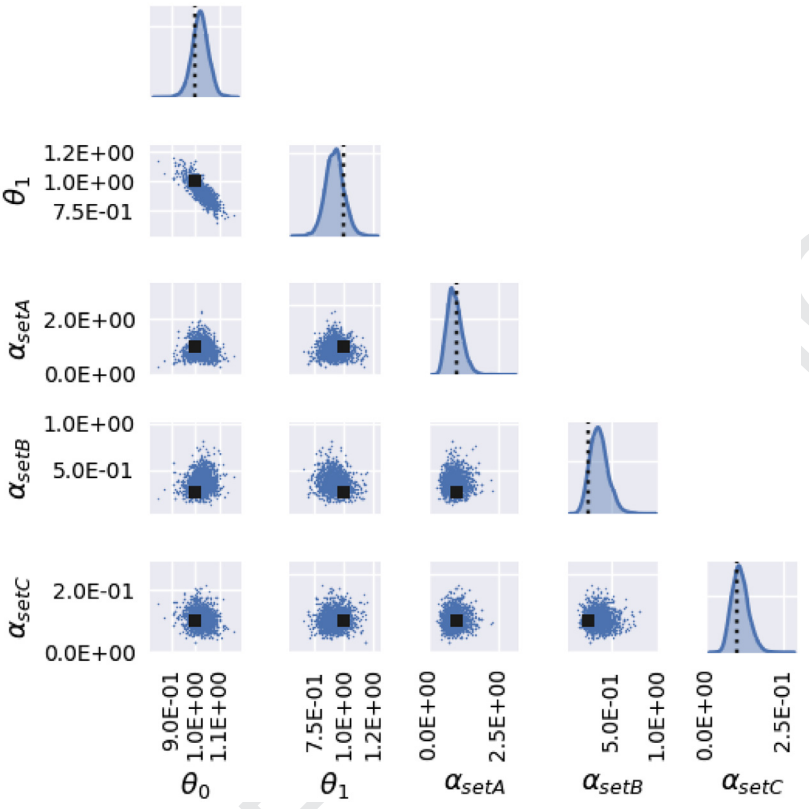


**Fig. 14.** The 2D marginal Posterior distributions obtained via the hyperparameter method. True parameter values are shown in a black dashed line.

## References

Adenstedt, H. K. (1952). Physical, thermal and electrical properties of Hafnium and high purity Zirconium. *Transactions of the American Society for Metals, 44*, 949–973.
Arblaster, J. W. (2013). Thermodynamic properties of Zirconium. *Calphad, 43*, 32–39. doi:10.1016/j.calphad.2013.07.015.
Arblaster, J. W. (2014). Thermodynamic properties of hafnium. *Journal of Phase Equilibria and Diffusion, 35*(4), 490–501. doi:10.1007/s11669-014-0319-5.
Arutyunov, A. V., Banchila, S. N., & Filippov, L. P. (1972). Thermal, electrical and emissive properties of Hf in the high-temperature region. *High Temperature, 10*(2), 375–377.
BIPM, I. E. C., IFCC, I., IUPAC, I., & ISO, O. (2008). Evaluation of measurement dataguide for the expression of uncertainty in measurement. JCGM 100: 2008. *Citado en las*, 167.
Blei, D. M., Kucukelbir, A., & McAuliffe, J. D. (2017). Variational inference: A review for statisticians. *Journal of the American Statistical Association, 112*(518), 859–877.
Buchner, J., Georgakakis, A., Nandra, K., Hsu, L., Rangel, C., Brightman, M., … Kocevski, D. (2014). X-ray spectral modelling of the AGN obscuring region in the CDFS: Bayesian model selection and catalogue. *\aap, 564*, A125. doi:10.1051/0004-6361/201322971.
Burk, D. L., Estermann, I., & Friedberg, S. A. (1958). The low temperature specific heats of Titanium, Zirconium and Hafnium. *Zhurnal Physical Chemistry (Munich), 16*, 183–193.
Cagran, C., Hüpf, T., Wilthan, B., & Pottlacher, G. (2008). Selected thermophysical properties of Hf-3% Zr from 2200 K to 3500 K obtained by a fast pulse-heating technique.. *High Temperatures–High Pressures, 37*(3).
Cezairliyan, A., & McClure, J. L. (1975). Simultaneous measurements of specific heat, electrical resistivity, and hemispherical total emittance by a pulse heating technique: Hafnium–3 (wt.%) zirconium, 1500 to 2400 K. *Journal Research National Bureau Stand., A, 79*(2), 431–436.
Chase, M. W., Ansara, I., Dinsdale, A., Eriksson, G., Grimvall, G., Hoglund, L., & Yokokawa, H. (1995). Group 1: Heat capacity models for crystalline phases from 0 K to 6000 K. *Calphad: Computer Coupling of Phase Diagrams and Thermochemistry, 19*(4), 437–447.
Chatterjee, N. D., Krüger, R., Haller, G., & Olbricht, W. (1998). The Bayesian approach to an internally consistent thermodynamic database: Theory, database, and generation of phase diagrams. *Contributions to Mineralogy and Petrology, 133*(1), 149–168. doi:10.1007/s004100050444.
Chatterjee, N. D., Miller, C., & Olbricht, W. (1994). Bayes estimation: A novel approach to derivation of internally consistent thermodynamic data for minerals, their uncertainties, and correlations. Part II: Application. *Physics and Chemistry of Minerals, 21*(1–2), 50–62.
Chib, S., & Greenberg, E. (1995). Understanding the metropolis-Hastings algorithm. *The American Statistician, 49*(4), 327–335. doi:10.1080/00031305.1995.10476177.
Collings, E. W., & Ho, J. C. (1971). Magnetic-susceptibility and low-temperature specific-heat studies of Ti, Zr, and Hf. *Physical Review B, 4*(2), 349–356. doi:10.1103/PhysRevB.4.349.
Cristescu, S., & Simon, F. (1934). Die Spezifischen Wärmen von Beryllium, Germanium und Hafnium bei tiefen temperaturen. *Zeitschrift für Physikalische Chemie, 25*(1), 273–282.
Dinsdale, A. T. (1991). SGTE data for pure elements. *Calphad, 15*(4), 317–425. doi:10.1016/0364-5916(91)90030-N.
Duong, T. C., Hackenberg, R. E., Landa, A., Honarmandi, P., Talapatra, A., Volz, H. M., … Arróyave, R. (2016). Revisiting thermodynamics and kinetic diffusivities of uraniumniobium with Bayesian uncertainty analysis. *Calphad, 55*, 219–230. doi:10.1016/j.calphad.2016.09.006.
Farr, B., & Farr, W. (2015). kombine: a kernel-density-based, embarrassingly parallel ensemble sampler.
Feroz, F., Hobson, M. P., Cameron, E., & Pettitt, A. N. (2013). Importance nested sampling and the MultiNest algorithm. *arXiv preprint arXiv:1306.2144*.

Please cite this article as: N.H. Paulson, E. Jennings and M. Stan, Bayesian strategies for uncertainty quantification of the thermodynamic properties of materials, International Journal of Engineering Science, https://doi.org/10.1016/j.ijengsci.2019.05.011

Fieldhouse, I. B., & Lang, J. I. (1961). Measurement of thermal properties. *Technical Report*. Armour Research Foundation Chicago IL.
Filippov, L. P., & Yurchak, R. P. (1971). High-temperature investigations of the thermal properties of solids. *Journal of Engineering Physics, 21*(3), 1209–1220.
Foreman-Mackey, D., Hogg, D. W., Lang, D., & Goodman, J. (2013). emcee: The MCMC hammer. *Publications of the Astronomical Society of the Pacific, 125*(925), 306.
Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2013). *Bayesian data analysis*. CRC press.
Goldberg, R. N., & Weir, R. D. (1992). Conversion of temperatures and thermodynamic properties to the basis of the International Temperature Scale of 1990 (Technical Report). *Pure and Applied Chemistry, 64*(10), 1545–1562.
Golutvin, Y. A., & Maslennikova, E. G. (1970). The heat capacity of metalic Hafnium. *Russian Metallurgy Metally, 5*, 129–135.
Goodman, J., & Weare, J. (2010). Ensemble samplers with affine invariance. *Commununications Appllied Mathematics Computer Science, 5*(1), 65–80. doi:10.2140/camcos.2010.5.65.
Grimvall, G. (1999). *Thermophysical properties of materials*. Elsevier.
Hawkins, D. T., Onillon, M., & Orr, R. L. (1963). High-temperature heat content of Hafnium.. *Journal of Chemical and Engineering Data, 8*(4), 628–629.
Hegde, R. I., Triyoso, D. H., Samavedam, S. B., & White, B. E. J. (2007). Hafnium zirconate gate dielectric for advanced gate stack applications. *Journal of Applied Physics, 101*(7), 74113. doi:10.1063/1.2716399.
Honarmandi, P., Duong, T. C., Ghoreishi, S. F., Allaire, D., & Arroyave, R. (2019). Bayesian uncertainty quantification and information fusion in CALPHAD-based thermodynamic modeling. *Acta Materialia, 164*, 636–647.
Hou, J. S., Guo, J. T., Wu, Y. X., Zhou, L. Z., & Ye, H. Q. (2010). Effect of Hafnium on creep behavior of a corrosion resistant Nickel base superalloy. *Materials Science and Engineering: A, 527*(6), 1548–1554. doi:10.1016/j.msea.2009.11.008.
Jansson, B. (1984). *Evaluation of parameters in thermochemical models using different types of experimental data simultaneously*.
Kass, R. E., & Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association, 90*(430), 773–795. doi:10.1080/01621459.1995.10476572.
Kats, S. (1985). Thermophysical properties of Zirconium and Hafnium athigh temperatures. *Teplofiz. Vys. Temp., 23*(2), 395–397.
Kattner, U. R. (1997). The thermodynamic modeling of multicomponent phase equilibria. *JOM, 49*(12), 14–19. doi:10.1007/s11837-997-0024-5.
Keller, H. W., Shallenberger, J. M., Hollein, D. A., & Hott, A. C. (1982). Development of Hafnium and comparison with other pressurized water reactor control rod materials. *Nuclear Technology, 59*(3), 476–482. doi:10.13182/NT82-A33005.
Kneip Jr, G. D., Betterton Jr, J. O., & Scarbrough, J. O. (1963). Low-temperature specific heats of Titanium, Zirconium, and Hafnium. *Physical Review, 130*(5), 1687.
Königsberger, E. (1991). Improvement of excess parameters from thermodynamic and phase diagram data by a sequential Bayes algorithm. *Calphad, 15*(1), 69–78. doi:10.1016/0364-5916(91)90027-H.
Korobenko, V. N., Polyakova, O. A., & Savvatimskii, A. I. (2005). Heat capacity of liquid Hafnium from the melting point to the boiling point at atmospheric pressure. *High Temperature, 43*(1), 38–44. doi:10.1007/s10740-005-0044-1.
Kruschke, J. K. (2013). Bayesian estimation supersedes the *t* test.10.1037/a0029146
Lahav, O., Bridle, S. L., Hobson, M. P., Lasenby, A. N., & Sodré, J. L. (2000). Bayesian hyper-parameters' approach to joint estimation: The Hubble constant from CMB measurements. *Monthly Notices of the Royal Astronomical Society, 315*(4), L45–L49. doi:10.1046/j.1365-8711.2000.03633.x.
Ma, Y.-Z., & Berndsen, A. (2014). How to combine correlated data setsA Bayesian hyperparameter matrix method. *Astronomy and Computing, 5*, 45–56. doi:10.1016/j.ascom.2014.04.005.
Malakhov, D. V. (1997). Confidence intervals of calculated phase boundaries. *Calphad, 21*(3), 391–400. doi:10.1016/S0364-5916(97)00039-4.
McClaine, T. A. (1964). Thermodynamic and kinetic studies for a refractory materials program. *Technical Report*. LITTLE (ARTHUR D) INC CAMBRIDGE MA.
Milošević, N. D., & Maglić, K. D. (2006). Thermophysical properties of solid phase hafnium at high temperatures. *International Journal of Thermophysics, 27*(2), 530–553. doi:10.1007/s10765-006-0045-2.
Olbricht, W., Chatterjee, N. D., & Miller, K. (1994). Bayes estimation: A novel approach to derivation of internally consistent thermodynamic data for minerals, their uncertainties, and correlations. Part I: Theory. *Physics and Chemistry of Minerals, 21*(1), 36–49. doi:10.1007/BF00205214.
Otis, R. A., & Liu, Z.-K. (2017). High-Throughput thermodynamic modeling and uncertainty quantification for ICME. *JOM, 69*(5), 886–892. doi:10.1007/s11837-017-2318-6.
Paradis, P.-F., Ishikawa, T., & Yoda, S. (2003). Non-contact measurements of the Thermophysical properties of Hafnium-3 Mass% Zirconium at high temperature. *International Journal of Thermophysics, 24*(1), 239–258. doi:10.1023/A:1022326618592.
Peletskii, V. E., & Druzhinin, V. P. (1971). Experimental study of some physical properties of Hafnium at high temperatures. *Teplofizika vysokikh temperatur, 9*(3), 539–545.
Rohatgi, A. (2011). WebPlotDigitizer. http://arohatgi.info/WebPlotDigitizer/app.
Roine, A. (2002). Outokumpu HSC chemistry for windows: Chemical reaction and equilibrium software with extensive thermochemical database. *Pori: Outokumpu Research OY*.
Roslyakova, I., Sundman, B., Dette, H., Zhang, L., & Steinbach, I. (2016). Modeling of Gibbs energies of pure elements down to 0 K using segmented regression. *Calphad, 55*, 165–180. doi:10.1016/j.calphad.2016.09.001.
Rösner-Kuhn, M., Drewes, K., Franz, H., & Frohberg, M. G. (2001). Enthalpy measurements of the solid high-temperature $\beta$-phase of Titanium and Zirconium by levitation drop calorimetry. *Journal of Alloys and Compounds, 316*(1), 175–178. doi:10.1016/S0925-8388(00)01509-7.
Silverman, B. W. (2018). *Density estimation for statistics and data analysis*. Routledge.
Skilling, J. (2004). Nested sampling. *AIP Conference Proceedings, 735*(1), 395–405. doi:10.1063/1.1835238.
Stan, M., & Reardon, B. J. (2003). A Bayesian approach to evaluating the uncertainty of thermodynamic data and phase diagrams. *Calphad, 27*(3), 319–323. doi:10.1016/j.calphad.2003.11.002.
Verde, L. (2010). Statistical methods in cosmology. In *Lectures on cosmology* (pp. 147–177). Springer.
Wang, Y.-l., Xiong, X., Zhao, X.-j., Li, G.-d., Chen, Z.-k., & Sun, W. (2012). Structural evolution and ablation mechanism of a Hafnium carbide coating on a C/C composite in an oxyacetylene torch environment. *Corrosion Science, 61*, 156–161. doi:10.1016/j.corsci.2012.04.033.
Wolcott, N. M. (1957). The atomic heats of T, Zirconium and Hafnium. *The Philosophical Magazine: A Journal of Theoretical Experimental and Applied Physics, 2*(22), 1246–1254. doi:10.1080/14786435708244013.



Please cite this article as: N.H. Paulson, E. Jennings and M. Stan, Bayesian strategies for uncertainty quantification of the thermodynamic properties of materials, International Journal of Engineering Science, https://doi.org/10.1016/j.ijengsci.2019.05.011