

## Full length article

## Bayesian uncertainty quantification and information fusion in CALPHAD-based thermodynamic modeling

P. Honarmandi <sup>a, \*</sup>, T.C. Duong <sup>a, b</sup>, S.F. Ghoreishi <sup>b</sup>, D. Allaire <sup>b</sup>, R. Arroyave <sup>a, b</sup><sup>a</sup> Department of Materials Science and Engineering, Texas A&M University, College Station, TX, USA<sup>b</sup> Department of Mechanical Engineering, Texas A&M University, College Station, TX, USA

## ARTICLE INFO

## Article history:

Received 11 June 2018

Received in revised form

3 November 2018

Accepted 4 November 2018

Available online 12 November 2018

## Keywords:

CALPHAD

Bayes' theorem

Metropolis-Hastings Markov chain Monte Carlo

Information fusion

Bayesian model selection/averaging

## ABSTRACT

Calculation of phase diagrams is one of the fundamental tools in alloy design—more specifically under the framework of Integrated Computational Materials Engineering. Uncertainty quantification of phase diagrams is the first step required to provide confidence for decision making in property- or performance-based design. As a manner of illustration, a thorough probabilistic assessment of the CALPHAD model parameters is performed against the available data for a Hf–Si binary case study using a Markov Chain Monte Carlo sampling approach. The plausible optimum values and uncertainties of the parameters are thus obtained, which can be propagated to the resulting phase diagram. Using the parameter values obtained from deterministic optimization in a computational thermodynamic assessment tool (in this case Thermo-Calc) as the prior information for the parameter values and ranges in the sampling process is often necessary to achieve a reasonable cost for uncertainty quantification. This brings up the problem of finding an appropriate CALPHAD model with high-level of confidence which is a very hard and costly task that requires considerable expert skill. A Bayesian hypothesis testing based on Bayes' factors is proposed to fulfill the need of model selection in this case, which is applied to compare four recommended models for the Hf–Si system. However, it is demonstrated that information fusion approaches, i.e., Bayesian model averaging and an error correlation-based model fusion, can be used to combine the useful information existing in all the given models rather than just using the best selected model, which may lack some information about the system being modelled.

© 2018 Acta Materialia Inc. Published by Elsevier Ltd. All rights reserved.

## 1. Introduction

Uncertainty quantification (UQ) and its propagation (UP) across multi-scale model/experiment chains are key elements of decision-based materials design in the framework of Integrated Computational Materials Engineering (ICME), where databases, multi-scale modeling and simulation tools, and experiments are integrated with the aim of time reduction in design and manufacturing of materials/products [1]. In this context, the understanding and quantification of uncertainties can provide a confidence measure for the applicability of models for decision making in materials design.

Generally, UQ incorporates the detection of uncertainty sources and the development of corresponding appropriate mathematical

approaches to calculate the error bounds for any quantity of interest in models [2–4]. The uncertainty can arise from different sources, which are categorized as natural uncertainty (NU) due to the random nature of a physical system, model parameter uncertainty (MPU) resulting from the lack of sufficient and/or accurate data for model parameters, propagated uncertainty (PU) in the case of multi-scale modeling, and model structure uncertainty (MSU) owing to any simplifications, assumptions, and/or incomplete physics in the model [5]. UP also refers to the determination of model output uncertainties based on the uncertainties of its input variables for either individual or multi-scale models [2,6,7].

Despite the importance of UQ/UP, there are comparably few works applied to problems in (computational) materials science. Chernatynskiy et al. [3] have reviewed these efforts, focusing on UQ problems related to atomic-scale as well as multi-scale simulations. For over a decade, at the electronic/atomic scale, probabilistic parameterization in first principles calculations based on density function theory (DFT) [8,9] remains as some of the most significant works on UQ applied to materials problems. Aldegunde et al. [10]

\* Corresponding author. Texas A&M University, Department of Materials Science and Engineering, 3003 TAMU, College Station, TX, 7843-3003, USA.

E-mail address: [hona107@tamu.edu](mailto:hona107@tamu.edu) (P. Honarmandi).

have recently used a machine learning-informed Bayesian approach to quantify uncertainties associated to the prediction, via cluster expansions, of the thermodynamic properties of alloys. In that work, the resulting uncertainties are associated to the model parameters as well as the structure of the models themselves, which results from poorly converged cluster expansions due to lack of training data. In another work, the diffusivity error in molecular dynamic simulations of a Ni/Al nano-laminate bilayer has been assessed using a Markov Chain Monte Carlo (MCMC) method in the context of Bayesian statistics [11]. In regard to UQ in multi-scale simulations, Volker et al. [12] have used DFT results for a phase field model to provide a connection between atomic- and meso-scale descriptions of ferro-electric materials, and also applied a sensitivity analysis to identify the most significant parameters of the free energy functional and to quantify the uncertainties in the simulations. Moreover, Liu et al. [13] have utilized a Bayesian stochastic approach for probabilistic parameter calibration and a stochastic projection with polynomial chaos expansions for the propagation of their uncertainties across multi-scale constitutive models that link the microstructure, property, and performance of random heterogeneous composite materials. It is worth mentioning that there is also some UQ work for multi-scale modeling of plasticity (deformation) in poly-crystalline materials [14–16].

While there are emerging several efforts centered on UQ/UP applied to materials simulations, UQ applied to CALPHAD of PHASE Diagrams (CALPHAD) methodologies remains poorly explored. This is unfortunate as CALPHAD-based thermodynamic descriptions are the foundation of most proposed ICME/alloy design frameworks. UQ in the calculation of equilibrium of phase diagrams is crucial since any small variations in chemical composition and/or temperature due to their uncertainties can alter the (predicted) stability, volume fraction, and chemical composition of microstructural phases and may affect the materials properties considerably. For example, if the stability of a phase should be suppressed in the design to prohibit its detrimental effects on the final product, the design space for this particular materials system should lie outside the boundaries representing the onset of stability of such detrimental phases but also out of their uncertainty bounds in order to have more confidence in the design.

Focused work on UQ applied to CALPHAD-based thermodynamic assessments remains scarce [17–22]. However, Stan and Reardon [23] and Otis and Liu [24] have recently proposed approaches that constitute increasingly sophisticated approaches to UQ applied to CALPHAD. In the former study [23], a fuzzy logic weighted genetic algorithm (GA) was applied as a sampling tool in a Bayesian-based framework to find the free energy parameters' posterior probability distribution given some uncertain thermodynamic data, and then the parameters' uncertainties were propagated to the resulting phase diagram by sampling from the posterior probability distribution. In the later work [24], an ICME-directed multi-scale (or rather multi-level) modeling has been introduced that links high-throughput first-principles calculations to CALPHAD modeling. The most relevant parameters in the sublattice-based CALPHAD modeling of that work were determined using both Akaike Information Criterion (AIC) and F-test, and a Markov chain Monte Carlo (MCMC) sampler was used to quantify the parameters' posterior probability distribution.

Furthermore, Duong et al. [25] have analyzed the uncertainties in the case of a multi-scale modeling for pseudo-binary  $\text{Ti}^{2}\text{AlC}$ — $\text{Cr}^{2}\text{AlC}$  MAX phases, which links first principles calculations and the CALPHAD method. In that study, the overall uncertainties have been reflected on the calibrated thermodynamic parameters through an MCMC approach, and then propagated to the quasi-binary phase diagram. Earlier, Duong et al. [26] employed UQ to

investigate the effect of uncertainties in the phase stability of the U–Nb system. However, a detailed and thorough analysis of uncertainty and selection of models with appropriate parameters is still lacking in CALPHAD modeling.

In this work, a thorough evaluation of the uncertainty is performed for not only CALPHAD model parameters, but also for their resulting phase stability diagrams using the Hf–Si binary system as a case study. It should be noted that calculation of Hf–Si binary phase diagram and its uncertainties is of great importance since adding Hafnium to Niobium silicide based alloys (as promising turbine airfoil materials with high operating temperature) increases their strength, fracture toughness, and oxidation resistance significantly [27–29]. In this regard, an MCMC–Metropolis Hastings algorithm and a forward analysis of parameters' posterior samples are applied for the quantification of the parameters' uncertainties and their propagation to resulting phase diagrams, respectively. In this paper, it is also shown how to select the most relevant CALPHAD model given experimental data for a system based on Bayesian hypothesis testing. For this analysis, four independently generated CALPHAD models for the Hf–Si system are considered.

Since finding the appropriate number of parameters in CALPHAD models—or rather, the appropriate model parameter set—is a hard and skillful task, it is necessary to propose systematic approaches that can provide sufficiently reliable results with no need for the manual identification of the best predicting model. While we do not claim that the search for improved CALPHAD models as new experiments/simulations or model-forms become available is render no longer necessary, the proposed information fusion framework provides a robust strategy to continuously improve CALPHAD descriptions without necessarily completely discarding previous assessments.

In our work, Bayesian model averaging (BMA) and an error correlation-based model fusion (CMF) approach are used to combine all the given model results together in different ways with their own specific purposes. In BMA, each model has some probability of being true and the fused estimate is a weighted average of the models. This method is extremely useful in the case of model-building process based on a weighted average over the models' responses, and/or less risk (more confidence) in design based on broader uncertainty bands provided by a weighted average over the uncertainties of the models' responses. On the other hand, the information fusion technique based on the correlations between the model deviations is applied for the purpose of acquiring more precise estimations and lower uncertainties compared to results obtained from each applied individual model.

## 2. CALPHAD model description

Four sets of models describing Gibbs free energies of the binary system are considered in the current work. Here we consider a family of models that have been carefully assessed/fitted to available data to avoid artifacts that are not to be expected in this system, such as inverted miscibility gaps. In each of these sets, the intermetallic compounds are described using the line-compound formalism [30], which reads:

$$G_{\text{Hf}_k\text{Si}_l}^{\text{Hf-Si}} = \frac{k}{k+l} {}^0G_{\text{Hf}}^{\text{HCP}} + \frac{l}{k+l} {}^0G_{\text{Si}}^{\text{Diamond}} + a + bT, \quad (1)$$

where  $k$  and  $l$  are the compound numbers,  ${}^0G_{\text{Hf}}^{\text{HCP}}$  and  ${}^0G_{\text{Si}}^{\text{Diamond}}$  are the chosen energy references corresponding to the energies of pure HCP–Hf and diamond–Si as given in the SGTE database [31],  $a$  and  $b$  are model parameters, and  $T$  is temperature (in Kelvin).

The liquid phase is described using the sub-regular solid solution model as follows:

$$G_i^{Liq} = \sum_{i=1}^N x_i^0 G_i^{Liq} + RT \sum_{i=1}^N x_i \ln x_i + \sum_{i=1}^N x_i \sum_{j \neq i}^N x_j \sum_{n=0}^M {}^n L_{ij} (x_i - x_j)^n, \quad (2)$$

where  $x_i$  is the mole fraction of the constituent  $i$  (either Si or Hf),  $G_i^{Liq}$  is the constituent energy taken, again, from the SGTE database [31],  $T$  is temperature (in Kelvin),  $R$  is the gas constant, and  ${}^n L_{ij}$  is given as:

$${}^v L_{ij} = {}^v a_{ij} + {}^v b_{ij} T, \quad (3)$$

where  ${}^v a_{ij}$  and  ${}^v b_{ij}$  are model parameters, which describe the interactions between the constituents beyond those of ideal mixing.

The terminal phases, namely HCP–Hf and diamond–Si, are described by either the line-compound formalism [30] or the sub-regular solid solution model.

In the current work, four models with 17, 20, 28, and 30 parameters identified as models 1 through 4 are chosen for Hf–Si binary system by expert opinion to show how the model selection and model fusion approaches work in this case study. The parameters associated with the liquid, intermetallic compounds and the end members are listed in Table 1 for all the given models.

**Table 1**  
The parameters of each given model that describe the Gibbs free energy of the phases in the system (G and L denote the parameters for the line-compound and the sub-regular solid solution models, respectively).

Model	Phase	Parameter Specification	Model	Phase	Parameter Specification
1	HCP – Hf	${}^0 L_{Hf, Si}^{HCP} : {}^0 a_{Hf, Si}^{HCP}$	2	HCP – Hf	${}^0 L_{Hf, Si}^{HCP} : {}^0 a_{Hf, Si}^{HCP} - {}^0 b_{Hf, Si}^{HCP} T$
	BCC – Hf	${}^0 L_{Hf, Si}^{BCC} : {}^0 a_{Hf, Si}^{BCC}$		BCC – Hf	${}^0 L_{Hf, Si}^{BCC} : {}^0 a_{Hf, Si}^{BCC} - {}^0 b_{Hf, Si}^{BCC} T$
	Hf <sub>2</sub> Si	${}^0 C_{Hf, Si}^{Hf_2Si} : {}^0 a_{Hf, Si}^{Hf_2Si} - {}^0 b_{Hf, Si}^{Hf_2Si} T$		Hf <sub>2</sub> Si	${}^0 C_{Hf, Si}^{Hf_2Si} : {}^0 a_{Hf, Si}^{Hf_2Si} - {}^0 b_{Hf, Si}^{Hf_2Si} T$
	Hf <sub>3</sub> Si <sub>2</sub>	${}^0 C_{Hf, Si}^{Hf_3Si_2} : {}^0 a_{Hf, Si}^{Hf_3Si_2} - {}^0 b_{Hf, Si}^{Hf_3Si_2} T$		Hf <sub>3</sub> Si <sub>2</sub>	${}^0 C_{Hf, Si}^{Hf_3Si_2} : {}^0 a_{Hf, Si}^{Hf_3Si_2} - {}^0 b_{Hf, Si}^{Hf_3Si_2} T$
	Hf <sub>5</sub> Si <sub>3</sub>	${}^0 C_{Hf, Si}^{Hf_5Si_3} : {}^0 a_{Hf, Si}^{Hf_5Si_3} - {}^0 b_{Hf, Si}^{Hf_5Si_3} T$		Hf <sub>5</sub> Si <sub>3</sub>	${}^0 C_{Hf, Si}^{Hf_5Si_3} : {}^0 a_{Hf, Si}^{Hf_5Si_3} - {}^0 b_{Hf, Si}^{Hf_5Si_3} T$
	Hf <sub>5</sub> Si <sub>4</sub>	${}^0 C_{Hf, Si}^{Hf_5Si_4} : {}^0 a_{Hf, Si}^{Hf_5Si_4} - {}^0 b_{Hf, Si}^{Hf_5Si_4} T$		Hf <sub>5</sub> Si <sub>4</sub>	${}^0 C_{Hf, Si}^{Hf_5Si_4} : {}^0 a_{Hf, Si}^{Hf_5Si_4} - {}^0 b_{Hf, Si}^{Hf_5Si_4} T$
	HfSi	${}^0 C_{Hf, Si}^{HfSi} : {}^0 a_{Hf, Si}^{HfSi} - {}^0 b_{Hf, Si}^{HfSi} T$		HfSi	${}^0 C_{Hf, Si}^{HfSi} : {}^0 a_{Hf, Si}^{HfSi} - {}^0 b_{Hf, Si}^{HfSi} T$
	HfSi <sub>2</sub>	${}^0 C_{Hf, Si}^{HfSi_2} : {}^0 a_{Hf, Si}^{HfSi_2} - {}^0 b_{Hf, Si}^{HfSi_2} T$		HfSi <sub>2</sub>	${}^0 C_{Hf, Si}^{HfSi_2} : {}^0 a_{Hf, Si}^{HfSi_2} - {}^0 b_{Hf, Si}^{HfSi_2} T$
	liquid	${}^0 L_{Hf, Si}^{liq} : {}^0 a_{Hf, Si}^{liq} - {}^0 b_{Hf, Si}^{liq} T$		liquid	${}^0 L_{Hf, Si}^{liq} : {}^0 a_{Hf, Si}^{liq} - {}^0 b_{Hf, Si}^{liq} T$
	Si	${}^0 L_{Hf, Si}^{Si} : {}^0 a_{Hf, Si}^{Si}$		Si	${}^0 L_{Hf, Si}^{Si} : {}^0 a_{Hf, Si}^{Si} - {}^0 b_{Hf, Si}^{Si} T$
3	HCP – Hf	${}^0 L_{Hf, Si}^{HCP} : {}^0 a_{Hf, Si}^{HCP} - {}^0 b_{Hf, Si}^{HCP} T$	4	HCP – Hf	${}^0 L_{Hf, Si}^{HCP} : {}^0 a_{Hf, Si}^{HCP} - {}^0 b_{Hf, Si}^{HCP} T$
	BCC – Hf	${}^1 L_{Hf, Si}^{HCP} : {}^1 a_{Hf, Si}^{HCP} - {}^1 b_{Hf, Si}^{HCP} T$		BCC – Hf	${}^1 L_{Hf, Si}^{HCP} : {}^1 a_{Hf, Si}^{HCP} - {}^1 b_{Hf, Si}^{HCP} T$
	BCC – Hf	${}^0 L_{Hf, Si}^{BCC} : {}^0 a_{Hf, Si}^{BCC} - {}^0 b_{Hf, Si}^{BCC} T$		BCC – Hf	${}^0 L_{Hf, Si}^{BCC} : {}^0 a_{Hf, Si}^{BCC} - {}^0 b_{Hf, Si}^{BCC} T$
	Hf <sub>2</sub> Si	${}^1 L_{Hf, Si}^{BCC} : {}^1 a_{Hf, Si}^{BCC} - {}^1 b_{Hf, Si}^{BCC} T$		Hf <sub>2</sub> Si	${}^1 L_{Hf, Si}^{BCC} : {}^1 a_{Hf, Si}^{BCC} - {}^1 b_{Hf, Si}^{BCC} T$
	Hf <sub>3</sub> Si <sub>2</sub>	${}^0 C_{Hf, Si}^{Hf_2Si} : {}^0 a_{Hf, Si}^{Hf_2Si} - {}^0 b_{Hf, Si}^{Hf_2Si} T$		Hf <sub>3</sub> Si <sub>2</sub>	${}^0 C_{Hf, Si}^{Hf_2Si} : {}^0 a_{Hf, Si}^{Hf_2Si} - {}^0 b_{Hf, Si}^{Hf_2Si} T$
	Hf <sub>5</sub> Si <sub>3</sub>	${}^0 C_{Hf, Si}^{Hf_3Si_2} : {}^0 a_{Hf, Si}^{Hf_3Si_2} - {}^0 b_{Hf, Si}^{Hf_3Si_2} T$		Hf <sub>5</sub> Si <sub>3</sub>	${}^0 C_{Hf, Si}^{Hf_3Si_2} : {}^0 a_{Hf, Si}^{Hf_3Si_2} - {}^0 b_{Hf, Si}^{Hf_3Si_2} T$
	Hf <sub>5</sub> Si <sub>4</sub>	${}^0 C_{Hf, Si}^{Hf_5Si_3} : {}^0 a_{Hf, Si}^{Hf_5Si_3} - {}^0 b_{Hf, Si}^{Hf_5Si_3} T$		Hf <sub>5</sub> Si <sub>4</sub>	${}^0 C_{Hf, Si}^{Hf_5Si_3} : {}^0 a_{Hf, Si}^{Hf_5Si_3} - {}^0 b_{Hf, Si}^{Hf_5Si_3} T$
	HfSi	${}^0 C_{Hf, Si}^{Hf_5Si_4} : {}^0 a_{Hf, Si}^{Hf_5Si_4} - {}^0 b_{Hf, Si}^{Hf_5Si_4} T$		HfSi	${}^0 C_{Hf, Si}^{Hf_5Si_4} : {}^0 a_{Hf, Si}^{Hf_5Si_4} - {}^0 b_{Hf, Si}^{Hf_5Si_4} T$
	HfSi <sub>2</sub>	${}^0 C_{Hf, Si}^{HfSi} : {}^0 a_{Hf, Si}^{HfSi} - {}^0 b_{Hf, Si}^{HfSi} T$		HfSi	${}^0 C_{Hf, Si}^{HfSi} : {}^0 a_{Hf, Si}^{HfSi} - {}^0 b_{Hf, Si}^{HfSi} T$
	liquid	${}^0 C_{Hf, Si}^{HfSi_2} : {}^0 a_{Hf, Si}^{HfSi_2} - {}^0 b_{Hf, Si}^{HfSi_2} T$		HfSi <sub>2</sub>	${}^0 C_{Hf, Si}^{HfSi_2} : {}^0 a_{Hf, Si}^{HfSi_2} - {}^0 b_{Hf, Si}^{HfSi_2} T$
3	liquid	${}^0 L_{Hf, Si}^{liq} : {}^0 a_{Hf, Si}^{liq} - {}^0 b_{Hf, Si}^{liq} T$	4	liquid	${}^0 L_{Hf, Si}^{liq} : {}^0 a_{Hf, Si}^{liq} - {}^0 b_{Hf, Si}^{liq} T$
	Si	${}^1 L_{Hf, Si}^{liq} : {}^1 a_{Hf, Si}^{liq} - {}^1 b_{Hf, Si}^{liq} T$		Si	${}^1 L_{Hf, Si}^{liq} : {}^1 a_{Hf, Si}^{liq} - {}^1 b_{Hf, Si}^{liq} T$
	Si	${}^0 L_{Hf, Si}^{Si} : {}^0 a_{Hf, Si}^{Si} - {}^0 b_{Hf, Si}^{Si} T$		Si	${}^2 L_{Hf, Si}^{liq} : {}^2 a_{Hf, Si}^{liq} - {}^2 b_{Hf, Si}^{liq} T$
	Si	${}^1 L_{Hf, Si}^{Si} : {}^1 a_{Hf, Si}^{Si} - {}^1 b_{Hf, Si}^{Si} T$		Si	${}^0 L_{Hf, Si}^{Si} : {}^0 a_{Hf, Si}^{Si} - {}^0 b_{Hf, Si}^{Si} T$
	Si	${}^1 L_{Hf, Si}^{Si} : {}^1 a_{Hf, Si}^{Si} - {}^1 b_{Hf, Si}^{Si} T$		Si	${}^1 L_{Hf, Si}^{Si} : {}^1 a_{Hf, Si}^{Si} - {}^1 b_{Hf, Si}^{Si} T$
	Si	${}^1 L_{Hf, Si}^{Si} : {}^1 a_{Hf, Si}^{Si} - {}^1 b_{Hf, Si}^{Si} T$		Si	${}^1 L_{Hf, Si}^{Si} : {}^1 a_{Hf, Si}^{Si} - {}^1 b_{Hf, Si}^{Si} T$
	Si	${}^1 L_{Hf, Si}^{Si} : {}^1 a_{Hf, Si}^{Si} - {}^1 b_{Hf, Si}^{Si} T$		Si	${}^1 L_{Hf, Si}^{Si} : {}^1 a_{Hf, Si}^{Si} - {}^1 b_{Hf, Si}^{Si} T$
	Si	${}^1 L_{Hf, Si}^{Si} : {}^1 a_{Hf, Si}^{Si} - {}^1 b_{Hf, Si}^{Si} T$		Si	${}^1 L_{Hf, Si}^{Si} : {}^1 a_{Hf, Si}^{Si} - {}^1 b_{Hf, Si}^{Si} T$
	Si	${}^1 L_{Hf, Si}^{Si} : {}^1 a_{Hf, Si}^{Si} - {}^1 b_{Hf, Si}^{Si} T$		Si	${}^1 L_{Hf, Si}^{Si} : {}^1 a_{Hf, Si}^{Si} - {}^1 b_{Hf, Si}^{Si} T$
	Si	${}^1 L_{Hf, Si}^{Si} : {}^1 a_{Hf, Si}^{Si} - {}^1 b_{Hf, Si}^{Si} T$		Si	${}^1 L_{Hf, Si}^{Si} : {}^1 a_{Hf, Si}^{Si} - {}^1 b_{Hf, Si}^{Si} T$

### 3. Uncertainty quantification methodology

#### 3.1. Markov chain Monte Carlo-Metropolis Hastings algorithm

In our work, the MCMC Metropolis Hastings toolbox in Matlab has been utilized for probabilistic calibration of the parameters in the applied CALPHAD models. After the introduction of prior knowledge for parameters (initial values (vector  $\theta^0$ )), lower and upper bounds, and probability density functions (PDFs) in this algorithm, parameter vectors are randomly sampled from a non-stationary proposal posterior probability distribution, which is an arbitrary multivariate Gaussian distribution with a mean value at  $\theta^0$  for the first sampling or the last accepted parameter vector during the next parameter sampling. The sampled parameter vector in each iteration can be accepted or rejected through a criterion in the context of the Bayesian statistic, known as the Metropolis Hastings (M-H) ratio:

$$M - H = \frac{p(\theta^{cand}) p(D|\theta^{cand}) q(\theta^i|\theta^{cand})}{p(\theta^i) p(D|\theta^i) q(\theta^{cand}|\theta^i)}, \quad (4)$$

where  $\theta^i$ ,  $\theta^{cand}$ , and  $D$  are the last accepted sample of parameter

vector, the new sample of parameter vector as a candidate, and the given data, respectively. It should be noted that some synthetic data sampled from the Hf–Si calculated phase diagram in Ref. [27] has been used as  $D$  in this work.

Based on Bayes' theorem, the joint (posterior) probability in each case is proportional to the prior probability times the likelihood,  $p(\theta)p(D|\theta)$ . The likelihood is the conditional probability of the data,  $D$ , given the parameter vector,  $\theta$ , and in this work is considered as a Gaussian distribution centered at the data,  $D$ , with variances,  $\sigma^2$ , determined by the data errors. Here, the variances are assumed to be the same for all the data and updated as a hyper-parameter during MCMC sampling since the data error is unknown. In this case, the variance samples are generated from an inverse gamma posterior PDF resulting from the introduction of a non-informative inverse gamma prior PDF for the variance  $\sigma^2$  [26,32]. The second ratio in Equation (4) is the Hastings ratio, which considers the asymmetric effect of the proposal probability distribution in the acceptance/rejection criterion of the parameter vector.

After the calculation of the M-H ratio for each sampling iteration, if  $\min\{M - H, 1\} \times 100$  is greater than 1, the candidate for the parameter vector is accepted as the new sample; otherwise, the candidate may still be accepted with a probability of  $\min\{M - H, 1\} \times 100$  [33]. In the case of the candidate rejection, the last accepted parameter vector is repeated in the MCMC chain as the new sample. At the end of the MCMC sampling process, a chain of the parameter vectors is obtained as  $\{\theta^0, \dots, \theta^N\}$  whose mean values and variance-covariance matrix after removal of "burn-in period" demonstrate the plausible optimum values of the parameters and their overall reflected uncertainties, respectively. It should be noted that the parameters' convergence during MCMC sampling algorithm is defined as the convergence of the parameters' cumulative mean values towards almost constant values, which is equivalent to the fluctuation of the parameters' samples around their mean values across MCMC sampling [34].

### 3.2. Bayesian model selection/averaging

Jeffreys [35] proposed a hypothesis testing method in the context of Bayesian statistics that evaluates the evidence in favor of a scientific theory or hypothesis. Models are considered as hypotheses in the case of Bayesian model selection (BMS). In BMS, non-nested models (models with at least one uncommon parameter) can also be compared together, which is very difficult or sometimes impossible through frequentist approaches [36]. Generally, models/hypotheses with greater posteriors are more favored by the evidence and considered as better models/hypotheses. Accordingly, the Bayesian comparison criterion for models/hypotheses is defined as the posterior odds of each two applied models/hypotheses given data. For instance, posterior odds of model one ( $M_1$ ) to model two ( $M_2$ ) given the data ( $D$ ) is expressed as follows:

$$\frac{p(M_1|D)}{p(M_2|D)} = \frac{p(D|M_1)}{p(D|M_2)} \frac{p(M_1)}{p(M_2)} = B_{12}\lambda_{12}. \quad (5)$$

Bayes' factor is the ratio of the marginal likelihoods and usually suggests which model/hypothesis is more favored by the evidence (data):

$$p(D|M_k) = \int p(D|\theta_k, M_k) p(\theta_k|M_k) d\theta_k, \quad (6)$$

where  $\theta_k$  is the parameter vector under  $M_k$ . Therefore, the key element in BMS is the calculation of the marginal likelihoods offered by Equation (6), which is challenging in the case of high

dimensional  $\theta_k$ . There are different methods to approximate these integrals, which include Laplace's method, the Schwarz criterion, simple Monte Carlo, importance sampling, adaptive Gaussian quadrature, and simulating approaches from the parameters' posterior [36]. In our work, the MCMC sampling method is used to simulate from the parameters' posterior for the integral estimation.

In the Monte Carlo method, the integral in Equation (6) is approximated as:

$$p(D|M_k) \approx \frac{1}{N} \sum_{i=1}^N p(D|\theta_k^{(i)}, M_k), \quad (7)$$

where  $\theta_k^{(i)}$ s are samples from the parameters' prior  $p(\theta_k|M_k)$ . This approximation is equivalent to the average of the likelihood values obtained from the sampled parameters. However, this approach can be fairly inefficient in the case of a more concentrated posterior distribution compared to the prior distribution since the likelihood values for most of the sampled  $\theta_k^{(i)}$  will be very small and the approximation result will be considerably weighted by a few samples with large likelihood values. For this reason, importance sampling techniques are usually applied [36]:

$$p(D|M_k) \approx \frac{1}{N} \sum_{i=1}^N \frac{p(D|\theta_k^{(i)}, M_k) p(\theta_k^{(i)}|M_k)}{p^*(\theta_k^{(i)})}, \quad (8)$$

$$p(D|M_k) \approx \frac{1}{N} \sum_{i=1}^N \frac{p(D|\theta_k^{(i)}, M_k) w_k^{(i)}}{w_k^{(i)}}, \quad (9)$$

where  $p^*(\theta_k)$  is a suitable distribution that enables the sampling of the important parameter values for the sake of more efficient sampling. In Equation (9),  $w_k^{(i)} = \frac{p(\theta_k^{(i)}|M_k)}{p^*(\theta_k^{(i)})}$ , which can provide a weighted average of likelihood values obtained using the sampled parameters from  $p^*(\theta_k)$ .

In the case that the importance sampling function is proportional to the posterior probability density function of the parameters in model  $k$ , i.e.,  $p^*(\theta_k) = p(D|\theta_k, M_k)p(\theta_k|M_k)$ , Equation (9) is turned into:

$$p(D|M_k) \approx \left\{ \frac{1}{N} \sum_{i=1}^N p(D|\theta_k^{(i)}, M_k)^{-1} \right\}^{-1} \quad (10)$$

MCMC is usually used to sample  $\theta_k^{(i)}$ s from  $(p^*(\theta_k))$ . Equation (10) shows a harmonic mean of likelihood values obtained from the MCMC sampled parameters after the removal of the burn-in period.

In addition to model selection, Bayes' factors can be used to determine the posterior probability distribution of the competing models given data, which can be defined as their associated weight in the context of Bayesian model averaging (BMA). Generally, BMA can provide a combined inference from all the competing models that can be very useful for model-building process or less risky predictions in design. In this approach, the posterior density (weight) associated with each model can be obtained as follows:

$$p(M_k|D) = \frac{p(D|M_k)p(M_k)}{\sum_{i=1}^K p(D|M_i)p(M_i)}$$



$$= \frac{B_{k1}\lambda_{k1}}{\sum_{i=1}^K B_{i1}\lambda_{i1}} = \frac{B_{k1}}{\sum_{i=1}^K B_{i1}}, \quad (11)$$

Where  $K$  is the total number of models. In this equation,  $M_1$  is considered as reference for the calculation of all the Bayes' factors, and all the prior odds are also 1. The posterior densities of any quantity of interest ( $\Delta$ ) existing in all the competing models can be combined together through BMA as [36–38]:

$$p(\Delta|D) = \sum_{i=1}^K p(\Delta|D, M_i)p(M_i|D). \quad (12)$$

### 3.3. Applied error correlation-based model fusion

In the case of multiple uncertain sources of information (e.g., different models for the same problem), there is a need to integrate all the sources to produce more reliable results [39]. In practice, there are several approaches for fusing information from multiple models. BMA is a model fusion technique that has some benefits in robust design. Other available techniques are fusion under known correlation [40–43], and the covariance intersection method [44]. The key distinction of BMA over other model fusion approaches is the assumption of statistical independence among models, which may be incorrect in some cases and can lead to potentially serious misconceptions regarding confidence in quantity of interest estimates.

A fundamental claim in this work is that *any model can provide potentially useful information to a given task*. We thus seek to take into account all potential information any given model may provide and fuse unique information from the available models. Our fusion goal then is to identify dependencies, via estimated correlations, among the model discrepancies. With these estimated correlations, the models are fused following standard practice for the fusion of normally distributed data.

Under the case of known correlations between the discrepancies of models, the fused mean and variance at a design point,  $\mathbf{x}$ , are shown to be [42].

$$E[\hat{f}(\mathbf{x})] = \frac{\mathbf{e}^\top \hat{\Sigma}(\mathbf{x})^{-1} \mu(\mathbf{x})}{\mathbf{e}^\top \hat{\Sigma}(\mathbf{x})^{-1} \mathbf{e}}, \quad (13)$$

$$\text{Var}(\hat{f}(\mathbf{x})) = \frac{1}{\mathbf{e}^\top \hat{\Sigma}(\mathbf{x})^{-1} \mathbf{e}}, \quad (14)$$

where  $\mathbf{e} = [1, \dots, 1]^\top$ ,  $\mu(\mathbf{x}) = [\mu_1(\mathbf{x}), \dots, \mu_S(\mathbf{x})]^\top$  is the vector of mean values given  $S$  models, and  $\hat{\Sigma}(\mathbf{x})^{-1}$  is the inverse of the covariance matrix between the models given as

$$\hat{\Sigma} = \begin{bmatrix} \sigma_1^2 & \rho_{12} \sigma_1 \sigma_2 & \cdots & \rho_{1S} \sigma_1 \sigma_S \\ \rho_{12} \sigma_1 \sigma_2 & \sigma_2^2 & \cdots & \rho_{2S} \sigma_2 \sigma_S \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{1S} \sigma_1 \sigma_S & \rho_{2S} \sigma_2 \sigma_S & \cdots & \sigma_S^2 \end{bmatrix}, \quad (15)$$

where  $\rho_{ij}$  is the correlation coefficient between the deviations of models  $i$  and  $j$  at point  $\mathbf{x}$ , and  $\sigma_i^2$  is the variance of model  $i$  at point  $\mathbf{x}$ .

To estimate the correlations between the model deviations in the case that they are unknown, we use the reification process defined in Refs. [45,46], which refers to the process of treating each model, in turn, as ground truth. This means that we assume the data generated by the reified model represents the true quantity of interest. These data are used to estimate the correlation between the errors of the different models. The process is repeated for each

model. The detailed process of estimating the correlation between the errors of two models can be found in Refs. [45,46].

Following Thomison et al. [46], assuming model  $i$  is chosen to reify, the correlation coefficients between the models  $i$  and  $j$ , for  $j = 1, \dots, i-1, i+1, \dots, S$ , are given as:

$$\rho_{ij}(\mathbf{x}) = \frac{\sigma_i^2(\mathbf{x})}{\sigma_i(\mathbf{x})\sigma_j(\mathbf{x})} = \frac{\sigma_i(\mathbf{x})}{\sqrt{(\mu_i(\mathbf{x}) - \mu_j(\mathbf{x}))^2 + \sigma_i^2(\mathbf{x})}}, \quad (16)$$

where  $\mu_i(\mathbf{x})$  and  $\mu_j(\mathbf{x})$  are the mean values of models  $i$  and  $j$  respectively at design point  $\mathbf{x}$ , and  $\sigma_i^2(\mathbf{x})$  and  $\sigma_j^2(\mathbf{x})$  are the variances at point  $\mathbf{x}$ . The first subscript under the correlation coefficient denotes which model has been reified. Since the only information we have regarding which model we believe to be more realistic is the variance of each, in addition to reifying model  $i$  to estimate the correlation, we also reify information source  $j$  and estimate  $\rho_{ji}(\mathbf{x})$ . We then estimate the correlation between the errors as the variance-weighted average of the two correlation coefficients as follows:

$$\bar{\rho}_{ij}(\mathbf{x}) = \frac{\sigma_j^2(\mathbf{x})}{\sigma_i^2(\mathbf{x}) + \sigma_j^2(\mathbf{x})} \rho_{ij}(\mathbf{x}) + \frac{\sigma_i^2(\mathbf{x})}{\sigma_i^2(\mathbf{x}) + \sigma_j^2(\mathbf{x})} \rho_{ji}(\mathbf{x}). \quad (17)$$

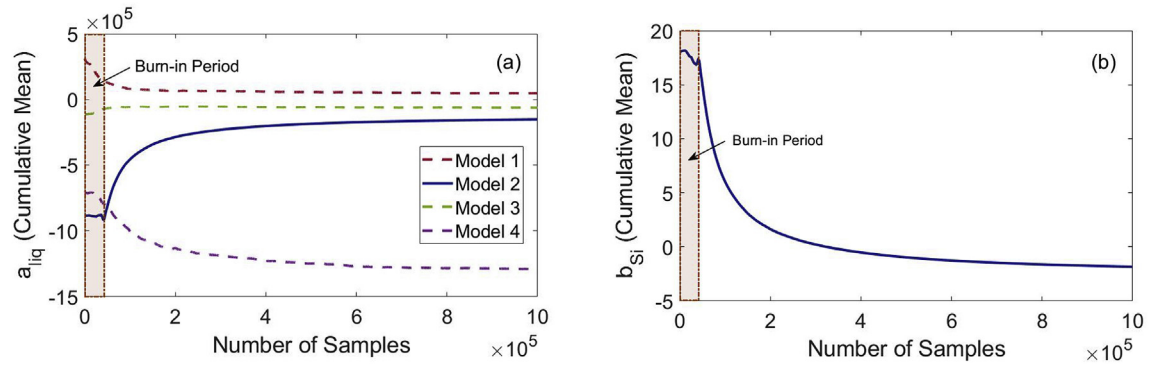
These correlations can then be used to estimate the mean and variance of the fused estimate from Equations (13) and (14).

## 4. Results and discussion

In CALPHAD modeling, it is a hard task, even for CALPHAD experts, to know what formalism should be used to describe the Gibbs free energy of each phase in the system and to what order the polynomial term in Equation (2) should be expanded in the case of the sub-regular solid solution model consideration in order to obtain the closest phase diagram to the available data after parameter optimization. For this reason, different models may be suggested for a system with different formalisms for the Gibbs free energy description of the phases and different orders of polynomial expansions for the sub-regular solid solution models. In the context of Bayesian statistics, the best model among the selected candidates according to the available data can be identified through the calculation of the Bayes' factor based on Equation (5).

Each of the models are probabilistically calibrated against the available calculated data for the phase diagram using one million MCMC samples of the parameter vectors. For each model, the deterministically optimized values of the parameters obtained from the PARROT module in Thermo-Calc software (through least squares) and  $\pm 3$  times these values are considered as the initial values ( $\theta_0$ ) and the ranges of the parameters, respectively. This prior information enables much faster convergence of the model parameters during MCMC sampling. However, it is assumed that the deterministic optimization in Thermo-Calc is global, which is not necessarily true as we discuss later. It should be noted that the compositions of stable phases estimated from the convex hull of the phase Gibbs energies at any given temperature are compared with the corresponding available data through the likelihood during the sampling process. Here, the results obtained for model 2 are discussed in detail to show how MCMC is used in this work for probabilistic calibration of the model parameters. The same approach is applied for the calibration of the other three model parameter sets.

During MCMC sampling, there are some homogeneous fluctuations around plausible optimum values of parameters after convergence. However, this means the cumulative means of parameter samples should converge to constant values. Therefore,



**Fig. 1.** Cumulative mean plots of the parameters a)  $a_{liq}$  and b)  $b_{Si}$  in model 2. The red shaded regions show the number of parameters belonging to the burn-in period. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

**Table 2**

Initial values for the parameters in model 2 and their plausible optimum values and overall uncertainties after MCMC calibration.

Phase	Parameter Specification	Deterministically Calibrated Initial Values	Probabilistically Calibrated Values after MCMC
HCP – Hf	${}^0G_{Hf, Si}^{HCP} : {}^0a_{Hf, Si}^{HCP} - {}^0b_{Hf, Si}^{HCP} T$	${}^0G_{Hf, Si}^{HCP} : -455420 + 221.79T$	${}^0G_{Hf, Si}^{HCP} : -29070 \pm 16377 + 15.08 \pm 8.10T$
BCC – Hf	${}^0G_{Hf, Si}^{BCC} : {}^0a_{Hf, Si}^{BCC} - {}^0b_{Hf, Si}^{BCC} T$	${}^0G_{Hf, Si}^{BCC} : -45254 + 18.56T$	${}^0G_{Hf, Si}^{BCC} : 10437 \pm 1845 - 6.17 \pm 1.10T$
Hf <sub>2</sub> Si	${}^0G_{Hf, Si}^{Hf_2Si} : {}^0a_{Hf, Si}^{Hf_2Si} - {}^0b_{Hf, Si}^{Hf_2Si} T$	${}^0G_{Hf, Si}^{Hf_2Si} : -454090 + 90.46T$	${}^0G_{Hf, Si}^{Hf_2Si} : -31675 \pm 6672 - 3.13 \pm 1.60T$
Hf <sub>3</sub> Si <sub>2</sub>	${}^0G_{Hf, Si}^{Hf_3Si_2} : {}^0a_{Hf, Si}^{Hf_3Si_2} - {}^0b_{Hf, Si}^{Hf_3Si_2} T$	${}^0G_{Hf, Si}^{Hf_3Si_2} : -481270 + 84.19T$	${}^0G_{Hf, Si}^{Hf_3Si_2} : -32102 \pm 7168 - 5.78 \pm 1.61T$
Hf <sub>5</sub> Si <sub>3</sub>	${}^0G_{Hf, Si}^{Hf_5Si_3} : {}^0a_{Hf, Si}^{Hf_5Si_3} - {}^0b_{Hf, Si}^{Hf_5Si_3} T$	${}^0G_{Hf, Si}^{Hf_5Si_3} : -454950 + 77.55T$	${}^0G_{Hf, Si}^{Hf_5Si_3} : -29246 \pm 6805 - 6.27 \pm 1.52T$
Hf <sub>5</sub> Si <sub>4</sub>	${}^0G_{Hf, Si}^{Hf_5Si_4} : {}^0a_{Hf, Si}^{Hf_5Si_4} - {}^0b_{Hf, Si}^{Hf_5Si_4} T$	${}^0G_{Hf, Si}^{Hf_5Si_4} : -522680 + 100.88T$	${}^0G_{Hf, Si}^{Hf_5Si_4} : -33133 \pm 7614 - 5.60 \pm 1.80T$
HfSi	${}^0G_{Hf, Si}^{HfSi} : {}^0a_{Hf, Si}^{HfSi} - {}^0b_{Hf, Si}^{HfSi} T$	${}^0G_{Hf, Si}^{HfSi} : -503250 + 98.26T$	${}^0G_{Hf, Si}^{HfSi} : -32910 \pm 7595 - 5.47 \pm 1.90T$
HfSi <sub>2</sub>	${}^0G_{Hf, Si}^{HfSi_2} : {}^0a_{Hf, Si}^{HfSi_2} - {}^0b_{Hf, Si}^{HfSi_2} T$	${}^0G_{Hf, Si}^{HfSi_2} : -468950 + 132.66T$	${}^0G_{Hf, Si}^{HfSi_2} : -32518 \pm 7259 + 0.36 \pm 2.43T$
liquid	${}^0L_{Hf, Si}^{liq} : {}^0a_{Hf, Si}^{liq} - {}^0b_{Hf, Si}^{liq} T$	${}^0L_{Hf, Si}^{liq} : -886370 - 9.47T$	${}^0L_{Hf, Si}^{liq} : -119005 \pm 13962 + 16.72 \pm 3.06T$
Si	${}^0G_{Hf, Si}^{Si} : {}^0a_{Hf, Si}^{Si} - {}^0b_{Hf, Si}^{Si} T$	${}^0G_{Hf, Si}^{Si} : -114810 + 68.22T$	${}^0G_{Hf, Si}^{Si} : 6842 \pm 794 - 2.72 \pm 0.32T$

plotting cumulative mean distributions of parameters can help identify the convergence regions. As examples, two of these plots are shown in Fig. 1. According to this figure, smooth changes towards constant values are observed for both parameters after 30,000 parameter generations, which correspond to the parameter convergence regions. Therefore, the first 30,000 generated samples are considered as the burn-in period (red shaded regions) and discarded from the total number of parameter samples. Since all the parameters generally converge at the same time, the burn-in period is assumed to be the same for the other model parameters (the other 18 parameters in model 2). To show that the parameters in other models also converge to their optimum values, the cumulative means of parameter  $a_{liq}$  are plotted for the other models as dotted lines in Fig. 1a.

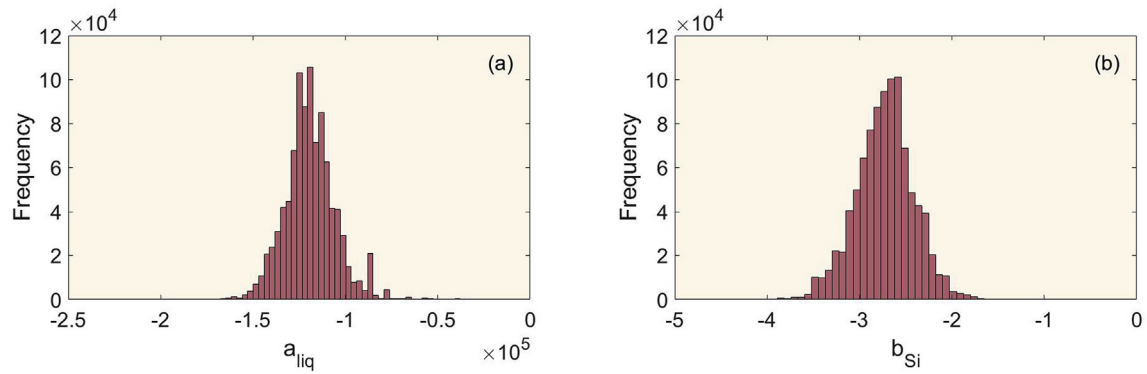
After the removal of the burn-in period, the mean values of the remaining parameter samples (970,000 samples) and the diagonal terms of their variance-covariance matrix are considered as the optimal plausible mean values and the overall reflected uncertainties (variances) for the model parameters, respectively. The initial values of the model parameters obtained through a deterministic optimization and their probabilistic values after MCMC calibration are listed in Table 2.

The marginal and joint posterior frequency distributions of the parameters can also be plotted to evaluate the convergence and dispersion of samples in the parameter space. Two examples of marginal posterior frequency histograms for parameters are plotted in Fig. 2. As can be observed in this figure, the parameter

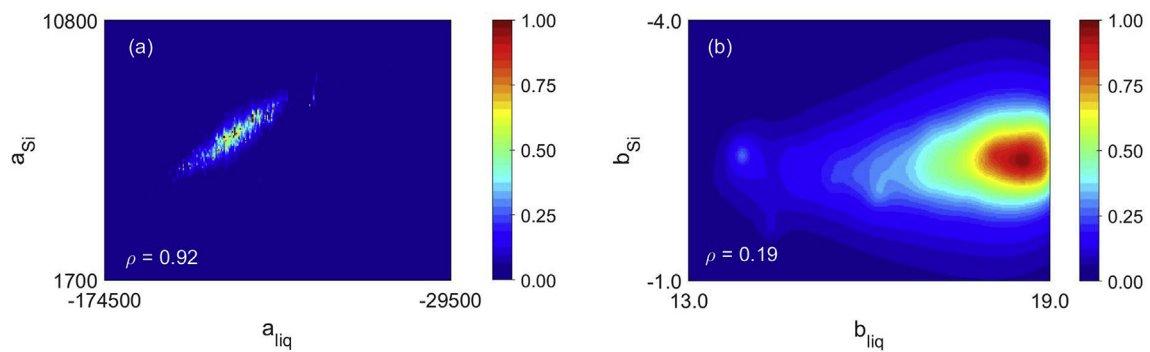
posteriors are almost in the form of Gaussian distributions. Since all the parameters show a similar form for their marginal posterior distributions, a distribution close to a multivariate Gaussian distribution is expected as the joint distribution for all the model parameters (20 parameters in the case of model 2). Fig. 3 shows the joint frequency distributions for two examples of a pair of parameters in model 2. Fig. 3 can also offer a qualitative representation of the correlation between the applied two parameters in each case. For example, the linearity and direction of the red regions show the strength and negativity/positivity of the correlation between parameters, quantified through the Pearson coefficient [47]:

$$\rho_{x,y} = \frac{cov(x,y)}{\sigma_x \sigma_y} \quad (18)$$

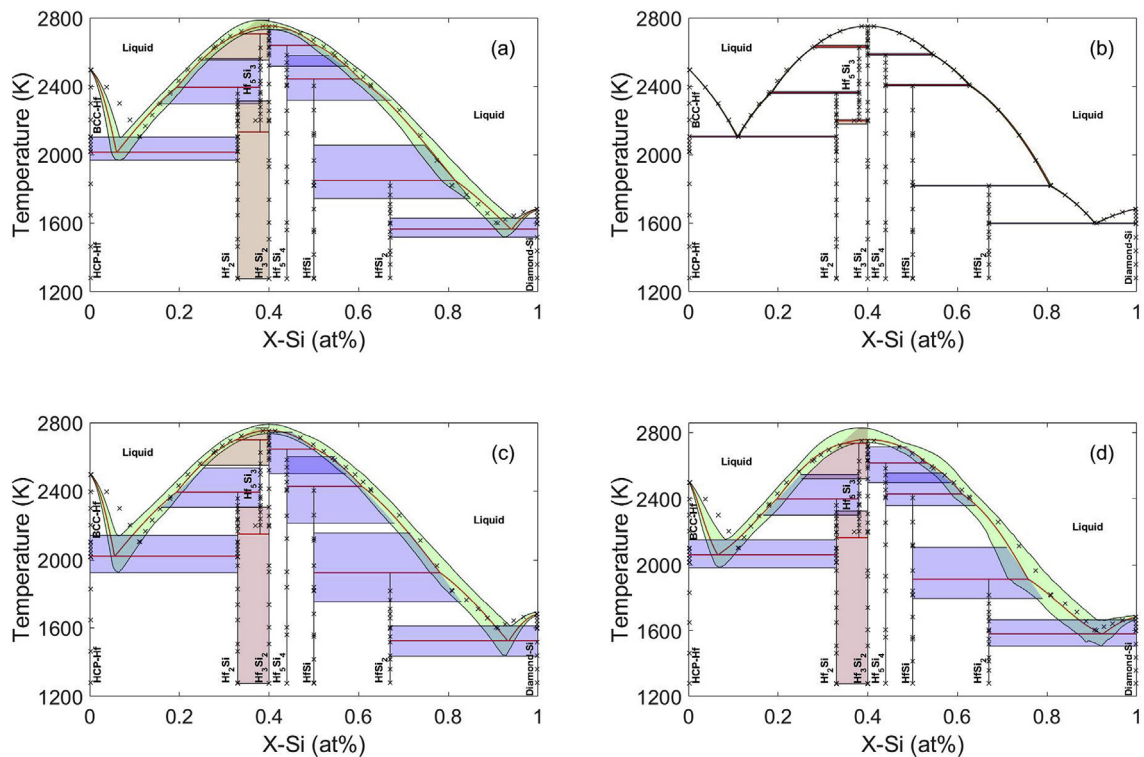
The same MCMC approach is applied to probabilistically calibrate the other three model parameter sets. Then, the overall uncertainties are propagated from parameters to phase diagram for each model through the model forward analysis. This process is performed using the forward calculation of the phase diagram by the last 5000 MCMC parameter samples as an ensemble of the whole convergence region. To find 95% Bayesian credible intervals (BCI), 2.5% of the resulting samples associated with liquidus and transformation lines are discarded from above and also below the total obtained uncertainty band at any specified composition. These results are shown for each model in Fig. 4. In this figure, red lines and shaded areas are the results obtained from the optimum



**Fig. 2.** Marginal posterior frequency distributions for the parameters a)  $a_{liq}$  and b)  $b_{Si}$  in model 2.



**Fig. 3.** Joint frequency distributions between the pair parameters a) ( $a_{Si}$ ,  $a_{liq}$ ) and b) ( $b_{Si}$ ,  $b_{liq}$ ) in model 2 besides their linear correlation coefficients and normalized colour bars. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)



**Fig. 4.** Optimum Hf–Si phase diagrams and their 95% BCI obtained from models 1–4 (a–d) after uncertainty propagation of the MCMC calibrated parameters in each case.

(mean) values of the parameters and 95% BCI for each model, respectively. It should be noted that these uncertainty bands are not sharp around eutectic points since the eutectic compositions are predicted slightly different for each parameter sample. A cross-sectional probability distribution can be achieved at any specific Si content along the green uncertainty interval, while there is a fixed (composition independent) cross-sectional probability distribution throughout any one of the blue or red uncertainty intervals. The red shaded regions in all models show unstable results with large uncertainties that result from the difficulty to predict the intermediate high temperature  $\text{Hf}^{\text{Si}}\text{Si}^3$  phase through CALPHAD modeling. Moreover, it is clear in Fig. 4 that there is a very good agreement between the results obtained from model 2 and the data with a very small uncertainty band.

For more clear demonstration of UP in this work, the propagation of uncertainty from the model parameters to the single-phase Gibbs free energies to phase diagram are illustrated in Fig. 5. In this figure, two isothermal examples of Gibbs free energy-Si composition curves are plotted for liquid and two different intermetallic phases in two models (i.e., models 1 and 2) to show how the uncertainties of single-phase Gibbs free energies and phase diagram can be related together. This relationship can clearly be observed in Fig. 5 through the consistency between the 95% BCIs of the phase equilibria obtained from parameter uncertainty propagations to Gibbs free energy of the single phases and resulting convex hulls at any specific temperature and their counterparts obtained from parameter uncertainty propagations to the corresponding phase diagram boundaries.

It is worth noting that the mean responses and uncertainty intervals in the Gibbs free energy graphs and phase diagrams have probably been affected by MSU. Generally, there are some statistical approaches such as Kennedy and OHagan approach [48] to quantify MSU in the presence of true data [5]. However, model 2 shows very small uncertainties and close mean value responses to the data after MCMC calibration that implies a very small MSU in this case. Therefore, larger models, i.e., models 3 and 4, are expected to have even less MSU due to fitting better to the data in the case that the global optimum for parameters are obtained during calibration (although the mean values obtained from models 3 and 4 are not close to the data in some parts of the phase diagrams due to the inability of the parameter global optimization in Thermo-calc software). For this reason, the large uncertainties for the phase diagrams obtained from models 3 and 4 are mostly attributed to MPU.

In the context of BMS/BMA, the weight of each one of the applied models can be obtained by calculating their corresponding posteriors given data using Equations (10) and (11). Based on these calculations, the weights associated with models 1 to 4 are 0.1352, 0.5938, 0.1331, and 0.1379, respectively. Model 2 thus has three times the weight of the other models, which otherwise have similar Bayesian importance, which is consistent with the phase diagram results in Fig. 4.

Traditionally, it has been generally expected that a better fit to experimental data can be achieved by simply increasing the number of parameters of the free energies. However, our results show that this is not always the case. The results hold, for example, when comparing model 1 to model 2, as increasing the complexity of the free energy functions clearly results in a much better fit to the data and a much narrower uncertainty bound. As the complexity of the models are increased, however, it is apparent that the uncertainty bounds become much worse. In this case, we have found that the most parsimonious model (i.e., simplest) that still has sufficient freedom to explain all the available data is clearly superior.

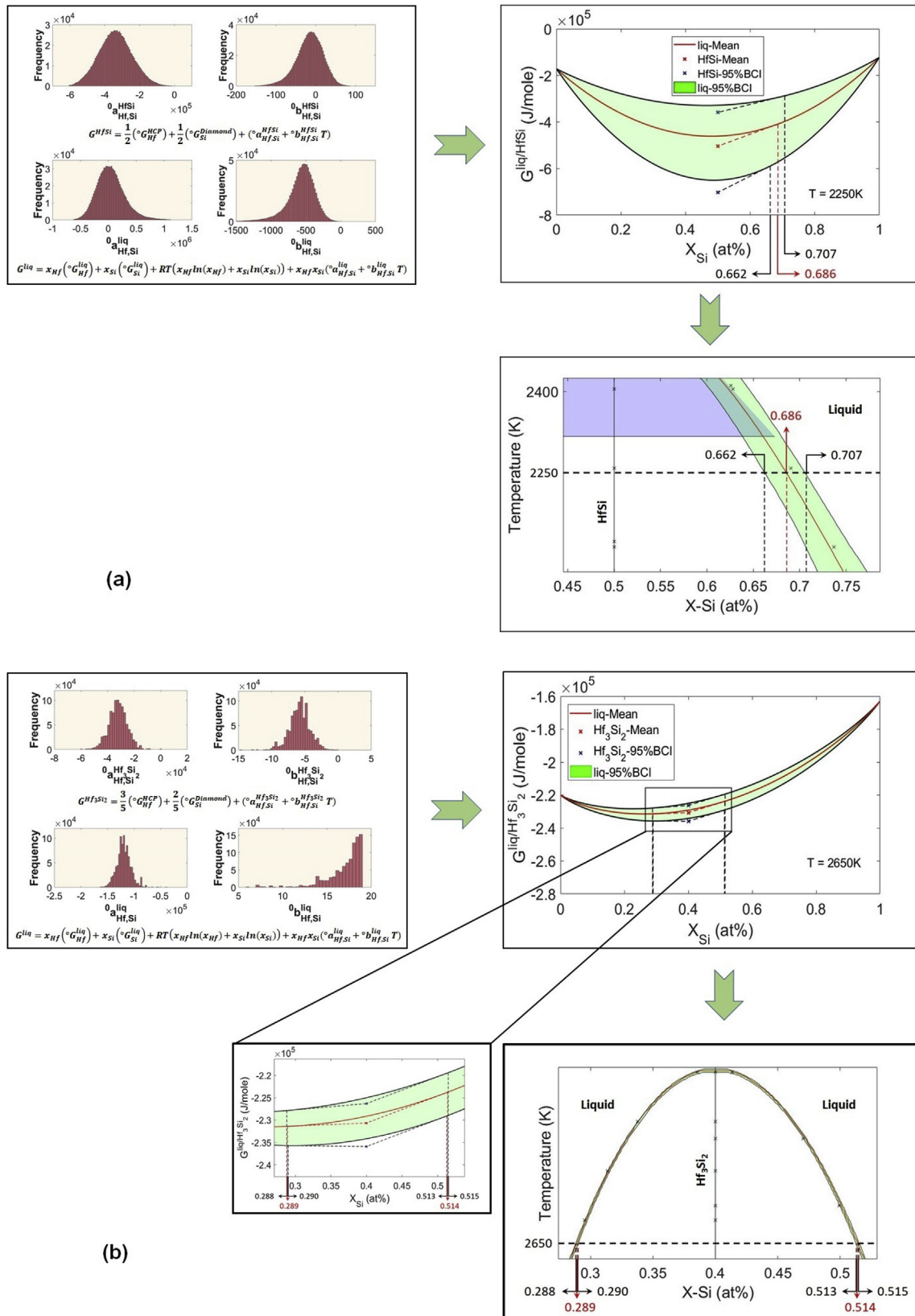
These results, although counter-intuitive to a degree, may be explained by imagining the fitness landscape of a particular model

as a multi-dimensional space, with many local minima corresponding to combinations of parameters that result in a fit with lower error relative to the neighborhood in model parameter space. Schematically, one can visualize this as shown in Fig. 6. The simplest model (blue) is depicted as having a relatively simple landscape with a sub-optimal fitness: the model is too simple to explain all the available data and the large uncertainty bounds are associated with the fitness error as a result of (small) variations in the values of individual parameters. Very complex models (red) have a high number of local minima as there are many combinations of parameters that result in somewhat-optimal fits. The fitting process, however, can get stuck in these local minima and since the MCMC results are inevitably biased by the parameter ranges defined based on the deterministically-attained values—to sample the model space at a reasonable cost—there is a significant probability that the global minimum is outside the parameter ranges, which results in an MCMC chain that converges to a sub-optimal region during MCMC sampling. The large uncertainty bounds in this case could be ascribed to the large sensitivity of such models to small deviations in the values of the parameters. The best model (green) has the right amount of complexity to fit all the available data without having too much model parameter uncertainty. There is thus a valid argument towards *model parsimony*: to make the model as simple as possible, but not more.

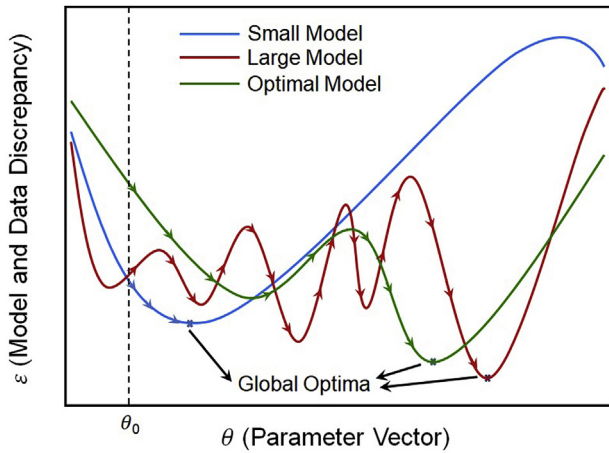
It must be stated, however, that while there is a clear superiority in model 2 in this case, the value of the inferior models should not be discounted outright as they may provide valuable information that can contribute to a more robust quantification of uncertainty. Based on Equation (12), weighted averages are taken over the cross-sectional probability distributions of all four phase diagrams in Fig. 4 at around 1000 Si compositions between 0 and 1 with a step size 0.001 for the green region and at just one arbitrary composition for each of the blue/red regions (since the cross-sectional probability distributions are fixed and composition independent along these regions) to calculate the average model and its uncertainty bands throughout the whole phase diagram, as shown in Fig. 7. For each individual model and weighted average model, the cross-sectional probability distributions obtained through kernel density estimation (KDE) are also demonstrated for all the transformation line uncertainty bands (the blue shaded regions with composition independent distributions) labeled from 1 to 8 and at some random compositions for the liquidus uncertainty band (across the green shaded region) labeled from 9 to 13 in supplementary material for Fig. 7 [49]. As can be observed, 95% BCIs of the average model (black lines in probability distribution plots) is broader than each one of the applied models and covers all the uncertainty ranges offered by the models 1 to 4. This is why BMA can provide more confidence in the results in the context of robust design.

As observed in Fig. 7, the phase diagram resulting from the mean value of probability distributions at different compositions and regions (red lines) leads to two unrealistic results (indentations) around the eutectic points at approximately  $X_{\text{Si}} = 0.1$  and 0.9. According to Fig. 7, this issue can be attributed to the difference of models in the prediction of eutectic point compositions, which results in averages of some hypo-eutectic points from model 2 with some hyper-eutectic points from models 1, 3 and 4 at small composition ranges. In order to solve this issue, the posterior modes of the probability distributions can be introduced as the optimum phase diagram rather than their mean values. The result is shown by red lines in Fig. 8. It is worth noting that the posterior mode of the probability distributions in the average model exactly corresponds to the posterior modes of the probability distributions in model 2, as indicated by black and purple lines in distribution plots of supplementary material for Fig. 7 [49], respectively.

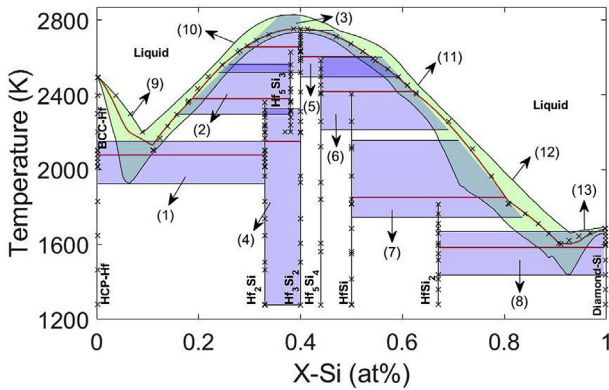




**Fig. 5.** Gibbs free energy-Si composition graphs with their 95% BCI obtained from the parameter uncertainty propagation for a) liquid and HfSi intermetallic phases in model 1 at 2250 K and b) liquid and Hf<sub>3</sub>Si<sub>2</sub> intermetallic phases in model 2 at 2650 K besides their corresponding phase diagram regions to show the uncertainty propagation from the parameters to single-phase Gibbs free energies to phase diagram.



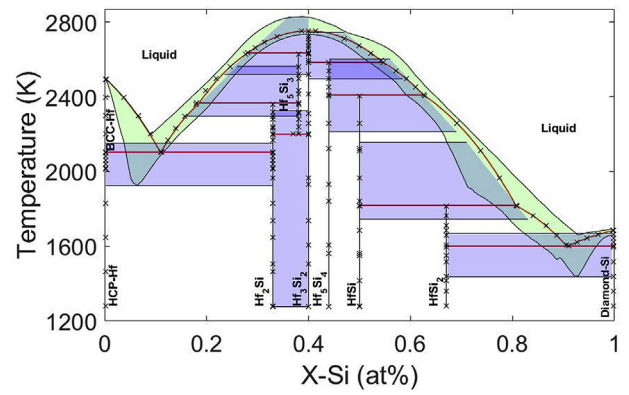
**Figure 6.** 2D schematic plots of the discrepancy between model and data in the parameter space for different model sizes (different number of parameters).



**Fig. 7.** Mean values and 95% BCIs at different compositions/regions in Hf–Si phase diagram obtained after BMA.

Therefore, the best model results can be considered as the optimum results for the average model, but with broader uncertainties, contributed by the inferior models.

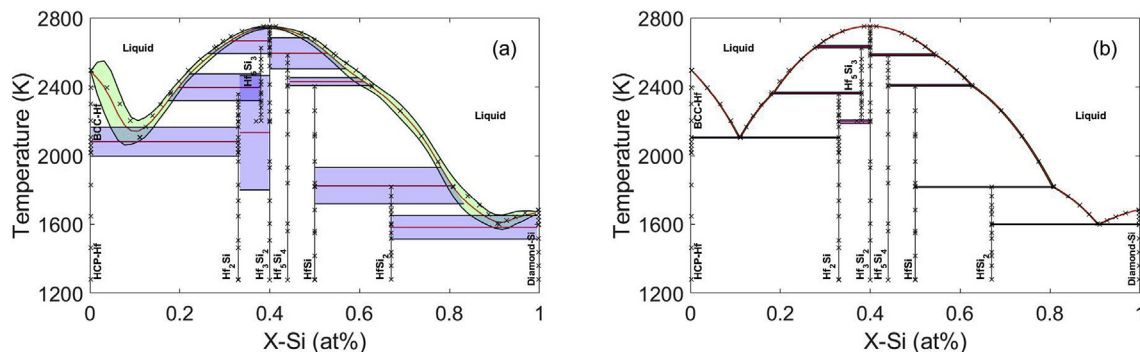
Generally, the dependency of the model precision to the number of parameters defined for the existing phases in the system has made CALPHAD modeling a hard task that requires a lot of experience. Therefore, the lack of a systematic approach towards thermodynamic assessments means that model search is essentially carried out in a trial-and-error mode or by relying on (considerable)



**Fig. 8.** Posterior modes and 95% BCIs at different compositions/regions in Hf–Si phase diagram obtained after BMA.

expert opinion. While the BMA approach already put forward in this work provides a robust estimate of the (weighed) output of a set of models, one drawback is that BMA assumes that models are statistically independent from each other. In this specific case, since the three suboptimal models have almost as much weight at the optimal model, the uncertainties carried out by the former overestimate (in a very conservative manner) the fused uncertainty. All models (at least in this case) have some degree of correlation as (i) they are describing the same underlying ground truth; (ii) have common model structures. We thus propose to exploit the statistical correlations between models as a strategy to arrive at an improved fused prediction of phase stability.

To perform this model fusion, the equivalent normal distributions at different compositions/regions are required to be calculated for each of the four models. Then, three and four models are fused based on the CMF approach, whose results are shown in Fig. 9. It should be noted that models 1, 3, and 4 with low precision and high uncertainties are first chosen for model fusion to examine whether the resulting fused model can be closer to the data and reduce the uncertainties. Fig. 9a shows that the approach can provide a phase diagram in much better agreement with data and less uncertainties compared to phase diagrams obtained from each one of the applied models individually. This result implies that random CALPHAD models can be fused together to find a reasonable estimation for phase diagram instead of trial-and-error to find the best predicting model. In addition, it is obvious that better predictions can be achieved as shown in Fig. 9b if model 2 (the best model) is also involved in the model fusion. However, it is very hard to compare the uncertainty of the fused model and model 2 based on the resulting phase diagrams. For this reason, the information



**Fig. 9.** Error correlation-based model fusions of a) three models (1, 3, and 4) and b) all four models.

**Table 3**  
Average entropy as a measure of uncertainty for each individual and fused model.

$H_M$	Model 1	Model 2	Model 3	Model 4	Bayesian Average Model	Fused Model from 3 sources	Fused Model from 4 sources
	1.8450	0.6553	1.9044	1.9876	1.4916	1.7923	0.6510

(Shannon) entropies are calculated to quantify the model uncertainties.

Shannon entropy can be utilized as a measure of uncertainty or missing information, which is determined as a weighted average (expected value) of information content gained from all the possible outcomes of an event. The information content obtained from an event outcome  $i$  is defined as the negative logarithm of its probability,  $-\log(p_i)$ . In this regard, it should be noted that event outcomes with lower probabilities convey more information since their occurrence are more surprising. In the case of  $r$  discrete possible outcomes for an event, entropy can be defined as the following summation [50]:

$$H = E[-\log(p_i)] = -\sum_{i=1}^r p_i \log(p_i). \quad (19)$$

Here, the transformation temperature resulting from each individual or fused model at a specific composition/region is considered as an event. For each individual or fused model, the average of all the entropies associated with probability distributions over 95% BCIs in different specified compositions/regions are introduced as the average entropy of that model:

$$H_M = \frac{1}{m} \sum_{j=1}^m H_j, \quad (20)$$

where  $m$  is the total number of probability distributions in different specified compositions/regions, which is the same for all the models. The average entropies of the models are listed in Table 3. It can be observed that model 2 has lower average entropy/uncertainty than model 1, 3 or 4, as expected.

The BMA fused model shows less average entropy/uncertainty compared to these three models. The lower uncertainty of the BMA fused model can be attributed to its concentrated distribution around the results of model 2, which gives more certainty to the total distributions over 95% BCIs. However, the average model still has broader uncertainty bounds compared to each of the individual models, which can provide more confidence in robust design. In other words, broader uncertainties can give more assurance about the presence of specific microstructural phases corresponding to a phase diagram region of interest by more shrinking the safe design space of composition-temperature in that region. In addition, it seems that the BMA model can incorporate some additional information about the uncertainties in the phase diagram compared to model 2 (the best model).

The incorporation of uncertainty from the individual models through CMF can be more optimal than BMA due to the consideration of model error correlations, as the given thermodynamic models clearly are statistically correlated. According to Table 3, it is clear that the fused models can yield better predictions with less average entropies/uncertainties than the individual models used in each fusion case.

## 5. Summary and conclusion

Due to the importance of uncertainty quantification in CALPHAD, an MCMC sampling approach is utilized in this work for the

probabilistic calibration of CALPHAD model parameters against the available data in the case of the Hf–Si binary system. Considering the vast high-dimensional parameter space in CALPHAD modeling, applying some prior information about the parameter values and ranges from Thermo-Calc optimization module is often required to achieve parameter convergence with a reasonable cost. However, choosing an appropriate CALPHAD model with a sufficient number of parameters is a challenging task.

Therefore, a systematic approach is required to find an optimal model rather than a trial-and-error approach. For this purpose, Bayesian hypothesis testing (or Bayesian model selection) based on Bayes' factors is proposed in this work and applied in a case study for the Hf–Si system to show how the best model can be chosen from a list of independently generated models. However, our work suggests to use information fusion approaches to smartly combine the given individual models into a fused model rather than just the application of the best model that may lose some useful information. BMA and an error correlation-based model fusion are used for our Hf–Si case study to show different beneficial purposes of these information fusion approaches.

The average model obtained from BMA shows larger 95% confidence intervals compared to any one of the individual models, which can provide more confidence for robust design but is likely too conservative. On the other hand, the error correlation-based technique can provide closer results to data with less uncertainties than the individual models used for the fusion. The uncertainty reductions through this fusion approach are also verified through the comparison of the average entropies (as a measure of uncertainty) obtained for the individual and fused models. Therefore, random CALPHAD models can smartly be fused together to find reasonable predictions for phase diagrams with no need to go through the cumbersome task of identifying the best CALPHAD models.

## Acknowledgement

Authors would like to thank the support of the National Science Foundation (NSF) [grant nos. CMMI-1534534, CMMI-1663130, DGE-1545403]. RA and DA also acknowledge the support of ARL through [grant No. W911NF-132-0018].

## Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.actamat.2018.11.007>.

## References

- [1] National Research Council, Integrated Computational Materials Engineering: a Transformational Discipline for Improved Competitiveness and National Security, National Academies Press, 2008.
- [2] J.H. Panchal, S.R. Kalidindi, D.L. McDowell, Key computational modeling issues in integrated computational materials engineering, *Comput. Aided Des.* 45 (1) (2013) 4–25.
- [3] A. Chernatynskiy, S.R. Phillpot, R. LeSar, Uncertainty quantification in multi-scale simulation of materials: a prospective, *Annu. Rev. Mater. Res.* 43 (1) (2013) 157–182.
- [4] S. Ghoreishi, D. Allaire, Adaptive uncertainty propagation for coupled multi-disciplinary systems, *AIAA J.* (2017) 1–11.
- [5] H.-J. Choi, D. McDowell, J. Allen, F. Mistree, An inductive design exploration

- method for hierarchical systems design under uncertainty, *Eng. Optim.* 40 (2008) 287–307.
- [6] S.F. Ghoreishi, Uncertainty Analysis for Coupled Multidisciplinary Systems Using Sequential Importance Resampling, Master's Thesis, Texas A&M University, 2016.
  - [7] S.F. Ghoreishi, D.L. Allaire, Compositional uncertainty analysis via importance weighted gibbs sampling for coupled multidisciplinary systems, in: 18th AIAA Non-deterministic Approaches Conference, 2016, p. 1443.
  - [8] J.J. Mortensen, K. Kaasbjerg, S.L. Frederiksen, J.K. Nørskov, J.P. Sethna, K.W. Jacobsen, Bayesian error estimation in density-functional theory, *Phys. Rev. Lett.* 95 (21) (2005) 216401.
  - [9] F. Hanke, Sensitivity analysis and uncertainty calculation for dispersion corrected density functional theory, *J. Comput. Chem.* 32 (7) (2011) 1424–1430.
  - [10] M. Aldegunde, N. Zabaras, J. Kristensen, Quantifying uncertainties in first-principles alloy thermodynamics using cluster expansions, *J. Comput. Phys.* 323 (2016) 17–44.
  - [11] F. Rizzi, M. Salloum, Y. Marzouk, R. Xu, M. Falk, T. Weihs, G. Fritz, O. Knio, Bayesian inference of atomic diffusivity in a binary Ni/Al system based on molecular dynamics, *Multiscale Model. Simul.* 9 (1) (2011) 486–512.
  - [12] B. Vliker, P. Marton, C. Elssner, M. Kamlah, Multiscale modeling for ferroelectric materials: a transition from the atomic level to phase-field modeling, *Continuum Mech. Therm.* 23 (5) (2011) 435–451.
  - [13] W.K. Liu, L. Siad, R. Tian, S. Lee, D. Lee, X. Yin, W. Chen, S. Chan, G.B. Olson, L.-E. Lindgren, M.F. Horstemeyer, Y.-S. Chang, J.-B. Choi, Y.J. Kim, Complexity science of multiscale materials via stochastic computations, *Int. J. Numer. Methods Eng.* 80 (6–7) (2009) 932–978.
  - [14] B. Kouchmeshky, N. Zabaras, Microstructure model reduction and uncertainty quantification in multiscale deformation processes, *Comput. Mater. Sci.* 48 (2) (2010) 213–227.
  - [15] M. Koslowski, A. Strachan, Uncertainty propagation in a multiscale model of nanocrystalline plasticity, *Reliab. Eng. Syst. Saf.* 96 (9) (2011) 1161–1170.
  - [16] S. Salehghaffari, M. Rais-Rohani, E.B. Marin, D.J. Bammann, A new approach for determination of material constants of internal state variable based plasticity models and their uncertainty quantification, *Comput. Mater. Sci.* 55 (2012) 237–244.
  - [17] E. Knipsberger, Improvement of excess parameters from thermodynamic and phase diagram data by a sequential Bayes algorithm, *Calphad* 15 (1) (1991) 69–78.
  - [18] W. Olbricht, N.D. Chatterjee, K. Miller, Bayes estimation: a novel approach to derivation of internally consistent thermodynamic data for minerals, their uncertainties, and correlations. Part I: Theory, *Phys. Chem. Miner.* 21 (1–2) (1994) 36–49.
  - [19] N.D. Chatterjee, K. Miller, W. Olbricht, Bayes estimation: a novel approach to derivation of internally consistent thermodynamic data for minerals, their uncertainties, and correlations. Part II: Application, *Phys. Chem. Miner.* 21 (1–2) (1994) 50–62.
  - [20] D.V. Malakhov, Confidence intervals of calculated phase boundaries, *Calphad* 21 (3) (1997) 391–400.
  - [21] N.D. Chatterjee, R. Krger, G. Haller, W. Olbricht, The Bayesian approach to an internally consistent thermodynamic database: theory, database, and generation of phase diagrams, *Contrib. Mineral. Petrol.* 133 (1–2) (1998) 149–168.
  - [22] J. Acker, K. Bohmhammel, Optimization of thermodynamic data of the NiSi system, *Thermochim. Acta* 337 (1) (1999) 187–193.
  - [23] M. Stan, B.J. Reardon, A Bayesian approach to evaluating the uncertainty of thermodynamic data and phase diagrams, *Calphad* 27 (3) (2003) 319–323.
  - [24] R.A. Otis, Z.-K. Liu, High-throughput thermodynamic modeling and uncertainty quantification for ICME, *JOM* 69 (5) (2017) 886–892.
  - [25] T.C. Duong, A. Talapatra, W. Son, M. Radovic, R. Arryave, On the stochastic phase stability of Ti 2 AlC–Cr 2 AlC, *Sci. Rep.* 7 (1) (2017) 5138.
  - [26] T.C. Duong, R.E. Hackenberg, A. Landa, P. Honarmandi, A. Talapatra, H.M. Volz, A. Lobet, A.I. Smith, G. King, S. Bajaj, A. Ruban, L. Vitos, P.E.A. Turchi, R. Arryave, Revisiting thermodynamics and kinetic diffusivities of uraniumniobium with Bayesian uncertainty analysis, *Calphad* 55 (2016) 219–230.
  - [27] J.-C. Zhao, B.P. Bewlay, M.R. Jackson, Q. Chen, Hf–Si binary phase diagram determination and thermodynamic modeling, *JPE* 21 (1) (2000) 40.
  - [28] Y. Yang, Y.A. Chang, J.C. Zhao, B.P. Bewlay, Thermodynamic modeling of the NbHfSi ternary system, *Intermetallics* 11 (5) (2003) 407–415.
  - [29] J.C. Zhao, B.P. Bewlay, M.R. Jackson, Determination of NbHfSi phase equilibria, *Intermetallics* 9 (8) (2001) 681–689.
  - [30] M. Hillert, The compound energy formalism, *J. Alloy. Comp.* 320 (2) (2001) 161–176.
  - [31] A. Dinsdale, Sgte data for pure elements, *Calphad* 15 (4) (1991) 317–425.
  - [32] A. Gelman, J.B. Carlin, H.S. Stern, D.B. Dunson, A. Vehtari, D.B. Rubin, *Bayesian Data Analysis*, vol. 2, CRC press, Boca Raton, FL, 2014.
  - [33] L.J. Billera, P. Diaconis, A geometric interpretation of the Metropolis-Hastings algorithm, *Stat. Sci.* 16 (4) (2001) 335–339.
  - [34] S.M. Lynch, *Introduction to Applied Bayesian Statistics and Estimation for Social Scientists*, Springer Science & Business Media, 2007.
  - [35] H. Jeffreys, Some tests of significance, treated by the theory of probability, *Math. Proc. Camb. Phil. Soc.* 31 (2) (1935) 203–222.
  - [36] R.E. Kass, A.E. Raftery, Bayes factors, *J. Am. Stat. Assoc.* 90 (430) (1995) 773–795.
  - [37] J.A. Hoeting, D. Madigan, A.E. Raftery, C.T. Volinsky, Bayesian model averaging: a tutorial, *Stat. Sci.* 14 (4) (1999) 382–401.
  - [38] T.M. Fragoso, W. Bertoli, F. Louzada, Bayesian model averaging: a systematic review and conceptual classification, *Int. Stat. Rev.* 86 (1) (2018) 1–28, <https://doi.org/10.1111/insr.12243>.
  - [39] T. Dasey, J. Braun, Information fusion and response guidance, *Linc. Lab. J.* 17 (1).
  - [40] S. Geisser, A Bayes approach for combining correlated estimates, *J. Am. Stat. Assoc.* 60 (310) (1965) 602–607.
  - [41] P.A. Morris, Combining expert judgments: a Bayesian approach, *Manag. Sci.* 23 (7) (1977) 679–693.
  - [42] R.L. Winkler, Combining probability distributions from dependent information sources, *Manag. Sci.* 27 (4) (1981) 479–488.
  - [43] S.F. Ghoreishi, D.L. Allaire, A fusion-based multi-information source optimization approach using knowledge gradient policies, in: 2018 AIAA/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference, 2018, p. 1159.
  - [44] S.J. Julier, J.K. Uhlmann, A non-divergent estimation algorithm in the presence of unknown correlations, in: *Proceedings of the 1997 American Control Conference* (Cat. No.97CH36041), vol. 4, 1997, pp. 2369–2373.
  - [45] D. Allaire, K. Willcox, Fusing information from multifidelity computer models of physical systems, in: 2012 15th International Conference on Information Fusion, 2012, pp. 2458–2465.
  - [46] W.D. Thomason, D.L. Allaire, A model reification approach to fusing information from multifidelity information sources, in: 19th AIAA Non-deterministic Approaches Conference, 2017, p. 1949.
  - [47] D.J. Battle, P. Gerstoft, W.S. Hodgkiss, W. Kuperman, P.L. Nielsen, Bayesian model selection applied to self-noise geoaoustic inversion, *J. Acoust. Soc. Am.* 116 (4) (2004) 2043–2056.
  - [48] M.C. Kennedy, A. O'Hagan, Predicting the output from a complex computer code when fast approximations are available, *Biometrika* 87 (1) (2000) 1–13.
  - [49] See Supplementary Material.
  - [50] J.F. Wellmann, K. Regenauer-Lieb, Uncertainties have a meaning: information entropy as a quality measure for 3-d geological models, *Tectonophysics* 526 (2012) 207–216.