Uncertainty quantification of phase boundary in a composition-phase map via Bayesian strategies

Biao Wu [1,2] Haihui Zhang,[1,2] Yuanxun Zhou [1,2] Lanting Zhang [1,2,3,*] and Hong Wang[1,2,3]
[1]School of Materials Science and Engineering, Shanghai Jiao Tong University, Shanghai 200240, China
[2]Materials Genome Initiative Center, Shanghai Jiao Tong University, Shanghai 200240, China
[3]Shanghai Key Laboratory of High-Temperature Materials and Precision Forming, Shanghai Jiao Tong University, Shanghai 200240, China

(Received 27 April 2022; accepted 26 January 2023; published 13 February 2023)

Phase boundary indicates the conditions of transition between phase regions, which is a key constituent of a phase diagram. We propose an approach to determine the phase boundary and its uncertainty in a phase-composition map based on the data from high-throughput experimentation. Bayesian logistics regression combined with the domain knowledge of phase diagrams was applied in the approach. For a typical ternary isothermal section, both the linear and nonlinear phase boundaries as well as the vertices of a ternary tie triangle were modeled to quantify the uncertainty with consideration of a couple of affecting factors such as data-point density, noise in the data, data coverage, etc. The effectiveness of the present approach was demonstrated by the Fe-Cr-Ni isothermal section data from database and the Fe-Co-Ni composition-phase map data by experiment. Moreover, the uncertainty of the phase boundary in the ternary system can be further reduced by incorporating the available data from the subbinary systems. The data-driven nature of the developed approach can further guide the efficient implementation of high-throughput experiments and provide confidence measures for decision-making in materials design and further Computer Coupling of Phase Diagrams and Thermochemistry (CALPHAD) method modeling.

DOI: 10.1103/PhysRevMaterials.7.025201

I. INTRODUCTION

Phase diagrams are crucial and fundamental for the development of materials [1]. A phase diagram is a graphical illustration of the substance phase regions in a space, either a composition space or a temperature-pressure space, etc. Experimental determination of a phase diagram is cumbersome for it requires determining the thermodynamic equilibrium state in the space point to point. Modeling by Computer Coupling of Phase Diagrams and Thermochemistry (CALPHAD) method [2] relies heavily on reliable experimental data. However, experimental data are limited, especially for the complicated alloy systems due to the time-consuming and labor-intensive nature of the work. Thus, efforts are devoted to accelerating the construction of the phase diagram, particularly for a complicated alloy system.

Early in the 1960s, Kennedy *et al.* [3] pioneered to determine isothermal sections of ternary alloy phase diagrams in a rapid way using deposited films with a composition spread. Other related attempts have been reviewed in Ref. [4] by Wong-Ng. One of the key techniques in these attempts is using a piece of thin film with a composition spread instead of a uniformly fixed composition. Such a piece of thin film is called a combinatorial materials chip or a materials library [5–12]. Combinatorial materials-chip method, which simply classifies phase region in the composition (X) space by one chip, has shown great potential in the experimental determination of phase diagrams. This method features synthesis and point-by-point high-throughput structure and chemical characterizations of a chip with a composition spread; thus, the phase boundaries can be determined from the dataset of characterization (Fig. 1). Phase boundary is the key geometric features of a phase diagram in 2D space, and a phase diagram can be constructed once all phase boundaries are determined. Unlike the conventional way, the combinatorial materials-chip method is data intensive so that machine-learning algorithms can be applied to make the best usage of data. In addition, data generated in such a way are comprehensive and systematic, which naturally suits the machine-learning algorithm. Recent works have reported a machine-learning–based classification approach for the efficient construction of phase diagrams [13–15]. Among various algorithms, Bayesian logistic regression can determine the phase boundary based on the data points located on both sides of the boundary, which are generated by the characterization of the combinatorial materials chip. Besides, it can provide the probability distribution of the phase boundary and thus provides critical information for the design and optimization of materials.

It is realized that there is a certain degree of uncertainty in any experimentally determined phase boundary. Uncertainty in the experimentally determined composition-phase map by combinatorial materials-chip method derives from several sources, including random and systematic errors in the x-ray-diffraction (XRD) and x-ray fluorescence (XRF) spectroscopy measurements. For instance, the determination limit of a phase in XRD is $\sim 3\%$, depending on the structure, grain size, signal-to-noise ratio, etc. Besides, the density of

*Corresponding author: lantingzh@sjtu.edu.cn

 ©2023 American Physical Society

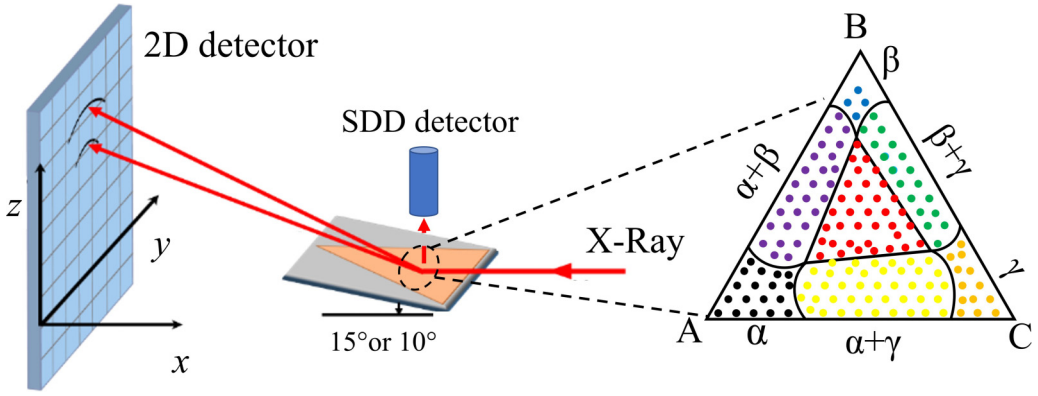

FIG. 1. Schematic diagram of constructing a composition-phase map of a ternary system by the combinatorial materials-chip method. By identifying the XRD patterns and XRF spectra point by point, a composition-phase map is obtained (shown on the right-hand side) (for details please refer to the previous work [7]).

data points during mapping, the unequal data coverage in the phase region, the form of phase boundary, and the accuracy of the machine-learning algorithm are all sources of uncertainty. While these uncertainties are known in the construction of a composition-phase map, quantification of the uncertainty of the phase boundary in a composition-phase map is infrequently addressed rigorously. Understanding and quantification of the uncertainties provide a confidence measure for decision-making in materials design and further CALPHAD modeling.

To quantify the uncertainty in the phase diagram, uncertainty sampling has been employed [13,14]. Paulson *et al.* [16] provided a method to quantify the uncertainty of the location of invariant points and the phase diagram. Besides, probability distributions for phase fractions, compositions, activities, sublattice site fractions, and Gibbs energies can all be accessed. In another work, Bayesian model averaging and an error-correlation–based model fusion approach have been used for probabilistic calibration of CALPHAD model parameters against the available data in the case of the Hf-Si binary system [17]. Besides, a method based on Markov chain Monte Carlo (MCMC) has been used to assess the diffusivity error in molecular dynamic simulations of a Ni/Al nanolaminate bilayer in the context of Bayesian statistics [18].

In this study, a method based on Bayesian statistics is developed to determine and evaluate the uncertainty of all possible phase boundaries under the principle of the Gibbs phase rule and cross rule in 2D space. MCMC Metropolis-Hastings (M-H) algorithm [19] was used to seek an approximate solution to the posterior distribution. Mathematical models based on Bayesian statistics have been established to determine the probability density of two different types of phase boundaries and the vertices of a ternary tie triangle. The determined probability distribution or credit interval of the phase boundary is a quantitative description of the uncertainty of the phase boundary. Even the points located far away from the boundary play a role in the probability-density distribution of the phase boundary.

This approach was first demonstrated using a known Fe-Cr-Ni isothermal section phase diagram where the data were generated using the THERMO-CALC software. Then, high-throughput experimental data of the Fe-Co-Ni ternary system were used as a case study of the approach. In addition, known binary phase-diagram data were incorporated into the approach to reduce the probability distribution of the phase boundaries. It is shown that the present approach is efficient in constructing the phase diagrams and offers fruitful information for further thermodynamic modeling and an active design of experiments.

II. METHODOLOGY

The Bayesian theorem states that given some data $\boldsymbol{D}$ and a model M, the relationship between the posterior distribution of the model parameter vector θ and the observed data is

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

$P(\theta|\boldsymbol{D},\, M)$ is the posterior distribution of the parameters. $P(\boldsymbol{D}|\theta,\, M)$ is the conditional probability of the data given the model and associated parameters. $P(\theta|,\, M)$ is the prior for the model parameters. $P(\boldsymbol{D}|M)$ is the marginal likelihood describing the probability of the data being generated by the model [20]. Though the Bayesian theorem gives a method to calculate the posterior distribution of the parameters from the priors, it is however difficult to evaluate the posterior distribution in an exact way for many problems. A numerical method—MCMC—-is thus employed to seek an approximate solution to the exact posterior distribution. Open-source PYMC3 PYTHON package was utilized for probabilistic calibration of the parameters in this work [21].

A. M-H algorithm

The M-H algorithm [22], which makes successive jumps in parameter space according to the proposal distribution, is a common method employed to sample the posterior distribution. Whether the sampled parameter vector in each iteration can be accepted or rejected depends on the M-H ratio r:

$$r = \min\left\{1,\, \frac{\pi(\theta')q(\theta',\, \theta)}{\pi(\theta)q(\theta,\, \theta')}\right\}. \tag{2}$$

Here, π is the target distribution, q is the conditional distribution, and θ and θ' are the model parameter vector. Sampled

parameter vector will be accepted if r exceeds a uniformly distributed random number between zero and one. When the parameter vectors are sampled by the MCMC M-H algorithm, qualitative measures of convergence of the chains for the parameters are conducted by the Gelman-Rubin statistic [23].

B. Gelman-Rubin statistic

In the context of Gelman and Rubin, it is assumed that h Markov chains have been simulated independently, each of length n. If the chains have achieved convergence, the inference is made by computing the sample mean and variance of each chain. In the case of convergence, the obtained h inferences should be similar. λ_{ij} is the function of interest evaluated at the jth observation from chain i.

$$\bar{\lambda}_i = \frac{1}{n}\sum_{(j=1)}^{n}\lambda_{ij},\ \bar{\lambda} = \frac{1}{h}\sum_{i=1}^{h}\bar{\lambda}_i. \tag{3}$$

$\bar{\lambda}_i$ is the mean of the ith chain. $\bar{\lambda}$ is the mean of all sampled parameters. The mean of the h within-sequence variances is given by

$$W = \frac{1}{h}\sum_{i=1}^{h}\frac{1}{n-1}\sum_{j=1}^{n}(\lambda_{ij}-\bar{\lambda}_i)^2. \tag{4}$$

The between-sequence variances are given by

$$B = \frac{n}{h-1}\sum_{i=1}^{h}(\bar{\lambda}-\bar{\lambda}_i)^2. \tag{5}$$

Then, the marginal posterior variance V is estimated by a weighted average of W and B:

$$\hat{V} = \frac{n-1}{n}W + \frac{1}{n}B. \tag{6}$$

In general, for any finite n, the within-sequence variances W should be an underestimate of V, as the individual sequences have not had time to range over all of the target distribution, resulting in less variability [24]. Then, the potential scale reduction is estimated by

$$\hat{R}_c = \sqrt{\frac{\hat{V}}{W}}\,. \tag{7}$$

As n increases, $\hat{V}$ decreases and W increases. The $\hat{R}_c$ the threshold for the Gelman-Rubin diagnostic to declare convergence is <1.1 for most cases.

III. MODELING OF A TYPICAL TERNARY ISOTHERMAL SECTION

In the combinatorial materials-chip method, the composition-phase map of a ternary system is constructed by point-to-point high-throughput structural and chemical characterizations and classification of the XRD patterns. While the techniques to classify the XRD patterns were discussed in detail elsewhere [25], the results of the classification show a map of the phase region (including order-disorder transition) of each data point. A typical composition-phase map of a ternary system or an isothermal section consists of three types of phase regions according to the Gibbs phase rule [degree of freedom $F = 3 - P$ (number of phases)]: single-phase region, two-phase region, and three-phase region (Fig. 2). Based on the Gibbs phase rule and the law of adjoining phase regions, there are thus two types of phase boundary: the linear boundary separating the two-phase and three-phase regions and the nonlinear boundary between the single-phase and two-phase regions [Fig. 2(a)]. Besides, there are also vertices of a ternary tie angle according to the cross rule, where two linear boundaries and two nonlinear boundaries meet [Fig. 2(a)]. Mathematical models are developed to evaluate the uncertainties of two types of phase boundaries as well as the vertices.

A. Linear boundary

To evaluate the uncertainty of the linear boundary, the boundary from the upper left corner to the lower right corner of Fig. 2(b) was modeled by 400 points evenly distributed in a unit square according to the following relationships:

$$\begin{aligned} T_i\,(x_i,\ y_i) &= 1 \quad \text{if} \quad x_i + y_i - 1 > 0 \\ T_i\,(x_i,\ y_i) &= 0 \quad \text{if} \quad x_i + y_i - 1 < 0, \end{aligned} \tag{8}$$

where (x_i, y_i) is the index Cartesian coordinates of the ith data point. Here, $T_i = 1$ or 0 stands for different phase regions. Independence of the data points is assumed and thus, the likelihood function L is the product of n Bernoulli trials as follows:

$$L = \prod_{\mathrm{i}=1}^{\mathrm{i}=\mathrm{n}} P_i^{T_i}(1-P_i)^{(1-T_i)}, \tag{9}$$

$$P_i = 1/[1+\exp(-Z_i)]\ , \tag{10}$$

$$Z_i = a_1 + a_2\mathrm{x_i} + a_3\mathrm{y_i}\ . \tag{11}$$

Z represents the model of boundary, which is chosen based on the available knowledge about this boundary, i.e., in the linear form [Eq. (11)]. Usually, the priors are chosen based on the best available knowledge about the parameters and the relationships between parameters. In this model, uniform distribution $\mathcal{U}$ was selected as there is no available knowledge about the parameters; uniform distribution $\mathcal{U}(-10,\ 10)$ was assigned for the set of parameters $(a_1,\ a_2,\ a_3)$.

Next, the MCMC M-H algorithm was employed to evaluate the posterior distributions of the parameter. Figure 3 displays the convergence behavior of the set of parameters $(a_1,\ a_2,\ a_3)$ for 20 chains with each chain recording 500 sampled values of the parameter. The paths of two chains are highlighted in black and red for the sake of visibility in Figs. 3(a)–3(c). The convergence of each chain was assessed by the Gelman-Rubin statistic ($\hat{R}_c$ threshold = 1.1). The $\hat{R}_c$ was estimated to be 1.023 78, 1.0115, and 1.046 96 for a_1, a_2, and a_3, respectively. To examine the posterior distribution of the parameters more clearly, kernel-density estimated (KDE) distributions of each parameter are plotted in Figs. 3(d)–3(f). Kernel-density estimation is the process of estimating an unknown probability-density function using a kernel function. Here, the Gaussian kernel was employed.

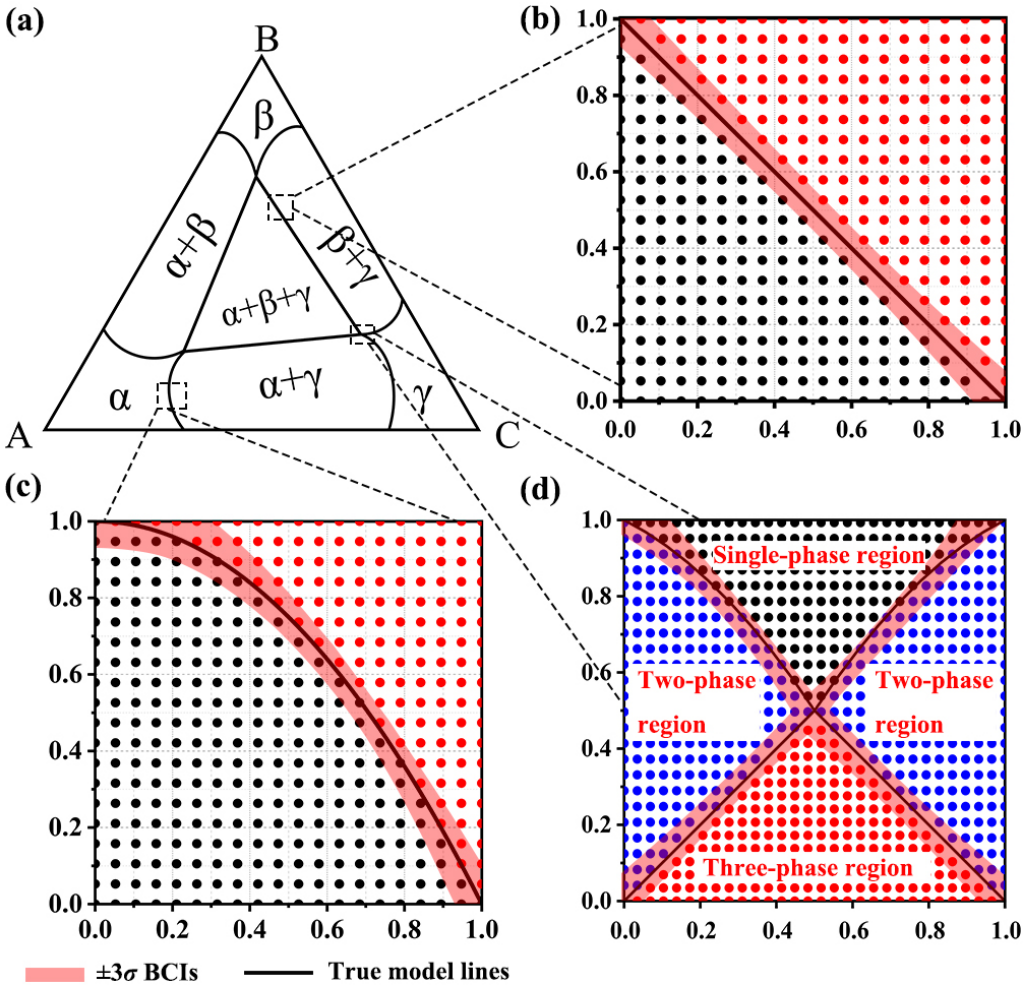

FIG. 2. (a) Isothermal section of a model ternary system. Bayesian credible intervals (BCIs) of the two types of boundaries simulated by 400 evenly distributed points: (b) linear boundary and (c) nonlinear boundary. (d) BCIs of the vertices of a ternary tie triangle (750 evenly distributed points).

The Bayesian credible intervals (BCIs) of the boundary were obtained by randomly sampling the posterior distributions of the model parameters and computing the desired model predictions. As shown in Fig. 2(b), the model fits the linear boundary well, and the prediction from the Bayesian statistical nearly reproduces the true boundary and exhibits tight uncertainty intervals, verifying the effectiveness of the method.

B. Nonlinear boundary

To model the nonlinear boundary, the evenly distributed 400 points in a unit square were generated according to the following relationships denoting different phase regions:

$$T_i\,(x_i,\ y_i) = 1 \quad \text{if} \quad x_i^2 + y_i - 1 > 0$$
$$T_i\,(x_i,\ y_i) = 0 \quad \text{if} \quad x_i^2 + y_i - 1 < 0. \qquad (12)$$

Similar to the linear boundary, the likelihood function L is the product of n Bernoulli trials [Eq. (9)], the only difference is that the form of the boundary should be changed as follows:

$$Z_i = a_1 + a_2 x_i + a_3 x_i^2 + a_4 y_i\ . \qquad (13)$$

When the boundary is nonlinear and no more knowledge can be applied, an initial fitting of the boundary should be conducted. For simplicity, a quadratic polynomial was then selected [Eq. (14)]. Uniform distributions $\mathcal{U}(-10,\ 10)$ were applied to the set of parameters $(a_1,\ a_2,\ a_3,\ a_4)$, similar to that in the linear boundary case [Eq. (11)].

According to Fig. 2(c), the BCIs of the nonlinear boundary coincided well with the true model, demonstrating the robustness of the method. Next, the effect of data-point density is discussed as follows.

C. Gaussian noise and mapping density

To simulate the random error in a real experiment, Gaussian noise of $\mathcal{N}(0,\ 0.05^2)$ was applied to the data points in the nonlinear boundary case. T_i stays the same as in Eq. (13) before noise has been applied for the Cartesian coordinate of each point. Figure 4(a) depicts the situation when the noise is applied, showing that the result from the Bayesian statistics coincides with the true boundary well in most parts of the boundary. This indicates the robustness of the method. Compared with the case without Gaussian noise, the uncertainty

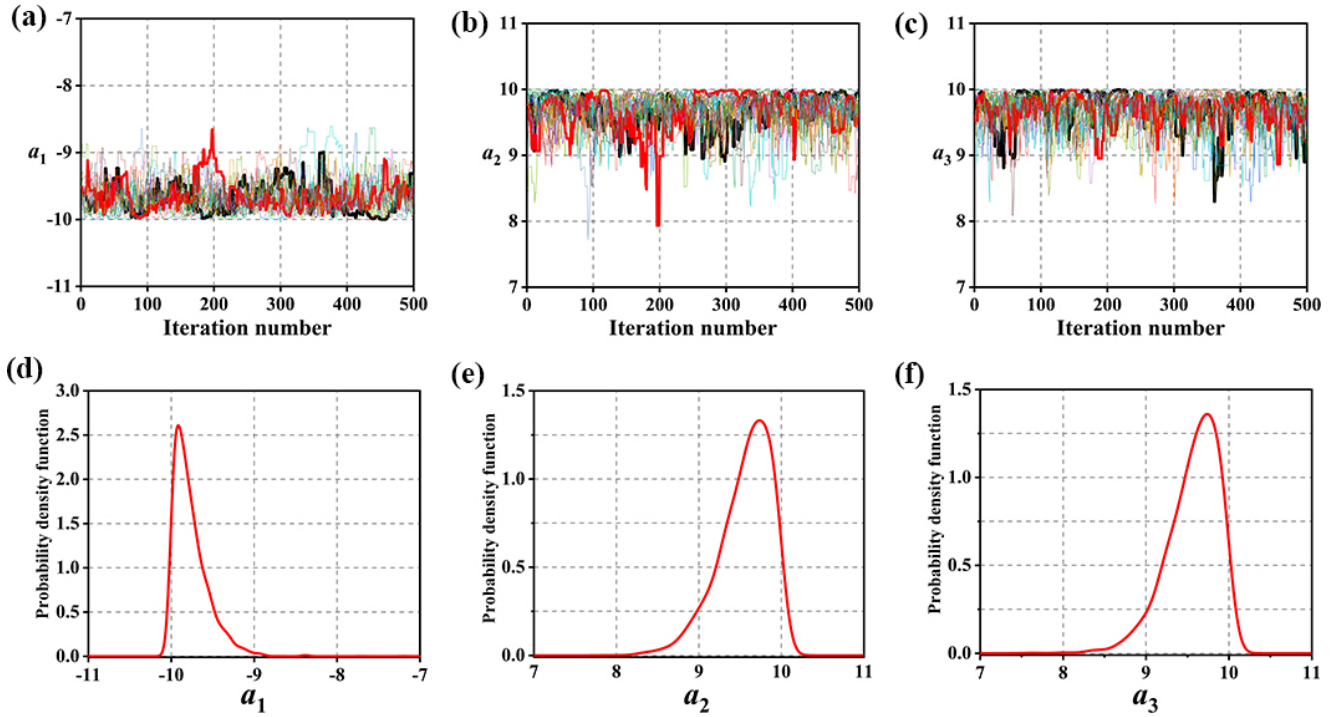

FIG. 3. Sequences of 20 chains from the M-H algorithm over 500 iterations for the parameters (a) a_1, (b) a_2, and (c) a_3 (two randomly selected chains are highlighted in black and red). KDE distributions for the parameters (d) a_1, (e) a_2, and (f) a_3.

increased slightly because the noise itself is a certain kind of uncertainty [Fig. 4(b)].

Since the composition-phase map is constructed by mapping a combinatorial materials chip point to point, the effect of mapping density on the uncertainty of the boundary is further studied. For simplicity, the linear boundary was considered. In the aforementioned section, 400 data points were evenly distributed in the unit square. In this section, the uncertainties of the linear boundary with 100 and 225 points per unit area were calculated and compared with that of 400 points per unit area. According to Figs. 5(a), 5(b), and 2(b), the BCIs of the boundary decrease as the mapping density increases. The distribution of 2σ in Fig. 5(c) shifts upward with decreasing mapping density, whose value is in reverse proportional to the square root of the mapping density. Since the model and priors for the parameter vector were the same regardless of the mapping density in the Bayesian analysis, it can be understood that the increased mapping density could offer more priors for the analysis, resulting in less uncertainty. Moreover, this result also indicates that in the process of the experiment, the

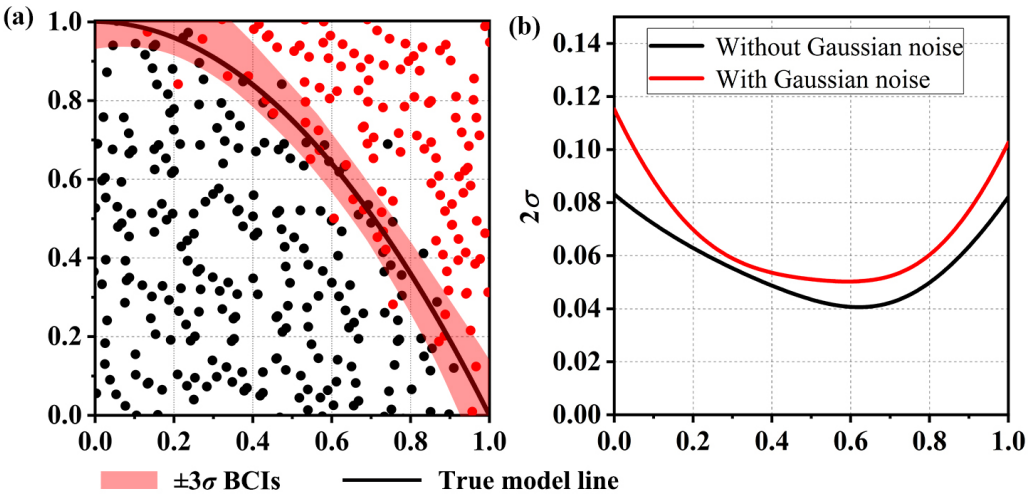

FIG. 4. (a) BCIs of the nonlinear boundary obtained from the data points with Gaussian noise. (b) Comparison of the 2σ distribution along the x axis with and without Gaussian noise.

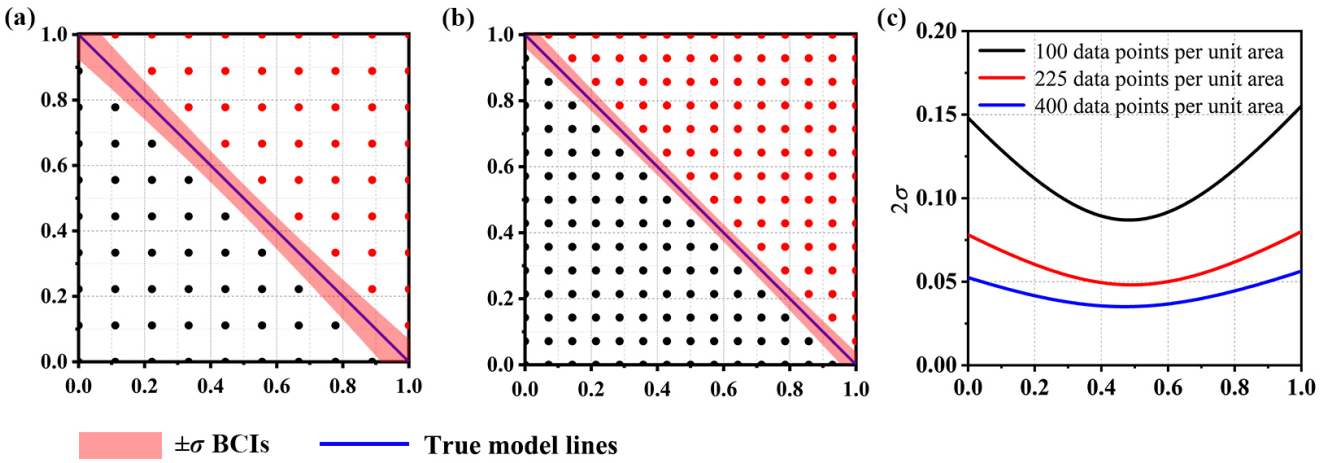

FIG. 5. BCIs of the boundary obtained from the data points with different mapping densities: (a) 100 points per unit area; (b) 225 points per unit area. (c) Distribution of the 2σ along the x axis.

uncertainty of the phase boundary in the composition-phase map could be controlled to an acceptable level by rationally controlling the mapping density.

D. Data-point coverage in the vicinity of the boundary

Compared with the traditional method where only a small number of data points can be acquired, a large number of data points can be collected on a small substrate by the combinatorial materials-chip approach. To further benchmark the advantage of the combinatorial materials-chip approach, the uncertainties of the boundaries with different data-point coverage were calculated in this section. The data-point coverage in the vicinity of the boundary is defined as the maximum distance normal to the boundary (D) [Fig. 6(a)]. For simplicity, the linear boundary was considered for demonstration purposes. The spacing between the data points was the same as in Fig. 2(b) (i.e., 400 points per unit area).

The effect of the D on the uncertainty of the boundary is obvious when D is within 0.141 [Figs. 6(a)–6(c)]. The 2σ shifts upward as D decreases [Fig. 6(f)], suggesting that all the data points are useful in determining the boundary even though some of them are some distance away from the boundary. However, the weight of the data point on the uncertainty of the boundary depends on its distance to the boundary. When the D is larger than 0.141, the change of σ is not obvious [Figs. 6(d) and 6(e)], suggesting that information on the data points near the boundary within a certain range is enough to determine the location of the boundary accurately and there is no need to know the distribution of data points in all the phase regions. Hence, this can guide the mapping of a combinatorial materials chip so that the data acquisition is both efficient and effective to determine a boundary.

E. Vertices of a ternary tie triangle

To model the vertices of a ternary tie triangle where four phase boundaries meet (Fig. 2), the 2D space of the unit square is partitioned into four regions: 1 three-phase region, 2 two-phase regions, and 1 single-phase region by 2 linear boundaries and 2 nonlinear boundaries. The 750 data points were classified into four regions [Fig. 2(d)] according to the following relationship:

$$T_i\,(x_i,\ y_i) = 1 \quad \text{if} \quad (y_i < x_i) \quad \text{and} \quad (y_i < 1 - x_i) \quad \text{else} \quad T_i\,(x_i,\ y_i) = 0$$
$$S_i\,(x_i,\ y_i) = 1 \quad \text{if} \quad \left(y_i > 1 - \frac{x_i}{2} - x_i^2\right) \quad \text{and} \quad (y_i > 2.5x_i - x_i^2 - 0.5) \quad \text{else} \quad S_i(x_i,\ y_i) = 0, \tag{14}$$

where $(x_i,\ y_i)$ is the index Cartesian coordinates of the ith data point. The region of each data point is determined by T_i and S_i according to the following table (Table I).

As shown in Fig. 2(d), the three-phase region is neighbored by the two-phase regions with two linear boundaries. The boundary between the single-phase region and the two-phase regions is nonlinear, which is similar to Fig. 2(c). The likelihood function L is changed into

$$L = \prod_{i=1}^{i=n} P_i^{T_i}(1-P_i)^{(1-T_i)} \prod_{i=1}^{i=n} Q_i^{S_i}(1-Q_i)^{(1-S_i)}, \tag{15}$$

$$P_i = 1/(1 + \exp\,(-\min\,(A_i,\ B_i))), \tag{16}$$
$$A_i = -\left(a_1x_1 + a_2x_1^2 + a_3y_1\right) + a_1x_i + a_2x_i^2 + a_3y_i, \tag{17}$$
$$B_i = -\left(a_4x_1 + a_5x_1^2 + a_6y_1\right) + a_4x_i + a_5x_i^2 + a_6y_i, \tag{18}$$
$$Q_i = 1/(1 + \exp(-\min(C_i,\ D_i))), \tag{19}$$
$$C_i = -(a_7x_1 + a_8y_1) + a_7x_i + a_8y_i, \tag{20}$$
$$D_i = -(a_9x_1 + a_{10}y_1) + a_9x_i + a_{10}y_i, \tag{21}$$

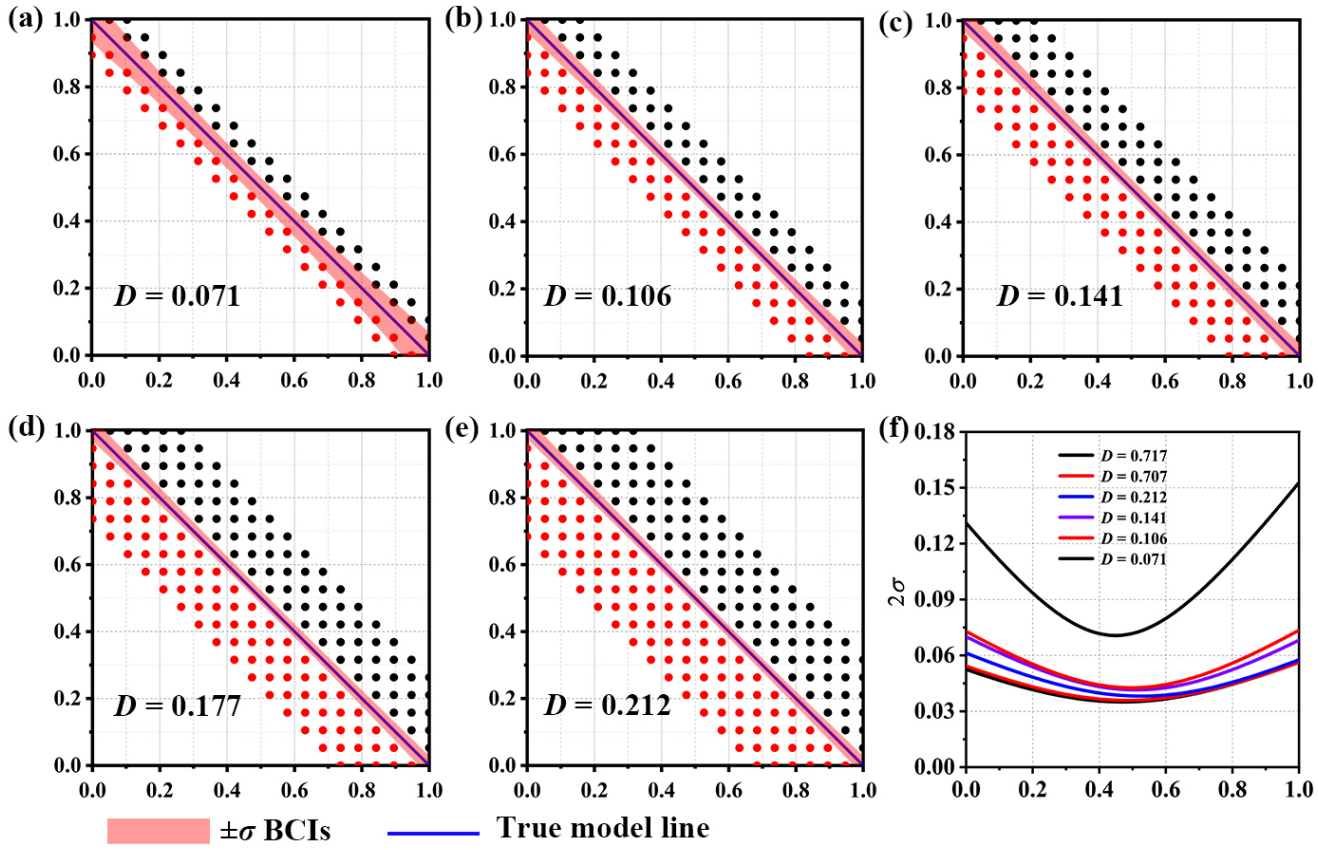

FIG. 6. (a)–(e) BCIs of the boundaries with different D (maximum distance normal to the boundary). (f) Distribution of 2σ along the x axis.

where n is the number of the data point; (x_1, y_1) is the vertex of the ternary tie triangle in Fig. 2(d). The model guarantees the boundaries meet at the same point (x_1, y_1). A_i and B_i represent the model of the two nonlinear boundaries in Fig. 2(d). The $\min(A_i, B_i)$ is to determine whether the ith data point locates above the two boundaries simultaneously. If A_i and B_i are both larger than 0, it can be speculated that the ith data point should locate above the two boundaries. C_i and D_i represent the model of the two linear boundaries in Fig. 2(d). Uniform priors $\mathcal{U}(-10, 10)$ were applied to the set of parameters $(a_1, a_2, a_3, a_4, a_5, a_6, a_7, a_8, a_9, a_{10})$ and uniform prior distributions $\mathcal{U}(0.3, 0.7)$ were applied to the set of parameters (x_1, y_1).

The calculated BCIs agree well with the ground truth of the phase boundaries [Fig. 2(d)]. Unlike the models in Figs. 2(b) and 2(c) where there is a single boundary in each case, the model here deals with four boundaries and one intersecting simultaneously. The significant discrepancy in Fig. 2(d) compared with the previous models is that there exists an inflexion point for the BCIs of the boundary. This inflexion point is the vertex of a ternary tie triangle. The probability-density distribution of the vertex can also be determined from the sampled (x_1, y_1) directly (Fig. 7). It is found that the ground truth of the vertex locates within the predicted BCIs, suggesting the effectiveness of the model.

TABLE I. Classification of data points according to T_i and S_i in Eqs. (17) and (18).

T_i	S_i	Region
0	1	Single-phase
0	0	Two-phase
1	0	Three-phase

IV. CASE STUDY AND DISCUSSION

In this section, the before-established method and model are further employed to quantify the uncertainty of the phase boundary in a ternary isothermal section of the Fe-Cr-Ni system and a Fe-Co-Ni composition-phase map experimentally determined by the combinatorial materials-chip method.

A. Fe-Cr-Ni ternary isothermal section

The Fe-Cr-Ni ternary isothermal section at 800 °C was obtained using the THERMO-CALC 2021A with the FEDEMO4.0 database. To simplify the problem, part of the phase diagram was selected, with Ni from 15 to 100%; Fe and Cr, from 0 to 85% [Fig. 8(a)]. In this portion, there are four phase regions (i.e., the fcc single-phase region, the body-centered-cubic (bcc) + fcc two-phase region, the fcc + σ two-phase region, and the fcc + σ + bcc three-phase region) with four boundaries plus a vertex of a ternary tie triangle. According to the phase regions calculated by the THERMO-CALC [Figs. 8(b) and 8(c)], 380 and 100 data points were uniformly generated

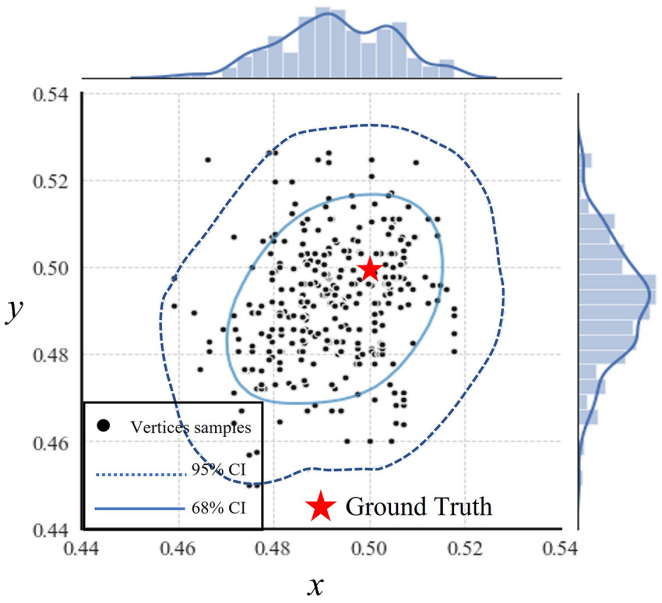

FIG. 7. The vertex point (x_1, y_1) plotted for 1000 sampled parameter sets with 68 and 95% BCIs.

in the composition space. BCIs of the boundaries and vertex can be determined from the model established in Eq. (16). Comparing Fig. 8(b) with Fig. 8(c), the BCIs of the boundaries increase with the decreasing number of data points even though the true boundaries locate within the determined BCIs. The BCIs are relatively large at the edge of the composition triangle due to the decreased effective data-point density, which is consistent with the situation described in Fig. 5. It is also found in Fig. 8(c) that the BCIs are so large that the three-phase region cannot be confidently identified. This illustrates a lower limit of the data-point density to distinguish a phase region. Similarly, the probability-density distribution of the vertex point exhibits the same trend: the distribution range of 68 and 95% uncertainty intervals increases with decreasing data points [Figs. 8(d) and 8(e)].

B. Fe-Co-Ni composition-phase map

The Fe-Co-Ni system and its subbinary systems relate to some key applications in materials science and engineering. There are three-phase regions in an experimentally determined composition-phase map of the Fe-Co-Ni chip heat treated at 700 °C: the bcc single-phase region, the bcc + fcc two-phase region, and the fcc single-phase region [Fig. 9(a)]. Details of the experimental procedure and the corresponding results were reported in a previous paper [7]. On one chip, 336 data points were characterized. In addition, a thermodynamic assessment of the Fe-Co-Ni system has been reported recently [26], which can be employed as a reference to illustrate the effectiveness of the method.

C. Phase-boundary determination

Unlike in the model systems, a suitable form of the formula for the boundary should be found before computing the BCIs. In many actual systems, analytical equations for the phase boundary between the single-phase region and the two-phase coexisting region cannot be established because of the excess Gibbs free energy. Polynomial models are then employed for cases where analytical equations do not exist. To further make use of the domain knowledge in materials science to guide the selection of the polynomial for the boundary, a simplified model was employed within the framework of thermodynamics [Fig. 9(b)]. First, the α phase is assumed to be an A-rich regular dilute solution and the β phase to be a regular solid solution, though there is significant solubility of Co in bcc-Fe below the Curie temperature. The results of this simplified model are only used to provide a reference for the selection of polynomials. The chemical potentials (partial molar quantity) μ of components A, B, and C for α solid-solution phase can be described as follows:

$$\mu_A^\alpha = {}^0G_A^\alpha + RT\ln x_A^\alpha, \tag{22}$$

$$\mu_B^\alpha = {}^0G_B^\alpha + I_{AB}^\alpha + RT\ln x_B^\alpha, \tag{23}$$

$$\mu_C^\alpha = {}^0G_C^\alpha + I_{AC}^\alpha + RT\ln x_c^\alpha, \tag{24}$$

where x_A^α, x_B^α, and x_c^α are the mole fraction of A, B, and C in the α phase, respectively. R is the gas constant and T is the absolute temperature. ${}^0G_A^\alpha$, ${}^0G_B^\alpha$, and ${}^0G_c^\alpha$ indicate the molar Gibbs energy of the pure components A, B, and C, respectively. I_{AB}^α and I_{AC}^α are the binary interaction parameters. Under the normal solution approximation, they are constants.

For β solid-solution phase, the chemical potentials μ of components A, B, and C are as follows:

$$\mu_A^\beta = {}^0G_A^\beta + RT\ln x_A^\beta + {}^E\mu_A^\beta, \tag{25}$$

$${}^E\mu_A = I_{AB}^\beta(1-x_A^\beta)x_B^\beta + I_{AC}^\beta(1-x_A^\beta)x_C^\beta - I_{BC}^\beta x_B^\beta x_C^\beta, \tag{26}$$

$$\mu_B^\beta = {}^0G_B^\beta + RT\ln x_B^\beta + {}^E\mu_B^\beta, \tag{27}$$

$${}^E\mu_B = I_{BC}^\beta(1-x_B^\beta)x_C^\beta + I_{BA}^\beta(1-x_B^\beta)x_A^\beta - I_{AC}^\beta x_A^\beta x_C^\beta, \tag{28}$$

$$\mu_C^\beta = {}^0G_C^\beta + RT\ln x_C^\beta + {}^E\mu_C^\beta, \tag{29}$$

$${}^E\mu_C = I_{CA}^\beta(1-x_C^\beta)x_A^\beta + I_{CB}^\beta(1-x_C^\beta)x_B^\beta - I_{AB}^\beta x_A^\beta x_B^\beta, \tag{30}$$

where the meaning of the parameters is similar to that described before. It is known that when the two phases are in equilibrium, the chemical potential of all the components in the α phase and β phase are equal. Then, the following equations can be derived:

$${}^0G_A^\beta + {}^E\mu_A^\beta + RT\ln x_A^\beta = {}^0G_A^\alpha + RT\ln x_A^\alpha, \tag{31}$$

$${}^0G_B^\alpha + I_{AB}^\alpha + RT\ln x_B^\alpha = {}^0G_B^\beta + RT\ln x_B^\beta + {}^E\mu_B^\beta, \tag{32}$$

$$K_B^{\alpha/\beta} = \frac{x_B^\alpha}{x_B^\beta} = \exp\frac{{}^0G_B^\beta - {}^0G_B^\alpha - I_{BA}^\alpha + {}^E\mu_B^\beta}{RT}, \tag{33}$$

$${}^0G_C^\alpha + I_{AC}^\alpha + RT\ln x_C^\alpha = {}^0G_C^\beta + RT\ln x_C^\beta + {}^E\mu_C^\beta, \tag{34}$$

$$K_C^{\alpha/\beta} = \frac{x_C^\alpha}{x_C^\beta} = \exp\frac{{}^0G_C^\beta - {}^0G_C^\alpha - I_{CA}^\alpha + {}^E\mu_B^\beta}{RT}. \tag{35}$$

For a dilute solution where $x_C^\alpha << 1$ and $x_B^\alpha << 1$, the following equations can be established:

$$\ln x_A^\alpha = \ln(1 - x_B^\alpha - x_C^\alpha) \approx -x_B^\alpha - x_C^\alpha. \tag{36}$$

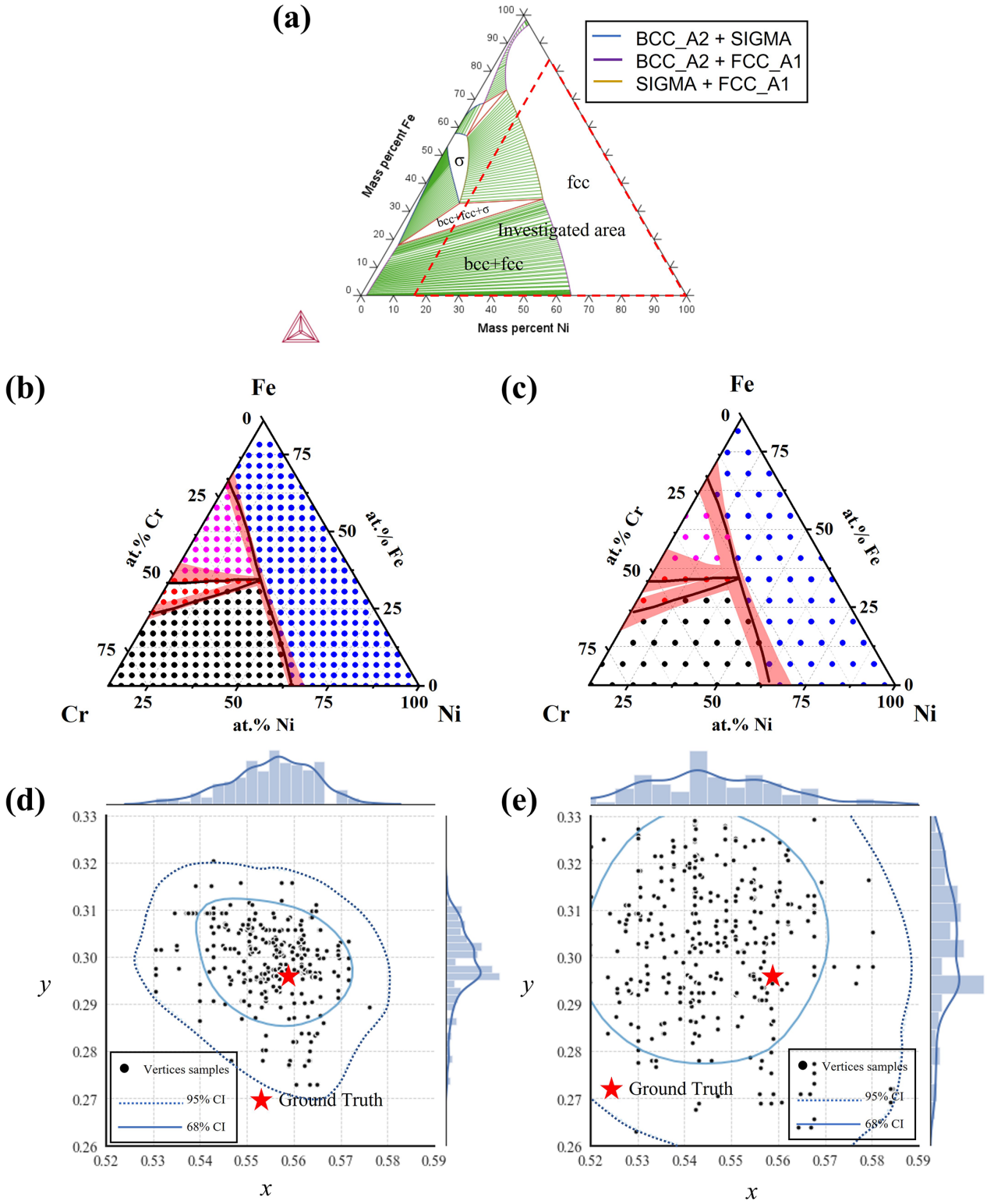

FIG. 8. (a) Fe-Cr-Ni ternary isothermal section at 800 °C generated by THERMO-CALC. Calculated BCIs of the phase boundary for the portion of the Fe-Cr-Ni system indicated by the dashed triangle in (a) with a different number of data points: (b) 100 and (c) 380 (the phase boundaries from THERMO-CALC are superimposed in the plot). (d), (e) The 68% and 95% BCIs of the vertex with 100 and 380 data points respectively.

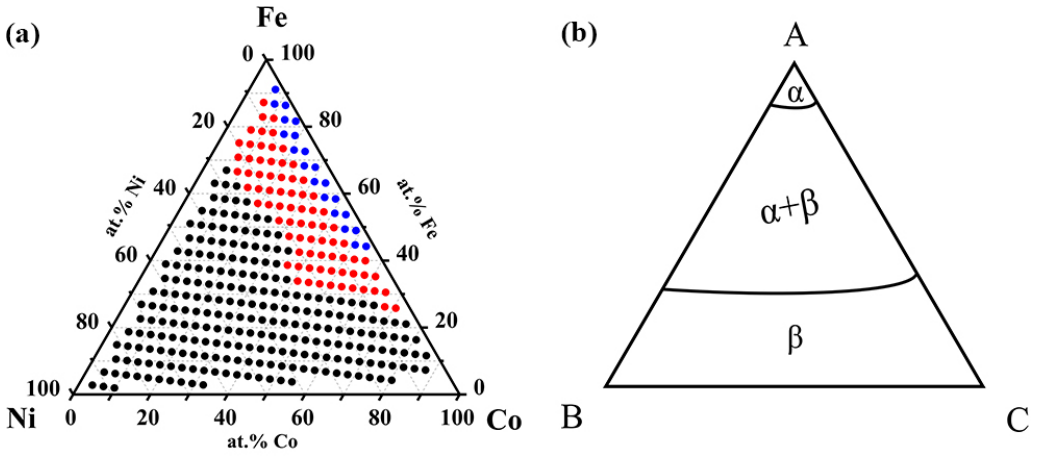

FIG. 9. (a) Manually constructed compositionphase map from the Fe–Co–Ni chip heat treated at 700 °C with 336 data points (blue dots: bcc; black dots: fcc; red dots: fcc+bcc). (b) A simplified thermodynamic model for the Fe–Co–Ni ternary system.

Then, based on the above equations, the formula for the phase boundary between the $\alpha+\beta$ two-phase coexisting region and the β single-phase region can be established. Here, only x_A^β , x_B^β , and x_C^β are the independent variable.

$$
\begin{aligned}
&{}^0G_A^\beta - {}^0G_A^\alpha + {}^E\mu_A^\beta + RT\ln x_A^\beta \\
&\quad = RT\left(-x_B^\beta\, e^{\frac{{}^0G_B^\beta - {}^0G_B^\alpha - I_{AB}^\alpha + {}^E\mu_B^\beta}{RT}} - x_C^\beta\, e^{\frac{{}^0G_C^\beta - {}^0G_C^\alpha - I_{AC}^\alpha + {}^E\mu_C^\beta}{RT}}\right).
\end{aligned}
\tag{37}
$$

Taylor expansion was employed to further simplify the above formula. The formula of the boundary can be summarized as follows:

$$
\begin{aligned}
&C_1 + {}^E\mu_A^\beta + RT\left(-x_B^\beta - x_C^\beta + \frac{\left(x_B^\beta + x_C^\beta\right)^2}{2}\right) \\
&\quad = RT\left(-C_2\left(1 + \frac{{}^E\mu_B^\beta}{RT}\right)x_B^\beta - C_3\left(1 + \frac{{}^E\mu_C^\beta}{RT}\right)x_C^\beta\right),
\end{aligned}
\tag{38}
$$

$$
C_1 = {}^0G_A^\beta - {}^0G_A^\alpha\,, \tag{39}
$$

$$
C_2 = e^{\frac{{}^0G_B^\beta - {}^0G_B^\alpha - I_{AB}^\alpha}{RT}}\,, \tag{40}
$$

$$
C_3 = e^{\frac{{}^0G_C^\beta - {}^0G_C^\alpha - I_{AC}^\alpha}{RT}}. \tag{41}
$$

For the isothermal section of a ternary system, C_1, C_2, T, and C_3 are all constant. Thus, the formula of the boundary between the β single-phase region and the $\alpha+\beta$ two-phase region can be expressed as the polynomial of the x_B^β and x_C^β .

Based on the above analysis, the form of the phase boundary is as follows:

$$
\begin{aligned}
Z_i = {} & a_1 + a_2x_i + a_3x_i^2 + a_4y_i + a_5y_i^2 + a_6x_iy_i + a_7x_i^3 \\
& + a_8x_i^2\,y_i + a_9x_iy_i^2 + a_{10}y_i^3.
\end{aligned}
\tag{42}
$$

Similar to the previous analysis, x_i and y_i index the position of the data point. $T_i = 1$ if the data point belongs to fcc single-phase region; otherwise, $T_i = 0$. Here, uniform priors $\mathcal{U}(-10,\ 10)$ were assigned for the parameters. Parameters were all diagnosed to converge.

Figure 10(a) shows the calculated BCIs of the boundary between the fcc single-phase region and the bcc + fcc two-phase region. Similarly, the BCIs of the boundary between the bcc single-phase region and the bcc + fcc two-phase region can also be computed in this way. The calculated intervals agree fairly well with the manual identification results.

D. Incorporating the data from binary systems

Meanwhile, Fig. 10(a) shows that the BCIs of two different phase boundaries are relatively large, especially at the edge of the composition triangle. This is due to lack of data in the experiment. To fix this deficiency, additional effort is made to reduce the BCI of the phase boundary. The difference between the classical statistical models and Bayesian statistical models lies in whether prior information is employed [27]. Bayesian statistical model attaches great importance not only to the use of sample information but also to the prior information. Then, it is expected that the deviation between the prediction and the actual end of truth would decrease if more priors about the phase diagram could be added to the model. It is realized that the ternary phase diagram should match the binary phase diagrams at the edge of the composition triangle. Therefore, the phase-boundary points of the binary systems available in the ASM ALLOY PHASE DIAGRAM DATABASE™ were used to help determine the location of the phase boundary of the ternary composition-phase map in the present case.

Though the phase-boundary points of the binary systems are available in the ASM database, the results of different investigations show a certain diversity. Probability-density distribution was employed to describe the location of the phase boundary accurately. Assuming the position of the phase boundary in the ASM database follows a uniform distribution, all the available data were collected from the database. For example, in the Fe-Ni system at 700 °C, the Fe contents at the boundary between the fcc single-phase region and the fcc + bcc two-phase region are 8.43% [28], 12.00% [29],

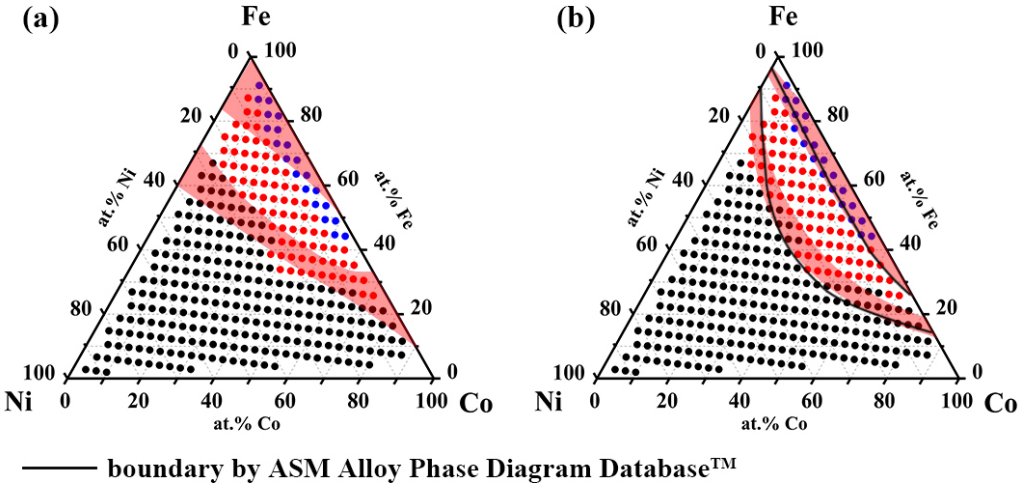

FIG. 10. (a) Calculated BCIs of the phase boundaries based on the experiment data points. (b) Calculated BCIs of the phase boundaries based on the experiment data points and the known binary phase-diagram data from the database (the phase boundaries from the ASM ALLOY PHASE DIAGRAM DATABASE™ are superimposed to the maps).

9.80% [30], 10.01% [31], 11.30% [32], 10.60% [33], 10.40% [34], and 10.10% [35], respectively. Then, parameters for the uniform distribution can be calculated as follows:

$$\hat{a} = \frac{nN_n - M_n}{n-1}\hat{b} = \frac{nM_n - N_n}{n-1}\,, \tag{43}$$

$$N_n = \min(x_1, x_2, , , x_n) M_n = \max(x_1, \ x_2, , , x_n)\,, \tag{44}$$

where $\hat{a}$ and $\hat{b}$ are the unbiased estimation of a and b; n is the length of the collected number. N_n and M_n are the minimum and maximum value for the collected data, respectively. Based on the above formula, $\hat{a}$ and $\hat{b}$ are calculated to be 0.07 and 0.12, respectively. Similarly, the position of the phase-boundary points located on the Fe-Co edge in the ternary composition-phase map can also be determined in a similar way, where $\hat{a}$ and $\hat{b}$ are estimated to be 0.82 and 0.86, respectively [36–40].

Applying the phase-boundary distribution information on the Fe-Ni and Fe-Co edges from the ASM database as prior, the two-phase boundaries in the Fe-Co-Ni ternary composition-phase map are much closer to the reported results in the ASM database and locate within the largely reduced BCIs [Fig. 10(b)]. This also suggests a feasible method for how to estimate the phase diagram from a composition-phase map of a combinatorial materials chip, i.e., by introducing more known information of the related phase diagrams as prior knowledge into the model.

In much literature and handbooks, the phase boundary is expressed as a single line. The uncertainty range offered by the developed method is a more quantitative description of the position of the phase boundary. It is recognized that our knowledge or understanding of the position of the phase boundary can be easily restricted by many factors, such as incomplete degree of equilibrium, mapping density, unequal data coverage, noise in the characterization, etc. An uncertainty range offers more practical guidance compared with just one single line for decision-making in engineering. It should also be noticed that the BCIs of the boundary between the bcc single-phase region and bcc + fcc region almost cover all the regions of the bcc single-phase region in Fig. 10, which should be contributed to the unequal data coverage as described in the previous section. Based on previous analysis, this problem can be overcome by increasing the mapping density. Alternatively, reducing the composition gradient of the film in the geometric region can also attain this purpose. Under the same mapping density in the geometric space, a reduced composition gradient will cause an increased mapping density in the composition space. Moreover, the quantitative information related to the phase boundaries can be used for further CALPHAD modeling to optimize the model parameters.

V. CONCLUSIONS

A method based on Bayesian statistics is developed to determine and evaluate the uncertainty of two types of phase boundaries as well as the vertices of a ternary tie triangle in the isothermal section of the ternary system under the principle of the Gibbs phase rule. The quantitative relationship between mapping density and uncertainty is established, and the result is of great significance for conducting high-throughput experiments. Other factors that affect the uncertainty, such as noise in the data and data coverage, are also investigated. Moreover, the uncertainty of the vertices of a ternary tie triangle can be quantified within the framework of the model. Fe-Cr-Ni and Fe-Co-Ni ternary systems are used as two typical examples showing that the phase boundaries and their BCIs can be determined satisfactorily from the classified data points. The uncertainty of the phase boundary in the ternary isothermal section can be further reduced by incorporating the data from the corresponding binary systems as priors. The present approach can be applied to other kinds of other high-throughput characterization methods, like optical characterization or electrical characterization, in a straightforward manner to establish the composition-structure or property relationship.

ACKNOWLEDGMENTS

Financial support from the National Key Research and Development Program of China (Grants No. 2021YFB3702102 and No. 2017YFB0701900) and the Major Science and Technology Project of Yunnan Province "Genome Engineering of Rare and Precious Metal Materials in Yunnan Province (Phase One 2020)" (Grant No. 202002AB080001-1) is gratefully acknowledged.

The authors declare no competing interests.

[1] R. Schmid-Fetzer, Phase diagrams: The beginning of wisdom, J. Phase Equilib. Diffus. **35**, 735 (2014).
[2] P. J. Spencer, A brief history of CALPHAD, Calphad **32**, 1 (2008).
[3] K. Kennedy, T. Stefansky, G. Davy, V. F. Zackay, and E. R. Parker, Rapid method for determining ternary-alloy phase diagrams, J. Appl. Phys. **36**, 3808 (1965).
[4] W. Wong-Ng, Evaluation of the current status of the combinatorial approach for the study of phase diagrams, J. Res. Natl. Inst. Stand. Technol. **117**, 304 (2012).
[5] C. J. Long, J. Hattrick-Simpers, M. Murakami, R. C. Srivastava, I. Takeuchi, V. L. Karen, and X. Li, Rapid structural mapping of ternary metallic alloy systems using the combinatorial approach and cluster analysis, Rev. Sci. Instrum. **78**, 072217 (2007).
[6] A. G. Kusne, T. Gao, A. Mehta, L. Ke, M. C. Nguyen, K.-M. Ho, V. Antropov, C.-Z. Wang, M. J. Kramer, C. Long, and I. Takeuchi, On-the-fly machine-learning for high-throughput experiments: Search for rare-earth-free permanent magnets, Sci. Rep. **4**, 6367 (2014).
[7] H. Xing, B. Zhao, Y. Wang, X. Zhang, Y. Ren, N. Yan, T. Gao, J. Li, L. Zhang, and H. Wang, Rapid construction of Fe–Co–Ni composition–phase map by combinatorial materials chip approach, ACS Comb. Sci. **20**, 127 (2018).
[8] X.-D. Xiang, G. Wang, X. Zhang, Y. Xiang, and H. Wang, Individualized pixel synthesis and characterization of combinatorial materials chips, Engineering **1**, 225 (2015).
[9] H. Koinuma and I. Takeuchi, Combinatorial solid-state chemistry of inorganic materials, Nat. Mater. **3**, 429 (2004).
[10] P. J. McGinn, Thin-film processing routes for combinatorial materials investigations—A review, ACS Comb. Sci. **21**, 501 (2019).
[11] E. Danielson, J. H. Golden, E. W. McFarland, C. M. Reaves, W. H. Weinberg, and X. D. Wu, A combinatorial approach to the discovery and optimization of luminescent materials, Nature (London) **389**, 944 (1997).
[12] X.-D. Xiang, Combinatorial materials synthesis and screening: An integrated materials chip approach to discovery and optimization of functional materials, Annu. Rev. Mater. Sci. **29**, 149 (1999).
[13] K. Terayama, R. Tamura, Y. Nose, H. Hiramatsu, H. Hosono, Y. Okuno, and K. Tsuda, Efficient construction method for phase diagrams using uncertainty sampling, Phys. Rev. Mater. **3**, 033802 (2019).
[14] K. Terayama, K. Han, R. Katsube, I. Ohnuma, T. Abe, Y. Nose, and R. Tamura, Acceleration of phase diagram construction by machine learning incorporating Gibbs' phase rule, Scr. Mater. **208**, 114335 (2022).
[15] G. Deffrennes, K. Terayama, T. Abe, and R. Tamura, A machine learning–based classification approach for phase diagram prediction, Mater. Des. **215**, 110497 (2022).
[16] N. H. Paulson, B. J. Bocklund, R. A. Otis, Z.-K. Liu, and M. Stan, Quantified uncertainty in thermodynamic modeling for materials design, Acta Mater. **174**, 9 (2019).
[17] P. Honarmandi, T. C. Duong, S. F. Ghoreishi, D. Allaire, and R. Arroyave, Bayesian uncertainty quantification and information fusion in CALPHAD-based thermodynamic modeling, Acta Mater. **164**, 636 (2019).
[18] F. Rizzi, M. Salloum, Y. M. Marzouk, R.-G. Xu, M. L. Falk, T. P. Weihs, G. Fritz, and O. M. Knio, Bayesian inference of atomic diffusivity in a binary Ni/Al system based on molecular dynamics, Multiscale Model. Simul. **9**, 486 (2011).
[19] S. Chib and E. Greenberg, Understanding the metropolis–hastings algorithm, Am. Stat. **49**, 327 (1995).
[20] N. H. Paulson, E. Jennings, and M. Stan, Bayesian strategies for uncertainty quantification of the thermodynamic properties of materials, Int. J. Eng. Sci. **142**, 74 (2019).
[21] J. Salvatier, T. V. Wiecki, and C. Fonnesbeck, Probabilistic programming in python using $PyMC_3$, PeerJ Comput. Sci. **2**, e55 (2016).
[22] G. Arminger and B. O. Muthén, A Bayesian approach to nonlinear latent variable models using the Gibbs sampler and the Metropolis–Hastings algorithm, Psychometrika **63**, 271 (1998).
[23] S. El Adlouni, A.-C. Favre, and B. Bobée, Comparison of methodologies to assess the convergence of Markov chain Monte Carlo methods, Comput. Stat. Data. Anal. **50**, 2685 (2006).
[24] A. Gelman and D. B. Rubin, Inference from iterative simulation using multiple sequences, Stat. Sci. **7**, 457 (1992).
[25] Y. Iwasaki, A. G. Kusne, and I. Takeuchi, Comparison of dissimilarity measures for cluster analysis of X-ray diffraction data from combinatorial libraries, npj Comput. Mater. **3**, 4 (2017).
[26] C.-H. Xia, Y. Wang, J.-J. Wang, X.-G. Lu, and L. Zhang, Thermodynamic assessment of the Co–Fe–Ni system and diffusion study of its fcc phase, J. Alloys Compd. **853**, 157165 (2021).
[27] Y. Liu, T. Zhao, W. Ju, and S. Shi, Materials discovery and design using machine learning, J. Materiomics **3**, 159 (2017).
[28] W. Rostoker, Analytical representation of certain phase boundaries, JOM **3**, 1203 (1951).
[29] K. Ono, Y. Ueda, A. Yamaguchi, and J. Moriyama, Thermodynamic study of Fe–Ni solid solution, Trans. Jpn. Inst. Met. **18**, 610 (1977).
[30] Y.-Y. Chuang, K.-C. Hsieh, and Y. A. Chang, A thermodynamic analysis of the phase equilibria of the Fe–Ni system above 1200 K, Metall. Trans. A **17**, 1373 (1986).
[31] F. Lihl, Die Amalgame als Hilfsmittel in der metallkundlichen Forschung, Int. J. Mater. Res. **46**, 434 (1955).
[32] L. Swartzendruber, V. Itkin, and C. Alcock, The Fe−Ni (iron−nickel) system, J. Phase Equilib. **12**, 288 (1991).
[33] G. Raynor and V. Rivlin, The Fe−Ni (Iron−Nickel) system, J. Phase Equilib. **2**, 102 (1981).

[34] A. D. Romig and J. Goldstein, Determination of the Fe–Ni and Fe–Ni–P phase diagrams at low temperatures (700 to 300 °C), Metall. Trans. A **11**, 1151 (1980).

[35] C.-W. Yang, D. Williams, and J. Goldstein, A revision of the Fe–Ni phase diagram at low temperatures (< 400 °C), J. Phase Equilib. **17**, 522 (1996).

[36] Y. I. Ustinovshikov, B. E. Pushkarev, and I. V. Sapegina, Phase separation and ordering in the Fe–Co system, Inorg. Mater. **42**, 354 (2006).

[37] T. Nishizawa and K. Ishida, The Co−Fe (cobalt−iron) system, Bull. Alloy Phase Diagr. **5**, 250 (1984).

[38] A. Fernández Guillermet, Critical evaluation of the thermodynamic properties of cobalt, Int. J. Thermophys **8**, 481 (1987).

[39] I. Ohnuma, H. Enoki, O. Ikeda, R. Kainuma, H. Ohtani, B. Sundman, and K. Ishida, Phase equilibria in the Fe–Co binary system, Acta Mater. **50**, 379 (2002).

[40] H. Okamoto, Co–Fe (cobalt-iron), J. Phase Equilib. Diffus. **29**, 383 (2008).