







## Data assimilation method for experimental and first-principles data: Finite-temperature magnetization of $(\text{Nd, Pr, La, Ce})_2(\text{Fe, Co, Ni})_{14}\text{B}$

Yosuke Harashima <sup>1,\*</sup>, Keiichi Tamai <sup>2</sup>, Shotaro Doi,<sup>1</sup> Munehisa Matsumoto <sup>2</sup>, Hisazumi Akai <sup>2</sup>, Naoki Kawashima,<sup>2</sup> Masaaki Ito,<sup>3</sup> Noritsugu Sakuma,<sup>3</sup> Akira Kato,<sup>3</sup> Tetsuya Shoji,<sup>3</sup> and Takashi Miyake <sup>1</sup>

<sup>1</sup>Research Center for Computational Design of Advanced Functional Materials, National Institute of Advanced Industrial Science and Technology, Tsukuba, Ibaraki 305-8568, Japan

<sup>2</sup>The Institute for Solid State Physics, The University of Tokyo, Kashiwa, Chiba 277-8581, Japan

<sup>3</sup>Higashifuji Technical Center, Toyota Motor Corporation, Susono, Shizuoka 410-1193, Japan

 (Received 28 July 2020; revised 26 October 2020; accepted 12 January 2021; published 26 January 2021)

We propose a data assimilation method for evaluating the finite-temperature magnetization of a permanent magnet over a high-dimensional composition space. Based on a general framework for constructing a predictor from two data sets including missing values, a practical scheme for magnetic materials is formulated in which a small number of experimental data in limited composition space are integrated with a larger number of first-principles calculation data. We apply the scheme to  $(\text{Nd}_{1-\alpha-\beta-\gamma}\text{Pr}_\alpha\text{La}_\beta\text{Ce}_\gamma)_2(\text{Fe}_{1-\delta-\zeta}\text{Co}_\delta\text{Ni}_\zeta)_{14}\text{B}$ . The magnetization in the whole  $(\alpha, \beta, \gamma, \delta, \zeta)$  space at arbitrary temperature is obtained. It is shown that the Co doping does not enhance the magnetization at low temperatures, whereas the magnetization increases with increasing  $\delta$  above 320 K.

DOI: [10.1103/PhysRevMaterials.5.013806](https://doi.org/10.1103/PhysRevMaterials.5.013806)

### I. INTRODUCTION

Even more than 30 years after the development of neodymium magnet [1,2], there has still been continuing effort in developing rare-earth permanent magnets. One of the central incentives for the development is the resource criticality issue. Rare-earth permanent magnets of current industrial use heavily rely on certain rare-earth resources. The main phase of the neodymium magnet is  $\text{Nd}_2\text{Fe}_{14}\text{B}$  which contains substantial amount of neodymium. More critical dysprosium is added to enhance the coercivity, while other rare earths such as lanthanum and cerium are abundant. Hence, more balanced utilization of them is an important issue, especially given geologic scarcity and political volatility of the former. While  $\text{Nd}_2\text{Fe}_{14}\text{B}$  is famous for its strong saturation magnetization at room temperature, it is also known to have relatively low Curie temperature compared to more traditional permanent rare-earth magnets such as  $\text{Sm}_2\text{Co}_{17}$ . Finding a rare-earth magnet with better balance between saturation magnetization and heat resistance thus deserves extensive efforts [3].

One of the most common approaches toward improvement of rare-earth permanent magnets has been partial substitution, where practitioners replace some elements of a particular mother compound (e.g.,  $\text{Sm}_2\text{Co}_{17}$  and  $\text{Nd}_2\text{Fe}_{14}\text{B}$ ) with other ones [4]. Even though the “full” substitution (e.g., from  $\text{Nd}_2\text{Fe}_{14}\text{B}$  to  $\text{Ce}_2\text{Fe}_{14}\text{B}$ ) does not work quite well due to the inferior magnetic properties of the resulting compound [5], there still is possibility that better performance compared to the mother compound may be achieved by properly setting the ratio of the substitution. This is evidenced by the celebrated

Slater-Pauling curve in 3d transition metal alloys. An example in rare-earth magnets is a recent report for  $\text{Sm}(\text{Fe}_{1-x}\text{Co}_x)_{12}$ , where the magnetization at and above room temperature increases with increasing cobalt concentration [6]. These results indicate the importance of searching the optimal chemical composition in a widespread landscape.

The brute-force strategy, however, quickly runs into difficulty as the number of candidates for substituents becomes large. This is a typical manifestation of the notorious “curse of dimensionality” [7]; the number of samples needed for uniform sampling scales exponentially with respect to the number of the candidates. Meanwhile, experimental preparation of a permanent magnet involves many time-consuming processes (such as hydrogen decrepitation and annealing), each of which takes one to tens of hours. Thus, it is obviously infeasible to experimentally perform uniform and dense study over the entire parameter space in question. In order to deal with the curse, practitioners usually restrict their investigations on a rather tiny subset of the parameter space, typically by restricting the number of constituent elements [8] or by fixing the ratio between elements [9]. While such a treatment is quite useful for studying certain aspects of magnetic compounds, it comes at a cost of sampling bias and thereby at a risk of overlooking a truly optimal compound even within the predetermined parameter space. Hence, a less biased but still manageable (in terms of the number of experimental trials) approach is highly desirable.

On the other hand, recent development of high-throughput calculation techniques enables us to perform first-principles calculation of thousands of rare-earth magnet compounds [10–12]. This is a powerful tool to capture trends over wide parameter space. However, it contains a systematic error. One remedy we consider in this work is to perform so-called multitask learning on the data obtained from experiments and

\*Present address: Center for Computational Sciences, University of Tsukuba, Tsukuba, Ibaraki 305-8577, Japan; harashima@ccs.tsukuba.ac.jp

those from the first-principles calculations. Multitask learning is an approach to improve prediction capability by learning multiple tasks simultaneously (not separately as classical machine-learning frameworks do). By doing so, one can utilize the hidden relationships among the tasks at hand. Given that the computational results are expected to be strongly correlated with experimental ones, it is natural to argue that the multitask learning may also work in this case, although the first-principles calculations do involve simplifying approximations [13] and hence validity of them should be examined with care.

In this work, we overcome the challenge by assimilating a limited number of experimental magnetization data and systematic first-principles calculation. The experimental small data are supplemented by the first-principles calculation data, whereas systematic error contained in the latter is corrected by the former. While we focus on Nd<sub>2</sub>Fe<sub>14</sub>B type compounds [or, more specifically, (Nd<sub>1- $\alpha$ - $\beta$ - $\gamma$ Pr <sub>$\alpha$</sub> La <sub>$\beta$</sub> Ce <sub>$\gamma$</sub> )<sub>2</sub>(Fe<sub>1- $\delta$ - $\zeta$ Co <sub>$\delta$</sub> Ni <sub>$\zeta$</sub> )<sub>14</sub>B] because of their practical relevance, we expect that the approach can be easily applied to other types of compounds. Apart from the methodology, we present analysis of the saturation magnetization at various temperatures in the entire ( $\alpha, \beta, \gamma, \delta, \zeta$ ) space. We show that the magnetization of partially substituted systems considerably deviates from the value linearly interpolated from end points. In particular, increase in cobalt concentration enhances the magnetization above 320 K, and we discuss the origin of the enhancement focusing on magnetization at zero temperature and Curie temperature.</sub></sub>

The rest of this paper is organized as follows: In Sec. II, we formulate the present framework of the data assimilation. This section also includes the test of the framework using some toy data. In Sec. III, we apply the present framework to magnet compound and discuss the implication of the results. We conclude the paper in Sec. IV.

## II. DATA ASSIMILATION: FORMALISM

We begin this section by introducing some notations and clarifying the objectives. Suppose we would like to model the target variable  $\mathbf{y} \in \mathbb{R}^q$  as a function of the descriptor  $\mathbf{x} \in \mathbb{R}^p$  with some noise  $\boldsymbol{\varepsilon}$ :

$$\mathbf{y} = \mathbf{f}(\mathbf{x}) + \boldsymbol{\varepsilon}. \quad (1)$$

For example, in the cases discussed in the following sections, an element of  $\mathbf{x}$  is a monomial up to the second-order power in the concentrations of component elements, whereas elements of  $\mathbf{y}$  represent the computational and the experimental values of either the magnetization or the Curie temperature. In the present context,  $p$  may or may not be larger than one, but  $q$  must, because  $\mathbf{y}$  is supposed to contain both experimental and computational results. As for the model  $\mathbf{f}$ , we consider a problem of linear regression, on which the multitask learning has been extensively studied [14]:

$$\mathbf{f}(\mathbf{x}) = \mathbf{W}\mathbf{x} = \begin{pmatrix} \mathbf{w}_1 \cdot \mathbf{x} \\ \vdots \\ \mathbf{w}_q \cdot \mathbf{x} \end{pmatrix}. \quad (2)$$

Now, the problem is to estimate the coefficient matrix  $\mathbf{W} \in \mathbb{R}^{q \times p}$  from given  $q$  sets of empirical data  $\{(\mathbf{x}_{n,i}, y_{n,i})\}_{n=1}^{N_i}\}_{i=1}^q$ .

In general, values of sampled data  $(\mathbf{x}_{n,i}, y_{n,i})$  and the number of samples  $N_i$  are not identical among  $q$  components. One can construct an ordinary least-square (OLS) estimator for each  $\mathbf{w}_i$  ( $i = 1, \dots, q$ ), provided that it exists:

$$\hat{\mathbf{w}}_{\text{OLS};i} = (\mathbf{X}_i^T \mathbf{X}_i)^{-1} \mathbf{X}_i^T \mathbf{Y}_i, \quad (3)$$

where

$$\mathbf{X}_i = \begin{pmatrix} \mathbf{x}_{1,i}^T \\ \vdots \\ \mathbf{x}_{N_i,i}^T \end{pmatrix}, \quad \mathbf{Y}_i = \begin{pmatrix} y_{1,i} \\ \vdots \\ y_{N_i,i} \end{pmatrix}. \quad (4)$$

It is also important to note that the OLS estimator is known to be the best linear unbiased estimator: that is, it has the smallest variance among all linear unbiased estimates [15]. However, restricting our interest to unbiased estimates is not necessarily the wisest decision we can make. In the present case, the target variables are one physical quantity obtained by different ways: experimental and theoretical approaches. Ideally,  $\mathbf{w}_i$  are equivalent among these ways, but they are not in practice. This discrepancy can be resolved by assimilating those data. We could construct a slightly biased estimator, with an aid of correlation between multiple outputs  $y_i$  ( $i = 1, \dots, q$ ), to achieve a better tradeoff between bias and variance. The focus of the rest of this section is on the construction of such an estimator.

In order to formulate our approach, we hereafter assume that the descriptor  $\mathbf{x}$  and the target  $\mathbf{y}$  jointly follow the multivariate Gaussian distribution:

$$p(\mathbf{y}, \mathbf{x}; \Sigma) = \frac{1}{\sqrt{(2\pi)^d |\Sigma|}} \exp\left(-\frac{1}{2} \mathbf{z}^T \Sigma^{-1} \mathbf{z}\right), \quad (5)$$

where  $\Sigma$  is a positive-definite (and thereby symmetric) covariance matrix and  $\mathbf{z}^T := (\mathbf{y}^T, \mathbf{x}^T)$ . For later convenience, we use a precision matrix  $\Lambda := \Sigma^{-1}$ . The component of  $\Lambda$  is represented as

$$\Lambda = \begin{pmatrix} \Lambda_{yy} & \Lambda_{yx} \\ \Lambda_{yx}^T & \Lambda_{xx} \end{pmatrix}. \quad (6)$$

Then, the probability distribution of  $\mathbf{y}$  conditional on  $\mathbf{x}$  can be easily found:

$$p(\mathbf{y}|\mathbf{x}; \Lambda) = \sqrt{\frac{|\Lambda_{yy}|}{(2\pi)^d}} \exp\left(-\frac{1}{2} (\mathbf{y} - \boldsymbol{\mu})^T \Lambda_{yy} (\mathbf{y} - \boldsymbol{\mu})\right), \quad (7)$$

where

$$\boldsymbol{\mu} = -(\Lambda_{yy})^{-1} \Lambda_{yx} \mathbf{x}. \quad (8)$$

Since Eq. (8) defines the regression coefficient matrix of  $\mathbf{y}$  with respect to  $\mathbf{x}$ , the central task here is to estimate the matrix from the data. In this work, we employ maximum likelihood estimation: that is, we consider a problem of maximization of the following log-likelihood function  $L(\Lambda | \{(\mathbf{x}_n, \mathbf{y}_n)\}_{n=1}^N)$ :

$$L(\Lambda_{yy}, \Lambda_{yx} | \{(\mathbf{x}_n, \mathbf{y}_n)\}_{n=1}^N) = \sum_n \log p(\mathbf{y}_n | \mathbf{x}_n; \Lambda). \quad (9)$$

Here, we note that, even though the original precision matrix  $\Lambda$  has  $(p+q)(p+q+1)/2$  independent elements, optimization of only  $\Lambda_{yy}$  and  $\Lambda_{yx}$  suffices as far as estimation of  $\boldsymbol{\mu}$  is concerned.

Until here, the present formulation is fairly general [under the assumption of (5), at least], and we did not assume any further relations between the target variables. Although combining multiple measurements may be beneficial for a better estimation (in terms of combined error) even when underlying mechanisms of those outputs are completely independent [16,17], it is usually advisable to incorporate a relation between the outputs during the estimation process (when reasonably possible). This is particularly true when data for some target variables are substantially harder to gather than the others, but one has a good guess over the relation between these two.

In order to introduce a bias (correlation) to our estimator, we postulate that the experimental values  $y_{\text{expt}}(\mathbf{x})$  can be represented by a scalar multiplication of the prediction  $y_{\text{comp}}(\mathbf{x})$  derived from computational data and a residual part  $R(\mathbf{x})$  which involves substantially fewer terms than those originally considered for modeling the computational results (hereafter we assume  $q = 2$ , and refer to  $y_1$  as computational output  $y_{\text{comp}}$  and  $y_2$  as experimental output  $y_{\text{expt}}$ )

$$y_{\text{expt}}(\mathbf{x}) = C y_{\text{comp}}(\mathbf{x}) + R(\mathbf{x}), \quad (10)$$

where

$$R(\mathbf{x}) := \mathbf{w}_{\text{res}} \cdot \mathbf{x}_{\text{proj}}. \quad (11)$$

Here,  $\mathbf{w}_{\text{res}} \in \mathbb{R}^r$  ( $r < p$ ) and  $\mathbf{x}_{\text{proj}}$  is a natural projection of  $\mathbf{x}$  onto the parameter space of relevant descriptors. This assumption can be easily implemented on the present formulation by enforcing

$$\Lambda_{y_{\text{expt}},i} = 0 \quad \text{for } i \in \text{irrelevant descriptors}. \quad (12)$$

This indicates that  $y_{\text{expt}}$  only indirectly correlates with the  $i$ th descriptor through other variables. In this case,  $C$  in Eq. (10) can be expressed in terms of  $\Lambda_{yy}$ :

$$C = \Lambda_{y_{\text{expt}},y_{\text{comp}}} / \Lambda_{y_{\text{comp}},y_{\text{comp}}}. \quad (13)$$

While the description of the present framework is conceptually complete up to here, we have to face with one more complication in practice. A central concern here is that we usually have the smaller number  $N_{\text{expt}}$  of experimental data than that  $N_{\text{comp}}$  of computational ones, thus, some data have a missing value in either  $y_{\text{expt}}$  or  $y_{\text{comp}}$ . Although one may simply discard such missing pairs during the analysis, one could not benefit from abundance of computational data in this approach and hence the estimates could not be very efficient.

In order to address this issue, we use the direct likelihood. The idea behind the direct likelihood is that we integrate out the missing variable. The direct likelihood is written as

$$\begin{aligned} L(\Lambda_{yy}, \Lambda_{yx} | \{(\mathbf{x}_n, \mathbf{y}_n)\}_{n=1}^N) &= \sum_{n \in \Omega_{\text{comp,expt}}} \log p(\mathbf{y}_n | \mathbf{x}_n; \Lambda) \\ &+ \sum_{n \in \Omega_{\text{comp}}} \log p_{\text{comp}}(y_{\text{comp},n} | \mathbf{x}_n; \Lambda) \\ &+ \sum_{n \in \Omega_{\text{expt}}} \log p_{\text{expt}}(y_{\text{expt},n} | \mathbf{x}_n; \Lambda) \end{aligned} \quad (14)$$

by decomposing a sample set into three subsets  $\Omega_{\text{comp,expt}}$  where both the experimental and computational data are

available, and  $\Omega_{\text{expt}}, \Omega_{\text{comp}}$  where either of the two is missing (for example, samples in  $\Omega_{\text{expt}}$  only contain experimental data). The distributions for the missing data are

$$p_{\Gamma}(y_{\Gamma} | \mathbf{x}; \Lambda) = \int dy_{\bar{\Gamma}} p(y_{\Gamma}, y_{\bar{\Gamma}} | \mathbf{x}; \Lambda) \quad (15)$$

$$= \sqrt{\frac{|\bar{\Lambda}_{\Gamma\Gamma}|}{2\pi}} \exp\left(-\frac{1}{2}(y_{\Gamma} - \mu_{\Gamma}) \bar{\Lambda}_{\Gamma\Gamma} (y_{\Gamma} - \mu_{\Gamma})\right), \quad (16)$$

where

$$\bar{\Lambda}_{\Gamma\Gamma} := \Lambda_{\Gamma\Gamma} - \Lambda_{\Gamma\bar{\Gamma}} \Lambda_{\bar{\Gamma}\bar{\Gamma}}^{-1} \Lambda_{\bar{\Gamma}\Gamma}, \quad (17)$$

and a pair of  $(\Gamma, \bar{\Gamma})$  represents either (expt, comp) or (comp, expt). Use of the direct likelihood is justified by the fact that the absence of data is simply a matter of design of the measurements in this case and hence unrelated to the values that have been missed [in other words, data are missing at random (albeit not completely) in the sense of Rubin [18]].

Then, we optimize  $\Lambda_{yy}$  and the relevant part of  $\Lambda_{yx}$  for Eq. (14). The optimization problem can be computationally solved using the limited-memory Broyden-Fletcher-Goldfarb-Shanno algorithm [19,20] with simple box constraints (L-BFGS-B) [21].

To see features of the method, we applied it to a toy data. The toy data were generated from two true models

$$f_i(x) = 2 - x + 5(x - 0.7)^2 + 20(x - 0.5)^3, \quad (18)$$

$$f_{ii}(x) = f_i(x) - 1.5, \quad (19)$$

where  $x$  is an input parameter and the descriptor consists of  $(1, x, x^2, x^3)$ . The data (i) and (ii) mimic computational and experimental data, respectively. 10 samples for model (i) and 5 for model (ii) were sampled randomly. Noise was applied with widths 0.03 for  $y_i$  and 0.1 for  $y_{ii}$ , respectively. An example sample data and a result of the assimilation are shown in Fig. 1. Although the sampled values of descriptors do not match between the two models, the prediction well agree with the true model. We also compared with a simple least-square fitting for these models separately. For the model (i), the OLS fit agrees with the true model as well as the assimilation. For the model (ii), however, the OLS fit disagrees with the true model for  $0.5 < x < 1.0$ , whereas, the assimilation well predicts the true model even for that region.

The remaining issue is to determine the list of candidates for descriptors. Since, as we will see in Sec. III, the behavior of the quantities of interest (e.g., magnetization, Curie temperature) is rather simple and we do not expect to encounter a singular point where these quantities diverge, it suffices to model them by polynomial functions of the concentration of each element. More specifically, we modeled the first-principles results by a second-order polynomial function:

$$y_{\text{comp}} = \sum_{i_{\alpha} + i_{\beta} + i_{\gamma} + i_{\delta} + i_{\zeta} \leq 2} c_{i_{\alpha} i_{\beta} i_{\gamma} i_{\delta} i_{\zeta}} \alpha^{i_{\alpha}} \beta^{i_{\beta}} \gamma^{i_{\gamma}} \delta^{i_{\delta}} \zeta^{i_{\zeta}}, \quad (20)$$

where  $\alpha, \beta, \gamma, \delta$ , and  $\zeta$  denote, respectively, the concentration of Pr, La, Ce, Co, and Ni. As for the list of candidates for relevant descriptors for  $R$  in Eq. (10), we considered a constant term for the magnetization, and constant and  $\delta$  linear terms for the Curie temperature.

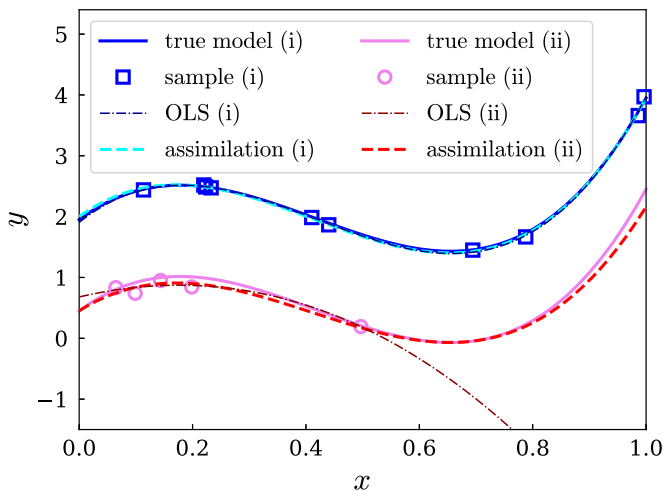


FIG. 1. An example for the data assimilation. The true models are given by hand as (i) (blue solid curve) and (ii) (violet solid curve). Sample data generated from the true models are shown as blue squares and violet points. The prediction from the data assimilation is shown as red for the model (i) and cyan for the model (ii). The prediction from the OLS fit is also shown as a dotted-dashed line for comparison.

### III. MAGNETIZATION OF $(\text{Nd,Pr,La,Ce})_2(\text{Fe,Co,Ni})_{14}\text{B}$

#### A. Data assimilation for magnetization

Now that we have introduced a general procedure for the data assimilation, the next step is to apply the methodology to build up a flexible framework for predictions of the essential properties of the magnetic compounds. By “flexibility” we mean the ability to predict the properties at an arbitrary temperature, in addition to one for arbitrary combination of the doping concentrations.

A problem here is that both the experiments and first-principles calculations suffer from their limitations. On one hand, experiments are hard to perform at an arbitrary condition due to, e.g., difficulty in synthesis and limitation in experimental facilities. On the other hand, density functional theory (DFT) [22,23], on which our first-principles calculation are based, rely on an approximation to the exchange-correlation functional in practical applications. Furthermore, DFT only address the ground-state property of the system (in other words, the system at 0 K). Finite-temperature magnetism such as Curie temperature can be evaluated by combining DFT with a mean field theory of classical spin dynamics [24,25]. However, quantitative agreement with experiments is limited. Magnetism is a consequence of quantum many-body effect, which requires sophisticated theoretical treatment. Therefore, the experiments and the calculations are in a sense complementary to each other.

To fix this problem, we introduce a practical scheme to estimate the finite-temperature magnetization. The scheme consists of the following four steps and its flowchart is illustrated in Fig. 2.

*Step 1-1.* We prepared 119 experimental samples of  $(\text{Nd}_{1-\alpha-\beta-\gamma}\text{Pr}_\alpha\text{La}_\beta\text{Ce}_\gamma)_2(\text{Fe}_{1-\delta-\zeta}\text{Co}_\delta\text{Ni}_\zeta)_{14}\text{B}$ , each of which has different chemical composition (this includes one sample value extracted from Refs. [26,27]). The

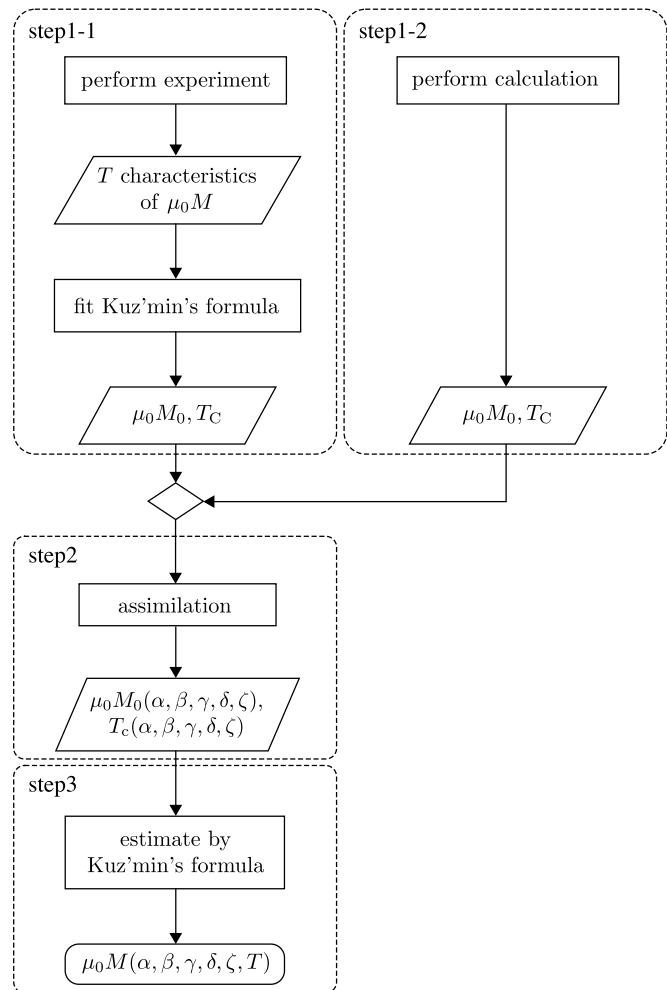


FIG. 2. Flowchart of data assimilation method for finite-temperature magnetization.

magnetization  $\mu_0 M(T)$  of each sample was measured at 3–7 temperatures within a range from 300 to 473 K. Experimental details can be found in Appendix A. We converted  $\mu_0 M(T)$  into the Curie temperature  $T_C$  and the magnetization  $\mu_0 M(T=0)$  at 0 K, which can then be directly compared to the computational results. To do this, we referred to Kuz'min's empirical formula for magnetization [28]:

$$\mu_0 M(T) = \mu_0 M_0 [1 - st^{\frac{3}{2}} - (1-s)t^{\frac{5}{2}}]^{\frac{1}{3}}, \quad (21)$$

where  $t := T/T_C$ ,  $\mu_0 M_0 := \mu_0 M(T=0)$ , and  $s$  is a phenomenological shape parameter specific to the compound [29,30]. For each chemical composition,  $\mu_0 M_0$  and  $T_C$  were obtained by the least-square fitting. Following Ref. [31], we fixed  $s$  to be 0.6 in this work. (To be more accurate,  $s$  could be fitted as well as  $\mu_0 M_0$  and  $T_C$ .) An example for  $\text{Nd}_2\text{Fe}_{14}\text{B}$  is shown in Fig. 3.

We note that the model to convert experimental data can be replaced with another model, e.g., a two-sublattice model [2,32,33]. We show an example for  $\text{Nd}_2\text{Fe}_{14}\text{B}$  in Appendix B. This will be important for fitting more complicated temperature dependence of magnetization.

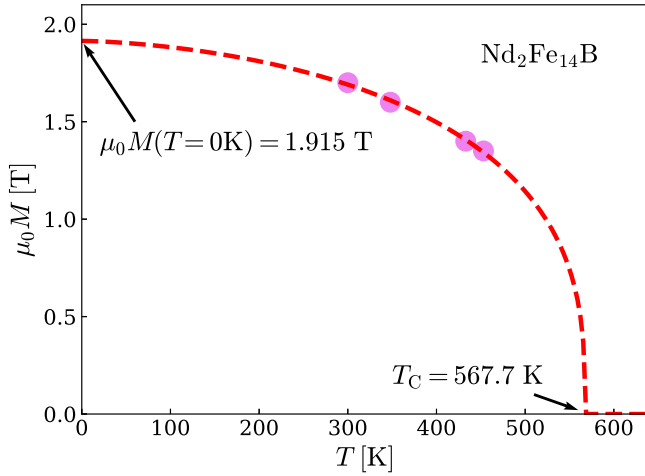


FIG. 3. The magnetization of  $\text{Nd}_2\text{Fe}_{14}\text{B}$  as a function of temperature. The violet points denote the experimental observation. The red dashed curve is a regression curve derived from a least-square fitting to Eq. (21). The obtained values for  $\mu_0 M(T=0\text{K})$  and  $T_C$  are also shown.

*Step 1-2.* On the theoretical side, we calculated the  $\mu_0 M_0$  and  $T_C$  following the method explained in Appendices C and D. The calculations were performed for 2869 compositions uniformly distributed in the  $(\alpha, \beta, \gamma, \delta, \zeta)$  space.

*Step 2.* We then applied the data assimilation method to the obtained  $\mu_0 M_0$  and  $T_C$ . We adopted the following regression models:

$$\mu_0 M_0^{\text{comp}}(\alpha, \beta, \gamma, \delta, \zeta) = \sum_I M_I^{\text{comp}} \alpha^{i_\alpha} \beta^{i_\beta} \gamma^{i_\gamma} \delta^{i_\delta} \zeta^{i_\zeta}, \quad (22)$$

$$T_C^{\text{comp}}(\alpha, \beta, \gamma, \delta, \zeta) = \sum_I T_I^{\text{comp}} \alpha^{i_\alpha} \beta^{i_\beta} \gamma^{i_\gamma} \delta^{i_\delta} \zeta^{i_\zeta}, \quad (23)$$

where  $I := (i_\alpha, i_\beta, i_\gamma, i_\delta, i_\zeta)$ . Both quantities were expanded up to the quadratic terms, namely  $0 \leq \sum_j i_j \leq 2$ . We assumed that the experimental data are correlated with the computational data as follows:

$$\mu_0 M_0^{\text{expt}}(\alpha, \beta, \gamma, \delta, \zeta) = C_m \mu_0 M_0^{\text{comp}}(\alpha, \beta, \gamma, \delta, \zeta) + m_0, \quad (24)$$

$$T_C^{\text{expt}}(\alpha, \beta, \gamma, \delta, \zeta) = C_t T_C^{\text{comp}}(\alpha, \beta, \gamma, \delta, \zeta) + t_0 + t_1 \delta. \quad (25)$$

The coefficients  $\{M_I^{\text{comp}}\}_I$ ,  $C_m$ ,  $m_0$  and  $\{T_I^{\text{comp}}\}_I$ ,  $C_t$ ,  $t_0$ ,  $t_1$  were determined by optimizing the likelihood (14) in terms of the parameters of the multivariate normal distribution  $\Lambda_{yy}$  and  $\Lambda_{yx}$ . Then, the coefficients were given by  $-(\Lambda_{yy})^{-1} \Lambda_{yx}$  as mentioned at Eq. (8).

In general, we may include higher-order terms to make a prediction model more flexible and accurate, which brings complexity in the determination of the coefficients in the prediction model. In the present case, we included the linear term of  $\delta$  in Eq. (25) and we will discuss that point later.

*Step 3.* The above data assimilation gives model functions for  $\mu_0 M_0 = \mu_0 M_0^{\text{expt}}$  and  $T_C = T_C^{\text{expt}}$  at arbitrary composition. Applying Kuz'min's formula (21) to these data, we estimate  $\mu_0 M$  at finite temperature. In this way, we can predict the

experimental magnetization at arbitrary point in the  $(5+1)$ -dimensional space spanned by  $\alpha$ ,  $\beta$ ,  $\gamma$ ,  $\delta$ ,  $\zeta$ , and  $T$ .

## B. Results and discussion

Figure 4 shows color maps of the magnetization on  $(\beta, \gamma)$ ,  $(\beta, \delta)$ ,  $(\gamma, \delta)$  planes at 0, 300, and 400 K, where other variables in  $\alpha$ ,  $\beta$ ,  $\gamma$ ,  $\delta$ ,  $\zeta$  are set to zero in each case. We see that the contour lines are not straight. This means that the magnetization varies nonlinearly in the composition space. The experimental data are unevenly distributed, which are interpolated and extrapolated over the wide composition space with the help of computational data by the data assimilation method.

Figure 5 is a comparison of  $\mu_0 M$  at  $T = 300$  K between measured values and prediction by data assimilation and OLS fit along the  $\beta$ ,  $\gamma$ , and  $\delta$  axes. Estimated root-mean-square errors between the predictions and the experimental data at 300 K are 0.085 T for the assimilation and 0.067 T for the OLS fit. It can be seen that both the predictions are able to predict the sampled experimental data, however, the OLS-fit model extremely decreases at the high- $\delta$  region where sample points are absent. This seems to be typical behavior of overfitted models in which sampled data are well described while prediction is sacrificed. On the other hand, the extreme decrease is not observed in the present assimilation method. The validity of prediction at the region can not be verified since there are no experimental sample data there. Instead, we next examine the generalizability of the predictions.

In order to examine the generalizability of the data assimilation method in comparison with the straightforward OLS fit, we performed 10-fold cross validation for the experimental  $\mu_0 M_0$  and  $T_C$ . The experimental data were randomly divided into 10 groups. One of the groups was taken as test data and the remaining were used for training. The data assimilation was applied to the training data along with the whole computational data, then the root-mean-square error between the predicted experimental value and the test data was evaluated. 10 error values were obtained by changing test data to another group. We iterated 10 sets of this procedure, and took an average of the errors. For the OLS fit, we used the training data only (without computational data). As demonstrated in the upper half of Table I, the average error of the present methodology was found to be significantly smaller than that of the OLS fit, indicating high generalizability of the former. The advantage of the data assimilation compared to the OLS

TABLE I. Root-mean-square error for the assimilated predictions  $T_C$  and  $\mu_0 M_0$ , corresponding to the components for the experiment. The error for both the 10-fold cross validation (CV) and the extrapolation are shown. In the extrapolation, the test data are the experimental data in the range of  $\alpha \geq 0.5$ . For comparison, results of the OLS fit are also shown.

			OLS fit	Data assimilation
10-fold CV	$T_C$	(K)	172.5	43.8
	$\mu_0 M_0$	(T)	0.719	0.092
Extrapolation	$T_C$	(K)	854.6	46.8
	$\mu_0 M_0$	(T)	6.180	0.097

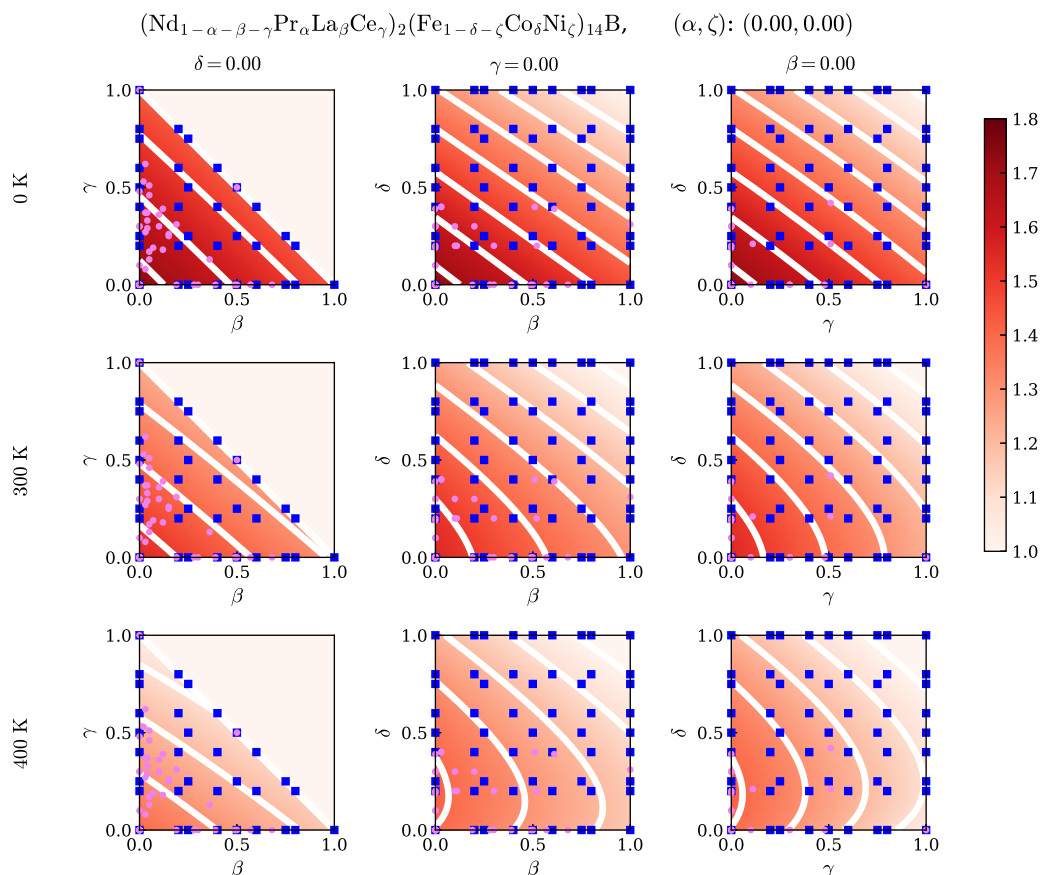


FIG. 4. Color map of the magnetization  $\mu_0 M(\alpha = 0, \beta, \gamma, \delta, \zeta = 0, T)$  at  $T = 0, 300,$  and  $400$  K. Experimental values, estimated by the data assimilation method, on  $(\beta, \gamma, \delta = 0)$ ,  $(\beta, \gamma = 0, \delta)$ , and  $(\beta = 0, \gamma, \delta)$  planes are shown. The values of the magnetization are described in Tesla. Blue squares are sampling points of the first-principles calculation data and violet points are those of the experimental data. Contour lines are also shown with interval of  $0.1$  T. Note that the upper triangle region on the  $(\beta, \gamma)$  plane is irrelevant because  $\beta + \gamma$  cannot exceed  $1$ .

fit is more drastic when the test data are distributed outside the region of training data. To see how the data assimilation works in the case of “extrapolation,” we extracted the experimental data with  $\alpha > 0.5$  as test data and used the remaining data as training data. The data assimilation and the OLS fit were applied to the training data as done for the cross validation. The estimated error is shown in the lower half of Table I. It indicates that the data assimilation extrapolates both  $\mu_0 M_0$  and  $T_C$  within permissible ranges of errors, while the OLS fit breaks down.

Coming back to Fig. 4, the predicted magnetization gives us useful information for Co substitution. The magnetization monotonically decreases with increasing  $\delta$  at  $T = 0$  and  $300$  K, whereas a Slater-Pauling-type peak is found as a function of  $\delta$  at  $400$  K. This is in sharp contrast with bcc-(Fe,Co), where Co doping up to  $\sim 30\%$  enhances the magnetization in the whole temperature region. This indicates that Co substitution enhances the magnetization at high temperatures via enhancement of the Curie temperature, but not via enhancement of the magnetization at zero temperature.

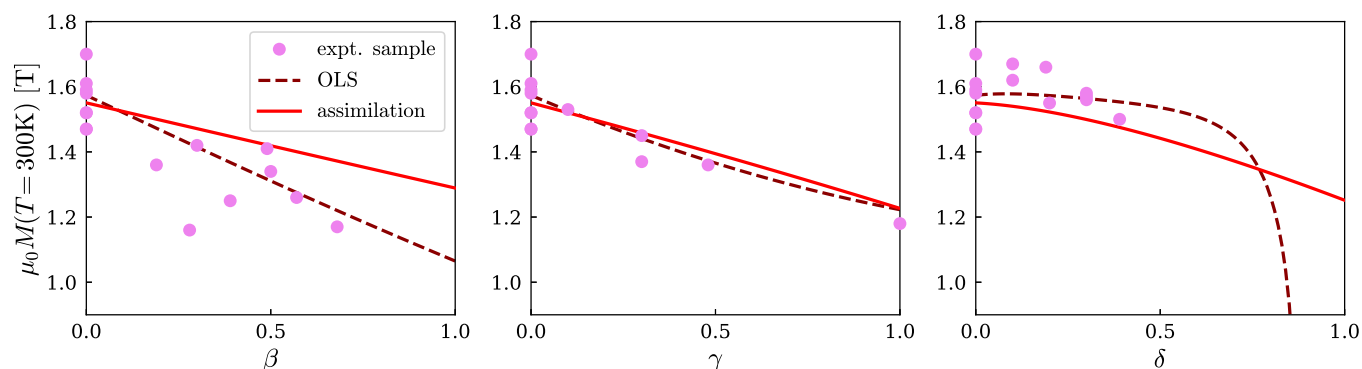


FIG. 5. Prediction (red solid lines) for the magnetization  $\mu_0 M$  at  $T = 300$  K is shown along  $(0, \beta, 0, 0, 0)$ ,  $(0, 0, \gamma, 0, 0)$ , and  $(0, 0, 0, \delta, 0)$ . Corresponding result for the OLS fit is shown as dark red dashed lines. Violet points indicate the experimental sample data observed at  $300$  K.

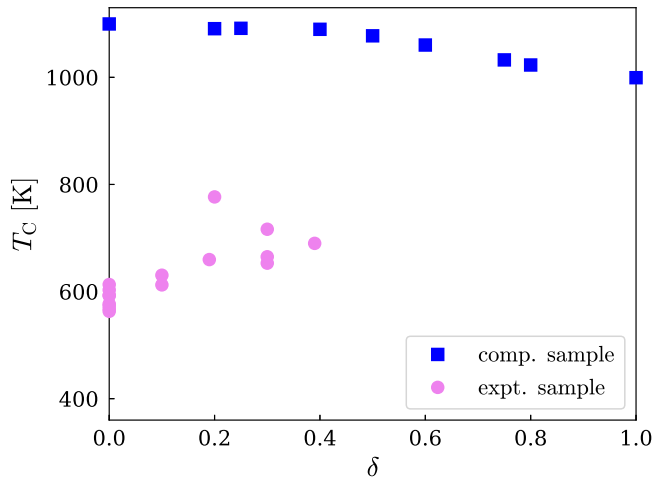


FIG. 6. Co concentration dependence of the Curie temperature of  $\text{Nd}_2(\text{Fe}_{1-\delta}\text{Co}_\delta)_{14}\text{B}$ . Blue squares denote the first-principles data and violet points denote the experimental data.

In  $\text{Sm}(\text{Fe}, \text{Co})_{12}$ , such reduction of the magnetization at zero temperature and enhancement of the Curie temperature by Co doping have been also observed in experiment [6].

This conclusion is based on the fact that the  $T_C$  increases with the Co concentration. However, it contains a tricky prob-

lem in the first-principles calculation. Figure 6 shows the Co concentration dependence of  $T_C$  obtained by the experiment (step 1-1 in Fig. 2) and by first-principles calculation (step 1-2). As seen in the figure, the calculated value decreases as the concentration increases, i.e., the trend is opposite from that of experiment. This discrepancy originates from theoretical errors in the first-principles data, which are systematic errors but not accidental ones. It is a hard task to reduce the errors by improving approximations contained in theoretical methods. However, such systematic errors can be complemented by the data assimilation if there is a correlation between computational data and experimental data, even though the correlation is negative. In our scheme, the negative correlation in  $T_C$  is expressed by the linear term in the right side of Eq. (25).

Figures 7(a) and 7(b) show scatter plots of the magnetization at 300 and 400 K, respectively, where the predicted magnetizations on uniform mesh points are plotted against the sum of Nd and Pr concentration ( $1 - \beta - \gamma$ ), which can be regarded as the ratio of critical rare earths in the present system. The dotted straight line connects the magnetizations of two systems:  $\text{La}_2\text{Fe}_{14}\text{B}$  and  $\text{Nd}_2\text{Fe}_{14}\text{B}$ . While all the points are below the dotted line at 300 K, some points appear above the line at 400 K. This is attributed to the enhancement of the magnetization at high temperatures by Co doping. The maximum magnetization at  $T = 400$  K is achieved in

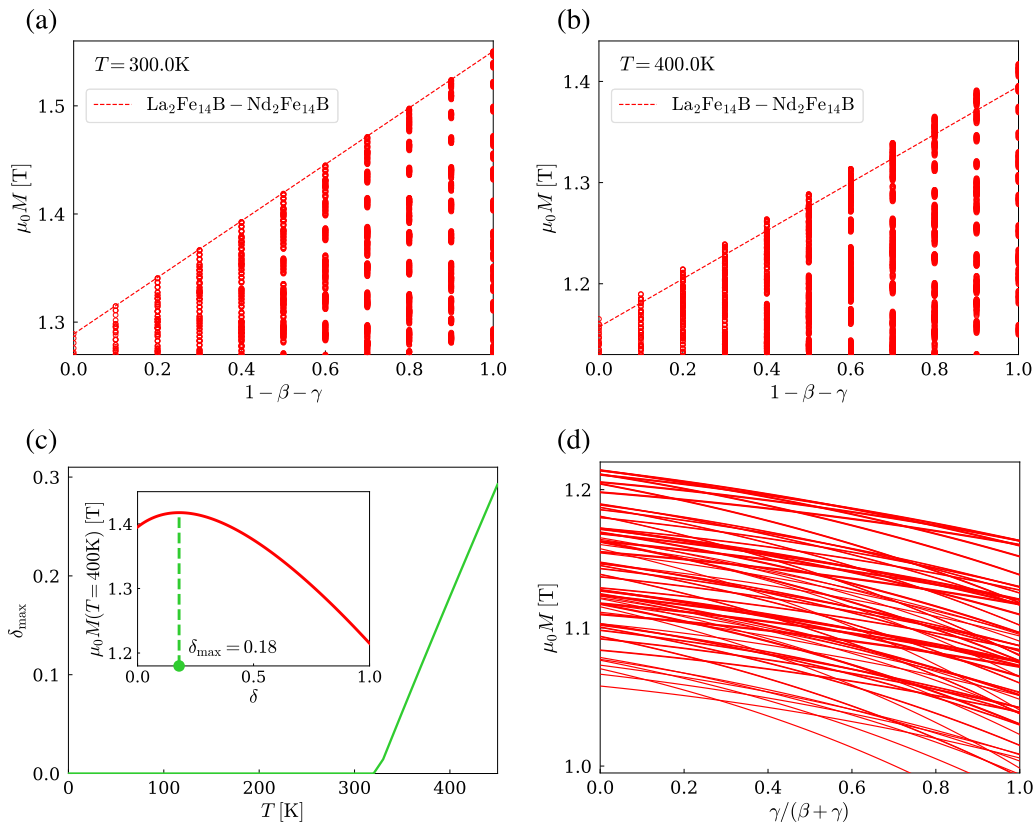


FIG. 7. Scatter plots of the magnetization at (a) 300 K and (b) 400 K as a function of Nd and Pr concentration ( $1 - \beta - \gamma$ ). The dotted straight line connects the magnetizations of two systems:  $\text{La}_2\text{Fe}_{14}\text{B}$  and  $\text{Nd}_2\text{Fe}_{14}\text{B}$ . (c) Cobalt concentration associated with chemical composition giving maximum magnetization. The horizontal axis is temperature in Kelvin. The inset is an example of the magnetization at 400 K of  $\text{Nd}_2(\text{Fe}_{1-\delta}\text{Co}_\delta)_{14}\text{B}$ . A green point denotes the  $\delta_{\text{max}}$  at the temperature. (d) Magnetization at 400 K as a function of  $\gamma/(\beta + \gamma)$ . Other variables are fixed in each line and sampled within the range of  $0.0 \leq \alpha \leq 0.2$ ,  $0.8 \leq \beta + \gamma \leq 1.0$ ,  $0.0 \leq \delta \leq 0.3$ , and  $0.0 \leq \zeta \leq 0.1$ .

$\text{Nd}_2(\text{Fe}, \text{Co})_{14}\text{B}$  with the cobalt concentration  $\delta_{\max} = 0.18$ . Figure 7(c) shows the temperature dependence of  $\delta_{\max}$ . The  $\delta_{\max}$  is nonzero above 320 K, and increases with raising the temperature. Finally, we analyze the magnetization as a function of Ce concentration. In Fig. 7(d), the magnetization is plotted as a function of  $\gamma/(\beta + \gamma)$  for fixed values of other variables  $\alpha, \beta + \gamma, \delta, \zeta$ . Each line shows the result for different  $(\alpha, \beta + \gamma, \delta, \zeta)$ . The magnetization is a convex upward function, and decreases with increasing the cerium concentration. This indicates that there is an optimum cerium concentration to effectively utilize abundant Ce element.

#### IV. CONCLUSION

We have proposed a data assimilation method in which a small number of accurate data and a large number of less accurate data are integrated. The method enables us to predict the behavior in the region where the former data are not available. Based on this method, we have developed a practical scheme to estimate the finite-temperature magnetization at arbitrary composition. We applied the scheme to  $(\text{Nd}_{1-\alpha-\beta-\gamma}\text{Pr}_\alpha\text{La}_\beta\text{Ce}_\gamma)_2(\text{Fe}_{1-\delta-\zeta}\text{Co}_\delta\text{Ni}_\zeta)_{14}\text{B}$  and obtained the magnetization in the five-dimensional composition space. Co addition enhances the magnetization above 320 K. It is noticed that this is solely due to the fact that the Curie temperature rises by the Co addition: the magnitude of magnetization at zero temperature cannot be enhanced.

#### ACKNOWLEDGMENTS

This work was supported by MEXT as ‘‘Program for Promoting Researches on the Supercomputer Fugaku’’ (DPMSD). Part of the computation in this work was performed using the facilities of the Supercomputer Center of the Institute for Solid State Physics at the University of Tokyo, and computational resources of the HPCI system through the HPCI System Research Project (Project No. ID:hp200125).

#### APPENDIX A: EXPERIMENTAL PROCEDURE

119 kinds of  $(\text{Nd}, \text{Ce}, \text{La}, \text{Pr})_{13.55}(\text{Fe}, \text{Co}, \text{Ni})_{80.54}\text{B}_{5.91}$  (at.%) alloys were prepared by arc melting. These alloys were annealed at 1373 K for 24 h in Ar atmosphere. These alloys were pulverized and sorted into particles with diameters of  $<20 \mu\text{m}$  in an inert atmosphere to make magnetically anisotropic powder. The powder density was determined using a pycnometer (Ultracyc1200e, Quantachrome Instruments, USA). The powder compositions were measured by ICP-AES (ICPS8100, Shimadzu, Japan) and the main phase, 2-14-1 phase, ratio was calculated from the obtained composition. The magnetic physical properties were measured by using a Vibrating Sample Magnetometer (VSM) (PPMS EverCool II, QuantumDesign, USA) at a maximum applied field of 9 T.

The powder was mixed with an epoxy resin in a Cu container and solidified in a magnetic field of 1 T to measure with VSM. The Magnetic Hysteresis (MH) curve of magnetically easy and hard direction was measured in the temperature range of 300 to 453 K. The anisotropy field ( $H_A$ ) was detected by singular point detection (SPD) method [34] from MH curve of hard direction. When the anisotropy field is

less than 9 T, less than the applied magnetic field of VSM,  $H_A$  can be detected by SPD. If the anisotropy field is over 9 T, the MH curves for both of easy and hard direction were extrapolated in high magnetic field to obtain the intersection point, and the magnetic field at the intersection was made into  $H_A$ . On the other hand, the saturation magnetization ( $J_s$ ) was estimated by the law of approach to saturation (LAS) [35–37] from the MH curve of easy direction. When  $H_A$  was higher than 9 T, which is the maximum applied field, Eq. (A1) was used, and when it was sufficiently lower than 9 T, Eqs. (A2) and (A3) were used.  $J_s$  is saturation magnetization and  $b$  and  $\chi_0$  are constants. The  $\chi_0 H$  term is often referred to as the so-called paramagnetismlike term [38]. Equations (A2) and (A3) dealing with the  $\chi_0$  term is more accurate for measuring saturation magnetization, but there is a condition that the applied magnetic field is sufficiently larger than the anisotropic magnetic field. The saturation magnetization estimated by the LAS was divided by the main phase ratio of powder to obtain the saturation magnetization of the 2-14-1 phase in these compositions, where the grain boundary phases other than the main phase were treated as paramagnetic phases:

$$J = J_s \left( 1 - \frac{b}{H^2} \right), \quad (\text{A1})$$

$$J = J_s \left( 1 - \frac{b}{H^2} \right) + \chi_0 H, \quad (\text{A2})$$

$$\frac{dJ}{dH} = J_s \left( \frac{2b}{H^3} \right) + \chi_0. \quad (\text{A3})$$

#### APPENDIX B: AN EXAMPLE OF TWO-SUBLATTICE MODEL

The scheme for finite-temperature magnetization shown by the flowchart of Fig. 2 includes the steps for modeling temperature dependence of magnetization. In this study, we use Kuz’min’s formula for the modeling. A two-sublattice model [2,32,33] is one of the other possible models to replace the steps. Then we demonstrate the fitting by using the two-sublattice model for  $\text{Nd}_2\text{Fe}_{14}\text{B}$ , which corresponds to Fig. 3 in the case of Kuz’min’s formula.

The two-sublattice model is defined as

$$\mu_R(T) = \mu_{R0} B_J(\mu_B \mu_{R0} H_R(\mu_R, \mu_T)/k_B T), \quad (\text{B1})$$

$$\mu_T(T) = \mu_{T0} B_S(\mu_B \mu_{T0} H_T(\mu_R, \mu_T)/k_B T), \quad (\text{B2})$$

where the subscripts  $R$  and  $T$  denote rare-earth and transition metal elements.  $B_J$  is the Brillouin function defined as

$$B_J(x) \equiv \frac{2J+1}{2J} \coth\left(\frac{2J+1}{2J}x\right) - \frac{1}{2J} \coth\left(\frac{1}{2J}x\right), \quad (\text{B3})$$

and molecular fields for rare-earth and transition metal atoms are

$$H_R(\mu_R, \mu_T) = d(2n_{RR}\mu_R + 14n_{RT}\mu_T), \quad (\text{B4})$$

$$H_T(\mu_R, \mu_T) = d(2n_{RT}\mu_R + 14n_{TT}\mu_T). \quad (\text{B5})$$

$\mu_R$  and  $\mu_T$  are the local magnetic moments of rare-earth and transition metal atoms.  $\mu_{R0}$  and  $\mu_{T0}$  are the corresponding values at zero temperature, and we here fix  $\mu_{R0}$  as  $3.3 \mu_B$



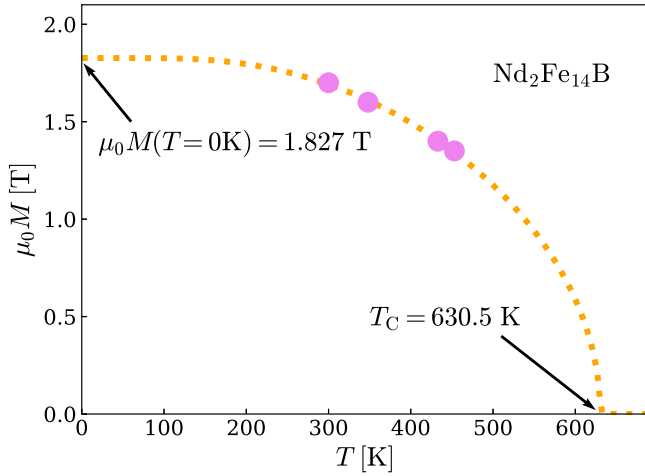


FIG. 8. The magnetization of  $\text{Nd}_2\text{Fe}_{14}\text{B}$  predicted by using the two-sublattice model, Eqs. (B1) and (B2).

for Nd.  $J$  and  $S$  are the free-ion moments of rare-earth and transition metal atoms, and thus,  $J = \frac{9}{2}$  and  $S = 1$ .  $\mu_B$  is the Bohr magneton.  $d \equiv \mu_0\mu_B/V$  is a constant to convert the unit of magnetic moment (Bohr magneton) to that of magnetization (Tesla).  $\mu_0$  is the vacuum permeability and  $V$  is a volume of  $\text{Nd}_2\text{Fe}_{14}\text{B}$  per formula unit.  $n_{RR}$ ,  $n_{TT}$ , and  $n_{RT}$  are the dimensionless molecular-field coefficients which represent magnetic couplings of  $R$ - $R$ ,  $T$ - $T$ , and  $R$ - $T$ , respectively. The magnetization is given as

$$\mu_0 M(T) = d[2\mu_R(T) + 14\mu_T(T)]. \quad (\text{B6})$$

To obtain  $\mu_R$  and  $\mu_T$  at a temperature, we solve the simultaneous Eqs. (B1), (B2), (B4), and (B5) self-consistently. We optimize  $\mu_{T0}$ ,  $n_{RR}$ ,  $n_{TT}$ , and  $n_{RT}$  such that the experimental data are well predicted. The result is shown in Fig. 8. In the present case, there is no qualitative difference from Kuz'min's formula of Fig. 3. When the temperature dependence will be more complicated, such as ferrimagnetism, this will be important.

#### APPENDIX C: FIRST-PRINCIPLES CALCULATION

We perform density functional theory calculation following Korringa-Kohn-Rostoker (KKR) [39,40] Green's function approach in the atomic sphere approximation (ASA) incorporating coherent potential approximation (CPA) [41,42] as implemented in AKAIKKR [43]. Continuous interpolation over  $0 \leq \alpha, \beta, \gamma, \delta, \zeta \leq 1$  for  $(\text{Nd}_{1-\alpha-\beta-\gamma}\text{Pr}_\alpha\text{La}_\beta\text{Ce}_\gamma)_2(\text{Fe}_{1-\delta-\zeta}\text{Co}_\delta\text{Ni}_\zeta)_{14}\text{B}$  is done on the basis of KKR-CPA. The local spin density approximation as parametrized by Moruzzi, Janak, and Williams [44] is adopted for the exchange-correlation energy functional. We set  $l_{\text{max}} = 2$  putting all  $4f$  electrons of the valence state on the basis of open-core approximation. We assume each configuration of the electrons as  $\text{Nd}^{3+}$ ,  $\text{Pr}^{3+}$ , and  $\text{Ce}^{4+}$ , respectively. Contributions from  $4f$  electrons of Nd and Pr to the magnetic moments are restored manually.

Intersite magnetic exchange couplings are calculated using the method developed by Liechtenstein *et al.* [45]. The Curie

temperature is evaluated by solving the derived Heisenberg model in the mean field approximation.

#### APPENDIX D: LATTICE PARAMETERS

We collected experimental lattice parameters of  $R_2T_{14}\text{B}$  ( $R = \text{Nd, La, Ce, Pr}$ ;  $T = \text{Fe, Co, Ni}$ ) from literature [2] [Table II(a)]. However, some of them are not available. In order to evaluate the lattice parameters for these missing points, we performed first-principles calculation based on density functional theory [22,23] in the generalized gradient approximation with Perdew-Burke-Ernzerhof formula [46] using VASP code [47] [Table II(b)]. We integrated the calculated lattice parameters  $a$  and  $c$  with the above experimental data assuming the following simple form:

$$a^{\text{comp}} = \sum_I a_I^{\text{comp}} \alpha^{i_\alpha} \beta^{i_\beta} \gamma^{i_\gamma} \delta^{i_\delta} \zeta^{i_\zeta}, \quad (\text{D1})$$

$$c^{\text{comp}} = \sum_I c_I^{\text{comp}} \alpha^{i_\alpha} \beta^{i_\beta} \gamma^{i_\gamma} \delta^{i_\delta} \zeta^{i_\zeta}, \quad (\text{D2})$$

$$a^{\text{expt}} = C_a a^{\text{comp}} + a_0, \quad (\text{D3})$$

$$c^{\text{expt}} = C_c c^{\text{comp}} + c_0, \quad (\text{D4})$$

where  $I := (i_\alpha, i_\beta, i_\gamma, i_\delta, i_\zeta)$  runs under the condition  $0 \leq \sum_j i_j \leq 1$ . Using the data assimilation method explained in Sec. II, we determined the coefficients  $\{a_I^{\text{comp}}\}$ ,  $\{c_I^{\text{comp}}\}$ ,  $C_a$ ,  $a_0$ ,  $C_c$ ,  $c_0$ , from which  $a^{\text{expt}}$  and  $c^{\text{expt}}$  for the missing points are estimated [Table II(c)]. We interpolated lattice parameters for nonstoichiometric systems according to Vegard's law by using the obtained lattice parameters for stoichiometric systems.

TABLE II. Lattice parameters ( $a$ ,  $c$ ) of  $R_2T_{14}\text{B}$  for  $R = \text{Nd, La, Ce, Pr}$  and  $T = \text{Fe, Co, Ni}$ . (a) Experimental values taken from Ref. [2], (b) computational results, and (c) values obtained by the data assimilation technique to complement the unavailable values of (a).

	Fe	Co	Ni
(a)			
Nd	(8.80, 12.20)	(8.64, 11.86)	
La	(8.82, 12.34)	(8.67, 12.01)	
Ce	(8.76, 12.11)		
Pr	(8.80, 12.23)	(8.63, 11.86)	
(b)			
	Fe	Co	Ni
Nd	(8.73, 12.07)	(8.58, 11.74)	(8.57, 11.75)
La	(8.75, 12.16)	(8.61, 11.80)	(8.59, 11.79)
Ce	(8.70, 11.96)	(8.55, 11.63)	(8.54, 11.62)
Pr	(8.75, 12.11)	(8.59, 11.77)	(8.59, 11.77)
(c)			
	Fe	Co	Ni
Nd			(8.62, 11.89)
La			(8.65, 11.93)
Ce		(8.61, 11.77)	(8.60, 11.75)
Pr			(8.64, 11.91)

- [1] M. Sagawa, S. Fujimura, N. Togawa, H. Yamamoto, and Y. Matsuura, *J. Appl. Phys.* **55**, 2083 (1984).
- [2] J. F. Herbst, *Rev. Mod. Phys.* **63**, 819 (1991).
- [3] T. Miyake and H. Akai, *J. Phys. Soc. Jpn.* **87**, 041009 (2018).
- [4] F. Kools, A. Morel, R. Grössinger, J. L. Breton, and P. Tenaud, *J. Magn. Magn. Mater.* **242–245**, 1270 (2002).
- [5] A. K. Pathak, M. Khan, K. A. Gschneidner Jr., R. W. McCallum, L. Zhou, K. Sun, K. W. Dennis, C. Zhou, F. E. Pinkerton, M. J. Kramer, and V. K. Pecharsky, *Adv. Mater.* **27**, 2663 (2015).
- [6] Y. Hirayama, Y. K. Takahashi, S. Hirose, and K. Hono, *Scr. Mater.* **138**, 62 (2017).
- [7] R. E. Bellman, *Adaptive Control Processes: A Guided Tour* (Princeton University Press, Princeton, 1961).
- [8] A. Alam, M. Khan, R. W. McCallum, and D. D. Johnson, *Appl. Phys. Lett.* **102**, 042402 (2013).
- [9] P. Tenaud, A. Morel, F. Kools, J. Le Breton, and L. Lechevallier, *J. Alloys Compd.* **370**, 331 (2004).
- [10] W. Körner, G. Krugel, and C. Elsässer, *Sci. Rep.* **6**, 24686 (2016).
- [11] P. Nieves, S. Arapan, J. Maudes-Raedo, R. Marticorena-Sánchez, N. Del Brío, A. Kovacs, C. Echevarria-Bonet, D. Salazar, J. Weischenberg, H. Zhang *et al.*, *Comput. Mater. Sci.* **168**, 188 (2019).
- [12] T. Fukazawa, Y. Harashima, Z. Hou, and T. Miyake, *Phys. Rev. Mater.* **3**, 053807 (2019).
- [13] N. Papanikolaou, R. Zeller, and P. H. Dederichs, *J. Phys.: Condens. Matter* **14**, 2799 (2002).
- [14] L. Breiman and J. H. Friedman, *J. R. Stat. Soc.: Ser. B Stat. Methodol.* **59**, 3 (1997).
- [15] T. Hastie, R. Tibshirani, and J. Friedman, in *The Elements of Statistical Learning: Data Mining, Inference, and Prediction* (Springer, New York, 2009), pp. 43–99.
- [16] C. Stein, in *Proceedings of the Third Berkeley Symposium on Math Statistics and Probability* (University of California Press, Berkeley, CA, 1956), Vol. 1, pp. 197–206.
- [17] B. Efron and C. Morris, *J. Am. Stat. Assoc.* **70**, 311 (1975).
- [18] D. B. Rubin, *Biometrika* **63**, 581 (1976).
- [19] C. G. Broyden, *IMA J. Appl. Math.* **6**, 76 (1970).
- [20] R. Fletcher, *Comput. J.* **13**, 317 (1970).
- [21] R. H. Byrd, P. Lu, J. Nocedal, and C. Zhu, *SIAM J. Sci. Comput.* **16**, 1190 (1995).
- [22] P. Hohenberg and W. Kohn, *Phys. Rev.* **136**, B864 (1964).
- [23] W. Kohn and L. J. Sham, *Phys. Rev.* **140**, A1133 (1965).
- [24] N. M. Rosengaard and B. Johansson, *Phys. Rev. B* **55**, 14975 (1997).
- [25] M. Pajda, J. Kudrnovský, I. Turek, V. Drchal, and P. Bruno, *Phys. Rev. B* **64**, 174402 (2001).
- [26] S. Hock, Ph.D. thesis, Universität Stuttgart, 1988.
- [27] R. Grössinger, X. K. Sun, R. Eibler, K. H. J. Buschow, and H. R. Kirchmayr, *J. Phys. Colloques* **46**, C6-221 (1985).
- [28] M. D. Kuz'min, *Phys. Rev. Lett.* **94**, 107204 (2005).
- [29] M. D. Kuz'min and A. M. Tishin, *Phys. Lett. A* **341**, 240 (2005).
- [30] M. D. Kuz'min, M. Richter, and A. N. Yaresko, *Phys. Rev. B* **73**, 100401(R) (2006).
- [31] M. Ito, M. Yano, N. M. Dempsey, and D. Givord, *J. Magn. Magn. Mater.* **400**, 379 (2016).
- [32] C. Fuerst, J. Herbst, and E. Alson, *J. Magn. Magn. Mater.* **54–57**, 567 (1986).
- [33] R. Skomski, *Simple Models of Magnetism* (Oxford University Press, Oxford, 2008).
- [34] R. Cabassi, *Measurement* **160**, 107830 (2020).
- [35] N. S. Akulov, *Z. Phys.* **69**, 822 (1931).
- [36] G. Hadjipanayis, D. J. Sellmyer, and B. Brandt, *Phys. Rev. B* **23**, 3349 (1981).
- [37] H. Kronmüller and M. Fähnle, *Micromagnetism and the Microstructure of Ferromagnetic Solids* (Cambridge University Press, Cambridge, 2003).
- [38] H. Zhang, D. Zeng, and Z. Liu, *J. Magn. Magn. Mater.* **322**, 2375 (2010).
- [39] J. Korringa, *Physica (Amsterdam)* **13**, 392 (1947).
- [40] W. Kohn and N. Rostoker, *Phys. Rev.* **94**, 1111 (1954).
- [41] H. Shiba, *Prog. Theor. Phys.* **46**, 77 (1971).
- [42] H. Akai, *Phys. B+C (Amsterdam)* **86–88**, 539 (1977).
- [43] H. Akai, AKAIKKR, <http://kkriissp.u-tokyo.ac.jp/>.
- [44] V. L. Moruzzi, J. F. Janak, and A. R. Williams, *Calculated Properties of Metals* (Pergamon Press, New York, 1978).
- [45] A. Liechtenstein, M. Katsnelson, V. Antropov, and V. Gubanov, *J. Magn. Magn. Mater.* **67**, 65 (1987).
- [46] J. P. Perdew, K. Burke, and M. Ernzerhof, *Phys. Rev. Lett.* **77**, 3865 (1996).
- [47] G. Kresse and D. Joubert, *Phys. Rev. B* **59**, 1758 (1999).