

Assessing Phase Diagram Accuracy

Axel van de Walle¹ · Qijun Hong¹

Submitted: 23 October 2018 / in revised form: 18 December 2018 / Published online: 31 January 2019
© The Author(s) 2019

Abstract Assessing the predictive power of any computational model requires the definition of an appropriate metric or figure-of-merit (e.g. mean square error, maximum error, etc). However, quantifying errors in an alloy phase diagram with a single figure-of-merit is a considerably more complex problem. The “distance” between phase boundaries is not a uniquely defined concept and different phase diagrams may differ in the possible stable phases which they predict, making it unclear which “distance” to measure. Given the difficulty associated with such metrics, we instead propose to use differences in predicted phase fractions between different phase diagrams as the basis of a suitable metric. We prove that our criterion satisfies all the properties of the mathematical notion of a norm or of a metric, in addition to other properties directly relevant to phase stability problems. We illustrate the use of such criterion to the study of the convergence of assessments performed on the same alloy system by different authors over time.

Keywords figure-of-merit · materials informatics · metric · norm

1 Introduction

One of the underlying assumption of the goal of achieving “predictive science” is the availability of suitable metric to quantify the accuracy of the predictions. Although mean square error-type quantities are often well-suited for this

purpose when the predicted quantities are simple functions or vectors, quantifying errors in a more general graph, such as phase diagram, in a single figure-of-merit represents a considerably more complex problem. Given this, it is perhaps not surprising that the area of research focusing on the construction of thermodynamic models [often referred to as CALculation of PHase Diagram (CALPHAD)]^[5,10–12,15] is currently lacking a theoretically justified and widely adopted figure-of-merit to quantify the discrepancies between two possible phase diagrams obtained via different routes. This paper intends to fill this gap by building upon earlier proposals.^[17]

There are multiple challenges associated with this task, as illustrated in Fig. 1. Phase boundaries cannot be considered single-valued functions: For instance, in a binary phase diagram, a phase boundary might cross a given vertical line (see label 1 on the figure) multiple times, thus precluding the use of simple “mean-square error along one axis” criteria. The same can occur for horizontal lines (see label 2 on the figure). More generally, no matter which “axes” one uses, the phase boundaries will generally be defined on different domains (label 3). The use of perpendicular distances between two curves is also ambiguous: To which of the two curves should the distance be perpendicular to (label 4)? What does “perpendicular” mean when the axes have different units so that their relative scaling is arbitrary (label 5)? For the same reason, the use of the Hausdorff metric^[2] is also affected by the fact that different axes may have different units. Boundary distance-based metrics are also unable to handle the fact that, sometimes, phases are entirely absent from one phase diagram but present in another (label 6). How should a figure-of-merit quantify this situation?

Given these issues, we instead propose to quantify the differences between two phase diagrams via differences in

✉ Axel van de Walle
avdw@alum.mit.edu

¹ School of Engineering, Brown University, Providence

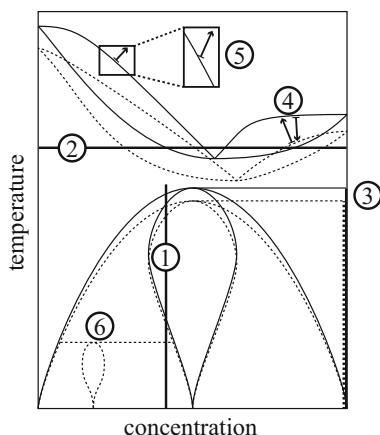


Fig. 1 Problems associated with metrics based on distance between phase boundaries. Solid and dashed curves distinguish two different phase diagrams while thicker curves mark potential problems: multiple crossing of a vertical (1) or horizontal (2) axis, different domains of definition (3) of the functions defining the boundaries, ambiguities in defining the orthogonal distance between 2 curves (4), ambiguity of the notion of orthogonality when axes have arbitrary relative scales due to differing units (5) and potential absence of a phase (6)

the predicted phase fractions of the corresponding phases. Phase fractions have been put forward as powerful and fundamental descriptors of phase equilibria.^[16] Phase fractions are scalar, dimensionless, everywhere defined and merely take the value 0 when a phase is not stable. These desirable properties solve all of the aforementioned problems. We prove that a criterion based on phase fractions satisfies all the properties of the mathematical notion of a norm or of a metric, in addition to other properties directly relevant to phase stability problems. We illustrate the use of such a criterion to the study of the convergence of assessments performed on the same alloy system by different authors over time.

2 Definition and Motivation

Let $f(c) = (f_1(c), \dots, f_p(c))^T$ denote a p -dimensional vector of the phase fractions of all possible phases in the system under conditions $c = (x, T)$, where T is temperature and x is a vector of overall compositions (omitting one composition, to avoid redundancy). One could also include pressure into the vector c , if desired. The knowledge of this vector-valued function over some region R fully defines the phase diagram over that region.

We propose to quantify the difference between two phase diagrams f^1, f^2 in a region R of interest via the following figure-of-merit:

$$\|f^1 - f^2\|_R \equiv \frac{\int_{c \in R} \sum_{\alpha=1}^p |f_{\alpha}^1(c) - f_{\alpha}^2(c)| dc}{\int_{c \in R} dc}, \tag{Eq 1}$$

where the integrals are multivariate, i.e. dc is a (hyper)volume element. This definition exhibits a number of desirable properties. It has the natural interpretation of the expected total absolute error in phase fraction for a condition chosen at random in region R .

Since this definition is mathematically equivalent to a so-called weighted L_1 norm^[18] defined on a vector-valued field, it automatically inherits all the natural properties of norm: It is zero if and only if the two phase diagrams f^1 and f^2 agree,¹ it is symmetric ($\|f^1 - f^2\|_R = \|f^2 - f^1\|_R$) and it obeys the triangular inequality $\|f^1 - f^2\|_R \leq \|f^1 - f^3\|_R + \|f^3 - f^2\|_R$, for 3 phase diagrams f^1, f^2, f^3 . (A norm also satisfies $\|af\| = |a|\|f\|$ but this property is not useful in this context, since the phase fractions must sum up to one.) Since a norm is a special case of a metric, our proposal also defines a proper metric.

Definition (1) provides a dimensionless quantity, which facilitates its interpretation. Another desirable property is that it naturally handles the case when one phase is simply missing in one phase diagram. This possibility is not uncommon when comparing experimental and ab initio phase diagrams. Also, in novel systems that are not yet well characterized, there may not be perfect knowledge of which phases are stable or metastable and it is useful to be able to quantify this type of discrepancy. It is not clear how missing phases could be handled with a figure-of-merit based on distances between phase boundaries. This definition also applies a less severe penalty in situations in which only one phase is in disagreement while the others agree. This makes sense, since this situation typically arises when two phases have very similar free energies and can easily be mispredicted without affecting the reliability of the phase diagram elsewhere.

Our approach also makes it simple to account for the fact that Gibbs triangles (or, more generally, Gibbs simplexes) are the most natural way to represent multiple composition axes. The fact that the axes are not orthogonal can be ignored in definition (1) because the same Jacobian terms appear in both the numerator and denominator. One can thus simply integrate over all but one composition using orthogonal axes, whether these axes are truly orthogonal or not in the phase diagram's representation.

The definition (1) is also computationally attractive since it can easily be calculated via Monte Carlo sampling. Let $c_i, i = 1, \dots, n$ denote independent random draws from a uniform distribution over the region R , then, by the Law of large numbers

¹ Except perhaps on an irrelevant set of zero measure.

$$\frac{1}{n} \sum_{i=1}^n \sum_{\alpha=1}^p |f_{\alpha}^1(c_i) - f_{\alpha}^2(c_i)| \quad (\text{Eq 2})$$

converges in probability to $\|f^1 - f^2\|_R$ as $n \rightarrow \infty$. In multicomponent systems, some care must be taken to ensure that the compositions are indeed drawn uniformly. For instance, in a ternary system, picking 3 composition uniformly on $[0, 1]$ and scaling them so their sum is one does *not* generate a uniform distribution. However, picking the 3 composition from an exponential distribution and then normalizing them to sum to 1, does generate a uniform distribution (see Ref 19). This Monte Carlo algorithm scales very favorably with the number of components in the system, unlike a standard grid integration.

The definition does exhibit some limitations. Most importantly, it is dependent on the choice of the region of interest R . This could be mitigated by agreeing on standardized regions. For instance, one can use a region including the entire composition range and temperatures from absolute zero (or room temperature) to the highest phase transformation temperature of all systems considered.

As an alternative, one could also use

$$\|f^1 - f^2\|_{\rho,R} \equiv \left(\frac{\int_{c \in R} \sum_{\alpha=1}^p |f_{\alpha}^1(c) - f_{\alpha}^2(c)|^{\rho} dc}{\int_{c \in R} dc} \right)^{1/\rho} \quad (\text{Eq 3})$$

for some $\rho \geq 1$ to obtain analogues of any of the familiar L_{ρ} norms.^[18] For single phase regions, this substitution has no effect since $0^{\rho} = 0$ and $1^{\rho} = 1$. For multiphase equilibria, the choice of ρ does matter, but we suggest keeping the $\rho = 1$ choice due to its ease of interpretation.

Another possible alternative to Definition (1), which is related to a previously proposed figure-of-merit,^[17] is to only compare the presence or absence of a phase, independently of its phase fraction and define the metric:

$$d_R(f^1, f^2) \equiv \frac{\int_{c \in R} \sum_{\alpha=1}^p |\mathbf{1}_{+}(f_{\alpha}^1(c)) - \mathbf{1}_{+}(f_{\alpha}^2(c))| dc}{\int_{c \in R} dc} \quad (\text{Eq 4})$$

where the function $\mathbf{1}_{+}(f)$ is equal to 1 if $f > 0$ and 0 otherwise. This definition has the interpretation of the expected number of mismatched phases for a point c picked at random in R . It also can be roughly interpreted as the fraction of the (hyper)volume of the phase diagram where disagreements between the nature of the stable phases exist, weighted by the number of mismatches [One could eliminate this weighting by replacing $\sum_{\alpha=1}^p$ by $\max_{\alpha \in \{1, \dots, p\}}$ in (4)]. The notation $d_R(f^1, f^2)$ (instead of $\|f^1 - f^2\|_R$) is used because Eq 4 does not define a norm when viewed as a function of the $f_{\alpha}^i(c)$, since it cannot be

written as a function of $f_{\alpha}^1(c) - f_{\alpha}^2(c)$. However, it is a norm when viewed as a function of the $\mathbf{1}_{+}(f_{\alpha}^i(c))$. This definition therefore still defines a proper metric $d_R(f^1, f^2)$, since the knowledge of $\mathbf{1}_{+}(f_{\alpha}(c))$ also fully characterizes the phase diagram. Definition (4) may be easier to implement if one only has access to the picture of the phase diagram (instead of its underlying thermodynamic model), because it avoids implementing the level rule to recover the phase fractions (which becomes difficult beyond binary systems). It is also in the same spirit as recent efforts to cast inverse problems in thermodynamics modeling as constraint satisfaction problems.^[8]

The presence of a miscibility gap leads to a subtle complication in Definition (1) [or (4)]. In this case, multiple phases exhibiting the same crystal structure but at different compositions could be in a multiphase equilibrium. We handle this by considering each phase (even with the same crystal structure) as distinct but when comparing the resulting phase fractions across two phase diagrams, we always re-order the phase fractions (among phases sharing the same crystal structure) so as to minimize (1). If the number of phases with the same crystal structure is different in the two phase diagrams, we then add the appropriate number of phases with a zero phase fraction. In the limit where the differences between the two phases diagrams are small, this simple rule yields differences in phase fraction between corresponding phases.

The proposed metrics have been implemented as the `phasenorm` command in the ATAT package^[23–25] and this implementation relies on OpenCalphad^[21,22] to compute phase equilibria. It has the following syntax:

```
phasenorm -tdb1=tdbfile1 -tdb2= tdbfile2
-e=element1,element2,... -n=nb of samples -T0=min
temperature -T1=max temperature [-01]
```

where

- *tdbfile1* and *tdbfile2* are the two thermodynamic database files (in the TDB format^[1]) of the assessments to be compared;
- *element1,element2,...* is a comma-separated list of the elements involved in the phase diagram of interest (which allows the user to extract a subsystem from the TDB files);
- *nb of samples* specifies the number of Monte Carlo sampling steps performed;
- *min temperature* and *max temperature* define the temperature range of the region R of interest (the full composition range is assumed);
- the optional `-01` switch instructs the use of Eq 4 instead of (1).

When using this tool, it is important to ensure that the two thermodynamic database files use the same naming conventions for the phases.

3 Application Example

The figure-of-merit proposed here enables instructive quantitative studies of the accuracy of phase diagrams. One natural question, for instance, is whether the assessments of an alloy system are actually converging, that is, becoming more accurate over time as more data become available and more researchers study the same system.

Since one never actually knows for sure what is the “true” phase diagram, it may not be obvious if any convergence of our knowledge of given system is really taking place. One needs to study convergence through an internal consistency criterion that does not require that the true phase diagram be perfectly known. Given a sequence s_1, s_2, \dots , how can one know if it converges without first calculating the limit $\lim_{n \rightarrow \infty} s_n$? The way out of this circular reasoning is to check if the sequence $\{s_n\}$ forms a *Cauchy sequence*, i.e., whether it has the property that

$$\lim_{n \rightarrow \infty} \max_{i,j \geq n} \|s_i - s_j\| = 0.$$

It can be shown that any such Cauchy sequence necessarily converges^[18] (under a technical condition known as completeness of the space in which the s_n live, which is typically satisfied for vector spaces commonly used to represent scientific data).

In our context, this amounts to checking if the distance between all phase diagrams $\{f^i\}$ published on a given system after some time t decreases significantly as time t progresses:

$$D_t = \max_{i,j \text{ such that } t_i \geq t \text{ and } t_j \geq t} \|f^i - f^j\|_R, \tag{Eq 5}$$

where t_i is the publication time of phase diagram i . Of course, by construction, the quantity D_t necessarily decreases with increasing t , but the rate at which this happens is instructive and the absolute magnitude of D_t is indicative of the expected remaining errors at time t .

Using the recently developed Thermodynamic DataBase DataBase (TDBDB),^[26] one can easily identify popular alloy systems that have been repeatedly assessed, so that the Cauchy property can be tested. We have selected the many assessments of the Fe-Ti system^[3,4,7,9,13,27] and of the Al-Cu system^[6,14,20,22,28] that are available in the TDBDB and that can be parsed by OpenCalphad. We have used the Monte Carlo algorithm (Eq 2) with 2000 draws and a region bounded by 300 and 2000 K that covers the entire composition range.

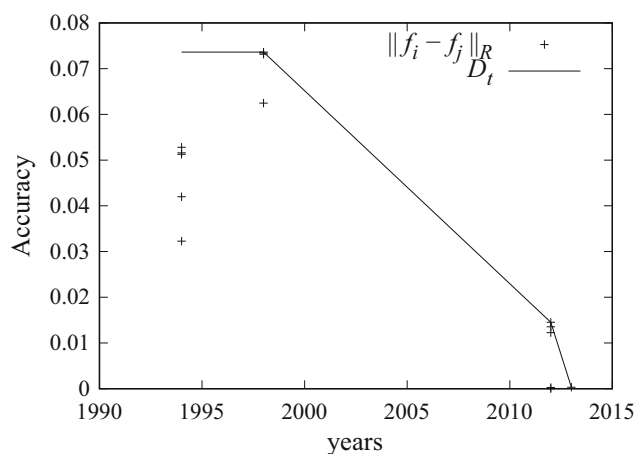


Fig. 2 Convergence of assessments of the Fe-Ti system. Each cross represents the distances between a phase diagram published at time t and a future (or concurrent) phase diagram. The line represents D_t , the largest distance between pairs of phase diagrams published at or after time t

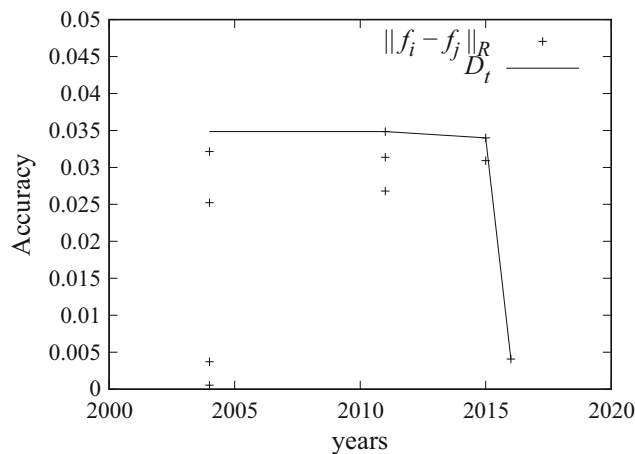


Fig. 3 Convergence of assessments of the Al-Cu system. Each cross represents the distances between a phase diagram published at time t and a future (or concurrent) phase diagram. The line represents D_t , the largest distance between pairs of phase diagrams published at or after time t

In Fig. 2, it can be seen that the distances between assessments do clearly decrease sharply over time, indicating that a consensus regarding the Fe-Ti phase diagram is steadily emerging. In contrast, in the Al-Cu system (shown in Fig. 3), it appears that disagreements have persisted for many years, although the two latest assessments reported in 2016^[20,22] do seem to show good mutual agreement. This analysis implicitly assumes that even if a recent study re-uses older assessments, its authors consider it as the current state-of-the-art, so that this data set inherits the “time stamp” of its most recent (re)use.

We can also use our metric to identify clusters of work that report mutually consistent results. Figure 4 and 5

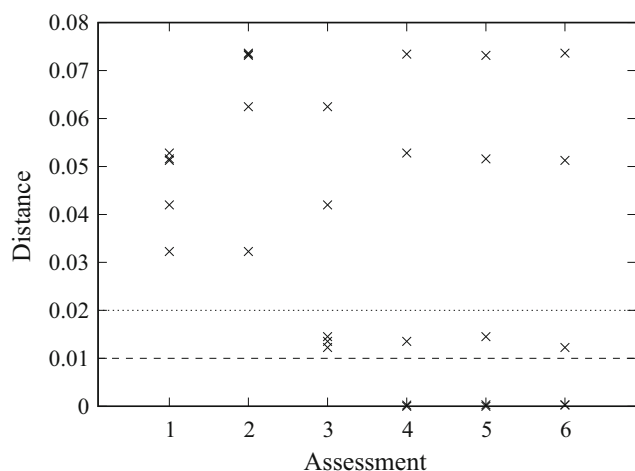


Fig. 4 Identification of “clusters” of assessments of the Fe-Ti system. The graph reports all distances between a given assessment and all other assessments. Assessments numbered 1 through 6 correspond to Ref 3, 4, 7, 9, 13 and 27, respectively. Assessments 4,5,6 are found to lie close to each other, within the 1% similarity threshold indicated a dashed line. A 2% threshold (dotted line) would also include assessment 3 in the cluster

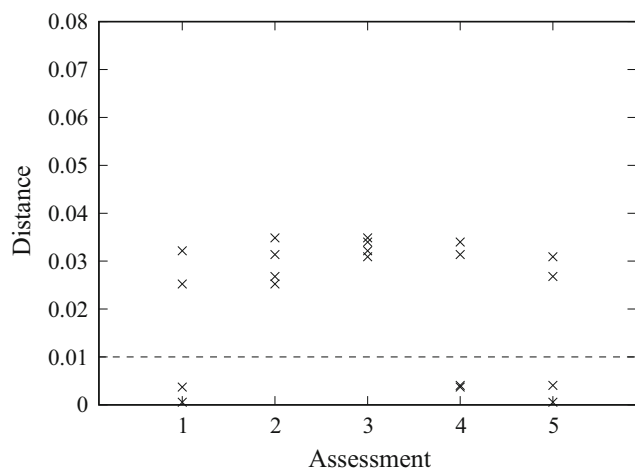


Fig. 5 Identification of “clusters” of assessments of the Al-Cu system. The graph reports all distances between a given assessment and all other assessments. Assessments numbered 1 through 5 correspond to Ref 6, 14, 20, 28 and 22, respectively. Assessments 1,4,5 are found to lie close to each other, within the 1% similarity threshold indicated a dashed line

report all pairwise distances between the assessments. For a given similarity threshold (here 1 or 2%), one can find groups of assessments that lie close to each other, within that threshold. Encouragingly, these clusters seem to primarily consist of recent publications thus again suggesting an emerging consensus.

4 Conclusion

We have described a formal methodology to quantify, in a single figure-of-merit, the level of agreement between two phase diagrams. Our proposal not only satisfies the mathematical requirements of a norm or a metric, but also has a sound physical basis, is invariant to scaling of the graph axes, and is easy to compute via Monte Carlo sampling, with or without access to the thermodynamic model underlying each phase diagram. We illustrate its usefulness in a meta-analysis of a set of thermodynamic assessments in popular alloy systems, in an effort to determine whether the most current assessments have reached a consensus. Our metric may find applications in other areas as well, for instance, to report how well phase diagrams generated purely via ab initio methods agree with the corresponding experiments-based thermodynamic assessments.

Acknowledgments This research was supported by ONR under Grant N00014-17-1-2202, and by Brown University through the use of the facilities at its Center for Computation and Visualization. This work uses the Extreme Science and Engineering Discovery Environment (XSEDE) resource Stampede 2 at the Texas Advanced Computing Center through allocation TG-DMR050013N, which is supported by National Science Foundation Grant Number ACI-1548562. The author would like to thank Kenny Lipkowitz for suggesting this line of research.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

- J.O. Andersson, T. Helander, L. Höglund, P.F. Shi, B. Sundman, Thermo-Calc & DICTRA, Computational Tools for Materials Science. *Calphad* **26**, 273 (2002)
- M. Barnsley, *Fractals Everywhere* (Academic Press Professional Inc., San Diego, 1988)
- H. Bo, L. Duarte, W. Zhu, L. Liu, H. Liu, Z. Jin, C. Leinenbach, Experimental Study and Thermodynamic Assessment of the Cu-Fe-Ti System. *Calphad* **40**, 24–33 (2013)
- H. Bo, J. Wang, L. Duarte, C. Leinebach, L. bin Liu, H. shan Liu, Z. peng Jin, Thermodynamic Re-assessment of Fe-Ti Binary System. *Trans. Nonferrous Met. Soc. China* **22**(9), 2204–2211 (2012)
- C. Campbell, U. Kattner, Z.K. Liu, File and Data Repositories for Next Generation CALPHAD. *Scr. Mater.* **70**, 7 (2014). <https://doi.org/10.1016/j.scriptamat.2013.06.013>
- C. Dai, H. Xu, S. Wang, M. Yin, S. Cui, L. Zhang, Y. Du, F. Zheng, Interdiffusivities and Atomic Mobilities in fcc Cu-Al-Fe Alloys. *Calphad* **35**(4), 556–561 (2011)
- L.F.S. Dumitrescu, M. Hillert, N. Sounders, Comparison of Fe-Ti Assessments. *J. Phase Equilibria* **19**(5), 441–448 (1998)

8. E. Galvan, R.J. Malak, S. Gibbons, R. Arroyave, A Constraint Satisfaction Algorithm for the Generalized Inverse Phase Stability Problem. *J. Mech. Des.* **139**, 011401 (2017)
9. C. Guo, C. Li, X. Zheng, Z. Du, Thermodynamic Modeling of the Fe-Ti-V System. *Calphad* **38**, 155–160 (2012)
10. T. Hickel, U.R. Kattner, S.G. Fries, Computational Thermodynamics: Recent Developments and Future Potential and Prospects. *Phys. Status Solidi B* **251**, 9 (2014)
11. U.R. Kattner, Thermodynamic Modeling of Multicomponent Phase Equilibria. *JOM J. Min. Met. Mat. S.* **49**, 14 (1997)
12. L. Kaufman, Computational Thermodynamics and Materials Design. *Calphad* **25**, 141 (2001)
13. K. Kumar, P. Wollaiits, L. Delaey, Thermodynamic Reassessment and Calculation of Fe-Ti Phase Diagram. *Calphad* **18**(2), 223–234 (1994)
14. S.M. Liang, R. Schmid-Fetzer, Thermodynamic Assessment of the Al-Cu-Zn System, Part II: Al-Cu Binary System. *Calphad* **51**, 252–260 (2015)
15. Z.K. Liu, First-Principles Calculations and CALPHAD Modeling of Thermodynamics. *J. Phase Equilib. Differ.* **30**, 517 (2009)
16. J. Morral, Two-Dimensional Phase Fraction Charts. *Scr. Metall.* **18**, 407 (1984)
17. J. Morral, H. Gupta, A Figure of Merit for Predicted Phase Diagrams. *J. Phase Equilib.* **13**, 373 (1992)
18. A.W. Naylor, G.R. Sell, *Linear Operator Theory in Engineering and Science* (Springer, Berlin, 1982)
19. R. Otis, M. Emelianenko, Z.K. Liu, An Improved Sampling Strategy for Global Energy Minimization of Multi-Component Systems. *Comput. Mater. Sci.* **130**, 282 (2017)
20. Y. Sun, H. Liu, Z. Xie, Z. Jin, Prediction of Interfacial Reaction Products Between Metals with Same Lattice Structure Through Thermodynamic Modeling. *Calphad* **52**, 180–185 (2016)
21. B. Sundman, U.R. Kattner, M. Palumbo, S.G. Fries, OpenCalphad—A Free Thermodynamic Software. *Integr. Mater. Manuf. Innov.* **4**, 1 (2015)
22. B. Sundman, U.R. Kattner, C. Sigli, M. Stratmann, R.L. Tellier, M. Palumbo, S.G. Fries, The OpenCalphad Thermodynamic Software Interface. *Comput. Mater. Sci.* **125**, 188 (2016)
23. A. van de Walle, Multicomponent Multisublattice Alloys, Non-configurational Entropy and Other Additions to the Alloy Theoretic Automated Toolkit. *Calphad* **33**, 266–278 (2009). <https://doi.org/10.1016/j.calphad.2008.12.005>
24. A. van de Walle, Methods for First-Principles Alloy Thermodynamics. *JOM J. Min. Met. Mat. S.* **65**, 1523–1532 (2013). <https://doi.org/10.1007/s11837-013-0764-3>
25. A. van de Walle, M.D. Asta, G. Ceder, The Alloy Theoretic Automated Toolkit: A user guide. *Calphad* **26**, 539–553 (2002). [https://doi.org/10.1016/S0364-5916\(02\)80006-2](https://doi.org/10.1016/S0364-5916(02)80006-2)
26. A. van de Walle, C. Nataraj, Z.K. Liu, The Thermodynamic Database. *Calphad* **61**, 173 (2018). <https://doi.org/10.1016/j.calphad.2018.04.003>
27. S. Wang, K. Wang, G. Chen, Z. Li, Z. Qin, X. Lu, C. Li, Thermodynamic Modeling of Ti-Fe-Cr Ternary System. *Calphad* **56**, 160–168 (2017)
28. V. Witusiewicz, U. Hecht, S. Fries, S. Rex, The Ag-Al-Cu System Part I: Reassessment of the Constituent Binaries on the Basis of New Experimental Data. *J. Alloys Compd.* **385**(1–2), 133–143 (2004)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.