

ТЕПЛОЕМКОСТЬ: ЭКСПЕРИМЕНТ И РАСЧЕТ

УДК 541.11

ПРОГРАММА SRFIT ДЛЯ АППРОКСИМАЦИИ ТЕПЛОЕМКОСТЕЙ И ТЕПЛОСОДЕРЖАНИЙ: НОВЫЕ ВОЗМОЖНОСТИ

© 2022 г. А. Л. Восков^{а,*}^аМосковский государственный университет им. М.В. Ломоносова, Химический факультет, Москва, Россия

*e-mail: voskova@my.msu.ru

Поступила в редакцию 20.03.2022 г.

После доработки 23.03.2022 г.

Принята к публикации 25.03.2022 г.

Созданная в 2016 г. в лаборатории химической термодинамики МГУ имени М.В. Ломоносова программа SrFit предназначена для аппроксимации экспериментальных данных по изобарным теплоемкостям и теплосодержаниям индивидуальных веществ моделями CALPHAD третьего поколения, в том числе включающих взвешенные суммы функций Эйнштейна. Она поддерживает аппроксимацию аномалий теплоемкости избыточными вкладами. В программу SrFit также были добавлены поддержка робастной регрессии и возможность задания пользователями произвольных аномальных вкладов в теплоемкость на интерпретируемом языке программирования Lua. SrFit – кросс-платформенное свободное программное обеспечение, распространяемое по лицензии GNU GPL 2.

Ключевые слова: аппроксимация, базы данных, термодинамические функции, теплоемкость, теплосодержание, SrFit, программа

DOI: 10.31857/S0044453722090308

ВВЕДЕНИЕ

Аппроксимация термодинамических функций индивидуальных веществ играет ключевую роль при создании баз данных. Нередко ее проводят полиномами, в том числе в базах данных CALPHAD 1 и 2 поколения, например, в SGTE Unary Database с термодинамическими функциями простых веществ [1]. Альтернативный подход – применение моделей CALPHAD третьего поколения, в которых наряду с полиномами используются функции Эйнштейна и/или Дебая [2]. Это позволяет аппроксимировать экспериментальные данные по изобарным теплоемкостям и теплосодержаниям в широком интервале температур вплоть до 0 К. К сожалению, программные пакеты для регрессионного анализа обычно не содержат их готовой реализации. Для восполнения этого пробела в 2016 г. в лаборатории химической термодинамики МГУ имени М.В. Ломоносова была создана программа SrFit [3]. Она стала активно использоваться для аппроксимации термодинамических свойств индивидуальных веществ. А за последние 5 лет в нее был добавлен ряд новых возможностей, краткому обзору которых и посвящена данная работа.

МОДЕЛИ И АЛГОРИТМЫ

Программа SrFit предназначена для аппроксимации экспериментальных данных по изобарной теплоемкости, теплосодержанию индивидуальных веществ моделями CALPHAD третьего поколения. Используемая в ней модель изобарной теплоемкости имеет следующий вид:

$$C_p(T) = \sum_{i=1}^{n_E} \alpha_i C_E \left(\frac{\theta_i}{T} \right) + R \sum_{i=1}^{n_p} a_i \left(\frac{T}{T_0} \right)^{p_i} + \sum_{k=1}^{n_{ex}} C_{p,k}^{ex}(\bar{b}_k, T), \quad \frac{C_E(x)}{R} = \frac{3x^2 e^x}{(e^x - 1)^2}, \quad (1)$$

где α_i , θ_i , a_i , p_i , \bar{b}_k , – параметры модели, $C_E(x)$ – функция Эйнштейна, R – универсальная газовая постоянная, $T_0 = 298.15$ К, n_E , n_p и n_{ex} – число функций Эйнштейна, степенных функций и аномальных вкладов соответственно. Первое слагаемое – взвешенная сумма функций Эйнштейна, ее использование независимо предложено в 2013 г. Ворониным и Куценком [4] и Jacobs и др. [5]. Второе слагаемое уравнения (1) – полиномиальная часть, добавленная в эту модель Bigdeli и др. [6]. Третье слагаемое – аномальные вклады в теплоемкость. Список встроенных в SrFit аномальных (избыточных) вкладов приведен в табл. 1. Также есть возможность (добавлена в SrFit 0.8 в 2018 г.) писать собственные выражения для C_p^{ex} на интер-

Таблица 1. Избыточные (аномальные) вклады в теплоемкость, реализованные в программе CrFit

Имя	Функция $C_p^{ex}(T)/R$	Вид пиков
Default	$b_1 \exp[b_2(b_3\Delta T - \Delta T)]$ $\Delta T = T - T_{tr}$	λ – образные симметричные с максимумом в T_{tr}
Empty	0	Исключение аномалий теплоемкости из оптимизации
Left exp.	$b_1 \exp[b_2(T - T_{max})]$	Левая часть асимметричного λ -образного пика
Left exp.2	$\sum_{i=1}^2 b_1^{(i)} \exp[b_2^{(i)}(T - T_{max})]$	Левая часть асимметричного λ -образного пика
Right exp.	$b_1 \exp[-b_2(T - T_{min})]$	Правая часть асимметричного λ -образного пика
Right exp.2	$\sum_{i=1}^2 b_1^{(i)} \exp[-b_2^{(i)}(T - T_{min})]$	Правая часть асимметричного λ -образного пика
Skewed	$\frac{b_1 e^{-x^2}}{1 + e^{-b_2 x}}; x = \frac{T - b_3}{b_4}$	Асимметричный колоколообразный пик
Two peaks	$\sum_{i=1}^2 b_1^{(i)} e^{b_2^{(i)}(b_3^{(i)}\Delta T_i - \Delta T_i)}$ $\Delta T_i = T - T_{tr}^{(i)}$	Частично перекрывающиеся λ -образные симметричные с максимумами в $T_{tr}^{(i)}$

претируемом языке программирования Lua. Выражения для теплосодержания и энтропии могут быть получены интегрированием уравнения (1).

Избыточные (аномальные) вклады в теплоемкость $C_{p,k}^{ex}(\vec{b}_k, T)$ использовались ранее для аппроксимации теплоемкостей UO_2 , ThO_2 [3], ряда цеолитов и других алюмосиликатов [7–9], а также Sr_2NiMoO_6 и Sr_2CoMoO_6 [10].

Для оптимизации параметров моделей в CrFit изначально использовался исключительно метод наименьших квадратов (МНК), но в 2021 г. была добавлена поддержка робастной (устойчивой к промахам) регрессии на основе M-оценок. Используемая в CrFit целевая функция имеет следующий вид:

$$F(\beta) = \sum_{k=1}^n \rho\left(\frac{W_k r_k(\beta)}{\hat{\sigma}}\right), \quad (2)$$

где Y_k – это $C_{p,k}$ или $H(T_{2,k}) - H(T_{1,k})$, β – параметры модели, $r_k = Y_k^{calc} - Y_k^{expt}$ – отклонение расчетного значения (calc) от экспериментального (expt), $\hat{\sigma}$ – оценка стандартной ошибки регрессии, $\rho(t)$ – функция потерь, W_k – статистические веса. В случае минимизации абсолютных погрешностей $W_k = \omega_k$, а при минимизации относительных $W_k = \omega_k / |Y_k^{expt}|$; $\omega_k \geq 0$ – задаваемые пользователем веса, по умолчанию $\omega_k = 1$. Для метода наименьших квадратов используется функция

потерь $\rho(t) = 0.5t^2$, а для метода наименьших модулей – $\rho(t) = |t|$.

Список реализованных в CrFit функций потерь приведен в табл. 2. Из них модуль, а также функции Хьюбера, Эндрюса и Коши являются робастными, т.е. устойчивыми к выбросам. В [9] приводится пример их использования для аппроксимации экспериментальных значений изобарной теплоемкости с промахами и аномалиями без их предварительного ручного отсева.

Для минимизации робастной целевой функции (уравнение (2)) в программе CrFit используется метод наименьших квадратов с итеративным пересчетом весов (англ. IRLS, Iterative Reweighted Least Squares method), независимо разработанный Мудровым, Кушко и др. [11], Schlossmacher [12] и Beaton и Tukey [13]. Он является обобщением метода Гаусса–Ньютона и детально описан в одной из предыдущих работ, посвященной программе CrFit и робастной регрессии [9]. Весовые функции $w(t)$ для IRLS, соответствующие различным функциям потерь $\rho(t)$, также приведены в табл. 2.

Для оценки доверительных интервалов коэффициентов применяются следующие оценки ковариационной матрицы C , основанные на работе Welsch [14]:

$$C = \hat{\sigma}^2 (\hat{J}^T \hat{J})^{-1}; \quad \hat{J}_{ij} = W_i \frac{\partial r_i}{\partial \beta_j}; \quad s(\beta_i) = \sqrt{C_{ii}}, \quad (3)$$

Таблица 2. Реализованные в программе CrFit функции потерь $\rho(t)$, соответствующие им весовые функции $w(t)$ и используемые значения параметра a для достижения 95%-ной асимптотической эффективности

Название	$\rho(t)$	$w(t) = \frac{1}{t} \frac{\partial \rho(t)}{\partial t}$	a
Квадрат	$0.5t^2$	1	—
Модуль	$ t $	$ t ^{-1}$	—
Функция Хьюбера	$\begin{cases} 0.5t^2, & t \leq a \\ a(t - 0.5a), & t > a \end{cases}$	$\begin{cases} 1, & t \leq a \\ t ^{-1}, & t > a \end{cases}$	1.345
Функция Эндрюса	$\begin{cases} a^2 \left(1 - \cos \frac{t}{a}\right), & t \leq a\pi \\ 2a^2, & t > a\pi \end{cases}$	$\begin{cases} \frac{a}{t} \sin \frac{t}{a}, & t \leq a\pi \\ 0, & t > a\pi \end{cases}$	1.339
Функция Коши	$\frac{a^2}{2} \ln \left[1 + \left(\frac{t}{a}\right)^2 \right]$	$\left[1 + \left(\frac{t}{a}\right)^2 \right]^{-1}$	2.385

где β_i – параметры модели, \hat{J} – матрица Якоби с учетом статистических весов, $\hat{\sigma}$ – оценка стандартной ошибки регрессии. В случае метода наименьших квадратов она рассчитывается следующим образом:

$$\hat{\sigma}^2 = \frac{1}{n - m} \sum_{k=1}^n r_k^2, \quad (4)$$

где n – число точек, m – количество параметров модели.

Для прочих M -оценок (метода наименьших модулей, функций Хьюбера, Эндрюса и Коши) используется другая, робастная оценка σ , основанная на медиане абсолютных значений отклонений:

$$\hat{\sigma} = 1.483 \operatorname{median}|r_k|. \quad (5)$$

Подробнее о нормировочном множителе 1.483, в том числе о методе его расчета на основе предположения о нормальном распределении ошибок, рассказывается в предыдущих публикациях о программе CrFit [7, 8].

Для оценки качества моделей в программу CrFit встроено четыре функции. Каждая из них может быть рассчитана для какой-то одной термодинамической величины Y , т.е. для изобарной теплоемкости C_p или теплосодержания $H(T_2) - H(T_1)$:

1. Стандартное абсолютное отклонение $s(\delta Y)$:

$$s(\delta Y) = \sqrt{\frac{1}{n} \sum_{k=1}^n r_k^2}. \quad (6)$$

2. Стандартное относительное отклонение $s(\epsilon Y)$:

$$s(\epsilon Y) = \sqrt{\frac{1}{n} \sum_{k=1}^n \left(\frac{r_k}{Y_k^{\text{expt}}} \right)^2}. \quad (7)$$

3. Робастная оценка стандартного абсолютного отклонения $s_{\text{MAD}}(\delta Y)$ на основе медианы абсолютных значений отклонений:

$$s_{\text{MAD}}(\delta Y) = 1.483 \operatorname{median}|r_k|. \quad (8)$$

4. Робастная оценка стандартного относительного отклонения $s_{\text{MAD}}(\epsilon Y)$ на основе медианы абсолютных значений отклонений:

$$s_{\text{MAD}}(\epsilon Y) = 1.483 \operatorname{median} \left| \frac{r_k}{Y_k^{\text{expt}}} \right|. \quad (9)$$

Эти четыре функции могут быть рассчитаны как для каждой выборки по отдельности, так и для всех данных по C_p и $H(T_2) - H(T_1)$ вместе. Основанные на медиане оценки стандартного отклонения являются робастными, т.е. устойчивыми к выбросам. Поэтому сильные различия между s и s_{MAD} могут указывать на наличие промахов во входных данных.

ПРОГРАММНАЯ РЕАЛИЗАЦИЯ

Как уже было сказано выше, разработка программы CrFit началась в лаборатории химической термодинамики МГУ имени М.В. Ломоносова в 2016 г. Уже первая общедоступная версия (CrFit 0.2) имела графический интерфейс пользователя. Все версии CrFit написаны на языке C++ и являются кроссплатформенными и свободным программным обеспечением, распространяемым по лицензии GNU GPL 2+. Программа до-

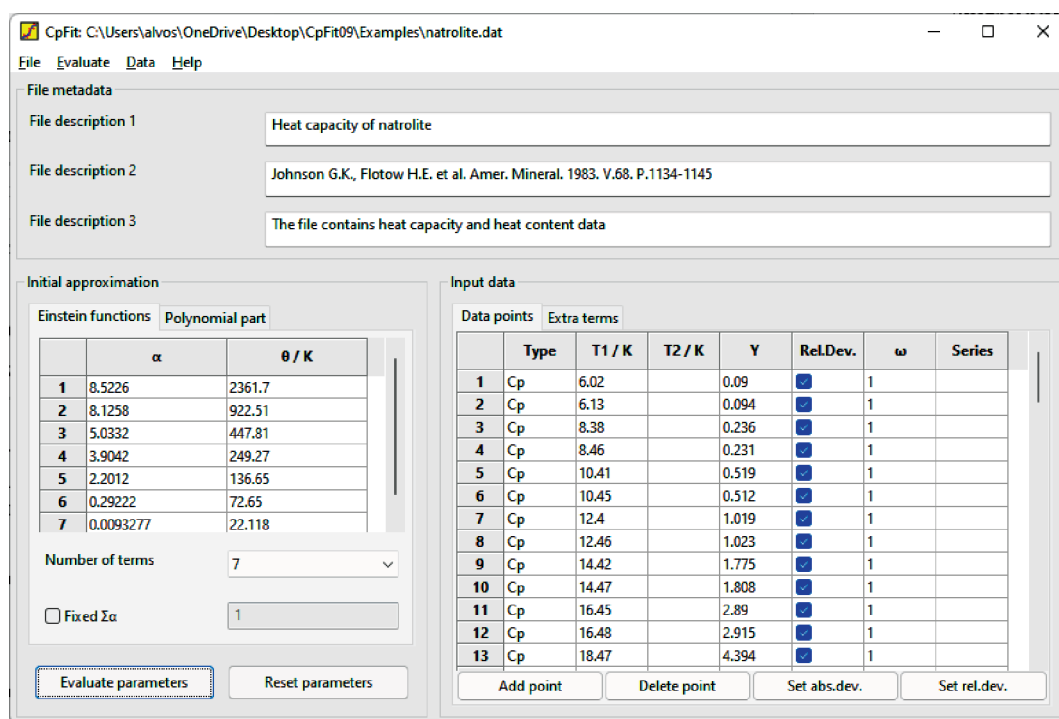


Рис. 1. Главное окно программы CpFit.

ступна на сайте лаборатории химической термодинамики <https://td.chem.msu.ru>.

По мере развития программы CpFit в нее добавлялся следующий функционал:

- 2016 г. CpFit 0.3: автоматический подбор количества функций Эйнштейна в уравнении (1) с помощью ступенчатой регрессии (англ. *stepwise regression*), улучшенная визуализация результатов оптимизации.

- 2017 г. CpFit 0.5: добавлена поддержка аппроксимации данных по теплосодержанию $H(T_2) - H(T_1)$ без их предварительного пересчета в C_p , а также нескольких серий экспериментальных данных. CpFit 0.6: поддержка избыточных вкладов в теплоемкость, а также экспорта графиков в формат wmf и скрипты MATLAB (GNU Octave).

- 2018 г. CpFit 0.8: поддержка задания пользовательских вкладов в теплоемкость на языке программирования Lua. Также добавлены робастные оценки для стандартных ошибок аппроксимации, в т.ч. их отдельный расчет для каждой выборки экспериментальных данных.

- 2022 г. CpFit 0.91: добавление поддержки робастной регрессии, а также возможности локализации (например, русификации) интерфейса. Робастная регрессия реализована в виде отдельной динамической библиотеки, вызываемой из ядра CpFit. При необходимости возможно написание пользователем собственных процедур оптимизации параметров моделей.

На рис. 1 показано главное окно программы CpFit, а на рис. 2а – окно с полученными в результате оптимизации параметрами модели. В качестве примера были взяты использованные в [7] экспериментальные данные по C_p и $H(T_2) - H(T_1)$ натролита $\text{Na}_2\text{Al}_2\text{Si}_3\text{O}_{10} \cdot 2\text{H}_2\text{O}$ и полученная на их основе термодинамическая модель. Помимо этих окон в CpFit реализовано построение различных графиков, облегчающих оценку качества проведенной оптимизации. Среди них – графики $C_p - T$, $\Delta H - T$, диаграммы рассеяния для абсолютных и относительных отклонений и т.п. Диаграммы рассеяния ошибок могут быть построены в координатах “нормальной вероятностной бумаги” (график “квантиль-квантиль”, англ. *Q-Q plot*), что может использоваться для глазомерной проверки на нормальность.

Помимо графического интерфейса у программы CpFit также есть и текстовый (консольный) пользовательский интерфейс. В обоих случаях для файлового ввода/вывода и оптимизации параметров моделей используются одно и то же ядро. Оно использует следующие библиотеки:

- *levmar* – реализация метода наименьших квадратов. Начиная с версии CpFit 0.91, вместо нее может быть загружена сторонняя библиотека, в том числе входящая в дистрибутив реализация робастной регрессии.

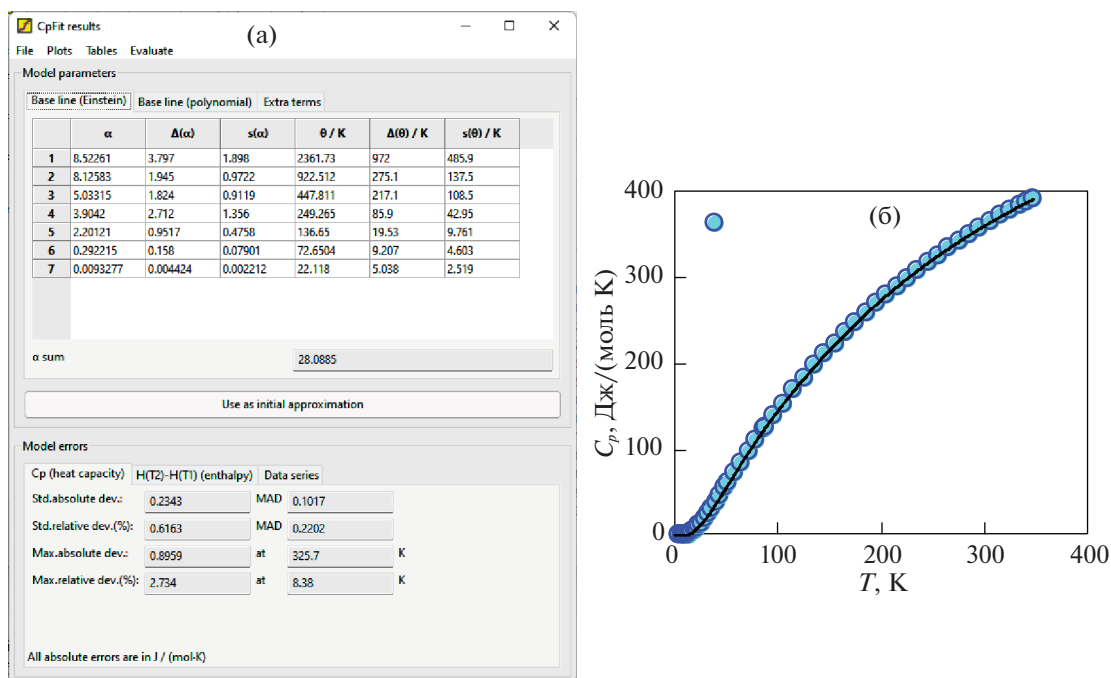


Рис. 2. Результаты оптимизации параметров моделей (уравнения (1)) в программе CpFit: (а) окно с оптимизированными параметрами; (б) модельная и экспериментальная изобарная теплоемкость. Использованы термодинамические функции натролита $\text{Na}_2\text{Al}_2\text{Si}_3\text{O}_{10} \cdot 2\text{H}_2\text{O}$ из работы [7].

- Lua – интерпретируемый язык для написания пользователями собственных аномальных вкладов в теплоемкости.

- wxWidgets – библиотека для создания графических интерфейсов. Используется также для файлового ввода/вывода.

Эти библиотеки являются свободным программным обеспечением (СПО), что облегчает создание и распространение модифицированных версий CpFit ее пользователями и сторонними разработчиками. Для ее компиляции из исходных текстов также достаточно СПО, а именно компилятора gcc и системы сборки CMake.

ИСПОЛЬЗОВАНИЕ ПРОГРАММЫ CpFit

В лаборатории химической термодинамики МГУ программа CpFit использовалась и используется для аппроксимации теплоемкостей ряда индивидуальных веществ: оксидов актиноидов UO_2 и ThO_2 [3], цеолитов [3, 7, 9] и иных алюмосиликатов [8]. Совместно с МИСиС и другими организациями были выполнены работы по аппроксимации термодинамических функций некоторых простых веществ моделями CALPHAD третьего поколения. Среди них – свинец [15], аллотропные модификации олова [16] и золото [17]. В [15–17] программа CpFit использовалась в основном для аппроксимации низкотемпературной

изобарной теплоемкости C_p при $T = 0–300$ К и расчета стандартной энтропии $S_{298.15}^\circ$.

Программа CpFit активно используется и за пределами МГУ. Примеры ее применения для аппроксимации термодинамических функций индивидуальных веществ:

- Сложные оксиды $\text{Sr}_2\text{NiMoO}_6$, $\text{Sr}_2\text{CoMoO}_6$, совместная аппроксимация C_p и $H(T_2) - H(T_1)$ при $T = 2–1273$ К [10]. В работе использованы возможности программы CpFit для аппроксимации аномалий теплоемкости.

- Протон-проводящие перовскиты $\text{BaCa}_{(1+y)/3}\text{Nb}_{(2-y)/3}\text{O}_{3-\delta}$, аппроксимация теплосодержаний при $T = 295–1273$ К [18].

- Орто tantalаты лантаноидов LnTaO_4 : аппроксимация теплоемкости при $T = 0–350$ К [19]. Сложные оксиды лантаноидов, тантала и железа $\text{Ln}_2\text{FeTaO}_7$ ($\text{Ln} = \text{Sm}, \text{Gd}$) [20]: использование встроенных в CpFit моделей для экстраполяции неаномальной части теплоемкости в область низких температур.

- Интерметаллид UZr_2 , аппроксимация C_p при $T = 2–270$ К [21].

- Диоксид кремния SiO_2 (несколько полиморфных модификаций), аппроксимация изобарной теплоемкости при $T = 0–300$ К с целью расчета стандартной энтропии $S_{298.15}^\circ$ [22].

• Кристаллическая и стекловидная модификации оксид бора B_2O_3 [23], аппроксимация C_p и расчет $S_{298.15}^\circ$ [23].

ЗАКЛЮЧЕНИЕ

Программа CpFit позволяет аппроксимировать изобарные теплоемкости и теплосодержания индивидуальных веществ моделями CALPHAD третьего поколения в широком интервале температур. В ходе ее развития в программу был добавлен ряд новых возможностей: поддержка полиномиальных вкладов, улучшенная работа с выборками из разных источников, задание пользователем моделей аномальной теплоемкости на языке Lua. Также расширились средства визуализации, экспорта и импорта данных, графиков и результатов оптимизации. В текущую версию (CpFit 0.91) была добавлена поддержка использования сторонних библиотек для оптимизации параметров моделей. Это сделало возможным добавление поддержки робастной регрессии, основанной на М-оценках: метода наименьших модулей, функций Хьюбера, Эндрюса, Коши.

Работа выполнена при поддержке Российского фонда фундаментальных исследований (код проекта № 20-03-00575) и частично в рамках темы “Химическая термодинамика” (121031300039-1).

СПИСОК ЛИТЕРАТУРЫ

1. *Dinsdale A.T.* // *Calphad*. 1991. V. 15. № 4. P. 317. [https://doi.org/10.1016/0364-5916\(91\)90030-N](https://doi.org/10.1016/0364-5916(91)90030-N)
2. *Chase M.W., Ansara I., Dinsdale A. et al.* // *Calphad*. 1995. V. 19. № 4. P. 437. [https://doi.org/10.1016/0364-5916\(96\)00002-8](https://doi.org/10.1016/0364-5916(96)00002-8)
3. *Voskov A.L., Kutsenok I.B., Voronin G.F.* // *Ibid*. 2018. V. 61. P. 50. <https://doi.org/10.1016/j.calphad.2018.02.001>
4. *Voronin G.F., Kutsenok I.B.* // *J. Chem. Eng. Data*. 2013. V. 58. № 7. P. 2083. <https://doi.org/10.1021/je400316m>
5. *Jacobs M.H.G., Schmid-Fetzer R., van den Berg A.P.* // *Phys. Chem. Minerals*. 2013. V. 40. № 3. P. 207. <https://doi.org/10.1007/s00269-012-0562-4>
6. *Bigdeli S., Chen Q., Selleby M.* // *J. Phase Equilib. Dif. fus.* 2018. V. 39. № 6. P. 832. <https://doi.org/10.1007/s11669-018-0679-3>
7. *Voskov A.L., Voronin G.F., Kutsenok I.B., Kozin N.Yu.* // *Calphad*. 2019. V. 66. ID 101623. <https://doi.org/10.1016/j.calphad.2019.04.008>
8. *Восков А.Л.* // *Журн. неорганической химии*. 2020. № 5. С. 694. <https://doi.org/10.31857/S0044457X20050268>
9. *Voskov A.L.* // *CEUR Workshop Proceedings*. 2021. V. 3036. P. 278.
10. *Sereda V.V., Tsvetkov D.S., Sednev A.L. et al.* // *Phys. Chem. Chem. Phys.* 2018. V. 20. № 30. P. 20108–20116. <https://doi.org/10.1039/c8cp03782e>
11. *Мудров В.И., Кушко В.Л., Михайлов В.И., Осовицкий Е.М.* // *Космические исследования*. 1968. Т. 6. № 4. С. 502.
12. *Schlossmacher E. J.* // *J. Am. Stat. Assoc.* 1973. V. 68. № 344. P. 857. <https://doi.org/10.1080/01621459.1973.10481436>
13. *Beaton A.E., Tukey J.W.* // *Technometrics*. 1974. V. 16. № 2. P. 147. <https://doi.org/10.1080/00401706.1974.10489171>
14. *Welsch R.E.* *Confidence Regions for Robust Regression* // *NBER Working Paper Series*. 1975. № 111. <https://doi.org/10.3386/w0111>
15. *Khvan A.V., Dinsdale A.T., Uspenskaya I.A. et al.* // *Calphad*. 2018. V. 60. P. 144–155. <https://doi.org/10.1016/j.calphad.2017.12.008>
16. *Khvan A.V., Babkina T., Dinsdale I.A. et al.* // *Ibid*. 2019. V. 65. P. 50–72. <https://doi.org/10.1016/j.calphad.2019.02.003>
17. *Khvan A.V., Uspenskaya I.A., Aristova N.M. et al.* // *Ibid*. 2020. V. 68. ID 101724. <https://doi.org/10.1016/j.calphad.2019.101724>
18. *Sereda V.V., Sednev-Lugovets A.L., Malyshkin D.A. et al.* // *J. Therm. Anal. Calorim.* 2020. V. 142. № 5. P. 1989. <https://doi.org/10.1007/s10973-020-09852-z>
19. *Guskov V.N., Gavrichev K.S.* // *Russ. J. Inorg. Chem.* 2021. V. 66. № 13. P. 1947. <https://doi.org/10.1134/S0036023621130088>
20. *Egorysheva A.V., Ellert O.G., Popova E.F. et al.* // *J. Chem. Thermodyn.* 2021. V. 161. ID 106565. <https://doi.org/10.1016/j.jct.2021.106565>
21. *Gyasi S., Benigni P., Rogez J., Barrachin M.* // *J. Nucl. Mater.* 2022. V. 562. ID 153580. <https://doi.org/10.1016/j.jnucmat.2022.153580>
22. *Bajenova I., Khvan A., Dinsdale A., Kondratiev A.* // *Calphad*. 2020. V. 68. ID 101716. <https://doi.org/10.1016/j.calphad.2019.101716>
23. *Benigni P.* // *Calphad*. 2021. V. 72. ID 102238. <https://doi.org/10.1016/j.calphad.2020.102238>