

Статистический подход к автоматическому взвешиванию данных по теплоемкости

Подготовил:
асп. 1 г/о Иванов Андрей

Проблема: если обрабатываемые данные взвешивать исключительно **вручную**, то большее качество описания экспериментальных данных требует большого объема знаний и **опыта (понимание методов и условий эксперимента)** исследователя, что обуславливают большую **субъективность** и **трату времени**.

Цель: обеспечить исследователя инструментом, **облегчающим** на начальном этапе ознакомления с данными **отсеивание сомнительных и малонадежных** данных, т.е. служащим **индикатором** необходимости дополнительной проверки качества данных (**чистота исследуемых образцов, ошибки оцифровки данных и т.п., что нельзя учесть автоматически**)

Объекты исследования:

Ni (ГЦК), Mg (ГПУ), Fe (ОЦК)

Решеточный

$$C_p^{Ein}(T, \theta_E) = 3R \left(\frac{\theta_E}{T} \right)^2 \frac{e^{\theta_E/T}}{(e^{\theta_E/T} - 1)^2},$$

$$C_p^{Deb}(T, \theta_D) = 9R \left(\frac{T}{\theta_D} \right)^3 \int_0^{\theta_D/T} \frac{x^4 e^x}{(e^x - 1)^2} dx,$$

Магнитный (если есть)

$$C_p^{magn}(T, \vec{\eta}^m) = RT \cdot g(\tau, p_m) \ln(\beta + 1),$$

$$g(\tau, p_m) = \begin{cases} \frac{0.63570895}{A} \left(\frac{1}{p_m} - 1 \right) \left(2\tau^3 + 2\frac{\tau^9}{3} + 2\frac{\tau^{15}}{5} + 2\frac{\tau^{21}}{7} \right), & \tau \leq 1 \\ \frac{1}{A} \left(2\tau^{-7} + 2\frac{\tau^{-21}}{3} + 2\frac{\tau^{-35}}{5} + 2\frac{\tau^{-49}}{7} \right), & \tau \geq 1 \end{cases},$$

where

$$A = 0.33471979 + 0.49649686 \left(\frac{1}{p_m} - 1 \right),$$

Ангармонический и другие

ИЛИ $C_p^{bcm}(T, a, b) = aT + bT^2$

ИЛИ $C_p^{bcm}(T, a, b) = aT + bT^4$

ИЛИ $C_p^{bcm}(T, \kappa_1, \kappa_2, \alpha, \gamma) = \begin{cases} \kappa_1 T, & T < \alpha - \gamma \\ \kappa_1 T + \frac{\kappa_2 (T - \alpha + \gamma)^2}{4\gamma}, & \alpha - \gamma \leq T \leq \alpha + \gamma, \\ \kappa_1 T + \kappa_2 T(\alpha - \gamma), & T > \alpha + \gamma \end{cases}$

$$C_p^{RW}(T, \bar{\xi}^{RW}, \bar{\eta}^m) = C_p^{Deb}(T, \theta_D) + aT + bT^2 + C_p^{magn}(T, \bar{\eta}^m),$$

RW = Ringberg Workshop model

$$C_p^{CS}(T, \bar{\xi}^{CS}, \bar{\eta}^m) = C_p^{Deb}(T, \theta_D) + aT + bT^4 + C_p^{magn}(T, \bar{\eta}^m),$$

CS = Chen-Sundman model

$$C_p^{SR}(T, \bar{\xi}^{SR}, \bar{\eta}^m) = C_p^{Deb}(T, \theta_D) + C_p^{bcm}(T, \kappa_1, \kappa_2, \alpha, \gamma) + C_p^{magn}(T, \bar{\eta}^m),$$

SR = Segmented regression model

- В отличие от модели SGTE, указанные три позволяют проводить описание от 0 К.
- Интеграл Дебая может рассчитываться напрямую на языке программирования **R**, иначе для аналитического разложения его нередко в приближении ЛК 2 ФПЭ.

Расчет параметров и GoF

$$y_i = M(x_i, \vec{\xi}) + \epsilon_i,$$

$$\vec{\xi} = \underset{\vec{\xi}' \in \Xi}{\operatorname{argmin}} \sum_{i=1}^n w_i (y_i - M(x_i, \vec{\xi}'))^2,$$

$$AIC = n \ln \sum_{i=1}^n (y_i - M(x_i, \vec{\xi}))^2 + p,$$

$$BIC = n \ln \sum_{i=1}^n (y_i - M(x_i, \vec{\xi}))^2 + p \ln n.$$

$$RSE = \sqrt{\frac{\sum_{i=1}^n (y_i - M(x_i, \vec{\xi}))^2}{n - p - 1}},$$

$$\bar{MSE} = \frac{1}{n} \sum_{i=1}^n (y_i - M(x_i, \vec{\xi}))^2,$$

$$CV_{score} = \frac{1}{k} \sum_{j=1}^k MSE^j. \quad (\text{Случайные и} \\ \text{Равнообъемные } k \\ \text{наборы})$$

Unequal K-Folded Cross-validation

Training set (рабочий набор)

- Оставшиеся $k-1$ наборов
- По нему находят минимум целевой функции

Validation set (проверочный набор)

- 1 исключенный набор
- По нему рассчитывают RSE_j

$$w_j = 1 - \frac{RSE_j - RSE_{min}}{RSE_{max}}; j = 1, \dots, k,$$

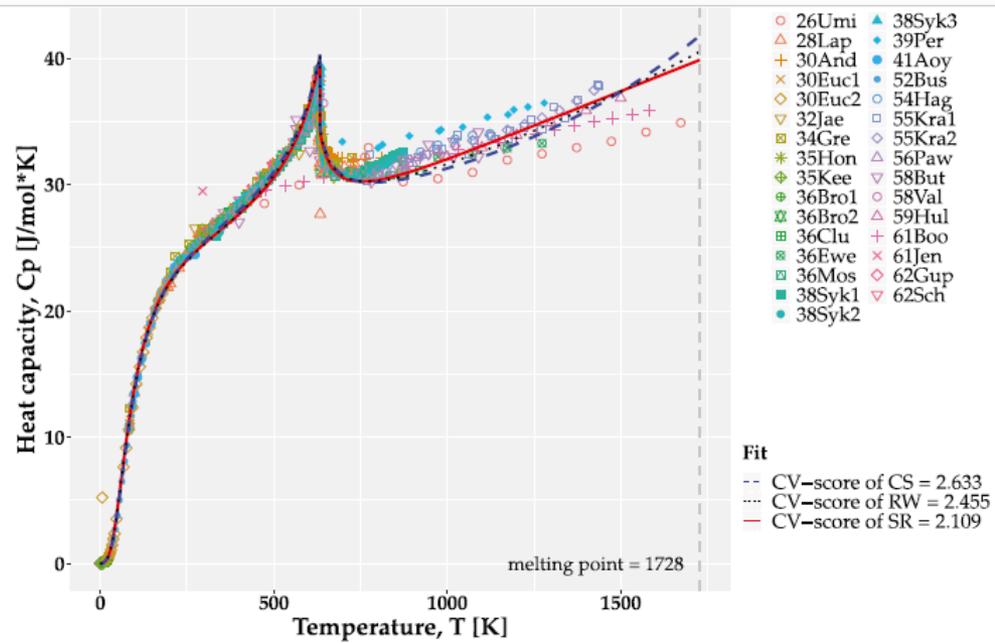


Fig. 1. Experimental points with fitted curves for RW, CS, and SR models including overview of calculated CV_{score} for fcc Ni.

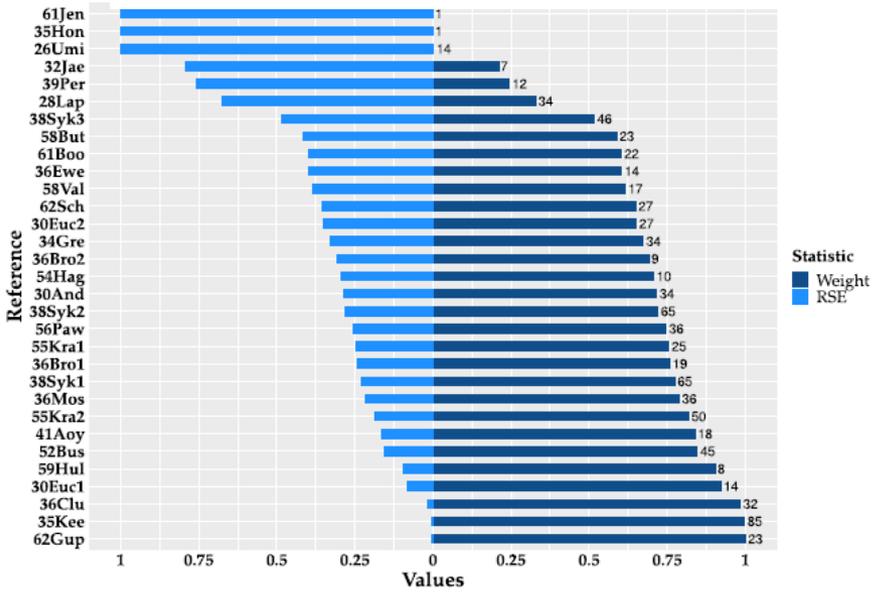
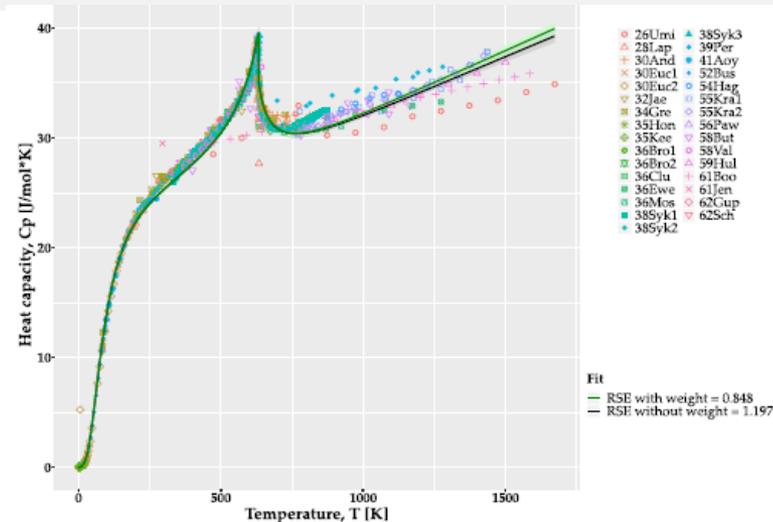
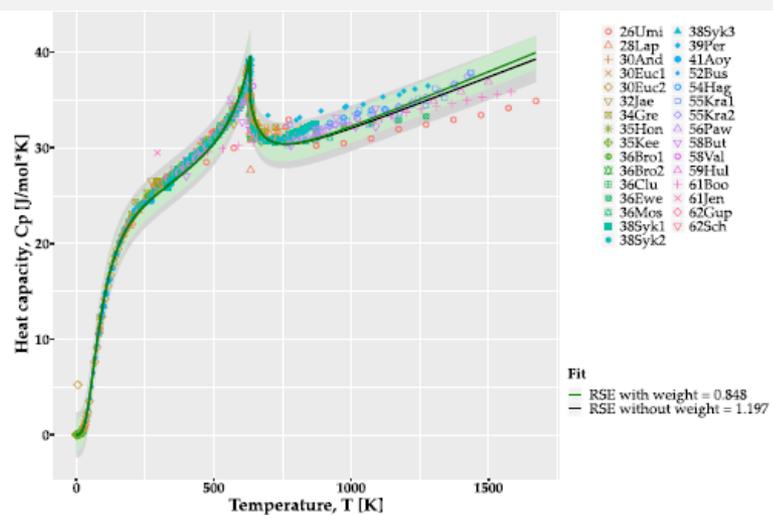
Table 2
The estimated parameters of RW, CS, and SR models with confidence interval (CI) for fcc Ni.

Model	Parameter	Value	2.5% CI	97.5% CI
RW	θ_D, K	369.8	360.2	379.3
	$a, \frac{J}{mol \cdot K^2}$	3.495×10^{-3}	3.062×10^{-3}	3.928×10^{-3}
	$b, \frac{J}{mol \cdot K^3}$	3.219×10^{-6}	2.765×10^{-6}	3.672×10^{-6}
CS	θ_D, K	381.0	371.3	390.8
	$a, \frac{J}{mol \cdot K^2}$	5.553×10^{-3}	5.339×10^{-3}	5.766×10^{-3}
	$b, \frac{J}{mol \cdot K^3}$	8.204×10^{-13}	6.739×10^{-13}	9.670×10^{-13}
SR	θ_D, K	371.3	361.7	380.8
	$\kappa_1, \frac{J}{mol \cdot K^2}$	4.723×10^{-3}	4.212×10^{-3}	5.233×10^{-3}

Table 4
The estimated parameters of SR model with confidence interval (CI) after weighting for fcc Ni.

Parameter	Value	2.5% CI	97.5% CI
θ_D, K	373.0	365.3	380.7
$\kappa_1, \frac{J}{mol \cdot K^2}$	4.758×10^{-3}	4.345×10^{-3}	5.170×10^{-3}
$\kappa_2, \frac{J}{mol \cdot K^3}$	6.790×10^{-3}	5.887×10^{-3}	7.692×10^{-3}
α, K	626.5	568.6	684.4
γ, K	150.0	5.3	294.7

Отработка методики на Ni



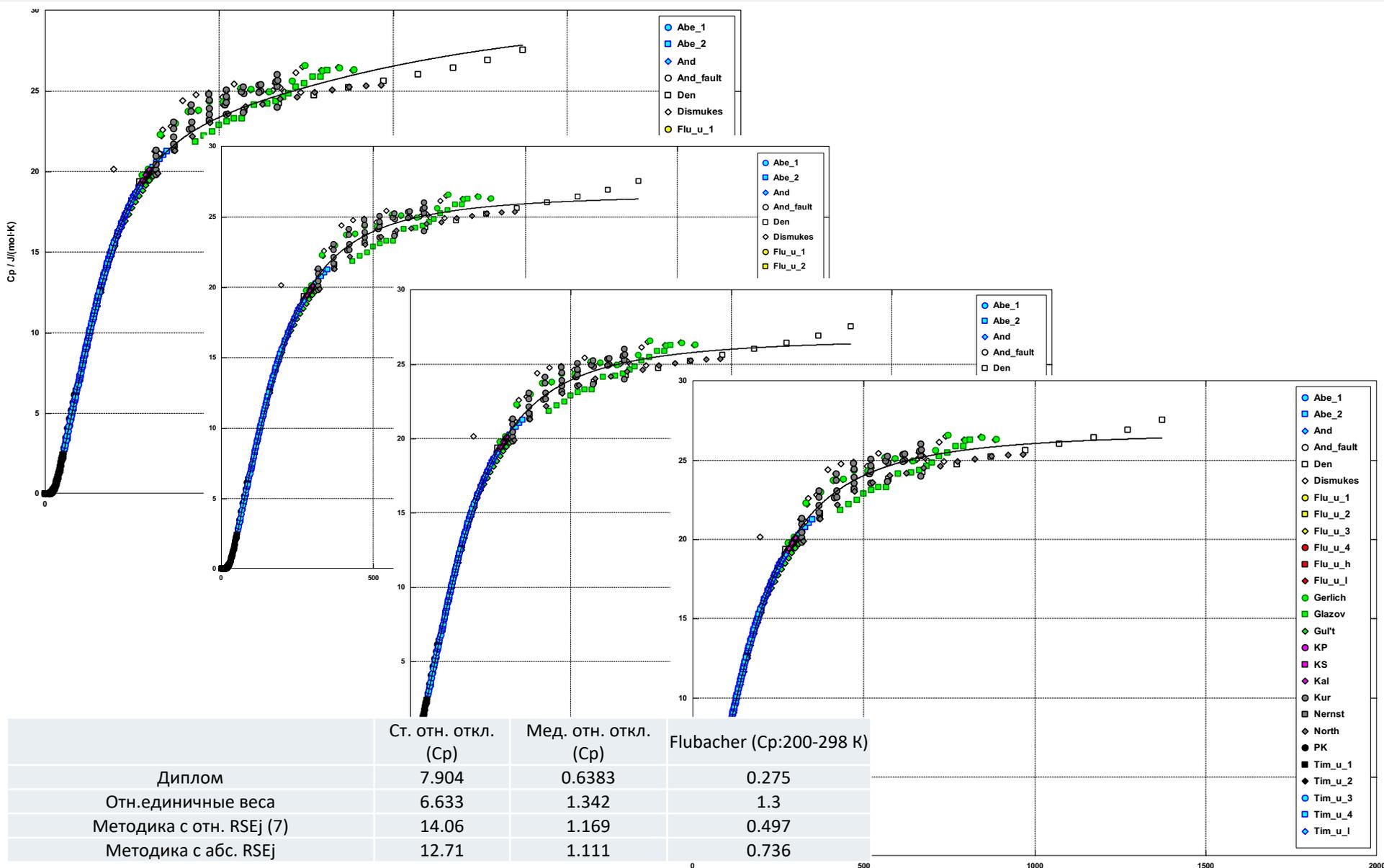
До:

Model	Parameter	Value	2.5% CI	97.5% CI
RW	θ_D, K	369.8	360.2	379.3
	$a, \frac{J}{mol \cdot K^2}$	3.495×10^{-3}	3.062×10^{-3}	3.928×10^{-3}
	$b, \frac{J}{mol \cdot K^3}$	3.219×10^{-6}	2.765×10^{-6}	3.672×10^{-6}
CS	θ_D, K	381.0	371.3	390.8
	$a, \frac{J}{mol \cdot K^2}$	5.553×10^{-3}	5.339×10^{-3}	5.766×10^{-3}
	$b, \frac{J}{mol \cdot K^3}$	8.204×10^{-13}	6.739×10^{-13}	9.670×10^{-13}
SR	θ_D, K	371.3	361.7	380.8
	$\kappa_1, \frac{J}{mol \cdot K^2}$	4.723×10^{-3}	4.212×10^{-3}	5.233×10^{-3}
	$\kappa_2, \frac{J}{mol \cdot K^3}$	6.107×10^{-3}	5.123×10^{-3}	7.091×10^{-3}
	α, K	610.9	538.1	683.6
	γ, K	150.0	-35.6	335.6

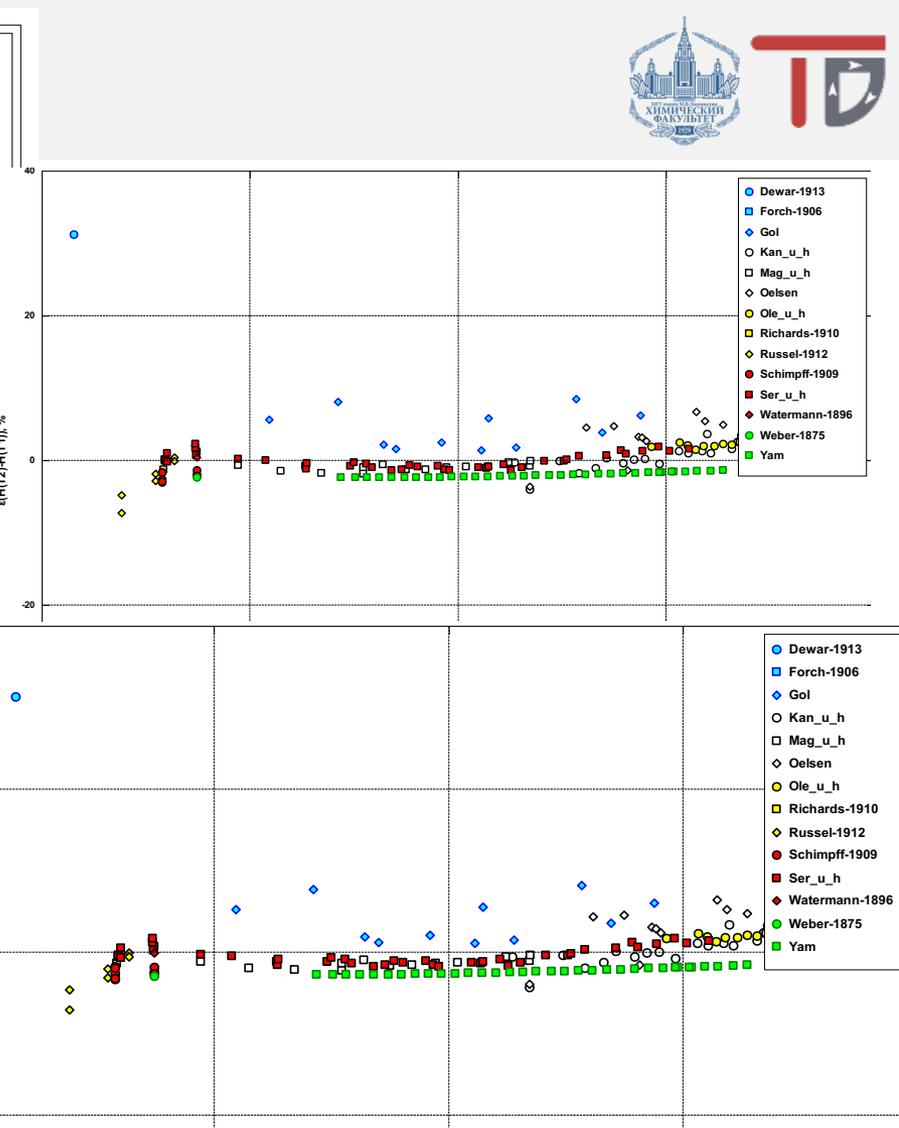
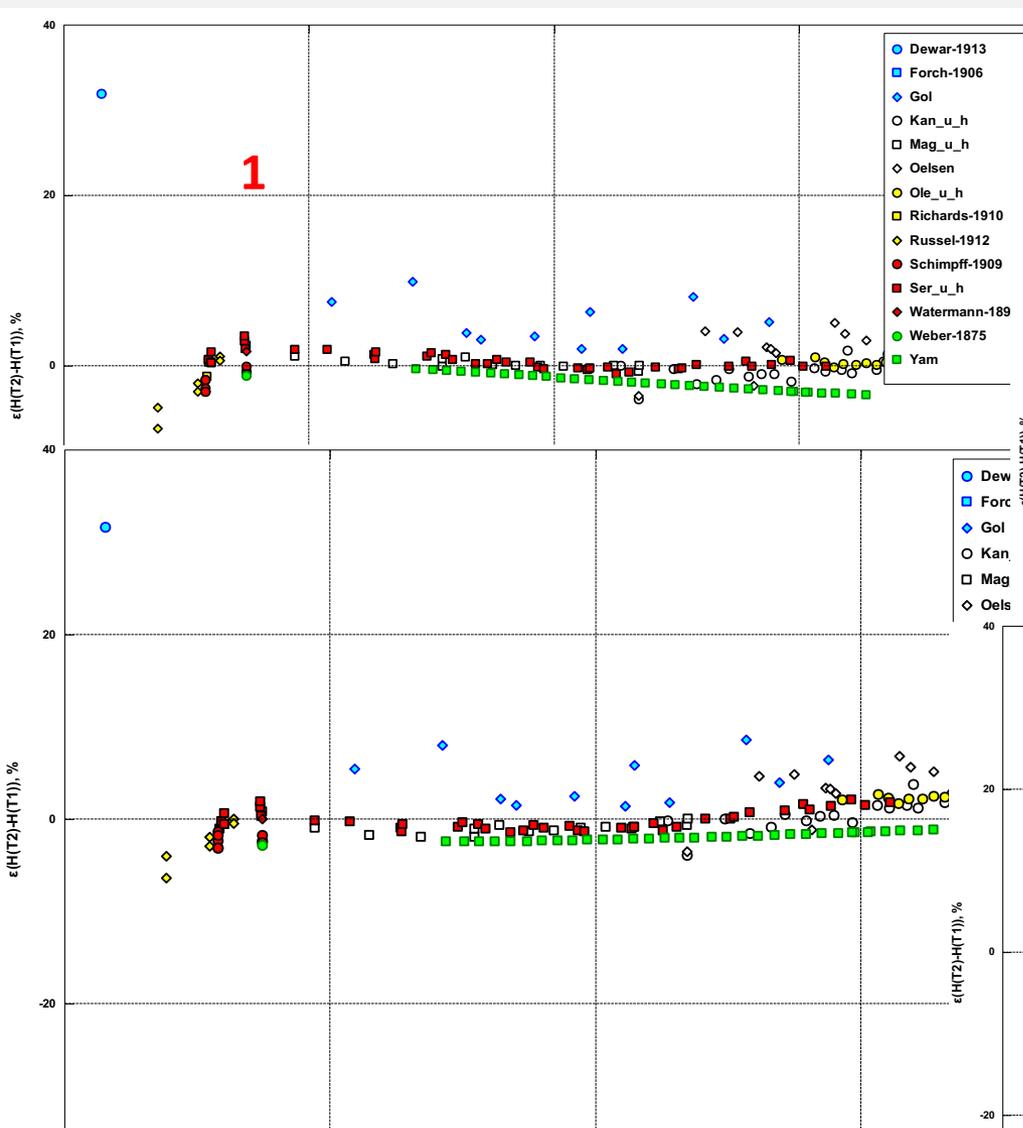
После:

Parameter	Value	2.5% CI	97.5% CI
θ_D, K	373.0	365.3	380.7
$\kappa_1, \frac{J}{mol \cdot K^2}$	4.758×10^{-3}	4.345×10^{-3}	5.170×10^{-3}
$\kappa_2, \frac{J}{mol \cdot K^3}$	6.790×10^{-3}	5.887×10^{-3}	7.692×10^{-3}
α, K	626.5	568.6	684.4
γ, K	150.0	5.3	294.7

Проверка на Si



	Ст. отн. откл. (Cp)	Мед. отн. откл. (Cp)	Flubacher (Cp:200-298 K)
Диплом	7.904	0.6383	0.275
Отн.единичные веса	6.633	1.342	1.3
Методика с отн. RSEj (7)	14.06	1.169	0.497
Методика с абс. RSEj	12.71	1.111	0.736



	Ст. отн. откл. (H-H0)	Мед. отн. откл. (H-H0)	Olette (H-H0:1500- 1700 K)
Диплом	4.478	1.146	0.493
Отн.единичные веса	4.482	1.55	2.35
Методика с отн. RSEj (7)	4.476	1.603	2.17
Методика с абс. RSEj	4.505	1.728	1.95

Преимущества и недостатки

Преимущества	Недостатки
Автоматически выделяет некоторые выбивающиеся "высотемпературные работы" (в предположении, что по крайней мере одна из двух разновыбивающихся работ - промах)	
Снизить влияние таких выбивающихся работ за счёт применения к ним меньшего веса	
Улучшать описание, с их доверительным и предсказательным интервалом по сравнению с единичными весами	
Автоматически выделяет работы с "недостаточным" числом точек (в предположении, что они являются промахами)	
Автоматически выделяет некоторые другие "выбивающиеся работы"	
Может быть использован для любого типа экспериментальных данных	
	Необходимость экспертной оценки (чистота, грубые ошибки, систематические отклонения приборов) для определения корректности результатов используемого метода
	Сглаженные данные могут восприниматься как надежные, не будучи такими
	Некорректность рассмотрения работ с систематическими погрешностями, и присваивание им весов
	Необходимость экспертной оценки для определения корректности результатов используемого метода