

**РЕГРЕССИОННЫЙ И ИНТЕРВАЛЬНЫЙ
ПОДХОДЫ В ЗАДАЧЕ
ВОССТАНОВЛЕНИЯ ЗАВИСИМОСТИ ПРИ
НЕСОВМЕСТНОЙ
ВЫБОРКЕ ИЗМЕРЕНИЙ
ЭЛЕКТРОХИМИЧЕСКОГО ПРОЦЕССА**

С. И. Кумков

Институт математики и механики им. Н.Н. Красовского
Уральского Отделения Российской Академии Наук,
Екатеринбург, Россия, kumkov@imm.uran.ru

Всероссийский web-семинар по интервальному анализу
Новосибирск, 31 октября 2022

Цель исследования – пример совместного применения
регрессионного подхода (МНК) и методов Интервального
Анализа в задаче восстановления зависимости
электрохимического процесса: ситуация неопределенности
и хаотических выбросов положительного знака
при несовместной выборке измерений

Исследование вязкости расплавленного электролита

– эксперимент дает зашумленную выборку измерений; выборка измерений имеет структуру

$$v(T_n) = v_n = v^*(T_n) + e_n + \chi_n, n = \overline{1, N}, \quad (1)$$

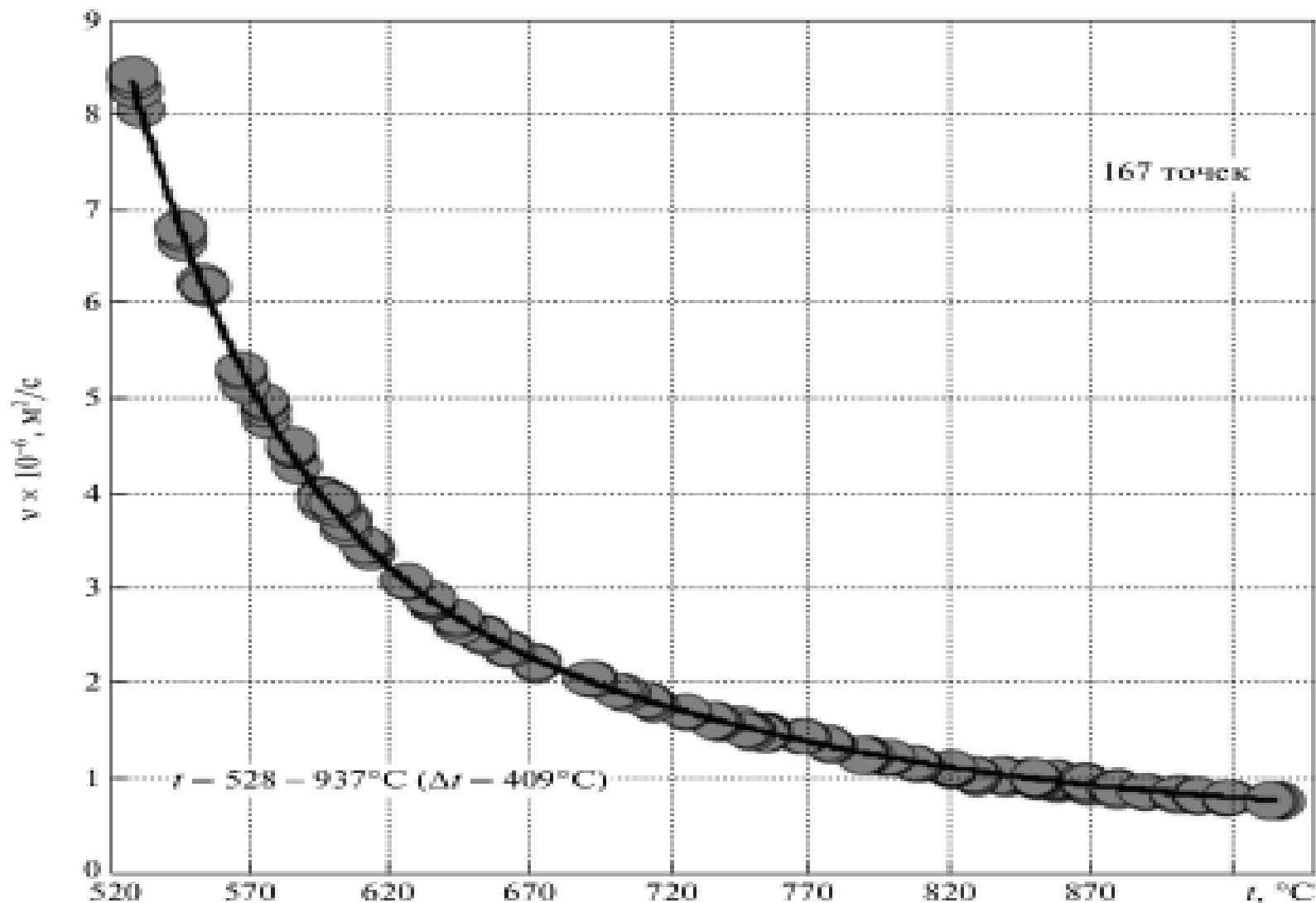
где N — объем выборки; v_n — измерение вязкости ($\times 10^{-6}$ м²/сек) при температуре T_n (градусы Цельсия); $v^*(T)$ — измеряемое истинное значение; e_n — погрешность измерения ($\times 10^{-6}$ м²/сек); χ_n — хаотическая компонента погрешности измерения ($\times 10^{-6}$ м²/сек); **специфика эксперимента:** зашумление $\chi_n > 0$ всегда положительно;

– истинный процесс надежно описывается функцией

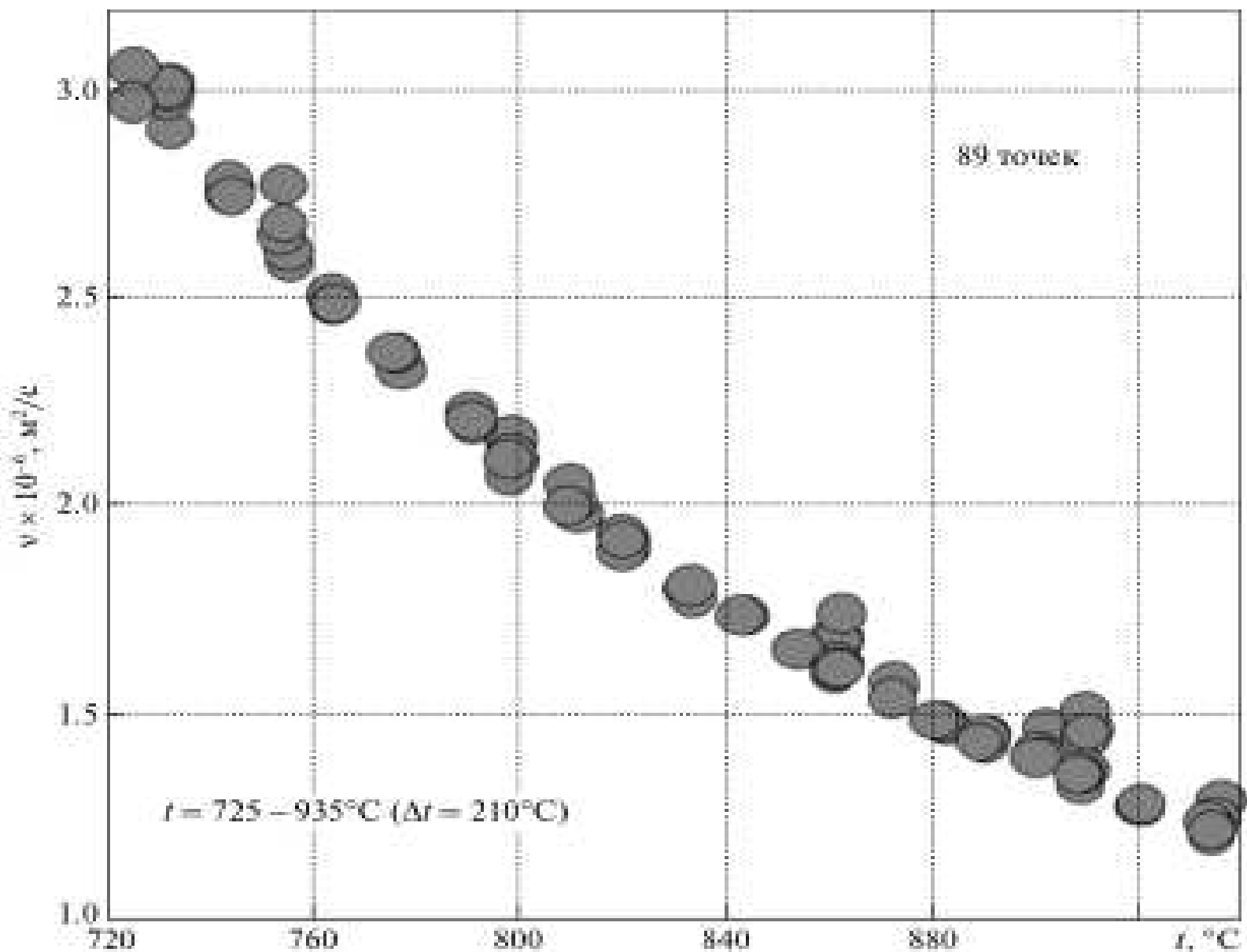
$$v^*(T) = v_0 \exp(E_A/RT). \quad (2)$$

Требуется: построить информационное множество $\Upsilon(v_0, E_A)$ параметров v_0, E_A и коридор совместных зависимостей.

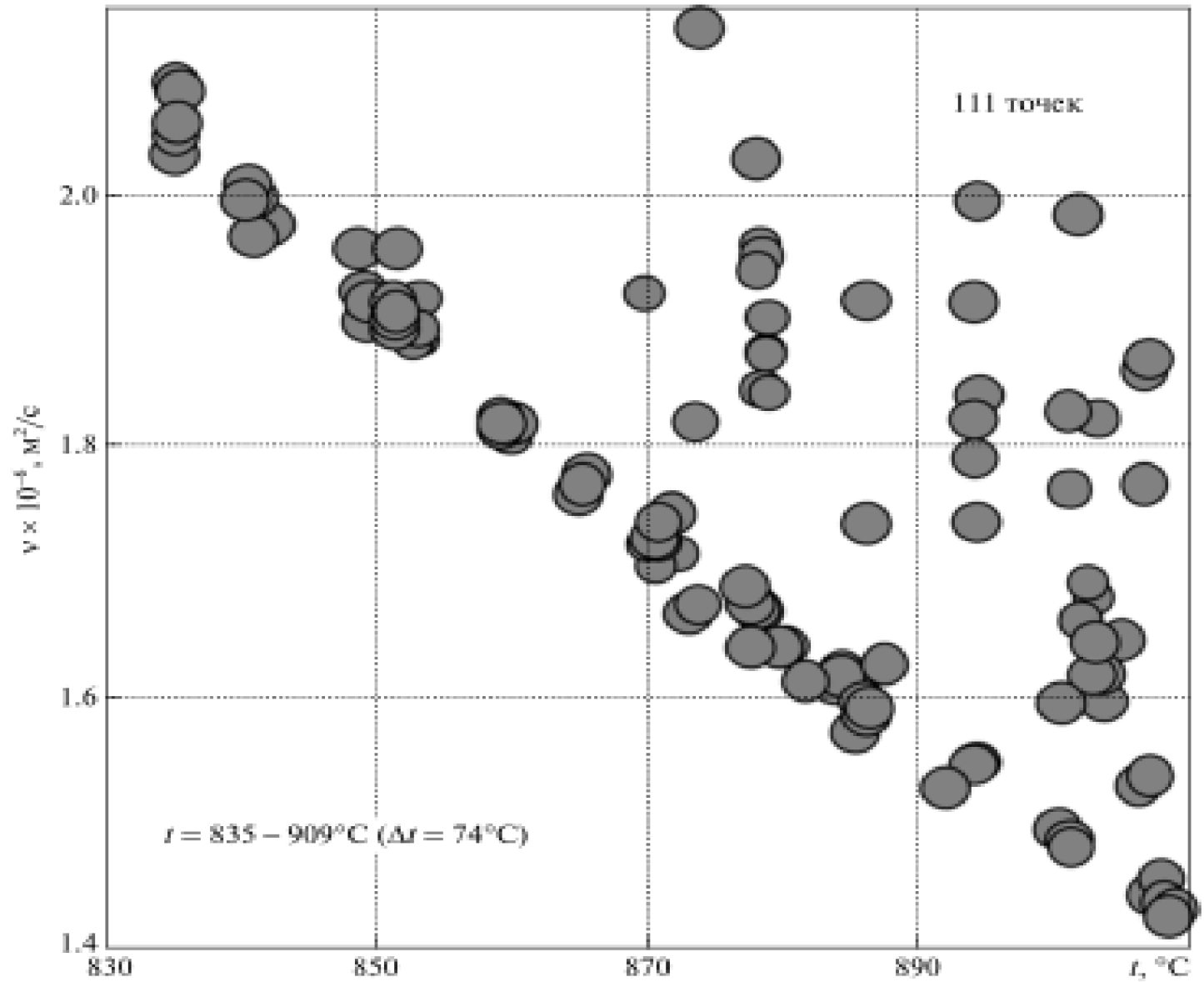
Пример 1: “хорошая” выборка без хаотического зашумления (рис. 1)



Пример 2: “умеренно” зашумленная выборка (рис. 2)



Пример 3: сильно зашумленная выборка (рис. 3)



Ситуация

Поясним принятый подход на примере обработки выборки измерений расплавленного электролита NbCl_3 , показанных на Рис. 2. Данный пример интересен тем, что хаотические искажения явно присутствуют практически во всем температурном диапазоне, но имеют малый пренебрежительный уровень, сравнимый с уровнем собственных измерительных (инструментальных и методических) погрешностей эксперимента. Это затрудняет “ручной” отсев таких выбросов.

Оценка величины погрешности (по выборке Рис. 2)

Основные процедуры:

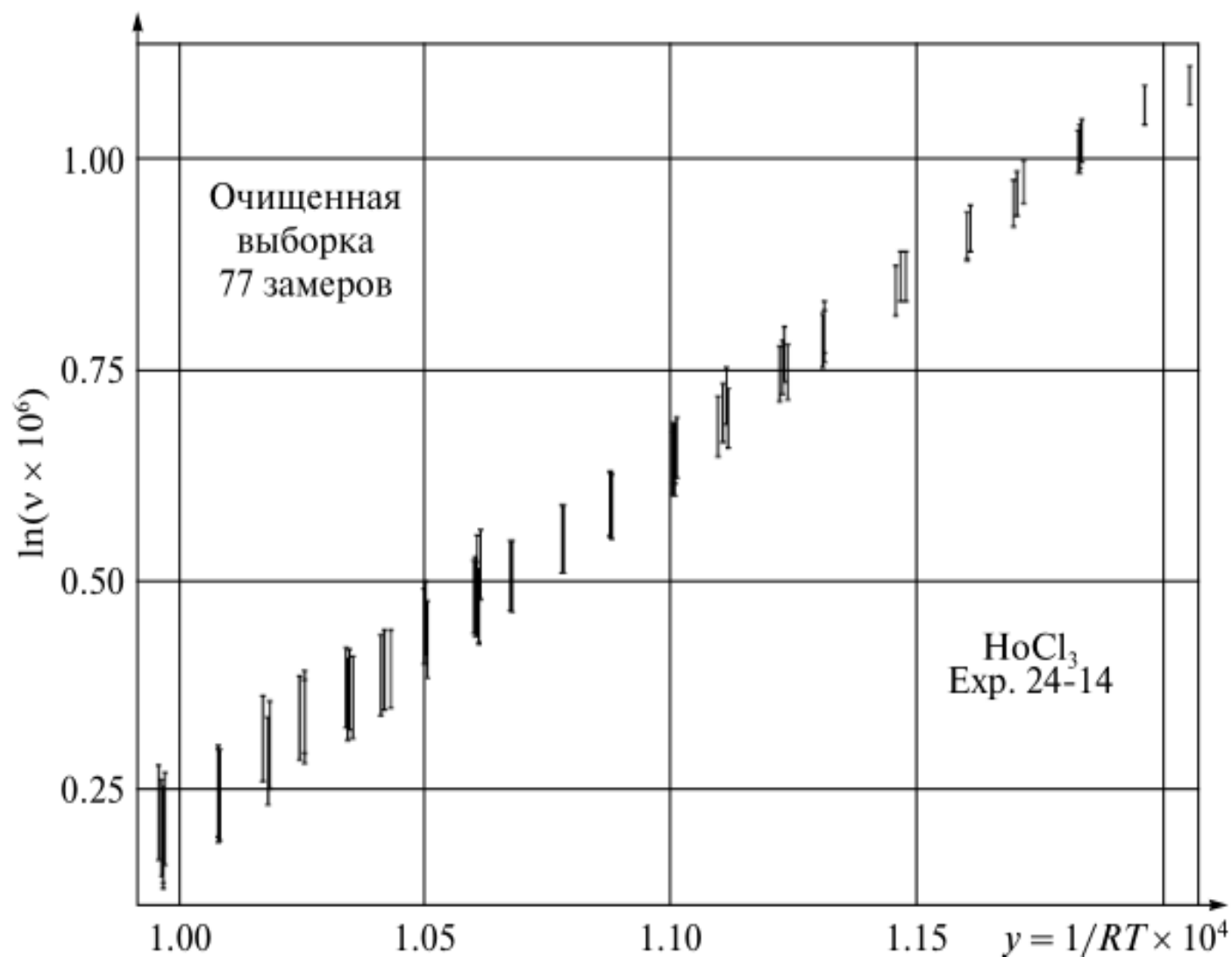
- используем “хороший” участок Рис. 2 (без явных выбросов) для оценки фактического уровня измерительной погрешности.
- задается некоторая с “запасом” исходная величина e_{\max} ограничения на погрешность измерений при которой выборка Рис. 2 заведомо является совместной, т.е. ее информационное множество параметров *непусто*;
- строится набор интервалов неопределенности $H_n = [v_n - e_{\max}, v_n + e_{\max}]$, $n = \overline{1, N}$ измерений;
- для простоты вычислительных процедур производится переход к логарифмическому масштабу интервалов неопределенности и вспомогательной линейной зависимости процесса от вспомогательной переменной y (Рис. 4);
- выполняется расчет соответствующего набора парциальных информационных множеств, и их пересечением находится непустое информационное множество $\Upsilon(e_{\max}, v_0, E_A)$ параметров.

Оценка величины погрешности (по выборке Рис. 2)

- путем вариации *вниз* величины e_{\max} (взятой с запасом) до вырождения информационного множества параметров $\Upsilon(e_{\max}, v_0, E_A)$ в точку находится минимальная величина e_{\max}^* , являющаяся оценкой снизу фактического уровня ограничения *измерительной* погрешности;
- отметим, эта величина e_{\max}^* определяется *в условиях отсутствия хаотических компонент*;
- для дальнейших расчетов экспериментатор на основании своего опыта задает величину этого ограничения с разумным “запасом” β :
 $e_{\max} = \beta e_{\max}^*$; практически, величина β задается на уровне 1.1 – т.е. с запасом в $\approx +10\%$.

При обработке случая Рис. 2 ограничение на максимальную величину инструментальной погрешности полагалось $e_{\max} = 0.045 \times 10^{-6}$ м²/сек.

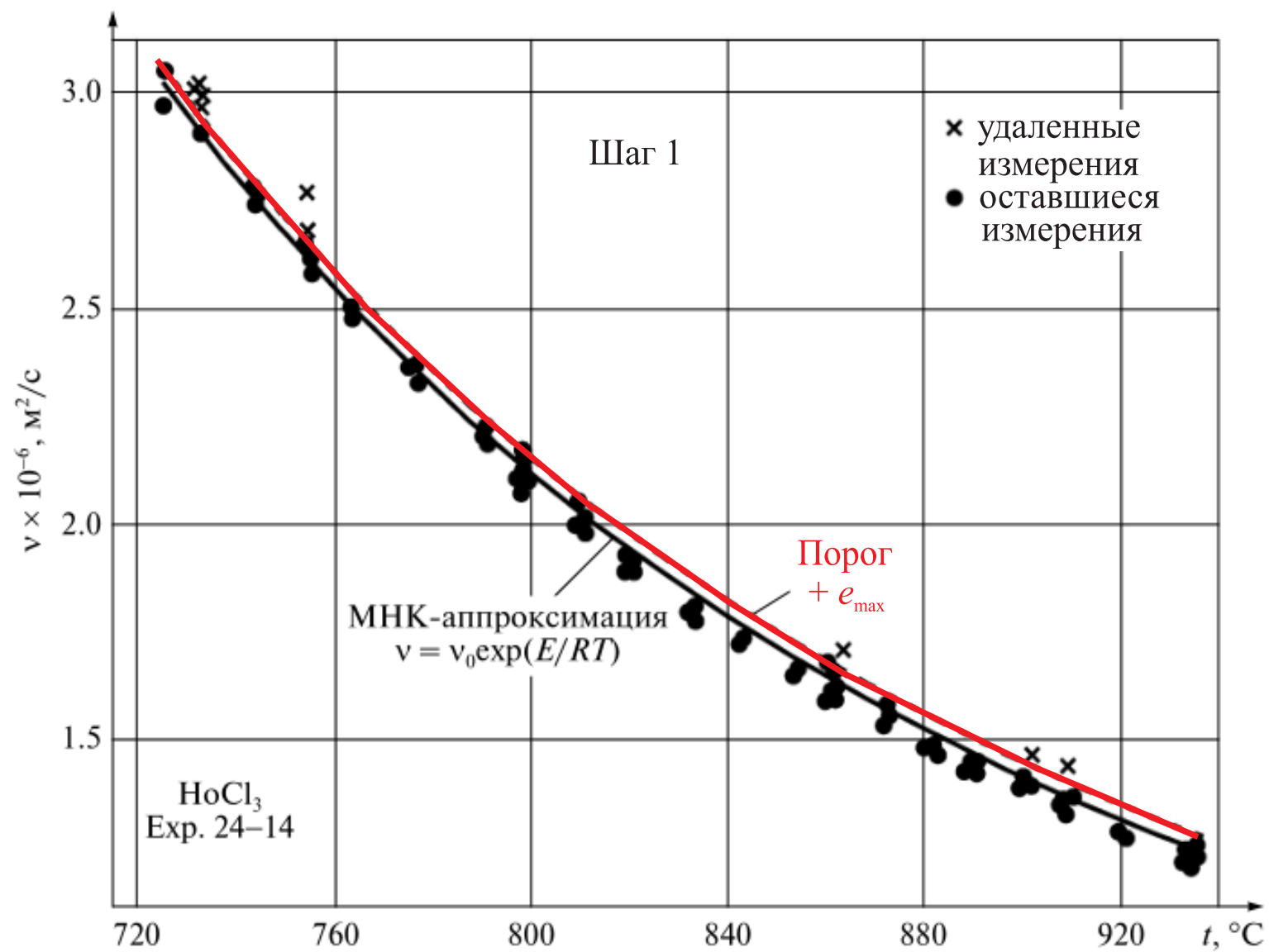
Упрощение вычислений: переход к логарифмическому масштабу, к преобразованной линейной зависимости и совокупности интервалов неопределенности (рис. 4)



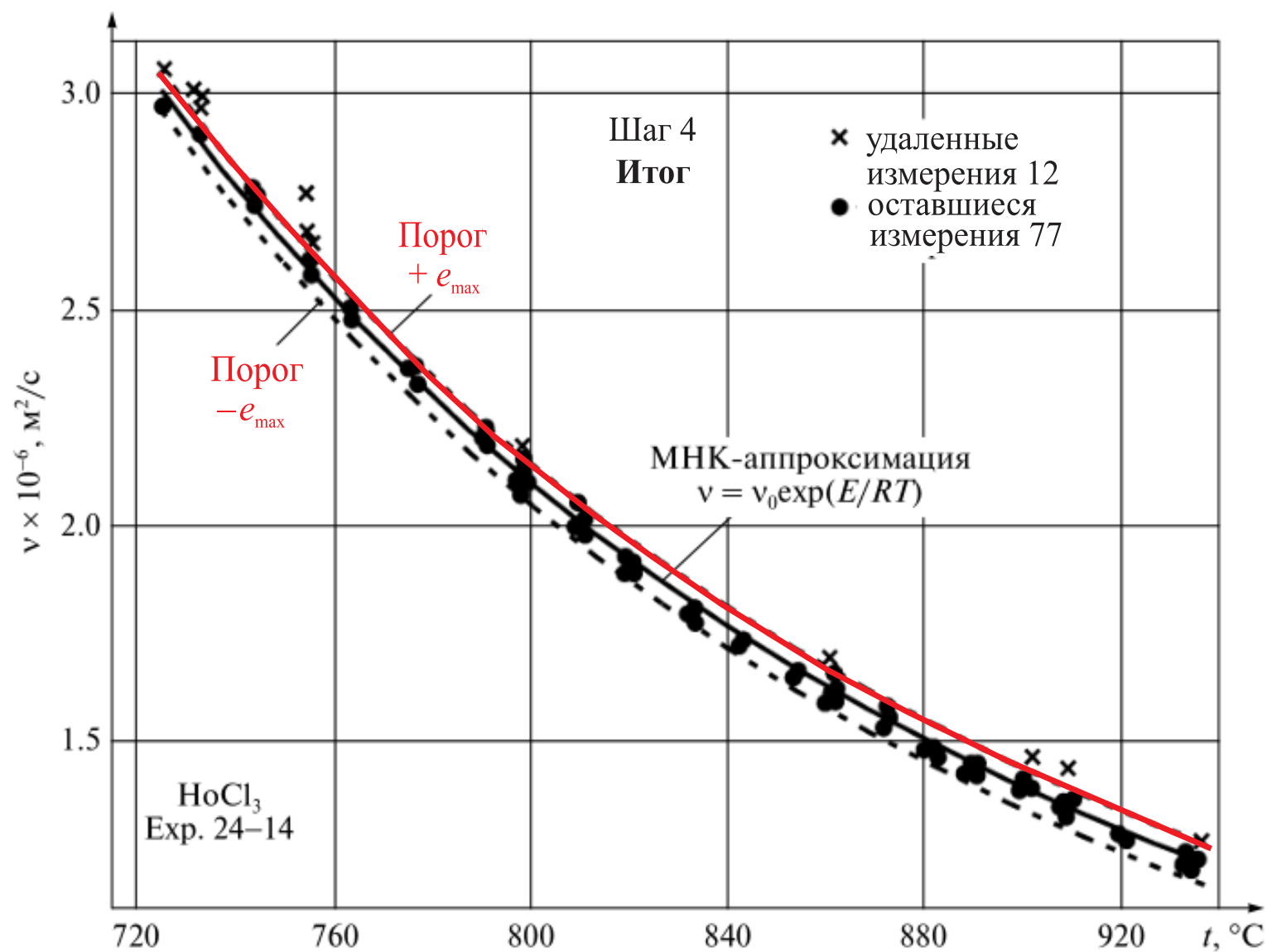
Процедуры выявления и удаления выбросов

- по текущей выборке рассчитывается её МНК-апроксимация;
- далее используется информация о положительности хаотических компонент погрешностей: по заданной величине e_{\max} , относительно данной кривой выставляется верхний порог $+e_{\max}$ (рис. 4);
- измерения, отклоняющиеся вверх от кривой МНК более чем на e_{\max} , объявляются выбросами и удаляются из текущей выборки;
- эта последовательность операций циклически повторяется до останова по отсутствию удаляемых выбросов.

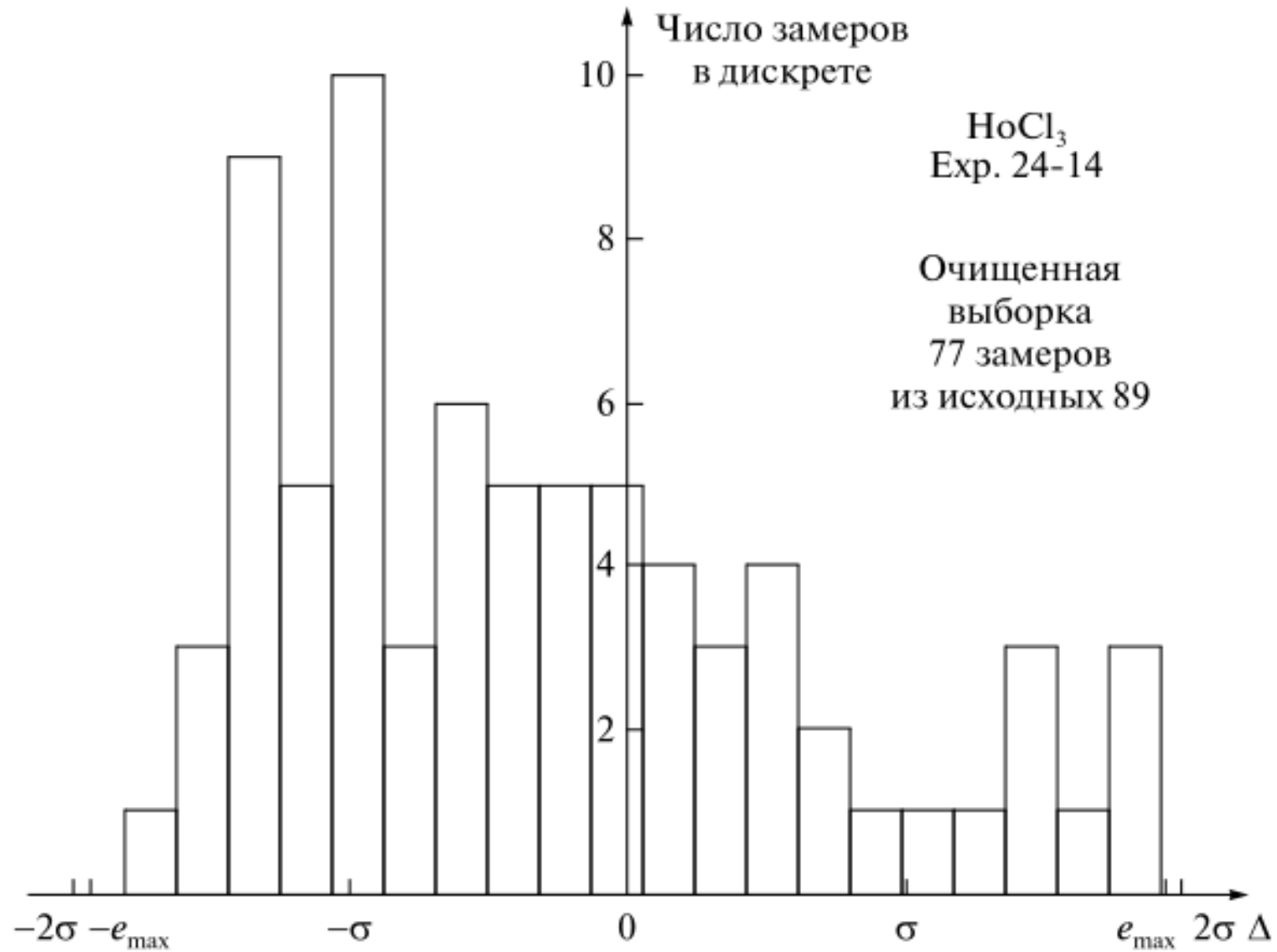
“Очищенная” выборка после 1-го шага (рис. 5)



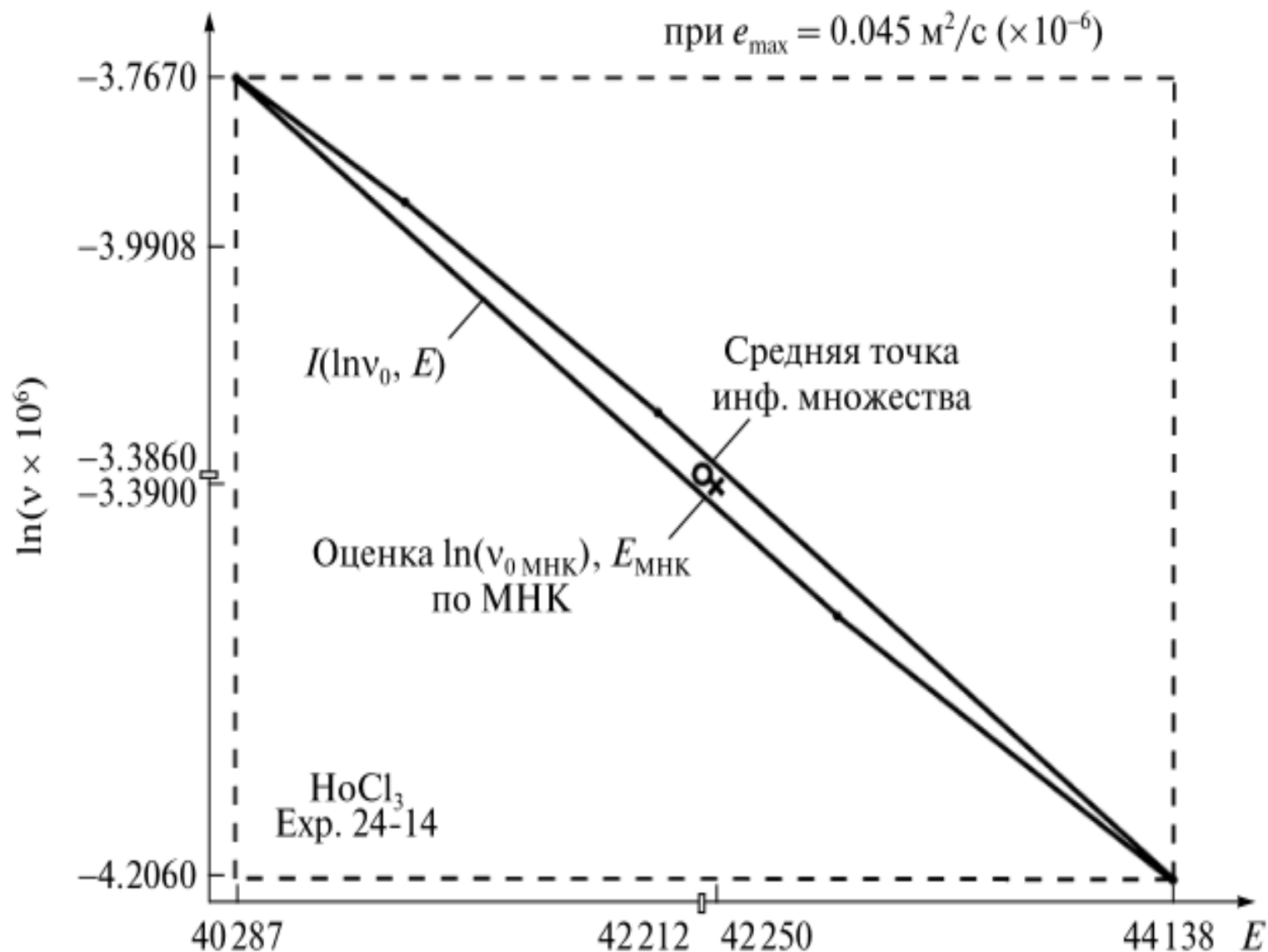
Итоговая “очищенная” выборка без выбросов после 4-х шагов их удаления (рис. 6)



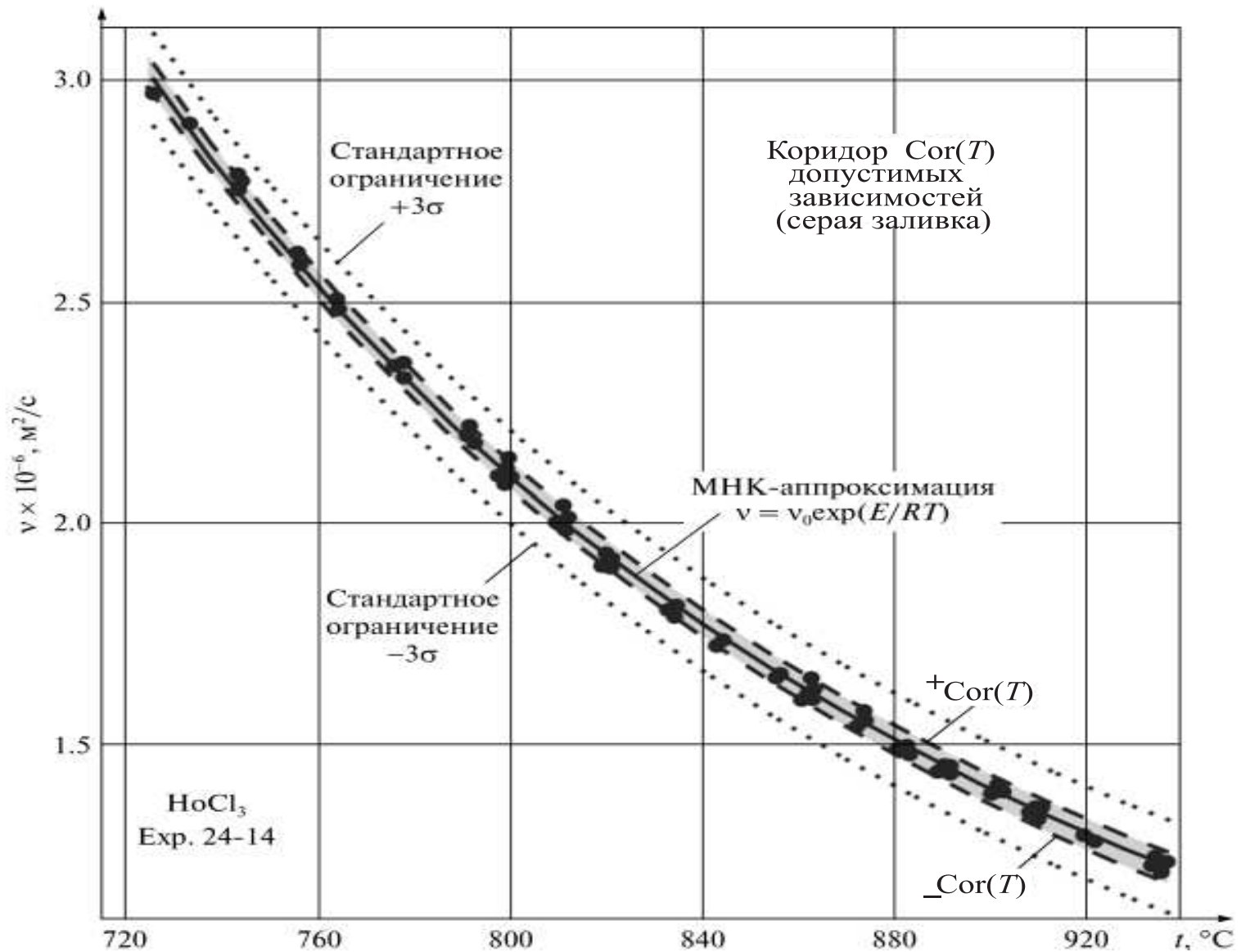
Гистограмма отклонений измерений в “очищенной” выборке (рис. 7)



Информационное множество параметров $\ln(v_0), E_A$ “очищенной” выборки (рис. 8)



Коридор совместных зависимостей “очищенной” выборки (рис. 9)



ИСТОЧНИК

1. А.М. Потапов, С.И. Кумков, Ю. Сато. Обработка экспериментальных данных по вязкости, содержащих случайные выбросы одного знака // Расплавы. 2010, 3, стр.48–63.