

УДК 519.25,537.533.2

## АНАЛИЗ ДОВЕРИТЕЛЬНЫХ ОБЛАСТЕЙ ДЛЯ ПАРАМЕТРОВ РЕГРЕССИОННОЙ МОДЕЛИ НА ОСНОВЕ ЗАКОНА ФАУЛера–НОРДГЕЙМА

© 2020 г. Н. В. Егоров<sup>а</sup>, \*, А. Ю. Антонов<sup>а</sup>, М. И. Вараюнь<sup>а</sup>

<sup>а</sup>Санкт-Петербургский государственный университет,  
Санкт-Петербург, 199034 Россия

\*e-mail: robby7@mail.ru

Поступила в редакцию 17.12.2019 г.

После доработки 22.01.2020 г.

Принята к публикации 24.01.2020 г.

В работе рассматривается двухпараметрическая регрессионная модель на основе классической для области полевой электронной эмиссии формулы Фаулера–Нордгейма. Оптимальные значения параметров определяются методом наименьших квадратов. В рамках статистического эксперимента проводится оценка характеристик доверительных областей. Уделяется внимание точности оценок параметров, границам индивидуальных доверительных интервалов, совместным доверительным эллипсам. Для значений напряжения используются два вида планов эксперимента. Погрешность измерений силы тока формируется с помощью модели шума специального вида. В ходе компьютерного эксперимента данные подвергаются тесту Шапиро–Уилка, проверяющего нормальность их распределения, а регрессионная модель проходит проверку на значимость. Выявляется процент выборок, которые могут не удовлетворять указанным критериям.

**Ключевые слова:** полевая электронная эмиссия, вольтамперная характеристика, регрессионная модель, метод наименьших квадратов, доверительные области.

DOI: 10.31857/S1028096020070055

### ВВЕДЕНИЕ

Приборы и устройства, одним из основных элементов которых является полевой электронный источник, продолжают рассматриваться как крайне перспективные [1]. Не в последнюю очередь этому способствует тот факт, что время запаздывания отклика на изменение напряжения  $V$  между катодом и анодом является ничтожно малым. Упомянутое обстоятельство позволяет рассматривать полевые эмиссионные катоды как близкие к совершенным системы управления.

Конечно, для работы систем в режиме полевой электронной эмиссии (ПЭЭ) необходимо выполнение некоторых условий, ограничивающих область их применения. Ясно, что помимо напряжения на отклик влияют и другие факторы, например, температура  $T$  и значение работы выхода  $\Phi$ . И если влиянием температуры часто пренебрегают, хотя она обеспечивает свой вклад в ширину энергетического спектра, то даже очень малых вариаций работы выхода вполне достаточно, чтобы отметить изменение силы тока  $I$  и даже провести косвенный анализ коэффициента прохождения электронами потенциального барьера на границе катода [2]. В связи с такой чувствительностью ра-

бочей поверхности полевого источника для его эксплуатации необходим высокий уровень вакуума. Есть основания считать, что новые материалы позволят расширить диапазон давления, при котором станет возможным стабильное и легко управляемое функционирование катодных узлов такого рода [3].

Основным законом, который используется для описания зависимости силы тока от напряжения, остается формула Фаулера–Нордгейма [4]:

$$I = AV^2 \exp\left[-\frac{B}{V}\right], \quad (1)$$

где параметры  $A$  и  $B$  можно считать постоянными величинами только для определенной, рассматриваемой в ходе эксперимента системы. Дело в том, что, являясь интегральными характеристиками,  $A$  и  $B$  зависят и от распределения работы выхода по поверхности, и от геометрии данной поверхности. Попытки придать данным параметрам вполне четкую связь с работой выхода  $\Phi$  [5], площадью эмиссии  $S$  [6] или геометрическим фактором поля  $\beta$  [7] являются одной из основных задач анализа вольт-амперных характеристик (ВАХ). Впрочем, трудно представить, что двухпа-

раметрическая модель (1) способна дать исчерпывающее описание источника, особенно после того, как эксперимент существенно отошел от предпосылок, заложенных в теорию Фаулера–Нордгейма. При этом отклонения от линейной зависимости выражения  $\lg(I/V^2)$ , как функции обратного напряжения  $1/V$  [7–9], не анализируются с использованием других моделей, например с трехпараметрической моделью, рассмотренной в [10 и 11].

Тем не менее, какая бы модель не применялась при анализе ВАХ, необходимо представлять, с какой точностью экспериментатор способен оценить ее параметры. Дать не только точечную, но и интервальную оценку позволяет математический аппарат, заложенный в регрессионном анализе. Но поскольку результат такой оценки зависит от конкретной выборки измеренных значений напряжения и силы тока, выводы, сделанные на их основе, могут касаться только рассматриваемого образца.

В связи с вышеизложенным, целью настоящей работы является выявление зависимости точности статистических оценок для параметров регрессионной модели на основе закона Фаулера–Нордгейма. Чтобы получить представление о поведении оценок в среднем, а не на основе одной серии наблюдений, для получения выборки значений напряжения и силы тока используется математическое моделирование и компьютерный эксперимент.

## МАТЕМАТИЧЕСКАЯ МОДЕЛЬ

Ранее [11] было показано, что задача параметрической идентификации сигнала полевой электронной эмиссии может быть сформулирована в рамках как линейной, так и нелинейной постановки. Здесь под сигналом понимается значение силы тока  $I$ , полученное при заданном значении напряжения  $V$ . Классическая двухпараметрическая ( $p = 2$ ) модель предлагается на основе (1) в виде:

$$I = I(V; \mathbf{q}) = I_0 A \left( \frac{V}{V_0} \right)^2 \exp \left[ -B \frac{V_0}{V} \right],$$

$$\mathbf{q} = (q_1, q_2) = (A, B) \in Q,$$

$$Q = \{(q_1, q_2) : q_1 > 0, q_2 > 0\},$$

где  $I_0$  и  $V_0$  – некоторые калибровочные значения, цель которых сделать параметры  $A$  и  $B$  безразмерными. Для дальнейшего анализа модель (2), как правило, приводится к линейному виду с помощью логарифмического преобразования.

Введем обозначения:

$$X = \frac{V_0}{V}, \quad Y = \lg \left[ \frac{I V_0^2}{I_0 V^2} \right]. \quad (3)$$

В регрессионном и дисперсионном анализе величина  $X$  называется фактором, а  $Y$  – откликом сигнала. Уже упоминалось, что преобразование (3) приводит к линейной зависимости между фактором и откликом (такая модель носит название простой [12]):

$$Y = \varphi(X; \boldsymbol{\theta}) = \boldsymbol{\theta}^T \mathbf{g}(X) = \sum_{j=1}^p \vartheta_j g_j(X) = A' + B' X,$$

$$\boldsymbol{\theta} = (\vartheta_1, \vartheta_2)^T = (A', B')^T \in \Theta,$$

$$\Theta = \{(\vartheta_1, \vartheta_2) : \vartheta_1 \in \mathbb{R}, \vartheta_2 < 0\}, \quad (4)$$

$$A' = \lg A, \quad B' = -B/\ln 10,$$

$$\mathbf{g}(X) = (g_1(X), g_2(X))^T = (1, X)^T.$$

Пусть имеются  $N$  наблюдений силы тока  $\tilde{I}_i$ , которые содержат неизбежные погрешности измерений  $\varepsilon_i$ . В предыдущем исследовании [11] было показано, что значения напряжения  $V$ , а значит и факторов  $X$ , можно, не умаляя общности, считать измеренными точно.

Цель работы может быть достигнута только после проведения серии статистических испытаний. Исходя из этого, будем считать зависимость (2) точной, а погрешности  $\varepsilon_i$  – гауссовскими случайными величинами с нулевым математическим ожиданием и среднеквадратичным отклонением, пропорциональным  $I$ :

$$\tilde{I}_i = I_i + \varepsilon_{i,i} = I_i + I_0 \varepsilon_i = I_i (1 + \delta \varepsilon'_i), \quad (5)$$

$$i = 1, 2, \dots, N,$$

где  $\delta$  – коэффициент пропорциональности, который будем называть уровнем шума, а  $\varepsilon'_i$  – стандартные нормально распределенные случайные величины. При моделировании отклика (5) значения параметров  $\mathbf{q}$  полагаются известными и равными  $\mathbf{q}^* = (A^*, B^*)$ , что после линеаризации дает значения  $\boldsymbol{\theta}^* = (\lg A^*, -B^*/\ln 10)^T$ . Здесь операция транспонирования подчеркивает, что  $\boldsymbol{\theta}^*$  рассматривается как вектор-столбец, что удобно для дальнейших выкладок. Преобразование (3) определяют при этом пары наблюдений  $\{X_i, \tilde{Y}_i\}_{i=1}^N$ . Ставится задача регрессионного анализа в общих обозначениях. Требуется отыскать на классе функций  $\varphi(X; \boldsymbol{\theta})$  значения параметров  $\hat{\boldsymbol{\theta}}$ , которые обеспечат минимум некоторому функционалу качества  $J(\boldsymbol{\theta})$ . Наиболее простым и эффективным в плане использования для линейной по параметрам регрессионной модели является функ-

ционал, связанный с методом наименьших квадратов (МНК) [12]:

$$J(\boldsymbol{\theta}) = \sqrt{\frac{1}{N} \sum_{i=1}^N (\tilde{Y}_i - f(X_i; \boldsymbol{\theta}))^2}.$$

Тогда

$$J(\hat{\boldsymbol{\theta}}) = \min_{\boldsymbol{\theta} \in \Theta} J(\boldsymbol{\theta}) = \sqrt{\frac{1}{N} \sum_{i=1}^N \hat{\epsilon}_i^2}, \quad \tilde{y}_i = \hat{y}_i + \hat{\epsilon}_i, \quad (6)$$

$$\hat{y}_i = f(x_i; \hat{\boldsymbol{\theta}}).$$

Остатки регрессионной модели могут обладать следующими свойствами:

1.  $\hat{\epsilon}_i$  являются независимыми случайными величинами;

2.  $\hat{\epsilon}_i$  одинаково распределены по нормальному закону с нулевым средним и постоянными дисперсиями.

В нашем случае (математическое моделирование) величины  $\hat{\epsilon}_i$  формируются посредством независимых реализаций  $\epsilon'_i$ , что дает основания считать первое свойство выполненным. В рамках натурального эксперимента наличие этого свойства подлежит анализу. Например, можно воспользоваться тестом Дарбина—Уотсона на проверку наличия или отсутствия автокорреляции [12]. Такое исследование вынесено за рамки настоящей работы и будет проведено позднее.

Второе свойство асимптотически выполнено при  $\delta \rightarrow 0$ , что установлено ранее [11], причем равенство нулю выборочного среднего для величин  $\hat{\epsilon}_i$  достигается согласно способу их построения. Но опять же единичная выборка наблюдений  $\{X_i, \tilde{Y}_i\}_{i=1}^N$  приводит к своей конкретной оптимальной оценке  $\hat{\boldsymbol{\theta}}$  (и, как следствие, к  $\hat{\mathbf{q}}$ ), являющейся случайной величиной, и конкретной совокупности остатков  $\hat{\epsilon}_i$ . Эта совокупность должна пройти тесты на принадлежность нормальному закону распределения и гомоскедастичность (равенство дисперсий при различных значениях фактора  $X$ ). Выбор критериев проверки остается за исследователем, но некоторые из них могут потребовать серий наблюдений отклика  $\tilde{Y}$  для каждого значения  $X$ . Построенные таким образом вольтамперные характеристики (ВАХ) трудно найти в литературе — как правило, наблюдения приводятся единичные (возможно, усредненные).

Если есть основания считать оба свойства для остатков  $\hat{\epsilon}_i$  выполненными, то оптимальная оценка  $\hat{\boldsymbol{\theta}}$  окажется случайной величиной с нормальным законом распределения и математическим ожиданием равным  $\boldsymbol{\theta}^*$ . Также она будет обеспечивать максимум функции правдоподобия [13].

С помощью указанных свойств гораздо удобнее становится процедура построения доверительных областей для значений параметра  $\hat{\boldsymbol{\theta}}$ . Если хотя бы одно требование не является выполненным, то квадратичный функционал перестает иметь какое-либо значительное преимущество перед другими, за исключением простоты.

### СТАТИСТИЧЕСКИЕ ОЦЕНКИ

Оценка МНК (6) для значений параметров линеаризованной модели (4) получается аналитически в виде:

$$\hat{\boldsymbol{\theta}} = \mathbf{M}^{-1} \mathbf{b}, \quad (7)$$

где матрица  $\mathbf{M}$  формируется на основе значений факторов анализируемой выборки и функций регрессионной модели  $g_j(X)$  по следующей формуле:

$$m_{jl} = \sum_{i=1}^N g_j(X_i) g_l(X_i),$$

а компоненты вектора  $\mathbf{b}$  имеют вид:

$$b_l = \sum_{i=1}^N g_l(X_i) \tilde{Y}_i;$$

индексы  $j$  и  $l$  изменяют свои значения от 1 до  $p$ .

Пусть остатки  $\hat{\epsilon}_i$  линеаризованной регрессионной модели удовлетворяют упомянутым в предыдущем разделе условиям. Тогда индивидуальные значения параметров  $\vartheta_j$  на уровне значимости  $\alpha$  укладываются в интервалы:

$$\begin{aligned} \vartheta_j^L &\leq \vartheta_j \leq \vartheta_j^R, \quad \vartheta_j^L = \hat{\vartheta}_j - \Delta \vartheta_j / 2, \\ \vartheta_j^R &= \hat{\vartheta}_j + \Delta \vartheta_j / 2, \quad \Delta \vartheta_j = \vartheta_j^R - \vartheta_j^L = \\ &= 2t_{1-\alpha/2}(k_2) \sqrt{s^2 \tilde{m}_{jj}}, \quad s^2 = \frac{1}{k_2} \sum_{i=1}^N \hat{\epsilon}_i^2, \quad (8) \\ k_2 &= N - p, \end{aligned}$$

где  $t_{1-\alpha/2}(k_2) = (1 - \alpha/2)$  — квантиль  $t$ -распределения Стьюдента с  $k_2$  степенями свободы [14]. Величины  $\tilde{m}_{jj}$  представляют собой диагональные элементы матрицы  $\mathbf{M}^{-1}$ , обратной к  $\mathbf{M}$ .

Параметры  $\vartheta_j$  не являются независимыми, поскольку оцениваются по одним и тем же измерениям. Совместная доверительная область для вектора  $\boldsymbol{\theta}$  представляет собой в общем случае внутренность  $p$ -мерного эллипсоида:

$$(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})^T \mathbf{M}(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}) \leq ps^2 F_{1-\alpha}(k_1, k_2), \quad k_1 = p - 1, \quad (9)$$

где  $F_{1-\alpha}(k_1, k_2) = (1 - \alpha)$ -квантиль  $F$ -распределения Фишера—Снедекора со степенями свободы  $k_1$  и  $k_2$  [12]. Для двухпараметрической модели

речь идет о доверительном эллипсе, который охватывает площадь:

$$S_{\theta} = \frac{\pi p s^2 F_{1-\alpha}(k_1, k_2)}{\sqrt{\det \mathbf{M}}}. \quad (10)$$

Переход к параметрам нелинейной модели (2) дает оценку  $\hat{\mathbf{q}}$  и границы для исходных значений, оцениваемых параметров  $q_j$ , шириной:

$$\Delta q_1 = 10^{\hat{v}_1^R} - 10^{\hat{v}_1^L}, \quad \Delta q_2 = \Delta \vartheta_2 \ln 10. \quad (11)$$

Это позволяет ввести в рассмотрение следующие показатели точности параметрической идентификации сигнала. Относительная погрешность определения значений параметров регрессионной модели вводится стандартным образом:

$$\gamma_j^{(1)} = \left| \frac{\hat{q}_j - q_j^*}{q_j^*} \right|. \quad (12)$$

Данная характеристика имеет смысл в рамках численного эксперимента, когда точные значения компонент  $\mathbf{q}^*$  известны (и отличны от нуля). Величину, описывающую с заданной доверительной вероятностью  $(1 - \alpha)$ , качество интервальной оценки, определим как

$$\gamma_j^{(2)} = \left| \frac{\Delta q_j}{\hat{q}_j} \right|. \quad (13)$$

Она может быть использована и в компьютерном, и в натурном эксперименте. Здесь надо помнить о том, что оценка  $\hat{q}_j$  вполне может оказаться близкой к нулю в отдельном эксперименте в силу разнообразных факторов. Также показателем, связанным со значениями  $\gamma_j^{(2)}$ , является площадь совместной доверительной области  $S_{\theta}$  – формула (10) – на множестве параметров  $\theta$  (для параметров  $\mathbf{q}$  доверительная область деформируется в силу преобразований и перестанет описываться эллипсом, ее площадь  $S_{\theta}$  не рассматривалась).

## ПОДГОТОВКА СТАТИСТИЧЕСКОГО ЭКСПЕРИМЕНТА

Как отмечалось ранее, точечные и интервальные оценки параметров регрессионной модели, а также любые статистики, связанные с единичной выборкой, являются случайными величинами. Чтобы установить их усредненное поведение в зависимости от уровня шума  $\delta$  и количества наблюдений  $N$ , необходимо провести серию экспериментов по анализу  $N_{\text{stat}}$  выборок  $\{X_i, \tilde{Y}_i\}_{i=1}^N$ . Авторы также решили обратить внимание на расположение значений факторов. Действительно, если положить измерения напряжения  $V_i$  равно-

отстоящими, то величины  $X_i$  в линеаризованной модели уже не будут таковыми. В связи с этим обстоятельством рассматривалось два способа. Первый способ (прямой, fV) подразумевал построение сетки  $V_i$  (плана эксперимента) с постоянным шагом – рис. 1:

$$V_i = V_{\min} + \frac{V_{\max} - V_{\min}}{N - 1} (i - 1), \quad i = 1, 2, \dots, N. \quad (14)$$

Второй способ (обратный, bV) реализовывал ситуацию, когда преобразованные безразмерные переменные  $X_i$  оказывались бы на одинаковом расстоянии друг от друга – рис. 2:

$$V_i = \frac{V_{\max} V_{\min} (N - 1)}{V_{\max} (N - i) + V_{\min} (i - 1)}, \quad i = 1, 2, \dots, N. \quad (15)$$

В ходе компьютерного эксперимента кроме средних показателей

$$\bar{\gamma}_j^{(1)} = \left| \frac{\langle \hat{q}_j \rangle - q_j^*}{q_j^*} \right| \neq \langle \gamma_j^{(1)} \rangle,$$

$$\bar{\gamma}_j^{(2)} = \langle \gamma_j^{(2)} \rangle = \left\langle \frac{\Delta q_j}{\hat{q}_j} \right\rangle, \quad \langle S_{\theta} \rangle$$

фиксировались также их минимальные и максимальные значения. Здесь под  $\langle x \rangle$  понимается усредненное по всем  $N_{\text{stat}}$  экспериментам значение величины  $x$ . Обратим внимание, что учету подлежали результаты только тех опытов, в которых остатки  $\hat{e}_i$  регрессионной модели (4) удовлетворяли критерию нормальности Шапиро–Уилка [15]. Данный тест имеет хорошую мощность против большого числа альтернативных гипотез. Так на основе упорядоченных по возрастанию значений остатков  $\hat{e}_{(i)}$  производится вычисление характеристики:

$$W = \frac{1}{s^2 (N - p)} \left( \sum_{i=1}^{N'} a_{N+1-i} (\hat{e}_{(N+1-i)} - \hat{e}_{(i)}) \right)^2, \quad (16)$$

$$N' = \left\lfloor \frac{N}{2} \right\rfloor,$$

где  $\lfloor x \rfloor$  – наибольшее целое число, меньшее или равное  $x$ . Коэффициенты  $a_i$  зависят от объема выборки. Если выполнялось условие  $W < W(\alpha)$ , где  $W(\alpha)$  – критическое значение статистики, то данная выборка не учитывалась. Величины  $a_i$  и  $W(\alpha)$  берутся из таблиц [15]. Другие способы проверки гипотезы о принадлежности выборки остатков нормальному закону распределения не рассматривались. Кроме причины экономии времени авторы решили избежать таким образом неоднозначных ситуаций, когда разные тесты дают основания прийти к разным выводам.

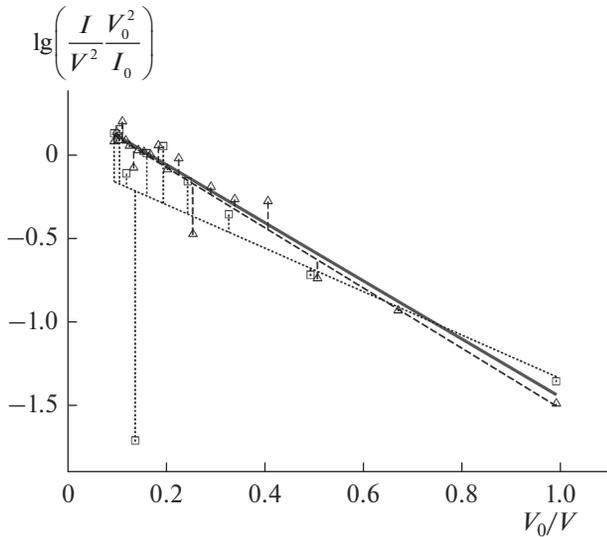


Рис. 1. Примеры ВАХ на сетке fv – (14).

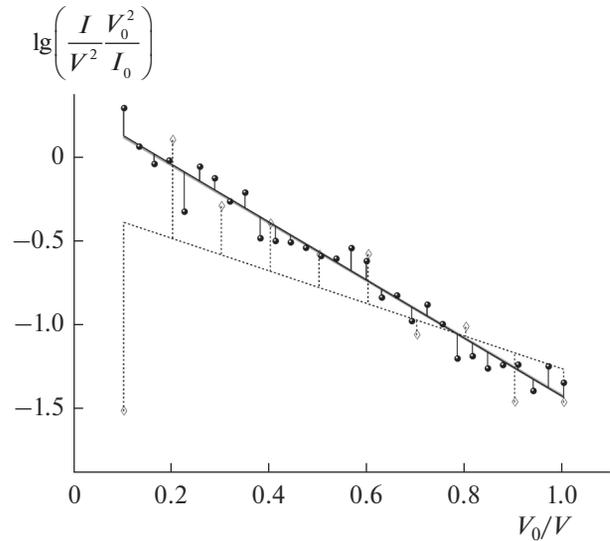


Рис. 2. Примеры ВАХ на сетке bV – (15).

В случае, когда значение уровня шума  $\delta$  является существенным, а объем анализируемой выборки достаточно мал, может реализоваться ситуация, при которой оценки  $\hat{\theta}$  выходят за пределы области возможных значений  $\Theta$ . Такое поведение обусловлено наличием так называемых “выбросов” – сильно отклоняющихся от среднего значения величин (рис. 1, 2). При визуальном анализе данных эксперимента ситуация с аномальными наблюдениями сразу же выделяется в отдельное рассмотрение. В ходе же статистического эксперимента из большого количества испытаний такой подход практически исключен. Здесь разумно допускать для формирования общей статистики те данные, регрессионные модели которых прошли проверку на значимость [14]. По этой причине проводилось вычисление статистики:

$$F = \frac{k_2 \sum_{i=1}^N (\bar{Y} - \hat{Y}_i)^2}{k_1 \sum_{i=1}^N (\tilde{Y}_i - \hat{Y}_i)^2}, \quad \bar{Y} = \frac{1}{N} \sum_{i=1}^N \tilde{Y}_i, \quad (17)$$

которая позднее сравнивалась с критическим значением  $F_{1-\alpha}(k_1, k_2)$ . В случае  $F < F_{1-\alpha}(k_1, k_2)$  данные отбраковывались (отсутствуют основания считать регрессионную модель значимой).

На рис. 1 треугольниками представлены данные ( $N = 20$ ), которые не прошли  $W$ -тест на нормальность величин  $\hat{\epsilon}_i$ . При этом  $F$ -тест не поставил под сомнение значимость линеаризованной модели (4). Квадратами на том же рисунке изображены отклики ( $N = 10$ ), которые не удовлетворили обоим критериям. На рис. 2 ромбами пока-

заны “измерения” ( $N = 10$ ), которые прошли  $W$ -тест, но регрессионная модель не была признана значимой. Наконец, оставшиеся точки (шарики) соответствуют примеру удовлетворительных данных ( $N = 30$ ), которые были использованы для дальнейшего анализа. При этом везде уровень шума составлял  $\delta = 20\%$ . Можно заключить, что  $F$ -тест позволяет обнаруживать если не все, то некоторые ситуации с аномальными наблюдениями.

### РЕЗУЛЬТАТЫ КОМПЬЮТЕРНОГО ЭКСПЕРИМЕНТА

Компьютерный статистический эксперимент предполагал анализ данных на основе  $N_{\text{stat}} = 10^7$  совокупности выборок, остатки которых прошли тест на нормальность в смысле величины (16) и регрессионные модели которых были признаны значимыми в смысле величины (17). Рассматривались три уровня шума  $\delta$ : 0.05, 0.1 и 0.2. Объемы анализируемых выборок  $N$  принимали значения: 10, 20, 30. При моделировании полагалось:  $V_0 = 1.0$ ,  $I_0 = 1.0$ ,  $V_{\text{min}} = 1.0$ ,  $V_{\text{max}} = 10.0$  (единицы измерения в нашем случае роли не играют). Значения напряжений  $V_i$  вычислялись по формулам (14) и (15). При формировании токового отклика  $I_i$  согласно (2) использовались значения  $A^* = 2.0$  и  $B^* = 4.0$ . Реализации нормально распределенных случайных величин  $\epsilon'_i$  проводились на основе генератора псевдослучайных величин Mersenne Twister MT19937 [16] с использованием алгоритма, приведенного в [17]. Можно отметить, что поскольку сила тока предполагается положитель-

Таблица 1. Статистические данные для сетки fV – (14)

$N$	$\delta$	$\langle \hat{\mathbf{q}} \rangle$	$[\langle \mathbf{q}^L \rangle, \langle \mathbf{q}^R \rangle]$	$\bar{Y}_j^{(1)}$	$\bar{Y}_j^{(2)}$	$\langle S_0 \rangle$	$W_{rej}, \%$
10	5	$\langle \hat{A} \rangle = 1.998$ $\langle \hat{B} \rangle = 4.000$	$[1.895, 2.107]$ $[3.865, 4.135]$	$9.756 \times 10^{-2}$ $7.958 \times 10^{-4}$	10.61 6.733	$6.018 \times 10^{-3}$	5.065
	10	$\langle \hat{A} \rangle = 1.992$ $\langle \hat{B} \rangle = 4.000$	$[1.792, 2.217]$ $[3.729, 4.271]$	$3.932 \times 10^{-1}$ $1.988 \times 10^{-3}$	21.37 13.56	$2.443 \times 10^{-2}$	5.621
	20	$\langle \hat{A} \rangle = 1.970$ $\langle \hat{B} \rangle = 4.001$	$[1.590, 2.449]$ $[3.447, 4.555]$	1.524 $1.693 \times 10^{-2}$	44.04 27.84	$1.033 \times 10^{-1}$	8.214
20	5	$\langle \hat{A} \rangle = 1.998$ $\langle \hat{B} \rangle = 4.000$	$[1.926, 2.072]$ $[3.896, 4.104]$	$1.097 \times 10^{-1}$ $3.326 \times 10^{-4}$	7.328 5.201	$2.943 \times 10^{-3}$	5.753
	10	$\langle \hat{A} \rangle = 1.991$ $\langle \hat{B} \rangle = 4.000$	$[1.850, 2.144]$ $[3.791, 4.209]$	$4.326 \times 10^{-1}$ $3.601 \times 10^{-3}$	14.75 10.47	$1.192 \times 10^{-2}$	7.439
	20	$\langle \hat{A} \rangle = 1.967$ $\langle \hat{B} \rangle = 4.002$	$[1.694, 2.285]$ $[3.576, 4.428]$	1.666 $5.169 \times 10^{-2}$	30.15 21.37	$4.973 \times 10^{-2}$	15.02
30	5	$\langle \hat{A} \rangle = 1.998$ $\langle \hat{B} \rangle = 4.000$	$[1.939, 2.059]$ $[3.912, 4.088]$	$1.147 \times 10^{-1}$ $4.154 \times 10^{-4}$	6.000 4.419	$1.980 \times 10^{-3}$	5.365
	10	$\langle \hat{A} \rangle = 1.991$ $\langle \hat{B} \rangle = 4.000$	$[1.874, 2.115]$ $[3.822, 4.178]$	$4.544 \times 10^{-1}$ $2.310 \times 10^{-3}$	12.07 8.891	$8.020 \times 10^{-3}$	7.907
	20	$\langle \hat{A} \rangle = 1.965$ $\langle \hat{B} \rangle = 4.002$	$[1.739, 2.222]$ $[3.640, 4.364]$	1.744 $4.401 \times 10^{-2}$	24.63 18.13	$3.336 \times 10^{-2}$	19.82

ной величиной, то моделирование ее флуктуаций требует усечения нормального распределения слева [18]. Альтернативный вариант генерирования шума может быть проведен на основе метода Монте-Карло, как в [19]. Проверка всех гипотез и построение доверительных областей проводились на уровне значимости  $\alpha = 0.05$  (с доверительной вероятностью 95%).

Значения точечных и интервальных оценок на основе (7), (8) и (11) для исходных параметров нелинейной модели (2) при равноотстоящих значениях напряжения (сетка fV) приведены в табл. 1. В ней же указаны усредненные значения показателей (10), (12) и (13). Видно, что увеличение количества наблюдений и уменьшение уровня шума позволяет уменьшать доверительные области определения параметров, что полностью согласуется со здравым смыслом. При этом средние значения для точечных оценок оказываются предельно близкими к данным, заложенным в модель сигнала. Это позволяет в дальнейшем рассчитывать на тот факт, что если оценки и получаются смещенными при рассматриваемой модели шума специального вида, то величина этого

смещения является сравнительно малой. Последняя колонка табл. 1 отражает процент выборок, которые были отбракованы на основе статистики (16). Здесь можно обратить внимание на тот факт, что характеристика  $W_{rej}$  коррелирует с уровнем шума, и чем меньше число наблюдений, тем лояльнее “отношение” критерия к гипотезе о нормальности. При этом процент наблюдений, которые не прошли  $F$ -тест, оказался настолько мал (менее десятой процента), что ему было решено не уделять внимания.

В табл. 2 сведены абсолютно те же характеристики статистического эксперимента, но при равноотстоящих значениях фактора  $X$  (сетка bV). Выводы относительно динамики их зависимости от числа наблюдений и уровня шума остаются неизменными – как и в предыдущем случае. Это же касается и процента данных, не прошедших тесты на нормальность остатков и значимость линеаризованной регрессионной модели (4). Важно отметить, что на сетке bV (15) точнее в смысле доверительного интервала осуществляется оценка параметра  $B$ , отвечающего за наклон ВАХ в

**Таблица 2.** Статистические данные для сетки bV – (15)

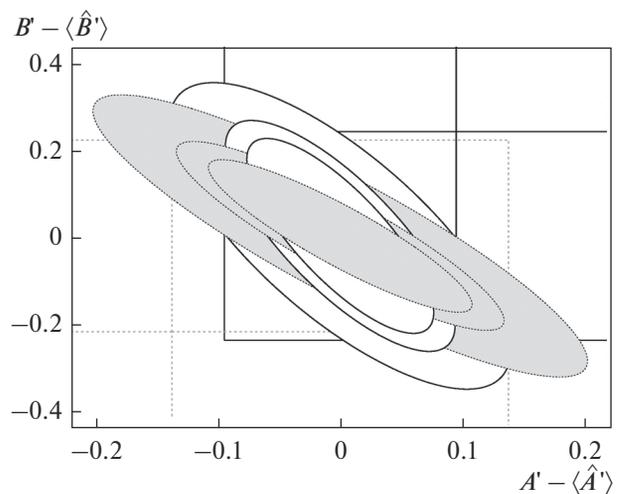
$N$	$\delta$	$\langle \hat{\mathbf{q}} \rangle$	$[\langle \mathbf{q}^L \rangle, \langle \mathbf{q}^R \rangle]$	$\bar{\gamma}_j^{(1)}$	$\bar{\gamma}_j^{(2)}$	$\langle S_\theta \rangle$	$W_{rej}, \%$
10	5	$\langle \hat{A} \rangle = 1.999$ $\langle \hat{B} \rangle = 4.000$	$[1.852, 2.158]$ $[3.877, 4.123]$	$6.502 \times 10^{-2}$ $7.681 \times 10^{-4}$	15.32 6.167	$5.514 \times 10^{-3}$	4.516
	10	$\langle \hat{A} \rangle = 1.995$ $\langle \hat{B} \rangle = 4.000$	$[1.712, 2.328]$ $[3.752, 4.248]$	$2.633 \times 10^{-1}$ $1.575 \times 10^{-3}$	30.95 12.42	$2.240 \times 10^{-2}$	4.936
	20	$\langle \hat{A} \rangle = 1.979$ $\langle \hat{B} \rangle = 4.000$	$[1.451, 2.721]$ $[3.752, 4.248]$	1.038 $1.180 \times 10^{-3}$	64.58 25.57	$9.557 \times 10^{-2}$	6.881
20	5	$\langle \hat{A} \rangle = 1.998$ $\langle \hat{B} \rangle = 4.000$	$[1.897, 2.105]$ $[3.915, 4.085]$	$9.321 \times 10^{-2}$ $6.335 \times 10^{-5}$	10.44 4.250	$2.405 \times 10^{-3}$	5.563
	10	$\langle \hat{A} \rangle = 1.993$ $\langle \hat{B} \rangle = 4.000$	$[1.794, 2.213]$ $[3.829, 4.171]$	$3.721 \times 10^{-1}$ $1.530 \times 10^{-4}$	21.04 8.550	$9.741 \times 10^{-3}$	7.131
	20	$\langle \hat{A} \rangle = 1.971$ $\langle \hat{B} \rangle = 4.000$	$[1.593, 2.442]$ $[3.651, 4.349]$	1.452 $7.109 \times 10^{-4}$	43.19 17.47	$4.075 \times 10^{-2}$	14.42
30	5	$\langle \hat{A} \rangle = 1.998$ $\langle \hat{B} \rangle = 4.000$	$[1.915, 2.084]$ $[3.931, 4.069]$	$1.031 \times 10^{-1}$ $2.773 \times 10^{-4}$	8.466 3.457	$1.549 \times 10^{-3}$	5.279
	10	$\langle \hat{A} \rangle = 1.992$ $\langle \hat{B} \rangle = 4.000$	$[1.829, 2.169]$ $[3.861, 4.139]$	$4.130 \times 10^{-1}$ $5.684 \times 10^{-5}$	17.05 6.956	$6.276 \times 10^{-3}$	7.767
	20	$\langle \hat{A} \rangle = 1.968$ $\langle \hat{B} \rangle = 4.000$	$[1.656, 2.340]$ $[3.717, 4.283]$	1.608 $5.238 \times 10^{-4}$	34.87 14.18	$2.612 \times 10^{-2}$	19.52

координатах Фаулера–Нордгейма, по которому чаще всего судят о величине работы выхода [4, 5]. Сетка fV (14), напротив, позволяет точнее судить о параметре  $A$ . Такой вывод, впрочем, можно было прогнозировать априори. Авторы акцентируют внимание на том, что планированию экспериментов [20] в области ПЭЭ можно уделить и большее внимание, чем это делается в современной литературе.

Совместные доверительные области (9) в поле параметров  $\theta = (A', B')^T$  представлены на рис. 3. Результаты приведены для количества наблюдений равных 10, 20 и 30 точкам с уровнем шума 20%. Сплошную заливку имеют эллипсы до сетки bV, работа с которой обеспечивает дополнительно меньшую среднюю площадь  $\langle S_\theta \rangle$ .

Из табл. 1 и 2 можно сделать вывод о том, что величины  $\bar{\gamma}_j^{(1)}$  также оказались меньшими для сетки наблюдений bV. Поэтому эволюция средних, минимальных и максимальных показателей качества параметрической идентификации сигнала ПЭЭ была построена только для равноотстоящих

значений факторов  $X$  – рис. 4–6. Впрочем, сетка fV дает схожие иллюстрации поведения  $\bar{\gamma}_j^{(1)}, \bar{\gamma}_j^{(2)}$ ,



**Рис. 3.** Совместные доверительные области.

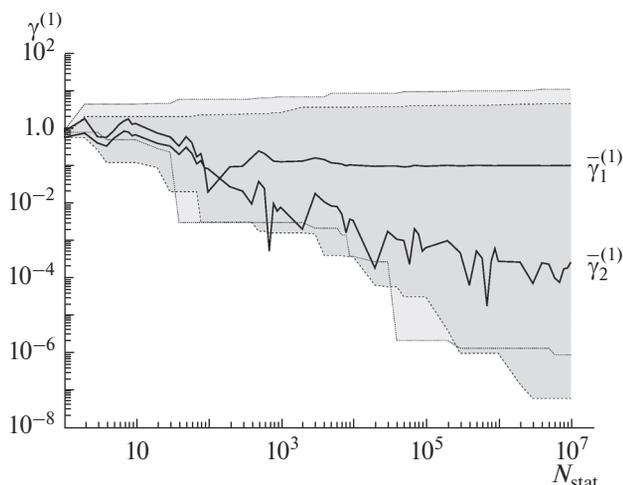


Рис. 4. Эволюция поведения  $\bar{\gamma}_j^{(1)}$ .

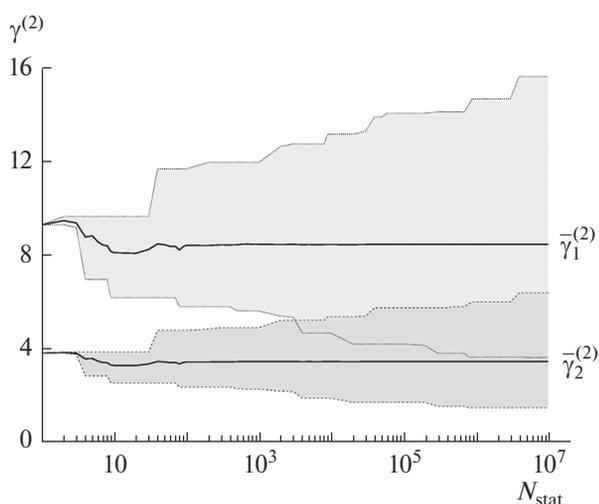


Рис. 5. Эволюция поведения  $\bar{\gamma}_j^{(2)}$ .

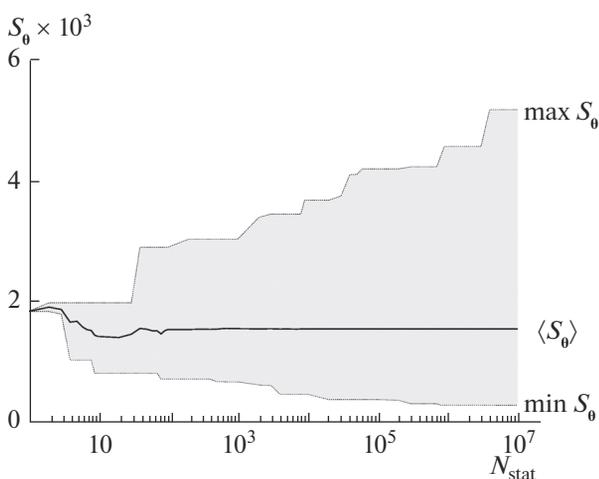


Рис. 6. Эволюция поведения  $\langle S_{\theta} \rangle$ .

$\langle S_{\theta} \rangle$  и их экстремальных значений в зависимости от количества статистических испытаний  $N_{stat}$ .

Рис. 4 демонстрирует не очень удачное поведение характеристики  $\bar{\gamma}_j^{(1)}$  в силу особенности ее построения. Можно наблюдать ситуацию, когда текущее значение относительной погрешности усредненного результата оказалось меньше минимальной из не усредненных. Впрочем, данная величина позиционировалась как имеющая смысл только для компьютерного эксперимента, что, возможно, позволяет простить ей такого рода недостатки.

На рис. 5 можно видеть прогресс характеристики  $\bar{\gamma}_j^{(2)}$ . Здесь усредненное должным образом отношение ширины индивидуального доверительного интервала к точечной оценке параметра ведет себя более предсказуемо. Можно отметить, что при  $N_{stat} \geq 10^2$  значение  $\bar{\gamma}_j^{(2)}$  перестает меняться ощутимо.

На рис. 6 представлена зависимость площади совместной доверительной области для параметров  $\theta = (A', B')^T$ . Выход на стабильное усредненное значение наблюдается при том же объеме статистических испытаний:  $N_{stat} \geq 10^2$ . По всей видимости, показатели  $\bar{\gamma}_j^{(2)}$  и  $\langle S_{\theta} \rangle$  лучше отражают результаты компьютерного эксперимента. Что касается экстремальных значений, то, как видно из представленных данных, для одной случайной выборки данных можно получить как очень хорошие, так и крайне далекие от исходных значений параметры оценки. С ростом  $N_{stat}$  вероятность наличия таких выборок, естественно, повышается.

### ЗАКЛЮЧЕНИЕ

В работе проанализирована регрессионная модель с двумя параметрами на основе закона Фаулера–Нордгейма. Для оценки усредненных характеристик доверительных областей при оценке параметров зависимости отклика от напряжения проведен статистический эксперимент. Построены границы индивидуальных доверительных интервалов и их центры, совместные доверительные области и их площади, относительная погрешность определения значений параметров, характеристика качества интервальной оценки. Для значений напряжения были использованы планы двух типов. В качестве погрешности значений силы тока была рассмотрена модель шума специального вида. Для формирования статистик были использованы только те данные, остатки которых прошли тест Шапиро–Уилка на принадлежность нормальному закону распределения, и регрессионные модели для которых можно считать значимыми. Все статистические

заклучения проведены на уровне значимости 5% (с доверительной вероятностью 95%). На основе полученных результатов можно сделать вывод о том, что для представления об усредненных значениях площади доверительной области (10) и качества интервальной оценки (13) вполне достаточно сотни испытаний. Также в рамках модели погрешности специального вида (для “постоянного” значения отношения интенсивностей сигнал/шум) можно ожидать формирования таких наблюдений, которые не будут удовлетворять тесту на нормальность. Процент таких выборок примерно соответствует уровню шума в откликах.

### СПИСОК ЛИТЕРАТУРЫ

1. *Егоров Н.В., Шешин Е.П.* // Поверхность. Рентген., синхротр. и нейтрон. исслед. 2017. № 3. С. 5.
2. *Antonov A.Yu., Denisov V.P.* // Physica Scripta. 1999. V. 59. P. 235.
3. *Шешин Е.П.* Структура поверхности и автоэмиссионные свойства углеродных материалов. Москва: Издательство МФТИ: Физматкнига, 2001. 288 с.
4. *Егоров Н.В., Шешин Е.П.* Автоэлектронная эмиссия. Принципы и приборы. Долгопрудный: Издательский Дом “Интеллект”, 2011. 704 с.
5. *Фоменко В.С.* Эмиссионные свойства материалов. Справочник. Изд. 4-е, перераб. и дополн. Киев: Наукова Думка, 1981. 340 с.
6. *Forbes R.G.* // J. Vac. Sci. Technol. B. 1999. V. 17. № 2. P. 526.
7. *Giubileo F., Di Bartolomeo A., Iemmo L. et al.* // Nanomaterials. 2017. V. 7. № 9. 275. P. 1.
8. *Rawal I., Kumar L., Tripathi R.K., Panwar O.S.* // ACS Omega. 2017. V. 2. P. 7515.
9. *Ramaneti R., Sankaran K.J., Korneychuk S. et al.* // APL Materials. 2017. V. 5. 066102. P. 1.
10. *Антонов А.Ю., Вараюнь М.И., Егоров Н.В.* // Нано- и микросистемная техника. 2019. Т. 21. № 2. С. 103.
11. *Егоров Н.В., Антонов А.Ю., Вараюнь М.И.* // Поверхность. Рентген., синхротр. и нейтрон. исслед. 2018. № 10. С. 72.
12. *Дрейпер Н., Смит Г.* Прикладной регрессионный анализ. Изд. 2-е, перераб. и дополн. М.: Финансы и статистика, 1986. 366 с.
13. *Демиденко Е.З.* Линейная и нелинейная регрессии. Москва: Финансы и статистика, 1981. 304 с.
14. *Буре В.М., Париллина Е. М., Седаков А.А.* Методы прикладной статистики в R и Excel. Санкт-Петербург: Лань, 2016. 152 с.
15. *Хан Г., Шапиро С.* Статистические модели в инженерных задачах / Хан Г., Шапиро С. Москва: Мир, 1969. 395 с.
16. *Matsumoto M., Nishimura T.* // ACM T. Model. Comp. S. 1998. V. 8. № 1. P. 3.
17. *Press W.H., Teukolsky S.A., Vetterling W.T., Flannery B.P.* Numerical Recipes. The Art of Scientific Computing. 3rd ed. Cambridge University Press, 2007. 1256 p.
18. *Вадзинский Р.Н.* Справочник по вероятностным распределениям. Санкт-Петербург: Наука, 2001. 295 с.
19. *Antonov A.Y., Rusina Y.I., Varayun' M.I.* // 14th International Baltic Conference on Atomic Layer Deposition (BALD). St. Petersburg, 2016. P. 38.
20. *Ермаков С.М., Жигляевский А.А.* Математическая теория оптимального эксперимента. Москва: Наука, 1987. 320 с.

## The Analysis of Confidence Regions for Regression Model Parameters Based on the Fowler–Nordheim Law

N. V. Egorov<sup>1,\*</sup>, A. Yu. Antonov<sup>1</sup>, M. I. Varayun'<sup>1</sup>

<sup>1</sup>St. Petersburg State University, St. Petersburg, 199034 Russia

\*e-mail: robby7@mail.ru

The two-parameter regression model based on the Fowler–Nordheim formula, which is classical for the field electron emission, is considered in the paper. Optimal values of parameters are determined by the least squares method. As part of the statistical experiment, characteristics of confidence regions are evaluated. The attention is paid to the accuracy of parameter estimates, boundaries of individual confidence intervals, joint confidence ellipses. For voltage values, two types of the experiment design are used. The errors in current measurements are generated using a noise model of a special type. During the computer experiment, the data is subjected to the Shapiro–Wilk test, which checks the normality of their distribution, and the regression model is tested for significance. The percentage of samples that may not meet the specified criteria is revealed.

**Keywords:** field electron emission, current–voltage characteristic, regression model, least-squares method, confidence regions.