УДК 543.552.054.1

АЛГОРИТМ ОБНАРУЖЕНИЯ СЕРИИ ВЫБРОСОВ ПО КРИТЕРИЮ ДИКСОНА В ИНВЕРСИОННОЙ ВОЛЬТАМПЕРОМЕТРИИ

Ларин С. Л., Кузнецов В. В., Романенко С. В.

Томский политехнический университет,
Российская Федерация, 634050, г. Томск, проспект Ленина, дом 30
kuvv@tpu.ru

Поступила в редакцию 25 октября 2013 г.,
после доработки – 24 июня 2014 г.

В вольтамперометрии большую роль играет выделение полезного сигнала в серии вольтамперных кривых, снятых в одних и тех же экспериментальных условиях. В силу влияния на сигнал различного рода помех в процессе регистрации вольтамперограмм, некоторые кривые могут быть сильно искажены и их необходимо исключить из дальнейшего рассмотрения. От правильности исключения вольтамперных кривых напрямую зависит результат определения массовой доли вещества в аналите, вычисляемый на основе способа «введено – найдено». В настоящий момент процесс исключения кривых производится оператором, который визуально оценивает, какие кривые необходимо исключить из серии, что бывает довольно сложно и не всегда надежно.

Целью работы является попытка выявления непригодных для дальнейшей обработки вольтамперных кривых. В статье предложен алгоритм применения критерия Диксона для обнаружения серии выбросов в инверсионной вольтамперометрии. Показана возможность применения предложенного алгоритма к выявлению аномальных кривых в серии измерений, на примере определения концентрации Cd в питьевой воде.

Используемый алгоритм был апробирован на данных, полученных с помощью прибора ТА-07 фирмы ООО «НПП Техноаналит» (г. Томск) и реализован в программе Valab Professional, поставляемой в комплекте с этим прибором.

***Ключевые слова*: с**татистика, критерий Диксона, аномальное значение, выбросы, инверсионная вольтамперометрия.

Кузнецов Виталий Владимирович – аспирант Томского политехнического университета.

Романенко Сергей Владимирович – доктор химических наук, профессор, заведующий кафедрой экологии и безопасности жизнедеятельности Томского политехнического университета.

Ларин Сергей Леонидович – кандидат химических наук, доцент Томского политехнического университета.

Введение

При решении задач статистической обработки результатов химического анализа часто возникает проблема выявления выбросов в полученных результатах (здесь и далее по тексту под выбросами понимаются аномальные значения в наборе экспериментальных данных). Выбросы возникают, как правило, из-за небрежного обращения с аппаратурой и сбоями в ее работе, скачками напряжения в электрической сети, кратковременными резкими изменениями условий измерений и многими другими причинами, зависящими от особенностей измерений. Присутствие выбросов может приводить к существенному искажению конечных концентраций компонентов, находимых в результате обработки таких выборок. В борьбе с выбросами используют два подхода [1]:

- исключение аномальных значений;
- использование робастных методов обработки.

Для увеличения вероятности выявления выбросов измерения одной и той же величины проводят многократно. В этом случае используют статистические критерии. Большинство существующих критериев отбраковки аномальных значений опираются на предположение о принадлежности наблюдаемых случайных величин закону нормального распределения, например: критерий Грабса [1-3], критерий Ирвина [5], критерий «трех сигм» [4, 5], критерий Романовского [5,6], критерий Шовине [6-9], критерий Пирсона [10]. Если же закон распределения случайной величины не известен, то воз-

можно применение критерия Диксона (*Q*-критерий) [5, 6, 11-17]. В работе [6] показано, что критерий Диксона не зависит от параметров распределения и его можно использовать для выявления аномальных значений при произвольном характере распределения исходных данных, а в работах [12-14] показано, что критерий Диксона можно использовать при малом числе экспериментальных данных (меньше 30).

В инверсионной вольтамперометрии при регистрации серии вольтамперных кривых возникает необходимость автоматизированного исключения невоспроизводимых (резко отличающихся от остальных) кривых вольтамперограмм. Аномальные вольтамперные кривые получаются, например, вследствие перемешивания раствора, когда на электродах образуются пузырьки воздуха. При этом могут появляться схожие по форме и высоте пика аномальные кривые в одной серии эксперимента. Число же получаемых вольтамперных кривых при проведении рутинных анализов, не превышает 10 (из-за длительности подготовительных этапов перед регистрированием каждой вольтамперограммы). Например, на рис. 1 показана типичная серия вольтамперных кривых (сплошные линии серого цвета), полученная при определении кадмия и свинца в реальных пробах питьевой воды по методике [18]. Из рисунка видно, что три вольтамперограммы являются аномальными по сравнению с остальными. Кроме того, все рассмотренные выше критерии применимы к одномерным выборкам, в то время как каждое вольтамперометрическое измерение представляет собой одномерный аналитический сигнал, что в ходе проведения анализа дает двумерный набор данных.

Таким образом, целью данной работы является выбор критерия сравнения вольтамперограмм и создание автоматизированной процедуры исключения одиночных и групповых аномальных измерений в вольтамперометрическом анализе на примере определении массовой концентрации кадмия в пробах пищевых продуктов.

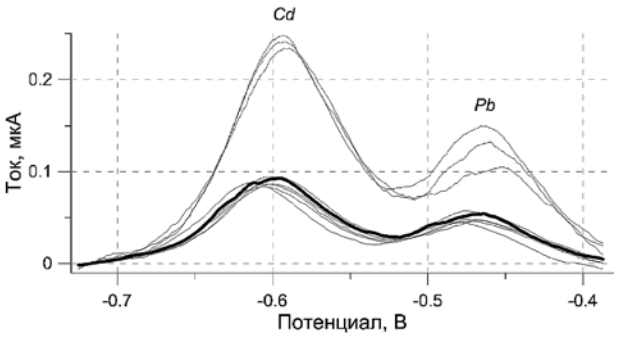

Рис. 1. Серия вольтамперограмм (тонкие кривые), полученная при определении кадмия и свинца в реальных пробах питьевой воды по методике [18], и медиана (жирная кривая)

Исключение аномальных вольтамперограмм в серии измерений

Для исключения аномальных вольтамперограмм необходимо для каждой рассчитать критерий, позволяющий количественно выразить степень ее отклонения от других в серии. Нами предлагается использовать критерий подобия *D*, который представляет собой усредненную сумму абсолютных отклонений вольтамперограммы относительно медианы $\tilde{I}_j$ (на рис. 1 медиана изображена жирной линией). Выбор медианы обусловлен ее меньшей чувствительностью к выбросам. Каждая точка кривой $\tilde{I}_j$ рассчитывается как медиана в *j*-ом отсчёте для всех вольтамперограмм.

Для каждой *i*-ой вольтамперограммы в серии критерий D_i рассчитывается следующим образом:

$$D_i = \frac{1}{n}\sum_{j=1}^{n}\left|I_{i,j} - \tilde{I}_j\right| \quad , \qquad (1)$$

где $I_{i,j}$ – *j*-ый отсчет *i*-ой вольтамперограммы, $\tilde{I}_j$ – *j*-ый отсчет медианы, *n* – число вольтамперограмм. Чем меньше значение критерия *D* для вольтамперограммы, тем меньше она отклоняется от медианы. Из формулы видно, что значение *D* находится в диапазоне [0, +∞].

На рис. 2 показан вектор значений *D,* вычисленных по формуле на основе исходных вольтамперных кривых (рис. 1). Из рис. 2 видно, что для трех аномальных вольтамперограмм значение критерия *D* существенно отличается от остальных значений.

Наиболее подходящим статистическим критерием для анализа выбросов в векторе *D* является критерий Диксона, т. к. он непараметрический и позволяет работать с малой выборкой [6, 11-17]. При использовании критерия Диксона полученные результаты записывают в вариационный возрастающий ряд:

$$D_1, D_2, \ldots D_n \; (D_1 < D_2 < \ldots < D_n) .$$

Особенностью критерия подобия формы *D* является то, что аномальными значениями могут

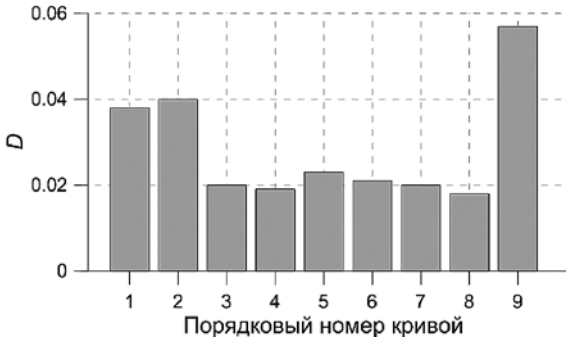

Рис. 2. Вектор подобия *D* для вольтамперограмм, полученных при определении массовых концентраций кадмия и свинца в реальных пробах питьевой воды по методике [18]

быть только крайние правые значения вариационного ряда. Крайние левые значения могут быть сколь угодно малыми и будут лишь означать высокую степень «похожести» вольтамперограмм на усреднённую кривую, полученную по формуле (1).

В этом случае значение одностороннего *Q*-критерия рассчитывается по формуле [13, 14]:

$$Q = \begin{cases} \left| \dfrac{D_n - D_{n-1}}{D_n - D_1} \right|, & n = 3 \ldots 7 \\ \left| \dfrac{D_n - D_{n-1}}{D_n - D_2} \right|, & n = 8 \ldots 10 \end{cases} \quad . \qquad (2)$$

Вычисленное значение *Q*-критерия сравнивается с критериальным значением $Q_{крит}$. Нулевая гипотеза об отсутствии грубой погрешности выполняется, если выполняется неравенство $Q < Q_{крит}$. Если $Q > Q_{крит}$, то последнее значение в вариационном ряду признается грубой погрешностью и исключается из дальнейшей обработки. Критериальные значения $Q_{крит}$ при заданной доверительной вероятности *p* и числе измерений *n* берутся из соответствующих таблиц [12, 13, 15].

Критерий Диксона, рассчитанный по формуле (2), работает только тогда, когда в выборке присутствует одиночный выброс. При повторном применении критерия к вариационному ряду можно исключить последовательность одиночных выбросов, как показано на рис. 3, *а*.

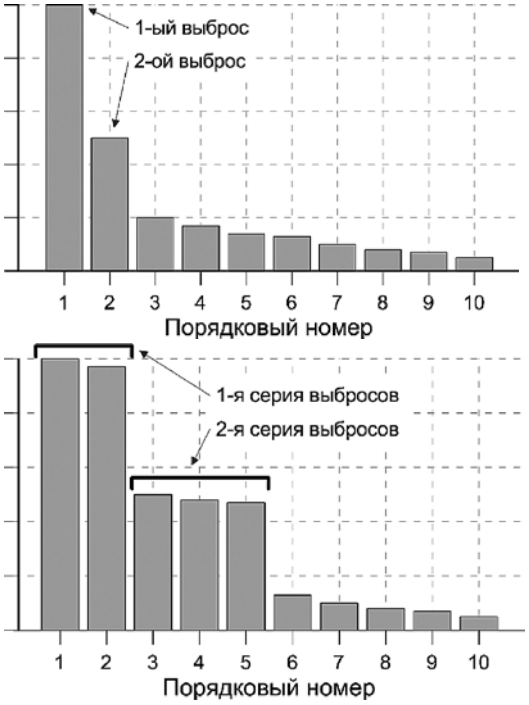

Рис. 3. Пример вариационного ряда с одиночными выбросами (а) и сериями выбросов (б)

Если в выборке есть последовательности выбросов с двумя и более значениями (рис. 3, *б*), то критерий Диксона, рассчитанный по формуле , может давать положительно ложный результат, поскольку сравниваются два соседних элемента (D_n

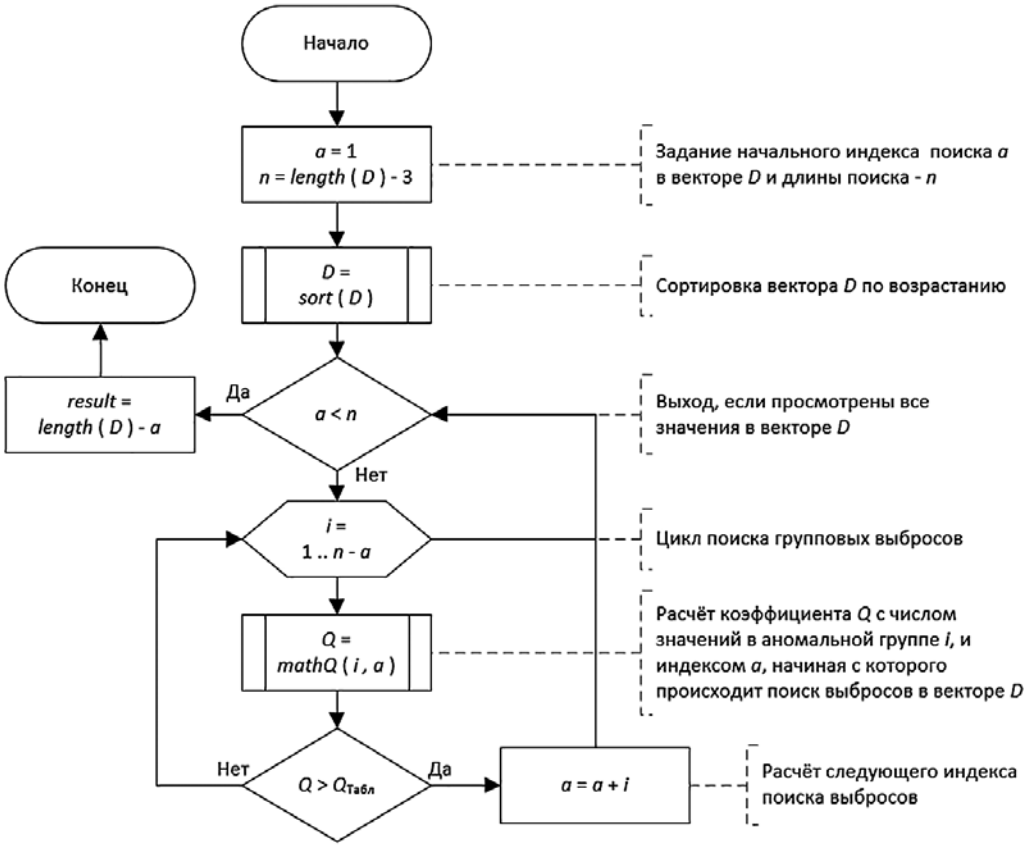

Рис. 4. Блок-схема алгоритма исключения аномальных значений в векторе *D*

и D_{n-1}) и их разность может быть мала, что приведёт к занижению критерия. Как показано в работе [6], формулу (2) можно модифицировать и для групповых выбросов, введя длину группы *L*. В результате, формула будет иметь вид:

$$Q_L = \begin{cases} \left| \dfrac{D_n - D_{n-1-L}}{D_n - D_L} \right|, & n = 3 \ldots 7 \\ \left| \dfrac{D_n - D_{n-1-L}}{D_n - D_{L+1}} \right|, & n = 8 \ldots 10 \end{cases}, \quad 0 < L < n-3. \qquad (3)$$

Таким образом, используя формулу , можно исключать серии выбросов. Применяя формулу многократно к вариационному ряду, можно устранить все серии выбросов. Блок-схема алгоритма выявления выбросов в серии кривых показана на рис. 4. При завершении алгоритма получаем вектор *D* без выбросов. После каждого исключения серии аномальных кривых необходимо перестраивать медиану, а, следовательно, и вектор *D*. На рис. 5 показан вектор *D*, рассчитанный для вольтамперограмм, представленных на рис. 1, после применения алгоритма исключения выбросов.

Результаты и обсуждение

Регистрация вольтамперограмм, обработка результатов и определение массовой концентрации кадмия проводили по методике [] на анализаторе ТА-07 (производства ООО «НПЦ Техноаналит», г. Томск). Для проверки алгоритма была проведена серия экспериментов в модельных растворах и в реальных пробах способом «введено - найдено». Определение проводили методом добавок при линейном учете базовой линии. В каждом эксперименте было получено по пять кривых фона, пробы и пробы с добавкой.

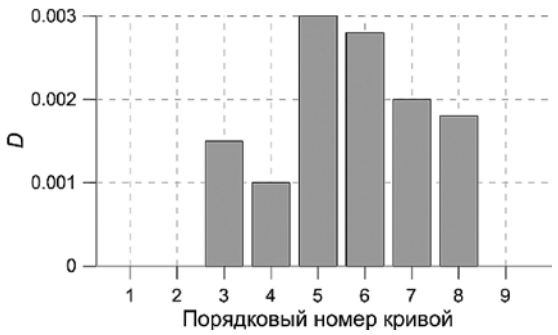

Рис. 5. Вектор подобия *D* для вольтамперограмм, полученных при определении массовых концентраций кадмия и свинца в реальных пробах питьевой воды по методике [18] после исключения аномальных значений

В таблице приведены значения концентраций ионов Cd в модельных растворах и в реальных пробах. Таблица содержит столбец без применения предложенного алгоритма исключения вольтамперограмм и с его применением. В таблице показаны только те результаты, где алгоритм обнаружил выбросы на любом этапе эксперимента (при регистрации фона, пробы или пробы с добавкой). В модельных растворах аномальные кривые были обнаружены в 5 экспериментах из 10. В реальных пробах - в 7 экспериментах из 10.

Как видно из таблицы, применение алгоритма дает существенное приближение результата к значению введенной концентрации. Причем результаты расчета концентрации без применения алгоритма исключения вольтамперограмм могут быть как занижены, так и завышены, т. к. выброс может происходить на любом этапе изменения потенциала при регистрации вольтамперограмм (фона, пробы или пробы с добавкой).

Выводы

Разработан автоматизированный алгоритм обнаружения серии выбросов с использованием критерия Диксона в инверсионной вольтамперометрии. Предложенный алгоритм позволяет проводить исключение одиночных и групповых аномальных измерений для серии вольтамперограмм, основываясь на степени подобия формы. Результаты применения алгоритма на примере определения содержания Cd в питьевой воде показали существенное приближение результата к значению введенной концентрации при наличии аномальных кривых в серии измерений. Алгоритм реализован в программе *Valab professional* для прибора ТА-07 фирмы ООО «НПП Техноаналит», г. Томск.

Таблица 1

Найденные методом стандартных добавок значения концентрации Cd в модельных растворах и в реальных пробах при введенной концентрации 0.0100, представленные в виде $(\bar{x} \pm \Delta x) \cdot 10^2$ мг/л. Значения рассчитаны без исключения кривых и с исключением кривых по критерию Диксона

Модельный раствор		Вода природная питьевая бутилированная «Афины» (не йодированная), производства ОАО «Томское пиво», г. Томск	
Без исключения кривых	С исключением кривых	Без исключения кривых	С исключением кривых
1.23 ± 0.22	1.13 ± 0.09	1.25 ± 0.31	1.13 ± 0.03
1.95 ± 0.09	1.14 ± 0.07	1.14 ± 0.27	1.06 ± 0.08
0.83 ± 0.12	0.92 ± 0.04	0.87 ± 0.11	0.96 ± 0.09
0.91 ± 0.21	0.99 ± 0.005	1.12 ± 0.17	1.10 ± 0.09
2.02 ± 0.38	1.22 ± 0.1	0.85 ± 0.11	0.98 ± 0.03

ЛИТЕРАТУРА

1. Лемешко Б.Ю. Робастные методы оценивания и отбраковка аномальных измерений // Заводская лаборатория. 1997. Т.63, № 5. С. 43-49.
2. Grubbs F.E. Sample Criteria for Testing Outlying observations // Annals of Mathematical Statistics. 1950. V. 63, № 1. P. 27-58.
3. Grubbs F.E. Procedures for Detecting Outlying Observations in Samples // Technometrics. 1969. V. 11, №. 1. P. 1-21.
4. Новицкий П.В., Зограф И.А. Оценка погрешностей результатов измерений. Л.: Энергоатомиздат, 1991. 304 с.
5. Третьяк Л. Н. Обработка результатов наблюдений: Учебное пособие. Оренбург: ГОУ ОГУ, 2004. 171 с.
6. Марчук В.И., Токарева С.В. Способы обнаружения аномальных значений при анализе нестационарных случайных процессов. Шахты.: ЮРГУС, 2009. 209 с.
7. Taylor J.R. An introduction to error analysis. The study of uncertainties in physical measurements. University Science Books. Mill Valley, CA.: 1982. 270 p.
8. Большев Л.Н., Убайдуллаева М. Критерий Шовенэ в классической теории ошибок // Теория вероятностей и ее применение. 1974. Т. XIX, вып. 4. С. 714-723.
9. Пагурова В.И. О критерии Шовэне для обнаружения нескольких выбросов // Теория вероятностей и ее применение. 1985. Т. 30, № 3. С. 558-561.
10. Pearson E.S., Chandra S.C. The efficiency of statistical tools and a criterion for rejection of outlying observations // Biometrika. 1936. V. 28, № ¾. P. 308-320.
11. Dixon W.J. Analysis of Extreme Values // Ann. Math. Stat. 1950. V. 21, № 4. P. 488-506.
12. Dixon W.J. Ratios Involving Extreme Values // Ann. Math. Stat. 1951. V. 22, № 1. P. 68-78.
13. Дёрффель К. Статистика в аналитической химии. Пер. с нем. М.: Мир, 1994. 286 с.
14. Dean R.B., Dixon W.J. Simplified Statistic of Small Numbers of Observations // Anal. Chem. 1951. №. 23. P. 636-638.
15. Rorabacher D.B. Statistical Treatment for Rejection of Deviant Values: Critical Values of Dixon Q Parameter and Related Subrange Ratios at the 95 percent Confidence Level // Anal. Chem. 1991. V. 63. P. 139-146.
16. Efstathiou C.E. Stochastic calculation of critical Q-test values for the detection of outliers in measurements // Journal of Chemical Education. 1992. V. 69. P. 733.
17. Efstathiou C. E. Estimation of type I error probability from experimental Dixon's "Q" parameter on testing for outliers within small size data sets // Talanta. 2006. V. 69. P. 1068-1071.
18. МУ 31-04/04. Количественный химический анализ проб пищевых продуктов, продовольственного сырья, кормов и продуктов их переработки. Методика выполнения измерений массовых концентраций цинка, кадмия, свинца и меди методом инверсионной вольтамперометрии на анализаторах типа ТА. Томск: ООО «Томьаналит», 2004. 19 с.

DETECTION ALGORITHM OF A SERIES OF RELEASES BY DIKSON CRITERION IN INVERSION VOLTAMMETRY

S.L. Larin, V.V. Kuznetsov, S.V. Romanenko

Tomsk Polytechnic University
ul. Lenina, 30 Tomsk, 634050, Russian Federation
kuvv@tpu.ru

The extraction of desired signal in the series of voltammetric curves, plotted in the same conditions, plays an important role in voltammetry. By virtue of effect on the signal of various noise, during the voltammogram registration, some curves can be greatly deformed and they must be excluded from the further consideration. The result of mass fraction of substance in the analyte, obtained by the method of "introduced-found", depends directly on the correctness of the voltammetric curves elimination. Currently, the laboratory assistant is responsible for the process of curves elimination, who estimates visually which curves should be excluded from the series, that it is quite difficult to do and this estimation depends on his qualification.

The aim of this work is the attempt to detect unsuitable for further analysis of the voltammetric curves. The article considers the algorithm of the application of Dixon test for outliers in stripping voltammetry. The proposed algorithm for the abnormal curves estimation in the series of experiments was used to determine the concentration of Cd in drinking water.

This algorithm was tested on the data, obtained by the analyzer TA-07, produced by the Tomsk Company OOO "NPP Tehnoanalit" and implemented in the program Valab Professional, supplied with this device.

Keywords: Statistics, criterion of Dixon, *Q*-test, outliers, stripping voltammetry.

REFERENCES

1. Lemeshko B. Iu. [Robust estimation and rejection of anomalous measurements]. *Zavodskaia laboratoriia* [Plant Laboratory], 1997, vol. 63, no. 5, pp. 43-49 (in Russian).
2. Grubbs F.E. Sample Criteria for Testing Outlying observations. *Ann. Math. Statist*., 1950, vol. 63, no. 1, pp. 27-58.
3. Grubbs F.E. Procedures for Detecting Outlying Observations in Samples. *Technometrics,* 1969, vol. 11, no. 1, pp. 1-21.
4. Novitskii P.V., Zograf I.A. *Otsenka pogreshnostei rezul'tatov izmerenii* [Estimation of errors of measurement results]. Leningrad: Energoatomizdat, 1991, 304 p. (in Russian).
5. Taylor J.R. An introduction to error analysis. The study of uncertainties in physical measurements. University Science Books. *Mill Valley, CA*., 1982, 270 p.
6. Marchuk V.I., Tokareva S.V. *Sposoby obnaruzheniia anomal'nykh znachenii pri analize nestatsionarnykh sluchainykh protsessov* [Methods for detecting the abnormal

value in the analysis of non-stationary random processes]. Shakhty: IuRGUS, 2009, 209 p. (in Russian).
7. Teilor Dzh. *Vvedenie v teoriiu oshibok* [Introduction to the theory of errors]. M.: Mir, 1985. 272 p. (in Russian).
8. Bol'shev L.H., Ubaidullaeva M. [The Chauvenet criterion in the classical theory of errors]. *Teoriia veroiatnostei i ee primenenie* [Theory of Probability and its Applications], 1974, vol. 19, no. 4. pp. 714-723 (in Russian).
9. Pagurova V.I. [About Shovene criteria for detecting several outliers]. *Teoriia veroiatnostei i ee primenenie* [Theory of Probability and its Applications], 1985, vol. 30, no 3, pp. 558-561 (in Russian).
10. Pearson E.S., Chandra S.C. The efficiency of statistical tools and a criterion for rejection of outlying observations. *Biometrika,* 1936, vol. 28, no. ¾, pp. 308-320.
11. Dixon W.J. Analysis of Extreme Values. *Ann. Math. Stat*., 1950, vol. 21, no. 4, pp. 488-506.
12. Dixon W.J. Ratios Involving Extreme Values. *Ann. Math. Stat*., 1951, vol. 22, no. 1, pp. 68-78.
13. Derffel' K. *Statistika v analiticheskoi khimii* [Statistics in Analytical Chemistry]. M.: Mir, 1994, 286 p. (in Russian).
14. Dean R.B., Dixon W.J. Simplified Statistic of Small Numbers of Observations. *Anal. Chem*., 1951, vol. 23, pp. 636-638.
15. Rorabacher D.B. Statistical Treatment for Rejection of Deviant Values: Critical Values of Dixon Q Parameter and Related Subrange Ratios at the 95 percent Confidence Level. *Anal. Chem*., 1991, vol. 63, pp. 139-146.
16. Efstathiou C.E. Stochastic calculation of critical Q-test values for the detection of outliers in measurements. *Journal of Chemical Education*., 1992. vol. 69, p. 733.
17. Efstathiou C. E. Estimation of type I error probability from experimental Dixon's "Q" parameter on testing for outliers within small size data sets. *Talanta*, 2006, vol. 69, pp. 1068-1071.
18. MU 31-04/04. *Kolichestvennyi khimicheskii analiz prob pishchevykh produktov, prodovol'stvennogo syr'ia, kormov i produktov ikh pererabotki. Metodika vypolneniia izmerenii massovykh kontsentratsii tsinka, kadmiia, svintsa i medi metodom inversionnoi vol'tamperometrii na analizatorakh tipa TA* [Quantitative chemical analysis of samples of food, food raw materials, feed and products. Methods for measuring mass concentrations of zinc, cadmium, lead and copper by anodic stripping voltammetry on the type of TA analyzers]. Tomsk. Tomanalyt Ltd. 2004. 19 p. (in Russian).