ВОССТАНОВЛЕНИЕ ПЛОТНОСТИ РАСПРЕДЕЛЕНИЯ ЧАСТИЦ ДИСПЕРСНЫХ МАТЕРИАЛОВ МЕТОДОМ ОКНА ПАРЗЕНА—РОЗЕНБЛАТТА

Е.В. Тарарушкин

Российский университет транспорта (МИИТ), 127994, г. Москва, ул. Образцова, д. 9, стр. 9

АННОТАЦИЯ: Предмет исследования: представлено описание и возможность применения метода окна Парзена—Розенблатта, относящегося к методам непараметрической статистики, для восстановления эмпирической плотности распределения частиц дисперсного материала — уплотненного микрокремнезема.

Цели: описание и реализация метода окна Парзена—Розенблатта для получения с его помощью эмпирической плотности распределения диаметров частиц уплотненного микрокремнезема, а также выполнение сравнительного анализа метода гистограмм с методом окна Парзена—Розенблатта при оценке распределения диаметров частиц уплотненного микрокремнезема.

Материалы и методы: микрокремнезем является побочным продуктом металлургической промышленности, и применяется в качестве пуццолановой добавки для изготовления различных видов бетона. Уплотненный микрокремнезем состоит из сферических частиц-кластеров, образованных из отдельных частиц микрокремнезема. Для реализации метода окна Парзена—Розенблатта в качестве весовых (ядерных) функций применяются гауссовы функции. Метод Шезора—Джоунса, относящийся к методам подстановки, используется для вычисления оптимальной ширины окна ядерных функций. Нелинейное уравнение в методе Шезора—Джоунса для поиска оптимальной ширины окна решается численно с помощью метода Ньютона. Реализации методов выполнены на языке программирования математической программы Matlab.

Результаты: с помощью полученной реализации метода окна Парзена—Розенблатта восстановлена эмпирическая плотность распределения диаметров частиц уплотненного микрокремнезема. Приведен сравнительный анализ метода окна Парзена—Розенблатта и метода гистограмм на примере восстановления плотности распределения диаметров частиц уплотненного микрокремнезема.

Выводы: применение метода окна Парзена—Розенблатта позволяет решить проблемы, которые возникают при использовании методов параметрической статистики и метода гистограмм при восстановлении эмпирической плотности распределения частиц дисперсных материалов. В частности, отпадает необходимость в назначении неизвестной статистики для методов параметрической статистики и определения количества интервалов для метода гистограмм. Плотности распределения, полученные методом окна Парзена—Розенблатта, можно применять для статистического моделирования физико-механических свойств строительных материалов.

КЛЮЧЕВЫЕ СЛОВА: непараметрическая статистика, плотность распределения, метод окна Парзена—Розенблатта, ядерные функции, диаметр частиц, уплотненный микрокремнезем

ДЛЯ ЦИТИРОВАНИЯ: Тарарушкин Е.В. Восстановление плотности распределения частиц дисперсных материалов методом окна Парзена—Розенблатта // Вестник МГСУ. 2018. Т. 13. Вып. 7 (118). С. 855–862. DOI: 10.22227/1997-0935.2018.7.855-862

RECONSTRUCTING DISTRIBUTION DENSITY OF PARTICLES FOR DISPERSE MATERIALS BY THE PARZEN—ROSENBLATT WINDOW METHOD

E.V. Tararushkin

Russian University of Transport (MIIT), building 9, 9 Obratsova st., Moscow, 127994, Russian Federation

ABSTRACT: Subject: the article contains the description and the possibility of using the Parzen—Rosenblatt window method which belongs to the methods of non-parametric statistics for estimation of empirical density of distribution of particles for disperse materials — compacted microsilica (silica fume). As a dispersed material, the compacted silica fume is considered. Microsilica is a byproduct of the metallurgical industry and is used as a pozzolanic additive for the manufacture of various types of concrete. The compacted silica fume consists of spherical particle-clusters formed of individual silica fume particles.

Research objectives: description and implementation of the Parzen—Rosenblatt window method for obtaining empirical density distribution function for the diameter of particles of compacted silica fume; comparison of the histogram method with the Parzen—Rosenblatt window method in estimating the distribution of the diameter of particles of compacted silica fume.

Materials and methods: the Gaussian weight (kernel) functions are used to implement the Parzen—Rosenblatt window method. The Sheather-Jones plug-in method is used to find the optimal bandwidth of the kernel functions. In the Sheather-Jones method, the non-linear equation for finding the optimal bandwidth is solved numerically using the Newton's method. Implementation of the methods is performed in the programming language of the numerical computing environment Matlab.

Results: the Parzen—Rosenblatt window method was described and implemented, and by implementing this method, the estimate of empirical density of distribution of the diameter of particles for compacted silica fume was obtained. A comparison of the Parzen—Rosenblatt window method and the histogram method is also given, for example, by reconstructing the density distribution of the diameter of particles of compacted silica fume.

Conclusions: application of the Parzen—Rosenblatt window method allows us to solve the problems that arise when using the methods of parametric statistics and the histogram method in estimating the empirical distribution density of particles of disperse materials. In particular, there is no need to assign unknown statistics for the methods of parametric statistics and determine the number of intervals for the histogram method. Density distributions obtained by the Parzen—Rosenblatt window method can be used for statistical modeling of physical and mechanical properties of building materials.

KEY WORDS: non-parametric statistics, probability density function, the Parzen—Rosenblatt window method, kernels, diameter of particles, compacted silica fume

FOR CITATION: Tararushkin E.V. Vosstanovlenie plotnosti raspredeleniya chastits dispersnykh materialov metodom okna Parzena-Rozenblatta [Reconstructing distribution density of particles for disperse materials by the Parzen—Rozenblatt window method]. Vestnik MGSU [Proceedings of the Moscow State University of Civil Engineering]. 2018, vol. 13, issue 7 (118), pp. 855–862. DOI: 10.22227/1997-0935.2018.7.855-862

ВВЕДЕНИЕ

Для восстановления эмпирической плотности распределения случайной величины используют параметрические и непараметрические методы математической статистики.

Основной задачей методов параметрической статистики является поиск неизвестного истинного параметра статистики θ на основе экспериментальных данных. При этом для поиска параметра θ необходимо выдвинуть гипотезу о виде статистики на основе выборочных данных, т.е. для восстановления плотности распределения необходимы априорные допущения о виде исследуемой статистической модели. Примерами методов параметрической статистики могут служить такие методы, как нахождение точечных и интервальных оценок неизвестных параметров предполагаемого закона распределения, метод максимального правдоподобия. Из-за необходимости выдвижения гипотезы у методов параметрической статистики существует ряд недостатков, таких как возможность совершения ошибки при назначении статистики, вариативность экспериментальных данных для исследуемого явления или процесса, строгая необходимость получения минимального размера выборки и др. Таким образом, методами параметрической статистики желательно пользоваться либо на известных и проверенных моделях, либо в совокупности с другими методами для вновь исследуемых экспериментальных данных.

Методы непараметрической статистики в целом нивелируют недостатки методов параметрической статистики и на сегодняшний день широко используются в таких дисциплинах, как анализ данных, машинное обучение, задачи прогнозирования. Недостатком метода непараметрической статистики является вариативность выбора методов для обработки экспериментальных данных, а также численные ошибки, накапливаемые при обработке данных. К методам непараметрической статистики можно отнести критерии согласия (критерий Колмогорова и др.), критерии независимости данных (критерий

Спирмена и др.); для восстановления же плотности распределения случайной величины используют метод гистограмм, метод окна Парзена—Розенблатта, метод гауссовых смесей и другие методы.

В данной работе описан метод окна Парзена—Розенблатта (или метод ядерного сглаживания), применяемый для восстановления эмпирической плотности распределения случайной величины. Применение данного метода является на сегодняшний день актуальным по следующим причинам: отсутствие необходимости назначения заранее неизвестной статистики для исследования экспериментальных данных; получение более качественной эмпирической плотности распределения случайной величины в условиях априорной неопределенности; возможность применения данного метода для выборки с малыми размерами.

В качестве приложения описанного метода проведено исследование по оценке плотности распределения диаметров частиц для такого дисперсного материала, как уплотненный микрокремнезем. Исследование заключалось в восстановлении плотностей распределения диаметров частиц методом окна Парзена—Розенблатта и методом гистограмм, а также в сравнении полученных плотностей распределения.

ОБЗОР ЛИТЕРАТУРЫ

Развитие метода окна Парзена—Розенблатта началось с середины XX в., а именно в работах М. Розенблатта [1] и Е. Парзена [2] были рассмотрены вопросы применения ядерных функций для определения эмпирических функций распределения и плотности распределения случайной величины. Они, по сути, заложили основы метода ядерного сглаживания. Е.А. Надарая и Г. Ватсон рассмотрели вопрос применения ядерных функций для оценки регрессии [3, 4]. М. Барлетт и В.А. Епанечников предложили один из наиболее удачных видов ядерной функции, именуемой на сегодняшний день ядром Епанечникова [5, 6].

Дальнейшие работы были посвящены сходимости метода и поиску оптимальной ширины окна. Б. Сильверман предложил определять ширину интервала из того условия, что случайная величина подчиняется нормальному закону распределения [7]. С. Шезор и М. Джоунс предложили свой подход для определения ширины интервала, основанного на минимизации интегрального среднего квадрата ошибки [8]. П. Холл и Д. Маррон использовали метод сглаженной кросс-валидации для определения ширины окна [9].

На сегодняшний день исследователи занимаются вопросами применения метода ядерного сглаживания для анализа больших данных [10–13], машинного обучения, в частности распознавания образов [14], для оценки распределений с «тяжелыми хвостами» [15]. Также проводятся исследования по поиску оптимальной ширины окна [13, 16–19].

Некоторые алгоритмы вычисления эмпирической плотности распределения случайной величины методом ядерного сглаживания реализованы в различных математических программах и языках программирования, таких как Matlab, R-Studio, Python (SciPy, Scikit-learn) и др.

МЕТОДЫ

Суть метода окна Парзена—Розенблатта заключается в следующем: в точках экспериментальных данных размещаются ядерные функции определенного вида и затем выполняется суммирование этих функций. На выходе получается эмпирическая плотность распределения случайной величины, сглаживающая показатели экспериментальных данных. В одномерном случае эмпирическая плотность распределения, вычисляемая с помощью ядерных функций, описывается следующей формулой:

$$\hat{p}(x) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x - x_i}{h}\right),$$

где n — размер выборки; K — ядерная функция; h — ширина окна; x — случайная выборка; x_i — i -я реализация случайной величины.

Выше приведенную оценку можно обобщить на многомерный случай по формуле

$$\hat{p}(x) = \frac{1}{n} \sum_{i=1}^n \prod_{j=1}^m \frac{1}{h^j} K\left(\frac{x^j - x_i^j}{h^j}\right),$$

где m — размер пространства.

Ядерные функции, используемые для восстановления эмпирической плотности распределения, должны удовлетворять следующим свойствам:

$$K(y) \geq 0,$$

$$K(y) = K(-y)$$

$$\int_{-\infty}^{+\infty} K(y) dy = 1,$$

$$\int_{-\infty}^{+\infty} y K(y) dy = 0$$

$$\int_{-\infty}^{+\infty} y^i K(y) dy = k_i(K) < \infty.$$

Иными словами, ядерные функции — это неотрицательные ограниченные симметричные вещественные функции, интеграл по которым равен единице. Последнее условие указывает на то, что статистические моменты любого порядка должны быть конечны.

Порядок ядерной ν функции равен порядку первого момента, который не равен нулю. Например, если $k_1(K) = 0$ и $k_2(K) > 0$, тогда K является ядром второго порядка ($\nu = 2$). Для практического применения чаще всего используют ядерные функции второго порядка, но также используют функции более высокого порядка.

На рис. 1 изображены примеры ядерных функций второго порядка, описываемые следующими выражениями:

- ядро Епанечникова (или параболическое ядро):

$$K(y) = \frac{3}{4}(1 - y^2);$$

- ядро Гаусса: $K(y) = \frac{1}{\sqrt{2\pi}} e^{-0.5y^2};$

- треугольное ядро: $K(y) = 1 - |y|;$

- биквадратное ядро: $K(y) = \frac{15}{16}(1 - y^2)^2.$

Для восстановления эмпирической плотности распределения с помощью метода окна Парзена—Розенблатта, неизвестным параметром является ширина окна, т.е. непосредственно перед определением эмпирической плотности распределения необходимо решить задачу поиска ширины окна, при этом задачу нужно решить так, чтобы найти оптимальную ширину окна. Для этого используют такие методы, как методы подстановки (метод Сильвермана, метод Шезора—Джоунса, метод Холла—Маррона и др.), принцип максимума правдоподобия с исключением объектов по одному (leave-one-out), метод кросс-проверки на основе метода наименьших квадратов и др. В представленной работе будет применен метод поиска оптимальной ширины окна, предложенный С. Шезором и М. Джоунсом [8].

Шезор и Джоунс предложили определять оптимальную ширину окна на основе минимизации интегрального среднего квадрата ошибки непараметрической оценки, используя следующее нелинейное уравнение:

$$\left(\frac{R(K)}{k_2^4(K) \hat{S}_D(\hat{\alpha}(h_{opt}))} \right)^{1/5} n^{-1/5} - h_{opt} = 0.$$

Рис. 1. Примеры ядерных функций

Figure 1. Examples of kernel functions

Параметры нелинейного уравнения определяются следующим образом:

$$R(K) = \int_{-\infty}^{+\infty} K(x)^2 dx,$$

$$\hat{S}_D(a) = n^{-1}(n-1)^{-1} a^{-5} \sum_{i=1}^n \sum_{j=1}^n K^{(4)}\left(\frac{x_i - x_j}{a}\right),$$

$$\hat{\alpha}(h_{opt}) = 1,357 \left(\frac{\hat{S}_D(a)}{\hat{T}_D(b)} \right)^{1/7} h_{opt}^{5/7},$$

$$\hat{T}_D(b) = -n^{-1}(n-1)^{-1} b^{-7} \sum_{i=1}^n \sum_{j=1}^n K^{(6)}\left(\frac{x_i - x_j}{b}\right),$$

$$a = 0,920 IQR n^{-1/7} \text{ и } b = 0,912 IQR n^{-1/9}.$$

где $K^{(4)}$ и $K^{(6)}$ — ядерные функции 4-го и 6-го порядков, IQR — межквартильный диапазон.

Автором настоящей работы были реализованы алгоритм поиска оптимальной ширины окна методом Шезора—Джоунса и алгоритм вычисления эмпирической плотности распределения методом окна Парзена—Розенблатта на языке программы Matlab. Нелинейное уравнение для поиска оптимальной ширины окна решалось методом Ньютона. За начальное приближение была принята ширина окна определенная по правилу Сильвермана [7].

РЕЗУЛЬТАТЫ ИССЛЕДОВАНИЯ

Рассмотрим пример восстановления эмпирической плотности распределения диаметров частиц для такого дисперсного материала, как уплотненный микрокремнезем. Микрокремнезем является побочным продуктом выплавки кремниевых сплавов в металлургической промышленности и ис-

пользуется в виде пуццолановой добавки для изготовления различных видов бетонов. Уплотненный микрокремнезем состоит из сферических частиц-кластеров, образованных отдельными частицами микрокремнезема.

Восстановление плотности распределения диаметров частиц уплотненного микрокремнезема выполнялось с помощью метода окна Парзена—Розенблатта по вышеописанной методике. Также была выполнена оценка плотности распределения методом гистограмм для сравнения двух методов. Для получения экспериментальных данных была выполнена фотосъемка частиц уплотненного микрокремнезема. На рис. 2 показана одна из полученных фотографий.

Обработка изображений и данных (бинаризация изображения, определение диаметров частиц) выполнялась с помощью математического пакета Matlab. Размер выборки составил 400 частиц. После получения данных выполнялась их обработка для

Рис. 2. Фото частиц уплотненного микрокремнезема

Figure 2. Photo of the particles of compacted microsilica

восстановлений плотностей вероятностей методом гистограмм и методом окна Парзена—Розенблатта.

Для метода гистограмм ширина интервала составила 14,0 мкм. Вычисление ширины интервала выполнялось по следующей формуле:

$$h = 3,73Sn^{-1/3},$$

где n — размер выборки; S — выборочное стандартное отклонение. Нужно отметить, что назначение ширины интервала гистограммы или количества интервалов гистограммы при использовании метода гистограмм может давать некорректный ответ при восстановлении плотности распределения случайной величины, т.е. присутствует возможность получить смещенную оценку. Также недостатком

метода можно назвать множество различных подходов к определению количества интервалов гистограммы [20].

Для восстановления плотности методом окна Парзена—Розенблатта применялись гауссовы ядра второго порядка. Вычисление оптимальной ширины окна выполнялось методом Шезора—Джоунса. Вычисленная ширина окна составила 7,6 мкм.

На рис. 3 и 4 показаны гистограммы и эмпирические плотности распределения диаметров частиц, полученные методом гистограмм и методом окна Парзена—Розенблатта.

На рис. 5 полученные эмпирические плотности распределения диаметров частиц совмещены на одном графике и заметно, что плотности распре-

Рис. 3. Гистограмма и плотность распределения, полученная методом гистограмм

Figure 3. Histogram and distribution density obtained by the histogram method

Рис. 4. Гистограмма и плотность распределения, полученная методом окна Парзена—Розенблатта

Figure 4. Histogram and distribution density obtained by the Parzen—Rosenblatt window method

Рис. 5. Плотности распределения диаметров частиц, полученные методом гистограмм и методом окна Парзена—Розенблатта

Figure 5. Distribution density of particles diameter obtained by the histogram method and Parzen—Rosenblatt window method

деления диаметров частиц отличаются друг от друга, особенно заметно отличие графиков на отрезке от 65 до 130 мкм.

ВЫВОДЫ

Метод окна Парзена—Розенблатта позволяет избежать проблем, которые возникают при восстановлении плотности распределения случайной величины методами параметрической статистики и методом гистограмм, а именно назначения заранее неизвестной статистики для методов параметрической статистики и определения количества интервалов для метода гистограмм. При этом метод окна

Парзена—Розенблатта не является точным методом, так как выбор ширины окна влияет на точность восстанавливаемой плотности распределения, поэтому необходимо уделять особое внимание методам определения ширины окна. На сегодняшний день метод окна Парзена—Розенблатта активно применяется в таких дисциплинах, как анализ данных, машинное обучение и задачи прогнозирования. Данный метод восстановления плотности распределения случайной величины можно применять для статистического моделирования в области строительного материаловедения и для вероятностных расчетов при проектировании конструкций зданий и сооружений.

ЛИТЕРАТУРА

1. *Rosenblatt M.* Remarks on some nonparametric estimates of a density function // *The Annals of Mathematical Statistics.* 1956. Vol. 27. No. 3. Pp. 832–837.
2. *Parzen E.* On estimation of a probability density function and mode // *The Annals of Mathematical Statistics.* 1962. Vol. 33. No. 3. Pp. 1065–1076.
3. *Надарая Э.А.* Об оценки регрессии // *Теория вероятностей и ее применения.* 1964. Т. 9. Вып. 1. С. 157–159.
4. *Watson G.S.* Smooth regression analysis // *Sankhya: The Indian Journal of Statistics.* 1964. Series A. Vol. 26. No. 4. Pp. 359–372.
5. *Bartlett M.S.* Statistical estimation of density functions // *Sankhya, Ser. A.* 1963. No. 25. Pp. 245–54.
6. *Епанечников В.А.* Непараметрическая оценка многомерной плотности вероятности // *Теория*

вероятностей и ее применения. 1969. Т. 14. Вып. 1. С. 156–161.

7. *Silverman B.W.* Density estimation for statistics and data analysis. London : Chapman & Hall/CRC. 1986. P. 48.

8. *Sheather S.J., Jones M.C.* A reliable data-based bandwidth selection method for kernel density estimation // *Journal of the Royal Statistical Society. Series B.* 1991. Vol. 53. No. 3. Pp. 683–690.

9. *Hall P., Marron J.S., Park B.U.* Smoothed cross-validation // *Probability Theory and Related Fields.* 1992. Vol. 92. Pp. 1–20.

10. *Wang S., Wang J., Chung F.* Kernel density estimation, kernel methods, and fast learning in large data sets // *IEEE Transactions on Cybernetics.* Jan. 2014. Vol. 44. Issue 1.

11. Zheng Y., Jests J., Phillips J.M., Li F. Quality and efficiency in kernel density estimates for large data // Proceedings of the 2013 ACM SIGMOD International Conference on Management of Data. 2013. Pp. 433–444.
12. Peherstorfer B., Pflüger D., Bungartz H.J. Density estimation with adaptive sparse grids for large data sets // Proceedings of the 2014 SIAM International Conference on Data Mining. 2014. Pp. 443–451.
13. Zheng Y., Phillips J.M. L_∞ error and bandwidth selection for kernel density estimates of large data // Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2015. Pp. 1533–1542.
14. Ланко А.В., Ланко В.А. Непараметрические алгоритмы распознавания образов при случайных значениях коэффициентов размытости ядерных функций // Автометрия. 2007. Т. 43. № 5. С. 47–55.
15. Buch-larsen T., Nielsen J.P., Guillén M., Boulancé C. Kernel density estimation for heavy-tailed distributions using the Champernowne transformation // Statistics: A Journal of Theoretical and Applied Statistics. 2005. Vol. 39. Issue 6. Pp. 503–516.
16. Добровидов А.В., Рудько И.М. Выбор ширины окна ядерной функции в непараметрической оценке производной плотности методом сглаженной кросс-валидации // Автоматика и телемеханика. 2010. Вып. 2. С. 42–58.
17. Altman N., Léger C. Bandwidth selection for kernel distribution function estimation // Journal of Statistical Planning and Inference. 1995. Vol. 46. Issue 2. Pp. 195–214.
18. Chen S. Optimal bandwidth selection for kernel density functional estimation // Hindawi. Journal of Probability and Statistics. 2015. Vol. 2015. Article ID 242683. 21 p.
19. Heidenreich N.-B., Schindler A., Sperlich S. Bandwidth selection for kernel density estimation: a review of fully automatic selectors // Advances in Statistical Analysis. October 2013. Vol. 97. Issue 4. Pp. 403–433.
20. Глаголев М.В., Сабреков А.Ф. О восстановлении плотности вероятности методом гистограмм в почвоведении и экологии // Динамика окружающей среды и глобальные изменения климата. 2008. № S1. С. 55–83.

Поступила в редакцию 13 марта 2017 г.

Принята в доработанном виде 23 апреля 2018 г.

Одобрена для публикации 30 мая 2018 г.

ОБ АВТОРЕ: Тарарушкин Евгений Викторович — ассистент кафедры строительных материалов и технологий, Российский университет транспорта (МИИТ), 127994, г. Москва, ул. Образцова, д. 9, стр. 9; tarmstu@yandex.ru.

REFERENCES

1. Rosenblatt M. Remarks on some nonparametric estimates of a density function. *The Annals of Mathematical Statistics*. 1956, vol. 27, no. 3, pp. 832–837.
2. Parzen E. On estimation of a probability density function and model. *The Annals of Mathematical Statistics*. 1962, vol. 33, no. 3, pp. 1065–1076.
3. Nadaraya E.A. Ob otsenki regressii [On the regression estimation]. *Teoriya veroyatnostey i ee primeneniya* [Probability Theory and its Applications]. 1964, vol. 9, issue 1, pp. 157–159. (In Russian)
4. Watson G.S. Smooth regression analysis. *Sankhya: The Indian Journal of Statistics. Series A*. 1964, vol. 26, no. 4, pp. 359–372.
5. Bartlett M.S. Statistical estimation of density functions. *Sankhya: The Indian Journal of Statistics. Series A*. 1963, no. 25, pp. 245–254.
6. Epanechnikov V.A. Neparаметricheskaya otsenka mnogomernoy plotnosti veroyatnosti [Nonparametric estimation of the multidimensional probability density]. *Teoriya veroyatnostey i ee primeneniya* [Probability Theory and its Applications]. 1969, vol. 14, issue 1, pp. 156–161. (In Russian)
7. Silverman B.W. *Density estimation for statistics and data analysis*. London, Chapman & Hall/CRC. 1986, p. 48.
8. Sheather S.J., Jones M.C. A reliable data-based bandwidth selection method for kernel density estimation. *Journal of the Royal Statistical Society. Series B*. 1991, vol. 53, no. 3, pp. 683–690.
9. Hall P., Marron J.S., Park B.U. Smoothed cross-validation. *Probability Theory and Related Fields*. 1992, vol. 92, pp. 1–20.
10. Wang S., Wang J., Chung F. Kernel density estimation, kernel methods, and fast learning in large data sets. *IEEE Transactions on Cybernetics*. Jan. 2014, vol. 44, issue 1.
11. Zheng Y., Jests J., Phillips J.M., Li F. Quality and efficiency in kernel density estimates for large data. *Proceedings of the 2013 ACM SIGMOD International Conference on Management of Data*. 2013, pp. 433–444.

12. Peherstorfer B., Pflüger D., Bungartz H.J. Density estimation with adaptive sparse grids for large data sets. *Proceedings of the 2014 SIAM International Conference on Data Mining*. 2014, pp. 443–451.

13. Zheng Y., Phillips J.M. L^∞ error and bandwidth selection for kernel density estimates of large data. *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 2015, pp. 1533–1542.

14. Lapko A.V., Lapko V.A. Neparametricheskie algoritmy raspoznavaniya obrazov pri sluchaynykh znacheniyakh koefitsientov razmytosti yadernykh funktsiy [Nonparametric algorithms for pattern recognition for random values of the coefficients of the smearing of nuclear functions]. *Avtometriya* [Optoelectronics, Instrumentation and Data Processing]. 2007, vol. 43, no. 5, pp. 47–55. (In Russian)

15. Buch-larsen T., Nielsen J.P., Guillén M., Bønlancé C. Kernel density estimation for heavy-tailed distributions using the Champernowne transformation. *Statistics: A Journal of Theoretical and Applied Statistics*. 2005, vol. 39, issue 6, pp. 503–516.

16. Dobrovidov A.V., Rud'ko I.M. Vybory shiriny okna yadernoy funktsii v neparametricheskoy otsenke proizvodnoy plotnosti metodom sglazhennoy kross-val-

idatsii [Choice of the width of the window of the nuclear function in the nonparametric estimation of the density derivative by the method of smoothed cross-validation]. *Avtomatika i telemekhanika* [Automation and Remote Control]. 2010, issue 2, pp. 42–58. (In Russian)

17. Altman N., Léger C. Bandwidth selection for kernel distribution function estimation. *Journal of Statistical Planning and Inference*. 1995, vol. 46, issue 2, pp. 195–214.

18. Chen S. Optimal bandwidth selection for kernel density functional estimation. *Hindawi. Journal of Probability and Statistics*, 2015, vol. 2015. Article ID 242683.

19. Heidenreich N.-B., Schindler A., Sperlich S. Bandwidth selection for kernel density estimation: a review of fully automatic selectors. *Advances in Statistical Analysis*. October 2013, vol. 97, issue 4, pp. 403–433.

20. Glagolev M.V., Sabrekov A.F. O vosstanovlenii plotnosti veroyatnosti metodom gistogramm v pochvovedenii i ehkologii [On the restoration of probability density by the method of histograms in soil science and ecology]. *Dinamika okruzhayushchey sredy i global'nye izmeneniya klimata* [Dynamics of the environment and global climate changes]. 2008, no. S1, pp. 55–83. (In Russian)

Received March 13, 2017.

Adopted in final form on April 23, 2018.

Approved for publication May 30, 2018.

ABOUT THE AUTHOR: **Tararushkin Evgeniy Victorovich** — Assistant, Department of the Building Materials and Technologies, **Russian university of transport (MIIT)**, 9 bldg 9 Obratsova str., Moscow, 127994, Russian Federation; tarmstu@yandex.ru.