

Ars Administrandi (Искусство управления). 2024. Т. 16, № 2. С. 198–216.

Ars Administrandi. 2024. Vol. 16, no. 2, pp. 198–216.

Научная статья

УДК 338:31.33

<https://doi.org/10.17072/2218-9173-2024-2-198-216>

**ОЦЕНКА КОЛИЧЕСТВА
МАЛЫХ И СРЕДНИХ ПРЕДПРИЯТИЙ
В ГРАНИЦАХ ГОРОДОВ И АГЛОМЕРАЦИЙ:
ПРИМЕНЕНИЕ МЕТОДОВ МАШИННОГО ОБУЧЕНИЯ
К ДАННЫМ РОССИИ**

Дарья Максимовна Радченко¹✉, Юрий Юрьевич Пономарев²,

Кирилл Владимирович Ростислав³

^{1, 2, 3} Институт прикладных экономических исследований РАНХиГС, Москва, Россия

² Институт экономической политики им. Е. Т. Гайдара, Москва, Россия

¹ radchenko-dm@ranepa.ru✉, <https://orcid.org/0000-0002-7457-5179>

² ponomarev@ranepa.ru, <https://orcid.org/0000-0002-1188-9293>

³ rostislav-kv@ranepa.ru, <https://orcid.org/0000-0003-1869-8171>

Аннотация. Введение: развитие агломераций, которые, как показывают исследования, становятся полюсами роста производительности, обозначено как один из приоритетов пространственной политики России. Несмотря на многочисленные обсуждения, в федеральном законодательстве отсутствует термин «городская агломерация», но при этом он широко используется в нормативно-правовой базе на других уровнях управления. Это приводит к отсутствию согласованной позиции в вопросах о структуре, размерах и параметрах агломераций и актуализирует формирование научно обоснованного подхода к определению экономических и географических границ агломераций. **Цель:** разработать подход к идентификации границ агломераций в России с учетом пространственного распределения экономической активности. **Методы:** машинное обучение, статистический



Контент доступен по лицензии Creative Commons Attribution 4.0 International License
The content is licensed under a Creative Commons Attribution 4.0 International License

анализ. Применение этих методов позволяет с высокой точностью идентифицировать границы агломерации на открытых данных по высотности застройки городской территории. **Результаты:** представлены результаты апробации алгоритма пространственной кластеризации (DBSCAN) на данных по жилой застройке в России, выделен перечень кластеров устойчивых муниципалитетов, которые определяют границы агломераций, и проведена оценка числа и выручки фирм, малых и средних предприятий (МСП) в границах крупнейших городов и агломераций, определенных аналитическим способом и формально в нормативных документах. Также на примере отдельных регионов и в среднем для России показано, что число фирм (общее и фирм из реестра МСП) в границах агломераций может недооцениваться более чем на треть, а совокупная выручка – почти на 60 %. **Выводы:** учет кластеров устойчивых муниципалитетов может повысить точность оценок агломерационных эффектов и деловой активности и, как следствие, способствовать формированию более точных мер поддержки со стороны государства.

Ключевые слова: делимитация границ, пространственная кластеризация, алгоритм DBSCAN

Благодарности: статья подготовлена в рамках выполнения научно-исследовательской работы по государственному заданию Российской академии народного хозяйства и государственной службы при Президенте Российской Федерации.

Для цитирования: Радченко Д. М., Пономарев Ю. Ю., Ростислав К. В. Оценка количества малых и средних предприятий в границах городов и агломераций: применение методов машинного обучения к данным России // *Ars Administrandi* (Искусство управления). 2024. Т. 16, № 2. С. 198–216. <https://doi.org/10.17072/2218-9173-2024-2-198-216>.

Original article

ESTIMATING SME NUMBERS WITHIN CITIES AND AGGLOMERATIONS: APPLYING MACHINE LEARNING METHODS TO RUSSIAN DATA

Daria M. Radchenko¹, Yuri Yu. Ponomarev², Kirill V. Rostislav³

^{1,2,3} Institute of Applied Economic Research of the RANEPa, Moscow, Russia

² Gaidar Institute for Economic Policy, Moscow, Russia

¹ radchenko-dm@ranepa.ru, <https://orcid.org/0000-0002-7457-5179>

² ponomarev@ranepa.ru, <https://orcid.org/0000-0002-1188-9293>

³ rostislav-kv@ranepa.ru, <https://orcid.org/0000-0003-1869-8171>

Abstract. Introduction: the development of agglomerations, which, as research shows, are becoming productivity growth poles, is one of the priorities in the spatial development policy in Russia. Despite numerous discussions, the term “urban agglomeration” is currently absent in federal legislation, but it is widely used in the regulatory framework at other levels of government. This leads to a lack of coordinated, accurate, and up-to-date information on the size and structure of agglomerations, thus bringing forward the need to formulate a well-based approach to identify their economic and geographical boundaries. **Objectives:** developing an approach to identify the boundaries of agglomerations in Russia, considering the spatial distribution of economic activity. **Methods:** machine learning, statistical analysis. The use of these methods makes it possible to identify the boundaries of agglomeration with high accuracy using open data on the height of the urban area. **Results:** the article presents the results of spatial clustering algorithm (DBSCAN) testing on residential development

data in Russia; a list of “sustainable municipalities clusters” defining the agglomerations boundaries has been identified and estimates have been made for the number and revenue of firms and SMEs within the largest cities and agglomerations boundaries defined both analytically and formally in regulatory documents. Using the example of individual regions and average data for Russia, it has been shown that the number of firms (total and those registered as SME) within agglomerations can be underestimated by more than one-third and total revenues – by almost 60 %. **Conclusions:** considering sustainable municipalities clusters can improve the accuracy of agglomeration effects and business activity estimates and contribute to the formation of more accurate government support measures.

Keywords: delimitation, spatial clustering, DBSCAN algorithm

Acknowledgments: the article was prepared as part of research work on the government fund to the Russian Presidential Academy of National Economy and Public Administration.

For citation: Radchenko, D. M., Ponomarev, Yu. Yu. and Rostislav, K. V. (2024), “Estimating SME numbers within cities and agglomerations: Applying machine learning methods to Russian data”, *Ars Administrandi*, vol. 16, no. 2, pp. 198–216, <https://doi.org/10.17072/2218-9173-2024-2-198-216>.

ВВЕДЕНИЕ

Задачи анализа агломерационных эффектов тесно связаны с проблемой определения границ агломераций, а понимание их границ, в свою очередь, необходимо для эффективного планирования и развития территорий. При этом существует некоторое количество методик определения границ агломераций, и они могут приводить к различным результатам, особенно в условиях недостатка открытых данных.

Границы городских агломераций в исследовательской и отраслевой литературе определяются наличием устойчивых пространственных социально-экономических связей. Определение конкретных параметров этих связей и их протяженности является предметом различного рода исследовательских, методических и аналитических построений. Кроме того, границы агломераций в разных странах могут устанавливаться в нормативных документах на государственном или региональном уровне. Второй вариант обычно оказывается вторичен по отношению к первому, однако он необходим для целей государственной политики и реализации механизмов управления.

Агломерационная практика в России опережает теорию и требует правового регулирования¹. По состоянию на 2023 год в отечественной нормативно-правовой базе отсутствует как само определение агломерации, так и порядок и критерии ее формирования, вследствие чего документы даже одного уровня оказываются не согласованы между собой (рис. 1).

¹ Развитие городских агломераций: аналит. обзор. Вып. 2 [Электронный ресурс] // Официальный сайт Рос. ин-та градостроительства и инвестиц. развития «Гипрогор». С. 7–8. URL: <https://web.archive.org/web/20150522043211/http://giprogor.ru/sites/default/files/Пазвитие%20городских%20агломераций%20-%20аналитический%20обзор.%20Выпуск%202.pdf> (дата обращения: 03.07.2023).

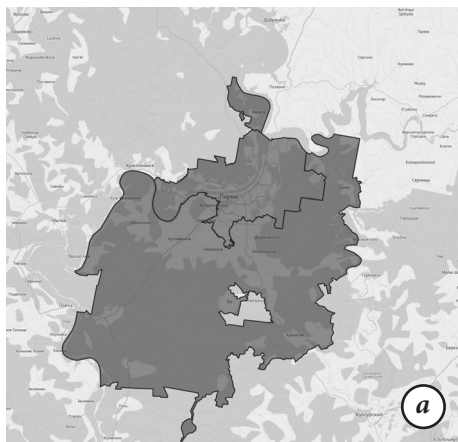
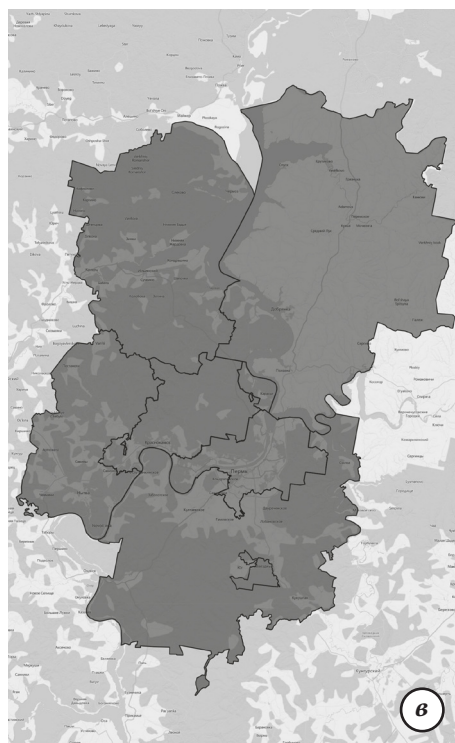


Рис. 1. Расхождение границ при разных официальных определениях агломераций
/ Fig. 1. Divergence of boundaries stemming from different official definitions of agglomerations

Примечание: выделены темно-серым цветом а) – г. Пермь, Пермский район (Соглашение о создании и совместном развитии Пермской городской агломерации²); б) – г. Пермь, Пермский, Краснокамский и Добрянский муниципальные районы (Программа комплексного развития транспортной инфраструктуры Пермской городской агломерации³); в) – г. Пермь, ЗАТО Звездный,



² Об установлении расходных обязательств Пермского муниципального района на реализацию мероприятий по разработке концепции социально-экономического развития Пермской городской агломерации [Электронный ресурс]: Постановление Администрации Перм. муницип. р-на от 14.09.2016 № 485. URL: https://permraion.ru/_res/fs/file30582.pdf (дата обращения: 03.07.2023).

³ О внесении изменений в постановление Правительства Пермского края от 29 марта 2017 г. № 156-п «Об утверждении Порядка предоставления и распределения в 2017 году иных межбюджетных трансфертов из бюджета Пермского края, передаваемых за счет средств федерального бюджета бюджетам муниципальных образований Пермского края на реализацию мероприятий Программы комплексного развития транспортной инфраструктуры Пермской городской агломерации приоритетного проекта

Пермский, Краснокамский, Нытвенский, Ильинский и Добрянский муниципальные районы (Схема территориального планирования Пермского края⁴).

Источник: составлено авторами на основе указанных нормативных документов.

Спектр аналитических подходов к определению границ агломераций весьма обширен, и выбор их обычно обусловлен наличием данных и страновой спецификой. К примеру, Организация экономического сотрудничества и развития (ОЭСР) широко применяет методы, базирующиеся на анализе показателей плотности населения. В отечественных и зарубежных исследованиях отмечаются подходы на основе близости экономических показателей территорий (производительность труда, отраслевая концентрация и иные) (Райсих, 2020; Коломак и Трубехина, 2013), ежедневных миграций (Райсих, 2020; Coombes, 2004), морфологических особенностей местности, интенсивности транспортных потоков и их направленности⁵; граф-аналитические, сетевые модели, методы на основе изохрон транспортной доступности (чаще всего автомобильной) (Guzik et al., 2017; Obaso et al., 2016)⁶, расчеты которых основываются и на временных, и на финансовых затратах (Монастырская и Песляк, 2017); а также иные методы и их различные комбинации. При этом вариация в результатах определения границ агломераций может критично влиять на оценку количества фирм и совокупного объема экономической активности в агломерации и, как следствие, на качество и полноту разрабатываемых в их отношении мер и инструментов поддержки.

В условиях нехватки статистических данных одним из возможных решений проблемы делимитации границ потенциальных агломераций могут стать методы машинного обучения и их комбинации с другими подходами, например с методом изохрон и методикой ОЭСР. Однако этому, по нашему мнению, уделяется недостаточное внимание в отечественной исследовательской литературе.

В настоящей статье описан поход к выделению границ агломераций с помощью методов машинного обучения, а именно алгоритма пространственной кластеризации (DBSCAN), представлены результаты его апробации на российских данных по жилой застройке в отдельных регионах. На основе полученных результатов оценено число малых и средних предприятий (МСП) в выявленных с помощью методов машинного обучения фактических границах агломераций и в их административно определенных границах, показаны и проанализированы соответствующие различия.

“Безопасные и качественные дороги”» [Электронный ресурс]: Постановление Правительства Перм. края от 19.09.2017 № 780-п. URL: <https://base.garant.ru/43072572/> (дата обращения: 03.07.2023).

⁴ *Схема территориального планирования* [Электронный ресурс] // Офиц. сайт Перм. муницип. округа. URL: <https://permraion.ru/build/stp/> (дата обращения: 03.07.2023).

⁵ Чугуевская Е. С. Городские агломерации в системе пространственного развития Российской Федерации [Электронный ресурс] // Офиц. сайт М-ва строительства Новосиб. обл. 2017. URL: https://minstroy.nso.ru/sites/minstroy.nso.ru/wodby_files/files/imce/chuguevskaya_-_gorodskie_aglomeracii_0.pdf (дата обращения: 04.07.2023).

⁶ См. также: *Definition of a methodology to delimitate metropolitan areas in Europe. Case study: Accessibility in Barcelona Metropolitan. Region by public and private transport* // ESPON website. 2013. 3 Apr. URL: https://www.espon.eu/sites/default/files/attachments/M4D_SIR_Annex6_Construction_FUA.pdf (Accessed July 4, 2023).

МЕТОДОЛОГИЯ (ТЕОРЕТИЧЕСКИЕ ОСНОВЫ) ИССЛЕДОВАНИЯ

Методы машинного обучения в целом и алгоритмы кластеризации в частности широко используются для решения задач по делимитации границ, например, в исследованиях для Испании (Arribas-Bel et al., 2021), Франции (de Bellefon et al., 2021), Индии (Galdo et al., 2021), Сербии (Gajović, 2013), Китая (Yang et al., 2019), Африки (Schug et al., 2018; Okwuashi and Ndehedehe, 2021), мировых макрорегионов (Moreno-Monroy et al., 2021; Uchida and Nelson, 2010; Hagenauer and Helbich, 2012). Однако их использование на больших пространственных базах данных будет неэффективно, если не выполнены следующие условия:

1. Применяющий обладает достаточными знаниями в предметной области для корректного определения диапазонов и величин входных параметров для работы алгоритма.

2. Алгоритм способен обнаруживать кластеры произвольной формы, поскольку в пространственных данных они могут быть любыми: эллипсоидными, вытянутыми, линейными и т. д.

3. Алгоритм обеспечивает высокую эффективность на больших объемах данных (от нескольких десятков тысяч объектов и более). Наборы пространственных данных, особенно когда речь идет о страновом масштабе, могут достигать десятков миллионов наблюдений.

Большая часть алгоритмов не удовлетворяет сразу всем трем требованиям, и на этом фоне выгодно выделяется алгоритм кластеризации DBSCAN, который первоначально описали М. Эстер и его соавторы в работе 1996 года (Ester et al., 1996). DBSCAN, или плотностный алгоритм пространственной кластеризации с шумами, представляет собой алгоритм машинного обучения без учителя, применяемый для классификации неразмеченных данных (в таком случае образцы данных, используемые для обучения модели, не имеют предопределенных категорий).

Как и в случае с большинством алгоритмов машинного обучения, поведение модели определяют несколько параметров (гиперпараметров)⁷:

- *Eps* – радиус окрестности вокруг точек-ядер;
- *MinPts* – наименьшее число точек, которое должно быть в окрестности радиуса ϵ , чтобы точка считалась ядерной (включая саму эту точку);
- метрика, используемая при расчете расстояния между точками в массиве (например, наиболее часто используемое евклидово расстояние).

Алгоритм начинает работу с вычисления расстояния между каждой точкой и всеми прочими точками. В итоге точка может быть отнесена к одной из трех категорий (см. пример на рис. 2):

- 1) точка-ядро – точка, вокруг которой есть *MinPts* точек, расстояние до которых относительно точки ниже порогового значения, определенного *Eps*;

- 2) пограничная точка – точка, которая не находится в непосредственной близости от всех точек в области *MinPts*, но близка к одной или нескольким точкам-ядрам. Пограничные точки включаются в кластер ближайшей точки-ядра;

⁷ Maklin C. DBSCAN Python example: The optimal value for epsilon (EPS) [Online] // Towards Data Science. 2019. 30 June. URL: <https://towardsdatascience.com/machine-learning-clustering-dbscan-determine-the-optimal-value-for-epsilon-eps-python-example-3100091cfbc> (Accessed June 28, 2023).

3) точка-шум – точка, расстояние от которой до точек-ядер больше Eps . Точки-шумы игнорируются и не являются частью какого-либо кластера.

Таким образом, кластер (агломерация) есть множество точек, расположенных в ϵ -окрестностях ядерных точек, чьи ϵ -окрестности пересекаются.

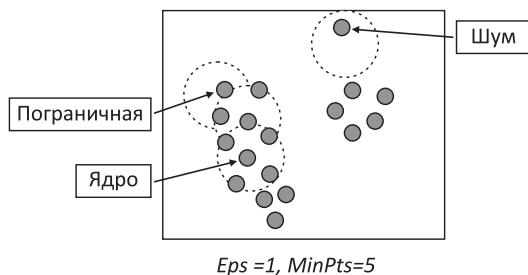


Рис. 2. Категоризация точек методом DBSCAN / Fig. 2. Points categorization using DBSCAN

Источник: доклад J. Gao⁸.

От того, как будут выбраны значения Eps и $MinPts$, зависит, насколько корректно и адекватно будут кластеризованы объекты. Н. Рахма и С. Ситанганг в 2016 году предложили подход к автоматическому определению оптимального значения Eps (Rahmah and Sitanggang, 2016). Первоначально рассчитывается расстояние между всеми точками по методу ближайшего соседа (число соседей обычно устанавливается равным 3–5). Коэффициент Eps определяется путем расчета наклона между точками с помощью уравнения $(y_2 - y_1)/(x_2 - x_1)$ таким образом, чтобы он вписывался в пороговое значение 1%. Оптимальное значение, следовательно, будет располагаться в точке максимальной кривизны. На рисунке 3 показан результат расчета расстояний между точками методом ближайшего соседа, его сортировка по возрастанию и оптимальное значение Eps .

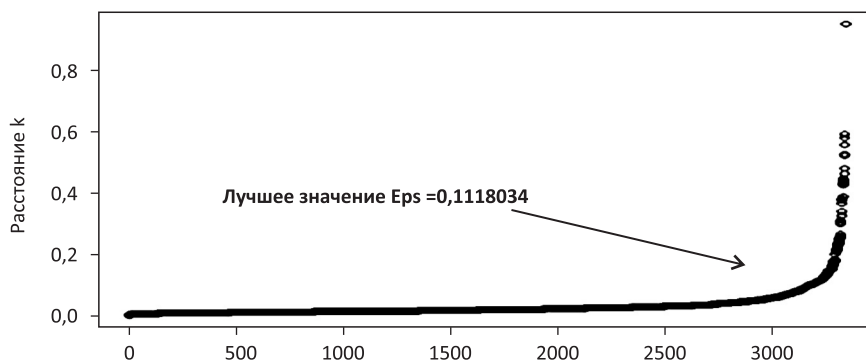


Рис. 3. Пример нахождения оптимального значения Eps / Fig. 3. Example of finding an Eps optimal value

Источник: (Rahmah and Sitanggang, 2016, p. 3).

⁸ Gao J. Lecture 4: Density-based methods [Online] // Website of the Department of Computer Science and Engineering, University at Buffalo. P. 5. URL: https://cse.buffalo.edu/~jing/cse601/fa12/materials/clustering_density.pdf (Accessed July 2, 2023).

Таким образом, суть DBSCAN состоит в определении того, достаточно ли близко расположено минимальное количество точек друг к другу, чтобы считаться частью одного кластера. При этом DBSCAN очень чувствителен к масштабу исходных данных, поскольку Eps фиксирован для всей выборки. Поэтому, например, для такой территориально дифференцированной страны, как Россия, имеет смысл разбивать данные на региональные подвыборки и проводить кластеризацию для каждого региона отдельно.

РЕЗУЛЬТАТЫ ИССЛЕДОВАНИЯ

За единицу кластеризации для проведения анализа нами было принято здание (жилое или нежилое). Положение зданий на карте и их тип взяты из открытой базы данных OpenStreetMap (OSM). Данные OSM собирались в границах страны, срез был произведен на 13 февраля 2022 года, собраны все данные по группам объектов с тегом “building=*”. Перед процедурой предварительной обработки количество точек составило около 6 млн; затем из них были исключены объекты, не представляющие интереса и вносящие шум. (Поскольку исследовательский интерес представляли только здания, расположение которых определяет развитие устойчивых социально-экономических связей, формирующих агломерации, такие объекты, как, например, диспетчерские вышки, столбы, трубопроводы, ограды, шлабгаумы и т. п., были исключены из выборки.)

Для коррекции проблемы чувствительности алгоритма к масштабу было решено отойти от использования заранее заданных входных параметров для каждого региона и, наоборот, рассмотреть широкий набор пар параметров и отобрать наиболее устойчивые кластеры, при этом не привязываясь к понятию «агломерация» и не ограничивая отбор численностью населения. Иначе говоря, далее будут исследоваться кластеры, полученные путем реализации алгоритма DBSCAN со всеми комбинациями $Eps = [5, 10, 15]$ км и $MinPts = [500, 1000, 1500, 2000, 2500, 3000]$ зданий, то есть будет рассмотрено 18 пар входящих параметров, 18 результатов кластеризации для каждого региона.

Поскольку агломерация представляется одной из важнейших географических единиц наблюдений для оценки различных эффектов (она позволяет учесть большую часть связей между хозяйственными единицами), для демонстрации возможных расхождений в оценках экономической активности при формальном и реальном (аналитическом, исходящем из фактических данных) определении границ агломераций дополнительно была решена задача оценки числа фирм и совокупных показателей их деятельности в административных границах выделенных кластеров и соответствующих им центральных городах. Для этого использовались выгрузка базы данных Федеральной налоговой службы России за 2022 год и реестр МСП по состоянию на конец 2021 года, из которых были отобраны действующие юридические лица с ненулевой выручкой в рассматриваемом году, при регистрации указавшие ОКТМО в пределах рассматриваемых муниципальных образований.

Все здания, попавшие в тот или иной кластер, соотносились с муниципальным образованием, в котором они расположены (уровень детализации –

муниципальные районы и городские округа). Чтобы выяснить, следует ли считать муниципальное образование частью кластера, требовалось оценить, какая доля точек в нем отнесена к данному кластеру. Для этого определялась доля точек, отнесенных к каждому кластеру внутри каждого муниципального образования, в котором есть точка кластера. И если эта доля превышала 50 %, то муниципальное образование проходило отбор. Получаемые таким образом кластеры могут состоять из любого числа муниципальных образований (даже из одного) и иметь любой состав (например, муниципальные районы без городских округов).

Для выявления групп наиболее устойчивых кластеров использовался следующий подход: если муниципальное образование проходит отбор в своем регионе в 75 % и более случаев (то есть как минимум для 13 из 18 пар параметров), то его можно отнести к устойчивым. Для исключения одиночных устойчивых муниципалитетов и приведения результатов алгоритма к сопоставимому виду были сохранены только кластеры (см. рис. 4), попадающие в зону полуторачасовой изохроны от наиболее крупного города (центра региона).



Рис. 4. Группы устойчивых муниципальных образований в зоне полуторачасовой изохроны / Fig. 4. Groups of stable municipalities in the 1.5 half isochrones zone

Источник: рассчитано авторами.

Метод изохрон применяется для идентификации границ агломераций в условиях серьезного дефицита пространственных и статистических данных. В России на его основе базируется методология Центрального научно-исследовательского и проектного института Министерства строительства и жилищно-коммунального хозяйства Российской Федерации (Листенбург, 1975). Однако использование изохрон может чрезмерно расширять границы агломерации (Райсх, 2020, с. 38–40), и для большей эффективности его следует комбинировать с прочими подходами (Coombes, 2004, р. 307). В нашей работе⁹ результаты делимитации были получены комбинацией

⁹ Радченко Д. М., Пономарев Ю. Ю. Разработка научно-методологических подходов к иден-

метода изохрон и методики ОЭСР и позволили скорректировать первоначальные результаты кластеризации и отсеять муниципалитеты, отдаленные от крупных городов и центров регионов.

Рассмотрим конкретный пример Пермского края (рис. 5). Можно видеть, насколько бывают велики различия в оценках численности фирм и их совокупной выручки, если рассматривать только центр региона – г. Пермь или потенциальные границы Пермской агломерации, в которую входит г. Пермь и шесть соседствующих муниципалитетов.

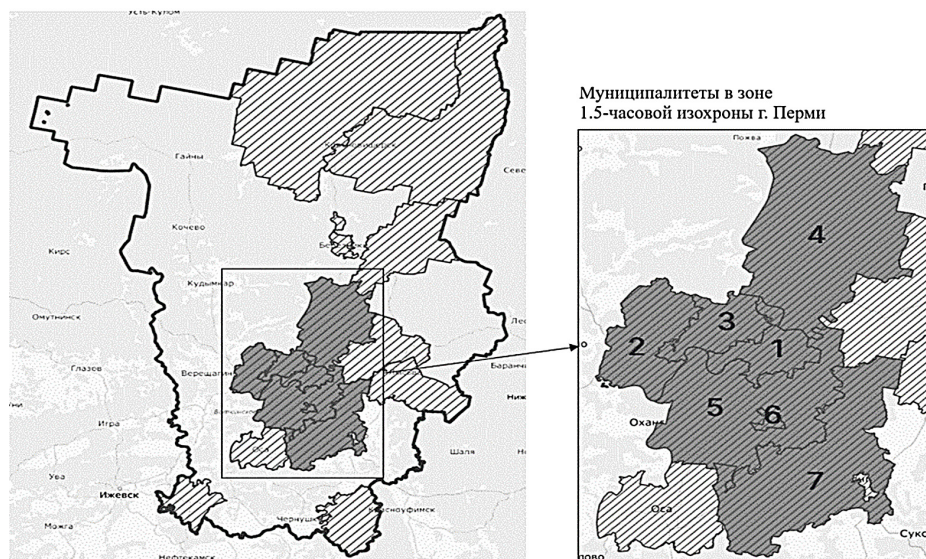


Рис. 5. Группы устойчивых муниципалитетов на примере Пермского края / Fig. 5. Groups of stable municipalities in Perm Territory

Примечание: 1 – г. Пермь, 2 – Нытвенский городской округ, 3 – Краснокамский городской округ, 4 – Добрянский городской округ, 5 – Пермский район, 6 – ЗАТО Звездный, 7 – Кунгурский район.

Источник: составлено авторами.

По данным ФНС России, в 2021 году в г. Перми зарегистрировано 18 484 фирмы с суммарной выручкой 1954,6 млрд рублей, в том числе 17 240 МСП с суммарной выручкой 799,1 млрд рублей. Включение в зону рассмотрения муниципалитетов, сформировавших устойчивый кластер, дает прибавку в 2 034 фирмы и 185,2 млрд рублей выручки, из которых 813 фирм и 77,9 млрд рублей приходится на МСП. Таким образом, недооценка по всем показателям может составить от 4,7 до 11 %.

тификации фактических границ агломераций в России с учетом пространственного распределения экономической активности. М.: РАНХиГС, 2020. 65 с. [Электронный ресурс]. URL: https://www.scipedia.com/wd/images/8/80/Draft_Семенов_664280003-5234-document.pdf (дата обращения: 02.07.2023).

Проведение аналогичной оценки в целом по России (табл.) показало, что недооценка числа фирм из реестра МСП составляет 32 %, общего числа фирм – 35 %, а их выручки – 59 %.

Таблица / Table

Разница между значениями показателей для административного центра региона и для кластера устойчивых муниципалитетов, включающих административный центр и попадающих в зону полуторачасовой изохроны от него / Indicators values difference for the region's administrative center and for the cluster of stable municipalities containing this center and falling within the 1.5 hour isochron zone from it

Регион	Число фирм в реестре МСП в 2022 г., %	Общее число фирм в 2021 г., %	Суммарная выручка фирм в 2021 г., %
Российская Федерация (среднее)	32	35	59
Республика Адыгея (Адыгея)	857	749	2917
Алтайский край	20	19	13
Амурская область	Кластер состоит только из административного центра		
Астраханская область	4	4	3
Республика Башкортостан	17	17	12
Белгородская область	14	14	15
Брянская область	16	17	19
Республика Бурятия	Кластер состоит только из административного центра		
Владимирская область	18	19	14
Волгоградская область	32	33	22
Вологодская область	71	65	173
Воронежская область	6	6	5
Республика Дагестан	1	1	4
Ивановская область	7	7	13
Иркутская область	26	25	14
Кабардино-Балкарская Республика	10	10	12
Калининградская область	15	15	13
Республика Калмыкия	Кластер состоит только из административного центра		
Калужская область	72	70	56
Республика Карелия	4	4	6
Кемеровская область – Кузбасс	78	83	132
Кировская область	Кластер состоит только из административного центра		
Костромская область			
Краснодарский край			
Красноярский край	5	4	3
Курская область	6	6	15

Регион	Число фирм в реестре МСП в 2022 г., %	Общее число фирм в 2021 г., %	Суммарная выручка фирм в 2021 г., %
Республика Марий Эл	13	6	10
Мурманская область	Кластер состоит только из административного центра		
Нижегородская область	32	33	26
Новгородская область	13	17	7
Новосибирская область	11	24	43
Омская область	Кластер состоит только из административного центра		
Оренбургская область	14	34	16
Орловская область	11	13	30
Пензенская область	8	10	3
Пермский край	11	9	9
Приморский край	6	6	3
Псковская область	10	6	1
Ростовская область	50	62	42
Рязанская область	7	3	6
Самарская область	65	50	37
Саратовская область	20	6	6
Свердловская область	22	23	12
Республика Северная Осетия – Алания	Кластер состоит только из административного центра		
Смоленская область	17	23	26
Ставропольский край	24	25	68
Тамбовская область	12	41	30
Республика Татарстан (Татарстан)	39	19	35
Тверская область	10	14	7
Томская область	13	17	7
Тульская область	29	25	43
Тюменская область	8	12	4
Удмуртская Республика	8	7	6
Ульяновская область	Кластер состоит только из административного центра		
Хабаровский край	16	14	12
Челябинская область	10	8	14
Чувашская Республика – Чувашия	12	12	10
Ярославская область	16	17	19

Примечание: приведены все регионы, в которых присутствуют кластеры устойчивых муниципалитетов, за исключением Москвы, Московской области, Санкт-Петербурга и Ленинградской области.

Источник: составлено авторами.

Более значительное расхождение в оценках для административного центра и его кластера наблюдается в регионах, где вне административного центра расположены крупные промышленные предприятия и иные точки притяжения экономической активности.

Таким образом, за счет более полного охвата связанных муниципалитетов точность оценок агломерационных эффектов, деловой активности, эффективности различных проектов может быть повышена, что, в свою очередь, будет способствовать формированию более точных выводов и повышению эффективности мер поддержки со стороны государства, особенно в части столь актуальных в текущих условиях мер поддержки МСП.

ЗАКЛЮЧЕНИЕ

Задача поиска источников и оценки агломерационных эффектов тесна связана с проблемой определения границ агломераций. Несмотря на многочисленные обсуждения, в федеральном законодательстве отсутствует термин «городская агломерация», хотя одновременно он широко используется в нормативно-правовой базе на других уровнях управления. В связи с этим обретает актуальность разработка подхода к определению экономических и географических границ агломераций. Решением могут стать методы машинного обучения и многокритериального анализа – современные подходы к идентификации границ агломерации.

На основе разбора зарубежных подходов к делимитации городских агломераций методами машинного обучения для применения в российских условиях адаптирован метод DBSCAN. Апробация метода на данных о расположении зданий (жилых и нежилых) в российских регионах показала, что его результаты сопоставимы с иными подходами, использовавшимися ранее, однако позволяют делать более точные выводы. Вычислен набор параметров, который обеспечивает результат, наиболее близкий к результатам применения методики ОЭСР¹⁰. При этом важно отметить, что данный набор неоднороден по регионам, что еще раз свидетельствует о невозможности использования унифицированной методики для всей территории России. На основе частоты попадания в разные пары параметров был сформирован перечень кластеров устойчивых муниципалитетов, учет которых может повысить точность оценок агломерационных эффектов и деловой активности и, как следствие, способствовать формированию более точных мер поддержки со стороны государства.

Для отдельных регионов и в среднем для России показано, что число фирм (общее и фирм из реестра МСП) в границах агломераций может недооцениваться более чем на треть, а совокупная выручка – почти на 60 %. Четкая идентификация границ агломераций позволит эффективнее и осознаннее принимать решения в сфере территориального планирования.

¹⁰ Там же. С. 43–47.

Например, это может помочь контролировать (к примеру, путем дальнейшего расширения программы «зеленых поясов») расползание или, наоборот, сжатие городов, основываясь не только на некоторых эвристиках или интуиции, но и на выявленных, подкрепленных данными хозяйственно-экономических связях.

В дальнейшем устойчивые муниципалитеты можно рассматривать как структуру, в рамках которой можно строить микроэкономические модели для эффективного городского планирования, исследовать плотность застройки и населенности агломераций и т. д.

Список источников

Коломак Е. А., Трубехина И. Е. Исследование агломерационных процессов на территории Новосибирской области // Регион: экономика и социология. 2013. № 3. С. 239–259.

Листенгурт Ф. М. Критерии выделения крупномасштабных агломераций в СССР // Известия Академии наук СССР. Серия географическая. 1975. № 1. С. 41–49.

Монастырская М. Е., Песляк О. А. Современные методы делимитации границ городских агломераций // Градостроительство и архитектура. 2017. Т. 7, № 3. С. 80–86. <https://doi.org/10.17673/Vestnik.2017.03.14>.

Райсик А. Э. К вопросу об определении границ городских агломераций: мировой опыт и формулировка проблемы [Электронный ресурс] // Демографическое обозрение. 2020. Т. 7, № 1. С. 27–53. <https://doi.org/10.17323/demreview.v7i1.10819>. URL: <https://demreview.hse.ru/article/view/10819/12136> (дата обращения: 04.09.2023).

Arribas-Bel D., Garcia-López M.-A., Viladecans-Marsal E. Building(s and) cities: Delineating urban areas with a machine learning algorithm // Journal of Urban Economics. 2021. Vol. 125. Art. № 103217. <https://doi.org/10.1016/j.jue.2019.103217>.

Coombes M. Multiple dimensions of settlement systems: Coping with complexity // New forms of urbanization: Beyond the urban–rural dichotomy / Ed. by T. Champion and G. Hugo. Aldershot: Ashgate Publishing Limited, 2004. P. 307–324.

De Bellefon M.-P., Combes P.-P., Duranton G. et al. Delineating urban areas using building density // Journal of Urban Economics. 2021. Vol. 125. Art. № 103226. P. 103–226. <https://doi.org/10.1016/j.jue.2019.103226>.

Ester M., Kriegel H.-P., Sander J. et al. A density-based algorithm for discovering clusters in large spatial databases with noise // Proceedings of the Second International Conference on Knowledge Discovery and Data Mining (KDD-96). Portland: AAAI Press, 1996. P. 226–231.

Gajović V. Comparative analysis of different methods and obtained results for delineation of functional urban areas // Spatium International Review. 2013. № 29. P. 8–15. <https://doi.org/10.2298/SPAT1329008G>.

Galdo V., Li Yu., Rama M. Identifying urban areas by combining human judgment and machine learning: An application to India // Journal of Urban Economics. 2021. Vol. 125. Art. № 103229. <https://doi.org/10.1016/j.jue.2019.103229>.

Guzik R., Kołós A., Gwosdz K. Interconnections in public transport as a method for delimiting urban functional areas and the settlement hierarchy in Poland // *Regional Statistics*. 2017. Vol. 7, № 1. P. 63–77. <https://doi.org/10.15196/RS07104>.

Hagenauer J., Helbich M. Mining urban land-use patterns from volunteered geographic information by means of genetic algorithms and artificial neural networks // *International Journal of Geographical Information Science*. 2012. Vol. 26, № 6. P. 963–982. <https://doi.org/10.1080/13658816.2011.619501>.

Moreno-Monroy A. I., Schiavina M., Veneri P. Metropolitan areas in the world. Delineation and population trends // *Journal of Urban Economics*. 2021. Vol. 125. Art. № 103242. <https://doi.org/10.1016/j.jue.2020.103242>.

Obaco M., Royuela V., Vitores X. Computing functional urban areas using a hierarchical travel time approach: An Applied Case in Ecuador // 56th Congress of the European Regional Science Association “Cities & Regions: Smart, Sustainable, Inclusive?”. Vienna, Louvain-la-Neuve: European Regional Science Association (ERSA), 2016. 21 p.

Okwuashi O., Ndehedehe Ch. E. Integrating machine learning with Markov chain and cellular automata models for modelling urban land use change // *Remote Sensing Applications: Society and Environment*. 2021. Vol. 21. Art. № 100461. <https://doi.org/10.1016/j.rsase.2020.100461>.

Rahmah N., Sitanggang I. S. Determination of optimal epsilon (Eps) value on DBSCAN Algorithm to clustering data on peatland hotspots in Sumatra // *IOP Conference Series: Earth and Environmental Science*. 2016. Vol. 31. Art. № 012012. <https://doi.org/10.1088/1755-1315/31/1/012012>.

Schug F., Okujeni A., Hauer J. et al. Mapping patterns of urban development in Ouagadougou, Burkina Faso, using machine learning regression modeling with bi-seasonal Landsat time series // *Remote Sensing of Environment*. 2018. Vol. 210. P. 217–228. <https://doi.org/10.1016/j.rse.2018.03.022>.

Uchida H., Nelson A. Agglomeration index: Towards a new measure of urban concentration // *Urbanization and development: Multidisciplinary perspectives* / Ed. by J. Beall, B. Guha-Khasnobis, R. Kanbur. Oxford: Oxford University Press, 2010. P. 1–16. <https://doi.org/10.1093/acprof:oso/9780199590148.003.0003>.

Yang J., Gong J., Tang W. et al. Delineation of urban growth boundaries using a patch-based cellular automata model under multiple spatial and socio-economic scenarios // *Sustainability*. 2019. Vol. 11, № 21. Art. № 6159. <https://doi.org/10.3390/su11216159>.

Информация об авторах

Д. М. Радченко – научный сотрудник Центра пространственной экономики Института прикладных экономических исследований ФГБОУ ВО «Российская академия народного хозяйства и государственной службы при Президенте Российской Федерации», 119571, Россия, Москва, пр-т Вернадского, 82

SPIN-код (РИНЦ): 1832-2477
AuthorID (РИНЦ): 902140
Web of Science ResearcherID: N-7848-2018
Scopus Author ID: 57225991715

Ю. Ю. Пономарев – кандидат экономических наук, руководитель Центра пространственной экономики Института прикладных экономических исследований ФГБОУ ВО «Российская академия народного хозяйства и государственной службы при Президенте Российской Федерации», 119571, Россия, Москва, пр-т Вернадского, 82; старший научный сотрудник фонда «Институт экономической политики им. Е. Т. Гайдара», 125009, Россия, Москва, Газетный пер., 3–5, стр. 1

SPIN-код (РИНЦ): 9934-4626
AuthorID (РИНЦ): 785681
Web of Science ResearcherID: P-6953-2014
Scopus Author ID: 57195243543

К. В. Ростислав – научный сотрудник Центра пространственной экономики Института прикладных экономических исследований ФГБОУ ВО «Российская академия народного хозяйства и государственной службы при Президенте Российской Федерации», 119571, Россия, Москва, пр-т Вернадского, 82

SPIN-код (РИНЦ): 3951-7222
AuthorID (РИНЦ): 1032315
Web of Science ResearcherID: W-4835-2019
Scopus Author ID: 57226001826

Вклад авторов: все авторы сделали эквивалентный вклад в подготовку публикации.

Авторы заявляют об отсутствии конфликта интересов.

Статья поступила в редакцию 25.09.2023; одобрена после рецензирования 12.03.2024; принята к публикации 12.03.2024.

References

- Kolomak, Ye. A. and Trubekkhina, I. Ye. (2013), “Analyzing the agglomeration processes in the Novosibirsk Oblast”, *Region: Economics and Sociology*, no. 3, pp. 239–259.
- Listengurt, F. M. (1975), “Criteria for the allocation of large-scale agglomerations in the USSR”, *Izvestiya Rossiiskoi Akademii Nauk. Seriya Geograficheskaya*, no. 1, pp. 48–50.
- Monastyrskaya, M. Y. and Peslyak, O. A. (2017), “Modern approaches to the delimitation of urban agglomerations”, *Urban Construction and Architecture*, vol. 7, no. 3, pp. 80 – 86, <https://doi.org/10.17673/Vestnik.2017.03.14>.

Raysikh, A. (2020), "Defining the boundaries of urban agglomerations: Problems, international experience, solutions and results", *Demographic Review*, vol. 7, no. 1, pp. 27–53, <https://doi.org/10.17323/demreview.v7i1.10819> [Online], available at: <https://demreview.hse.ru/article/view/10819/12136> (Accessed Sept. 4, 2023).

Arribas-Bel, D., Garcia-López, M.-A. and Viladecans-Marsal, E. (2021), "Building(s and) cities: Delineating urban areas with a machine learning algorithm", *Journal of Urban Economics*, vol. 125, art. no. 103217, <https://doi.org/10.1016/j.jue.2019.103217>.

Coombes, M. (2004), "Multiple dimensions of settlement systems: Coping with complexity", in Champion, T. and Hugo, G. (eds.), *New forms of urbanization: Beyond the urban–rural dichotomy*, Ashgate Publishing Limited, Aldershot, UK, pp. 307–324.

De Bellefon, M.-P., Combes, P.-P., Duranton, G. et al. (2021), "Delineating urban areas using building density", *Journal of Urban Economics*, vol. 125, art. no. 103226, pp. 103–226, <https://doi.org/10.1016/j.jue.2019.103226>.

Ester, M., Kriegel, H.-P., Sander, J. et al. (1996), "A density-based algorithm for discovering clusters in large spatial databases with noise", in *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining (KDD-96)*, AAAI Press, Portland, OR, US, pp. 226–231.

Gajović, V. (2013), "Comparative analysis of different methods and obtained results for delineation of functional urban areas", *Spatium International Review*, no. 29, pp. 8–15, <https://doi.org/10.2298/SPAT1329008G>.

Galdo, V., Li, Yu. and Rama, M. (2021), "Identifying urban areas by combining human judgment and machine learning: An application to India", *Journal of Urban Economics*, vol. 125, art. no. 103229, <https://doi.org/10.1016/j.jue.2019.103229>.

Guzik, R., Kołoś, A. and Gwosdz, K. (2017), "Interconnections in public transport as a method for delimiting urban functional areas and the settlement hierarchy in Poland", *Regional Statistics*, vol. 7, no. 1, pp. 63–77, <https://doi.org/10.15196/RS07104>.

Hagenauer, J. and Helbich, M. (2012), "Mining urban land-use patterns from volunteered geographic information by means of genetic algorithms and artificial neural networks", *International Journal of Geographical Information Science*, vol. 26, no. 6, pp. 963–982, <https://doi.org/10.1080/13658816.2011.619501>.

Moreno-Monroy, A. I., Schiavina, M. and Veneri, P. (2021), "Metropolitan areas in the world. Delineation and population trends", *Journal of Urban Economics*, vol. 125, art. no. 103242, <https://doi.org/10.1016/j.jue.2020.103242>.

Obaco, M., Royuela, V. and Vitores, X. (2016), "Computing functional urban areas using a hierarchical travel time approach: An Applied Case in Ecuador", in *56th Congress of the European Regional Science Association "Cities & Regions: Smart, Sustainable, Inclusive?"*, European Regional Science Association (ERSA), Vienna, Austria, Louvain-la-Neuve, Belgium, 21 p.

Okwuashi, O. and Ndehedehe, Ch. E. (2021), "Integrating machine learning with Markov chain and cellular automata models for modelling urban land

use change”, *Remote Sensing Applications: Society and Environment*, vol. 21, art. no. 100461, <https://doi.org/10.1016/j.rsase.2020.100461>.

Rahmah, N. and Sitanggang, I. S. (2016), “Determination of optimal epsilon (Eps) value on DBSCAN Algorithm to clustering data on peatland hotspots in Sumatra”, *IOP Conference Series: Earth and Environmental Science*, vol. 31, art. no. 012012, <https://doi.org/10.1088/1755-1315/31/1/012012>.

Schug, F., Okujeni, A., Hauer, J. et al. (2018), “Mapping patterns of urban development in Ouagadougou, Burkina Faso, using machine learning regression modeling with bi-seasonal Landsat time series”, *Remote Sensing of Environment*, vol. 210, pp. 217–228, <https://doi.org/10.1016/j.rse.2018.03.022>.

Uchida, H. and Nelson, A. (2010), “Agglomeration index: Towards a new measure of urban concentration”, in Beall, J. Guha-Khasnobis, B. and Kanbur, R. (eds.), *Urbanization and development: Multidisciplinary perspectives*, Oxford University Press, Oxford, UK, pp. 1–16, <https://doi.org/10.1093/acprof:oso/9780199590148.003.0003>.

Yang, J., Gong, J., Tang, W. et al. (2019), “Delineation of urban growth boundaries using a patch-based cellular automata model under multiple spatial and socio-economic scenarios”, *Sustainability*, vol. 11, no. 21, art. no. 6159, <https://doi.org/10.3390/su11216159>.

Information about the authors

D. M. Radchenko – Researcher of the Centre for Spatial Economics, Institute of Applied Economic Research, The Russian Presidential Academy of National Economy and Public Administration, 82 Vernadsky Ave., Moscow, Russia, 119571
 SPIN-code (RSCI): 1832-2477
 AuthorID (RSCI): 902140
 Web of Science ResearcherID: N-7848-2018
 Scopus Author ID: 57225991715

Yu. Yu. Ponomarev – Candidate of Economics, Head of the Centre for Spatial Economics, Institute of Applied Economic Research, The Russian Presidential Academy of National Economy and Public Administration, 82 Vernadsky Ave., Moscow, Russia, 119571; Senior Research of the Gaidar Institute for Economic Policy, 3–5 Gazetny Lane, Building 1, Moscow, Russia, 125009
 SPIN-code (RSCI): 9934-4626
 AuthorID (RSCI): 785681
 Web of Science ResearcherID: P-6953-2014
 Scopus Author ID: 57195243543

K. V. Rostislav – Researcher of the Centre for Spatial Economics, Institute of Applied Economic Research, The Russian Presidential Academy of National Economy and Public Administration, 82 Vernadsky Ave., Moscow, Russia, 119571
 SPIN-code (RSCI): 3951-7222
 AuthorID (RSCI): 1032315

Web of Science ResearcherID: W-4835-2019

Scopus Author ID: 57226001826

Contribution of the authors: the authors contributed equally to this article.

The authors declare no conflicts of interests.

*The article was submitted 25.09.2023; approved after reviewing 12.03.2024;
accepted for publication 12.03.2024.*