

# Геоинформационные методы поиска территорий повышенного риска распространения социально значимых инфекций в мегаполисах (на примере туберкулеза)

И.Б. Куликова<sup>1</sup>, И.С. Кузнецов<sup>2,3</sup>, В.Г. Коровка<sup>2</sup>, М.В. Бельтюков<sup>2,4</sup>,  
В.Б. Галкин<sup>2</sup>, О.П. Соколова<sup>2</sup>, Е.А. Паниди<sup>3</sup>, П.К. Яблонский<sup>2,3</sup>

<sup>1</sup>Министерство здравоохранения Российской Федерации, Москва

<sup>2</sup>Санкт-Петербургский научно-исследовательский институт фтизиопульмонологии

<sup>3</sup>Санкт-Петербургский государственный университет

<sup>4</sup>Санкт-Петербургский государственный электротехнический университет «ЛЭТИ»  
им. В.И. Ульянова (Ленина)

## Geoinformation search methods for increased risk areas for the spread of socially significant infections in metropolitan cities (through TB as an example)

I. Kulikova<sup>1</sup>, I. Kuznetsov<sup>2,3</sup>, V. Korovka<sup>2</sup>, M. Beltyukov<sup>2,4</sup>, V. Galkin<sup>2</sup>,  
O. Sokolova<sup>2</sup>, E. Panidi<sup>3</sup>, P. Yablonskiy<sup>2,3</sup>

<sup>1</sup>Ministry of Health of the Russian Federation, Moscow

<sup>2</sup>St. Petersburg State Research Institute of Phthisiopulmonology

<sup>3</sup>St. Petersburg State University

<sup>4</sup>St. Petersburg State Electrical Engineering University

© Коллектив авторов, 2023 г.

### Резюме

Для изучения распространения туберкулеза могут быть применены методы смежных наук. Картография может выступить в роли интеграционной науки, методы которой могут использоваться в организации эпидемиологической работы. **Цель исследования:** геопространственный анализ данных медицинской статистики при помощи методов ГИС с последующим созданием геоинформационной модели распространения заболевания. **Материалы и методы исследования.** Проведено пространственное исследование ретроспективных сведений о местах проживания

пациентов в двух мегаполисах, предложены методические решения для подготовки, анализа и визуализации данных. Использовались программные средства, входящие в состав открытого программного обеспечения QGIS, такие как метод агрегирования данных до отдельных домов, метод построения регулярных сеток, методы кластерного анализа (k-means) и построения «тепловых» карт (KDE). **Результаты исследования.** Методами пространственного анализа установлено, что в крупных мегаполисах существуют территории, на которых отмечается длительное устойчивое воспроизведение возникновения случаев забо-

левания. Методом агрегирования данных определено, что в 3,8% от всех жилых домов Санкт-Петербурга зарегистрировано 41,5% всех случаев туберкулеза, а в 1,4% жилого фонда Екатеринбурга зарегистрировано 19,7% всех случаев туберкулеза. По результатам построения регулярных сеток выявлено, что среднее число больных на отдельных территориях кратно превышало средний показатель по городу. В Санкт-Петербурге существуют от 34,2 до 40,3% территорий, где ранее отмечались случаи туберкулеза, а в Екатеринбурге — от 31,3 до 34,1%. **Заключение.** Впервые проанализированные пространственные многолетние данные в двух мегаполисах России позволяют по-иному взглянуть на проблему распределения зарегистрированных случаев туберкулеза при разных масштабах: от города в целом до отдельного дома, что в свою очередь дает новое знание для выработки доказательных противоэпидемических и профилактических мероприятий.

**Ключевые слова:** медицинские пространственные данные, медицинская картография, медицинская география, управление пространственными данными, социально значимые заболевания, туберкулез

### Summary

Methods of related sciences can be applied to study the spread of tuberculosis. Cartography may act as an integration science, the methods of which can be used in the organization of epidemiological work. **Purpose:** Use GIS methods to conduct a geospatial analysis of medical statistics' data, and then create a geoinformation model of the spread of the disease. **Materials and methods.** A spatial study of retrospective information about the places of residence of patients in two megacities was

conducted. Methodological solutions for data preparation, analysis and visualization have been proposed. The software tools from the QGIS open source software were used in the work. The research uses such methods as the method of aggregating data to individual houses, the method of constructing regular grids, methods of cluster analysis (k-means) and the construction of KDE maps. **Results.** Spatial analysis methods have established that in large megacities there are territories in which there is a long-term stable reproducibility of cases of the disease. The data aggregation method determined that 41.5% of all cases of tuberculosis were registered in 3.8% of all residential buildings in St. Petersburg, and 19.7% of all cases of tuberculosis were registered in 1.4% of the residential buildings in Yekaterinburg. According to the results after constructing regular grids, it was revealed that the average number of patients in certain territories was many times higher than the average for the city. In St Petersburg, there are from 34.2 to 40.3% of territories where cases of tuberculosis were previously reported, and in Yekaterinburg — from 31.3 to 34.1%. **Conclusion.** Long-term spatial data, which was analyzed for the first time in two megacities of Russia, allow us to take a different look at the problem of the distribution of registered cases of tuberculosis at different scales: from the city as a whole to a separate building, which in turn gives new knowledge for the development of evidence-based anti-epidemic and preventive measures.

**Keywords:** medical geospatial data, medical cartography, medical geography geospatial data management, socially valuable diseases, tuberculosis

### Введение

Поиск закономерностей распространения инфекционных заболеваний неизменно ведется с учетом механизма передачи возбудителя. Особый интерес представляют заболевания, передающиеся аэрогенным (воздушно-капельным) путем. Одно из наиболее опасных заболеваний с подобным механизмом заражения — туберкулез. На фоне стабильного снижения показателей заболеваемости и распространенности туберкулез является наиболее подходящей инфекцией, для изучения которой могут быть применены методы и приемы смежных естественных наук. Так, в ряде регионов, например, в Санкт-Петербурге, достигнут абсолютный минимум заболеваемости за всю историю наблюдений [1, 2]. Все это свидетельствует, что складывающиеся условия благоприятствуют соз-

данию и апробации новых эффективных механизмов управления эпидемическим процессом для достижения целевых показателей стратегии ВОЗ «Ликвидации туберкулеза» («END TB») по снижению заболеваемости туберкулезом к 2035 г. на 90%, а летальности — на 95% [3, 4]. Предпринятое исследование представляло интерес еще и потому, что его результаты в перспективе могут быть корректно экстраполированы на актуальные инфекции с аналогичным механизмом передачи (грипп, COVID-19, корь и т.д.).

Наши предыдущие исследования выявили значительное количество неединичных очагов, которые в силу принятой годовой системы учета и отчетности представлялись спорадическими, но при проведении многолетнего мониторинга с использованием геоинформационных приемов оказались эндемичными, что доказывалось множественными случаями

заболевания туберкулезом по одному адресу [5]. По этой причине именно медицинская картография и география выступили в роли интеграционной науки, результаты которой могут как использоваться в организации эпидемиологической работы в медицинских учреждениях, так и стать фундаментом дальнейших медико-географических исследований территорий.

В отечественной практике создание геоинформационных систем (ГИС) проводят на основе отчетно-популяционного анализа исключительно в рамках административных территорий разного уровня [6]. Переход на уровень отдельных домохозяйств в изучении туберкулеза ранее не предпринимался, однако именно он может стать основным при планировании противоэпидемических мероприятий [7]. В зарубежной практике одними из самых популярных методов пространственного анализа заболеваний на примере COVID-19 являются агрегирование данных на уровне региона и построение регулярных сеток для оценки заболеваемости, а наиболее популярными методами пространственной кластеризации остаются методы глобального и локального индекса Морана [8]. В ряде исследований методы геоинформационного анализа применяются для прогнозирования развития заболевания и связи с культурными особенностями местного населения [9]. Нередко объектами пространственного сравнения становятся сопутствующие заболевания, например, туберкулез, ВИЧ-инфекция, сахарный диабет, гепатиты В и С [10]. Однако эти исследования охватывают только крупные (в географическом плане) территории и не концентрируют внимание на отдельных проблемных территориях и мегаполисах.

С другой стороны, к настоящему времени успешно опробованы и внедрены в практику инструменты для веб-визуализации пространственной информации [11], по предложению территориальных противотуберкулезных служб ряда регионов совершенствуются способы трехмерной визуализации жилых строений с целью наиболее точного определения мест проживания пациентов и контингентов риска [12].

Внедрению этих методов способствует и действующая на территории Российской Федерации нормативно-правовая база. В соответствии с Указом Президента Российской Федерации от 11.03.2019 № 97 «Об основах государственной политики Российской Федерации в области обеспечения химической и биологической безопасности на период до 2025 года и дальнейшую перспективу» геоинформационные системы (ГИС) прямо обозначены в качестве средств обеспечения медицинских (медико-географических) исследований и мониторинга. Ключевым является приказ Министерства здравоохранения Российской Федерации от 21.03.2003 № 109 «О совершенствовании противотуберкулезных мероприятий в Россий-

ской Федерации» [13], в котором можно найти рекомендации по картографированию территории как одному из разделов противоэпидемической работы. Результатом такой работы предполагается получение карты-схемы, которая позволяет эффективно определять и верифицировать лиц, относящихся к различным категориям контактных.

### Цель исследования

Проведение геопространственного анализа данных медицинской статистики при помощи отдельных методов, которые способны обеспечить создание геоинформационной модели распространения заболевания, выявить территории риска возникновения новых случаев заболевания и подготовить материалы для совершенствования организации противоэпидемической работы на этих территориях.

### Материалы и методы исследования

Технологической базой для выполнения данного исследования послужила геоинформационная платформа QGIS.

В рамках данного исследования была выдвинута следующая гипотеза: в крупных мегаполисах существуют территории, в границах которых на протяжении ряда лет регистрируется множество случаев туберкулеза, что представляется признаком потенциального риска возникновения новых случаев заболевания. Эти территории могут быть установлены различными методами геоинформационного анализа.

В данном исследовании границы очага туберкулеза соответствовали адресу проживания/регистрации больного — фактически дом, строение или жилой корпус. Под очагом туберкулеза понимается место пребывания источника микобактерий туберкулеза вместе с окружающими его людьми и обстановкой в тех пределах пространства и времени, в которых возможно возникновение новых заражений и заболеваний [13].

Формализованная и использованная в настоящем исследовании методика определения территорий риска распространения туберкулеза включает:

- 1) формирование исходной базы медицинских данных;
- 2) геокодирование адресов регистрации случаев заболевания (очагов);
- 3) установление территорий риска заболевания следующими методами:
  - а) агрегирование данных в установленных границах;
  - б) построение гексагональных сеток;
  - в) построение «тепловой» карты на основе метода ядерной оценки плотности;

- г) построение карт на основе методов кластерного анализа (в частности, метода  $k$ -средних);
- 4) разграничение границ территорий риска, картографическая визуализация потенциально опасных территорий.

Адресная база данных о больных туберкулезом сформирована на основе их регистрации в системе автоматического эпидемиологического учета инфекционных больных на базе САУ «Инфекция» и Федеральном регистре больных туберкулезом (ФРБТ). Диагноз туберкулеза подтверждался СПб ГБУЗ «Городской противотуберкулезный диспансер» и ГБУЗ СО «Областной клинический медицинский центр фтизиопульмонологии и инфекционных заболеваний» г. Екатеринбурга. База данных включает сведения о больных туберкулезом в г. Санкт-Петербурге за 2007–2022 гг., в г. Екатеринбурге за 2017–2022 гг. Все сведения, представленные в работе, перед началом их использования были обезличены в соответствии с требованиями Федерального закона от 27.07.2006 № 152-ФЗ «О персональных данных».

Адресная информация была преобразована в соответствии с требованиями описания адресов, используемых в открытом API OSM Nominatim для определения географических координат. В ФБГУ «СПб НИИФ» Минздрава России в 2020 г. на основе открытого API разработана программа геокодирования данных под рабочим названием GeoMedica (свидетельство о государственной регистрации программы для ЭВМ №2020665320). Пространственные данные хранились в виде отдельных файлов в формате GeoPackage (GPKG). Система координат, в которой хранились данные, — WGS 84 (ESPG 3857). Информация о числе жилых строений, в том числе многоэтажных, для городов и районов была взята как из открытых источников официальных органов исполнительной власти, так и из сторонних городских веб-сайтов и карт. Территории, на которых невозможно установить или оценить численность населения, в расчетах не учитывались.

После формализации адресов была проведена процедура по формированию площадных и точечных слоев векторных объектов цифровой карты. На этапе обработки из базы данных исключены:

- 1) сведения о пациентах, не имеющих диагноза «туберкулез» в атрибутивной таблице;
- 2) сведения о некорректных и не существующих в городах адресах;
- 3) сведения о пациентах, адрес проживания которых установлен в другом субъекте.

Для Санкт-Петербурга из 33 936 записей о местах проживания пациентов нанесено на карту 33 395 записи. Относительная погрешность геокодирования ( $d_r$ ) составила 2,6%. Из 9135 записей о местах работы пациентов геокодированы 8684 адреса ( $d_r=4,9\%$ ).

Для г. Екатеринбурга географически определялись только сведения по месту проживания пациента. Из 3574 записей о местах проживания пациентов географически установлены 3424 адреса ( $d_r=4,2\%$ ).

Относительная погрешность геокодирования неравномерно распределена по районам городов и зависит от полноты представляемых данных (наличия адресов мест работы/учебы, адресов контактных и т.д.) и от корректности их ввода специалистами первичного звена противотуберкулезной службы.

Гипотеза о существовании территорий риска распространения туберкулеза в рамках мегаполисов была проверена несколькими различными методами.

Первый метод сводился к задаче агрегирования данных с множественными (неединичными) случаями заболевания по отдельному адресу. Задача сводилась к подсчету геометрических примитивов в виде точек внутри полигонов отдельных домов, муниципалитетов, районов. Полигоны домов, а также их количество были получены путем обращения к открытым данным OpenStreetMap. Группировка данных при агрегировании в ГИС выполнялась по методу естественных интервалов (метод Джентка), отдельно выделялись адреса с единичными случаями, адреса с множественными случаями классифицировались.

Критерии отбора адреса как очага с множественными случаями ( $\alpha \geq k$ , где  $k \in \mathbb{N}$ ) были установлены эмпирически как оптимальные для каждого из городов. Основным критерием отбора адреса в число адресов с множественными случаями туберкулеза являлся промежуток времени, за который были доступны медицинские данные. Дополнительными критериями отбора служили такие параметры, как число жителей по указанному адресу и плотность населения на этой территории, а также показатель заболеваемости. Расчет оптимальных параметров для определения значения величины  $\alpha$  в рамках данного исследования не проводился.

Значение величины  $\alpha$  соответствовало адресу, по которому в течение исследуемого периода регистрировалось: для Санкт-Петербурга — 5 и более случаев заболевания ( $\alpha \geq 5$ ), а для Екатеринбурга — 3 и более случаев заболевания ( $\alpha \geq 3$ ). Значение данной величины может меняться исходя из указанных выше критериев и требует уточнения в дальнейших исследованиях.

Вторым методом определения территорией риска распространения туберкулеза был выбран метод создания регулярных сеток. В настоящее время он является самым доступным средством классификации территорий в ГИС, давая возможность разделить территорию крупного мегаполиса на равные по площади объекты. В данной работе использовались гексагоны, поскольку шестиугольники позволяют уменьшить

смещение выборки из-за краевых эффектов, вызванных геометрией сетки, что связано с низким значением соотношения периметра к площади шестиугольника [14]. Построение шестиугольников производилось встроенными средствами QGIS, расчет длин ребер шестиугольников основан на данных библиотеки H3 от UBER (Официальная документация библиотеки h3 uber, 2023. <https://h3geo.org/docs/>). Масштабы слоя шестиугольников подобраны исходя из геометрии исследуемых городов. В границах каждого объекта определялась величина  $\alpha$  и ее среднее значения, определялись геометрии с наибольшими и наименьшими значениями, доля объектов с показателями, превышающими средние по городу. Объекты с нулевыми значениями исключались.

Третий метод основан на методе ядерной оценки плотности (KDE) с последующим построением тепловой карты. Основная идея метода заключается в создании сглаженной поверхности или растрового слоя, который представляет интенсивность точечных объектов в изучаемой области. Функция ядра, которая используется в методе, присваивает вес каждой точке на основании близости ее к центру ядра. Математическая формула, выражающая ядерную оценку плотности, выглядит следующим образом (1):

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

где  $K$  — неотрицательная функция;  $h > 0$  — сглаживающий параметр.

Визуальное представление этого метода будет сводиться к построению «тепловой карты» — графического изображения, при котором показатель отражается на карте интенсивностью его цветового обозначения. Наиболее важными параметрами в построении тепловой карты по методу ядерной оценки плотности будут служить вес точки, радиус окна поиска и тип ядерной функции (Kernel shape,  $K_u$ ), по которой будет определяться скорость изменения веса точки. Весом точки в данном исследовании выступит количество случаев заболевания в доме. Радиус окна поиска определялся с помощью эмпирического правила Сильвермана. Это правило учитывает и среднеквадратическую ошибку выборки, и размер выборки, что позволяет учесть возможное наличие пространственных выбросов. Типом ядерной функции  $K_u$  была определена квадратичная функция вида (2):

$$K_u = \frac{15}{16} (1 - u^2)^2, \quad (2)$$

где  $u \leq 1$

Построение «тепловых карт» производилось для различных промежутков времени, выбранных случайно.

Четвертым методом определения территорий риска распространения заболевания стал алгоритм

пространственной кластеризации K-means. В основе алгоритма лежит определение центра масс всех данных. Его применение на срезе ретроспективных данных позволяет установить границы возникновения очагов туберкулезной инфекции и подготовить аналитическую карту.

Метод K-means предполагает разбиение множества элементов векторного пространства (слоя) на заранее известное число кластеров  $K$ . В рамках данного исследования метод был применен как на ограниченной территории, так и для всего города. Для определения величины  $K$  использованы метод локтя (Elbow method) — определение числа кластеров через отношение дисперсии  $D(x)$  к подбираемому числу кластеров (оптимальным числом кластеров будет такое значение, которое может свести сумму внутрикластерной дисперсии каждого кластера к минимуму) и эмпирический метод (определение числа кластеров как функции от общего числа объектов).

$$D(x) = \frac{\sum (x - x_i)^2 \cdot f_i}{\sum f_i}, \quad (3)$$

где  $D(x)$  — величина дисперсии;  $i$  — анализируемый показатель (уникальный идентификатор кластера);  $x$  — среднее значение показателя;  $f_i$  — количество значений в анализируемой совокупности данных (случаев туберкулеза).

Для реализации метода локтя использован инструмент «Базовая статистика по полям», доступный в QGIS, который позволил определить величину среднеквадратического отклонения внутри кластера и, как следствие, величину дисперсии.

Для представления сформированных кластеров на карте территории потенциального риска распространения туберкулеза необходимо перейти от представления случаев в форме точек к отображению площадными объектами. Разграничение границ территорий потенциального риска проводилось двумя способами: ручной векторизацией полигонов на основании результатов работы всех методов и построением полигонов Вороного.

## Результаты

По результатам пространственного анализа многолетних данных о регистрации случаев по отдельным адресам установлено следующее.

1. В Санкт-Петербурге из 47 000 жилых строений случаи туберкулеза отмечались в 12 691 (27,0%; 95% ДИ 26,7–27,3%) жилом доме, соответственно в 34 309 (73,0%; 95% ДИ 72,7–73,3%) жилых домах за исследуемый период не регистрировалось ни одного случая туберкулеза (рис. 1). Из 12 691 жилого дома города, в которых отмечались случаи туберкулеза:

а) 10 882 (85,7% от пораженных) адресов отнесены к адресам со спорадической заболеваемостью ( $\alpha < 5$ ). По этим адресам проживало 19 520 (58,5%; 95% ДИ 58,0–58,9%) пациентов с диагнозом «туберкулез»;

б) 1809 (14,3% от пораженных) домов отнесены к адресам с множественными случаями ( $\alpha \geq 5$ ). По этим адресам проживало 13 875 (41,5%; 95% ДИ 41,1–42,0%) пациентов с диагнозом «туберкулез».

2. В Екатеринбурге за 6-летний период из 13 620 жилых строений [OSM] случаи туберкулеза отмечались в 2442 (17,9%; 95% ДИ 17,4–18,5%) жилых строениях. Следовательно, по 11 178 (82,1%; 95% ДИ 81,5–82,6%) адресам за исследуемый период не регистрировалось ни одного случая (рис. 2). Из 2442 адресов жилых домов города:

- 2246 (92,0% пораженных) адресов отнесены к адресам со спорадической заболеваемостью

( $\alpha < 3$ ). В этих домах проживало 2750 (80,3%; 95% ДИ 79,2–81,4%) больных с диагнозом «туберкулез»;

- 196 (8,0% пораженных) адресов отнесены к адресам с множественными случаями ( $\alpha \geq 3$ ). В них проживало 674 (19,7%; 95% ДИ 18,6–20,8%) больных с диагнозом «туберкулез».

Несмотря на статистически значимые ( $p < 0,001$ ) различия в рассмотренных мегаполисах по доле домов, где регистрировались случаи туберкулеза, в том числе множественные, проведенные расчеты показывают (табл. 1), что при ретроспективном пространственном анализе лишь незначительная часть (менее 4%) всего жилого фонда обоих мегаполисов дает значительную часть (20–40%) новых случаев туберкулеза.

Например, в 3,8% всех жилых домов (1809 домов) Санкт-Петербурга суммарно зарегистрировано 41,5% всех случаев туберкулеза, а в 1,4% (196 домов)

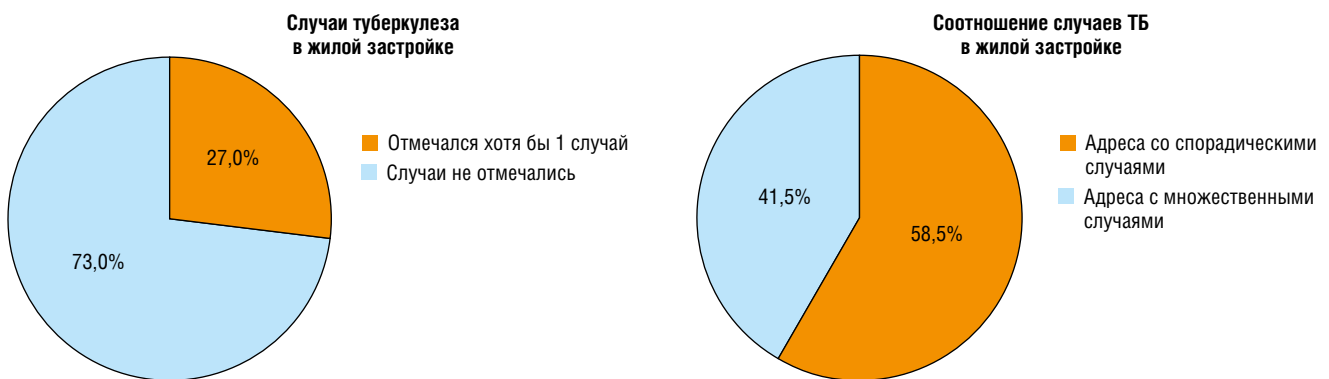


Рис. 1. Соотношение регистрации случаев туберкулеза в жилой застройке Санкт-Петербурга

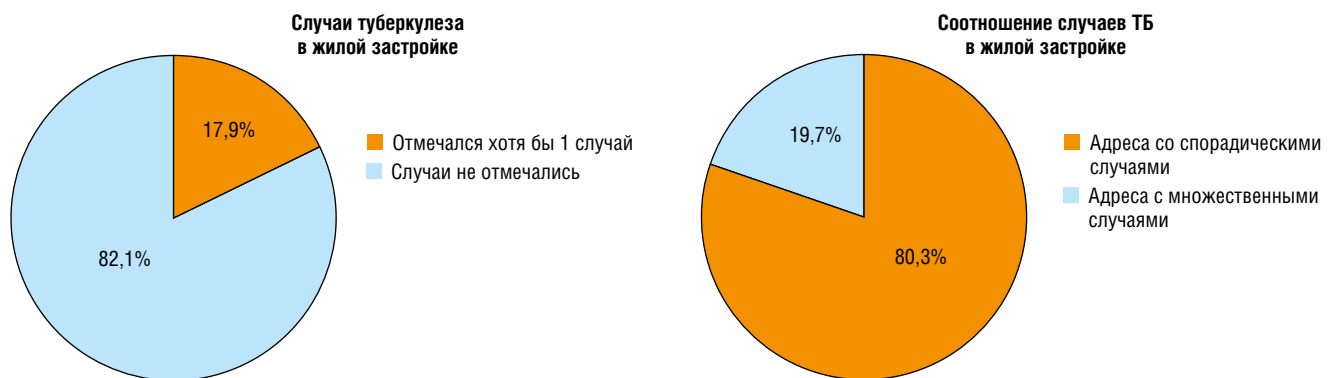


Рис. 2. Соотношение регистрации случаев туберкулеза в жилой застройке Екатеринбурга

Таблица 1

#### Количество домов со спорадическими и множественными случаями (%)

Город	Жилых домов	Отмечались случаи туберкулеза (%)	Адресов со спорадическими случаями туберкулеза (%)	Адресов с множественными случаями туберкулеза (%)	Больных в адресах с множественными случаями (% от всех случаев заболевания)
Санкт-Петербург	47 000	12 691 (27,0)	10 882 (23,2)	1809 (3,8)	13 875 (41,5)
Екатеринбург	13 620	2441 (17,9)	2246 (16,5)	196 (1,4)	674 (19,7)

жилого фонда Екатеринбурга зарегистрировано суммарно 19,7% всех случаев туберкулеза. Таким образом, показано, что на территории двух мегаполисов существуют адреса, дающие наибольшее число случаев заболевания.

На этапе планирования исследования предполагалось, что наибольшее число больных будут давать случайно расположенные отдельные дома, в которых регистрировалось наибольшее число случаев, и рост общего числа больных будет происходить именно в этих домах. Например, в Санкт-Петербурге установлено 19 объектов, в которых за время наблюдения выявлено более 20 случаев туберкулеза на адрес. Суммарно в 19 адресах зарегистрировано 492 больных. Изучена степень влияния таких аномально высоких значений на общую аналитическую картину.

Для установления конкретных групп домов, которые давали наибольшее количество всех случаев туберкулеза, выполнялась пространственная группировка агрегированных данных в ГИС методом естественных интервалов Дженкса. Это позволило составить таблицу отношения множественных случаев туберкулеза к числу адресов, их составляющих (табл. 2).

Из таблицы следует, что больше всего больных (89,5% в Санкт-Петербурге и 98,8% в Екатеринбурге) проживали по адресам, которые по результатам пространственной группировки находились в срединных группах. Наибольшее число больных из множественных очагов будет относиться к адресам с относительно небольшим числом случаев. Однако для достоверного подтверждения данного факта необходимо проводить пространственные группировки другими, более сложными методами анализа.

Границей очага может выступать не только отдельный дом, но чаще всего и подъезд многоквартирного дома, и придомовая территория, и группа домов. И поэтому для примерного определения территорий потенциального риска распространения туберкулеза были построены регулярные сетки в виде геометрических шестиугольных примитивов или объектов. Реляционные сетки выполнены для разных масштабов и разных длин ребер. Результат работы метода представлены в табл. 3.

Одновременно с этим на электронной карте средствами ГИС группировались и отмечались те объекты гексагональной сетки, значения показателей которых были равны среднему для установленного масштаба

Таблица 2

#### Группировка больных по их числу на один адрес методом естественных интервалов

Группировка больных по их числу на 1 адрес	Число пораженных строений	Доля домов с множественными случаями, %	Число больных в выборке	Доля больных из адресов с множественными случаями, %
<i>Санкт-Петербург, 2007–2022 гг.</i>				
5–10	1454	80,5	9169	66,3
10–15	283	15,1	3289	23,2
15–20	53	2,9	926	6,7
Больше 20	19	1,0	491	3,5
<i>Екатеринбург, 2017–2022 гг.</i>				
3–4	176	89,7	564	83,7
5–6	18	9,1	102	15,1
Больше 6	1	0,5	8	1,2

Таблица 3

#### Расчет числа больных в правильных геометрических примитивах

Длина ребра, м	Площадь, км <sup>2</sup>	Всего объектов	Среднее число больных в 1 объекте	Объектов с показателями выше среднего	Доля объектов от всех примитивов, %
<i>Санкт-Петербург, 2007–2022 гг.</i>					
3700	35,6	342	98,9	117	34,2
1400	5,1	1246	27,1	467	37,5
530	0,72	4458	7,6	1795	40,3
<i>Екатеринбург, 2017–2022 гг.</i>					
1400	5,1	360	9,5	123	34,1
530	0,72	1112	3,6	348	31,3
200	0,10	2192	1,6	733	33,4

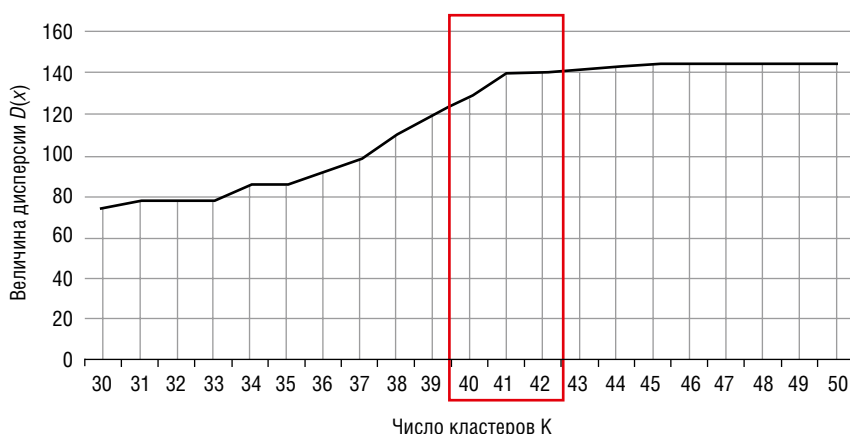


Рис. 3. График зависимости числа кластеров от величины внутрикластерной дисперсии на экспериментальной территории

или превышали в несколько раз (приложение 1 и 2). При построении разномасштабных объектов с разной длиной ребра отмечено, что с увеличением масштаба территории потенциального риска будут различимы визуально.

Альтернативой методу реляционных сетей может служить метод ядерной оценки плотности (KDE). На основе метода были определены территории скопления пациентов с диагнозом «туберкулез». При изучении результатов работы алгоритма на разных временных промежутках было замечено, что данные скопления больных в отдельных районах мегаполисов достаточно стабильны, хоть и меняются в зависимости от временного промежутка. Данный метод применен на территории Екатеринбурга. На примере Екатеринбурга видно, что в период с 2017 по 2018 г. (приложение 3б) новые случаи туберкулеза отмечались в разных районах города, тогда как в 2019–2020 гг. (приложение 3в) случаи отмечались на севере и северо-востоке города. Если же анализировать временной промежуток с 2017 по 2022 г. (приложение 3а), можно отметить скопление случаев туберкулеза в северо-восточной части города.

Исходя из результата анализа долей геометрических объектов, в которых среднее число больных превышало средний показатель, можно утверждать, что в Санкт-Петербурге существуют от 34,2 до 40,3% территорий, где ранее отмечались случаи туберкулеза, которые могут быть отнесены к территориям повышенного риска возникновения туберкулеза. В Екатеринбурге данный показатель ниже и составляет от 31,3 до 34,1% таковой (приложение 2). Данные факты подтверждают, что в каждом отдельном мегаполисе будут существовать территории, на которых риск развития заболевания значительно выше.

Следующим методом поиска территорий потенциального риска выступил метод кластерного анализа. С целью сокращения времени получения результатов было принято решение провести отработку метода на

ограниченной экспериментальной территории двух административных районов Санкт-Петербурга, данные для которой были формализованы одними из первых.

В результате анализа необходимого числа кластеров для исследуемой территории сформирован график зависимости числа кластеров от величины дисперсии (рис. 3). При достижении нужного числа кластеров величина дисперсии остается практически неизменной, следовательно, дальнейшее увеличение числа кластеров не требуется.

После определения числа кластеров геокодированные очаги туберкулеза в исследуемой области были разбиты на кластеры (приложение 4). При построении границ ареалов риска развития заболеваемости использованы полигоны Вороного. Отметим, что результат кластеризации не учитывал характеристики больных (бактериовыделение, ВИЧ-инфекция, гепатиты В и С, МЛУ/ШЛУ и т.д.), что может быть учтено при дальнейшей корректировке границ территорий потенциального риска распространения туберкулеза. Полученная таким образом тематическая медицинская карта (приложение 5) отображает потенциально опасные территории развития заболеваемости на основе многолетних ретроспективных данных и выявления эндемичных очагов, сформированных в недавнее время. В рамках двух административных районов методами геоинформационных систем были установлены наиболее пораженные территории. Далее метод определения территории потенциального риска распространения туберкулеза был экстраполирован до масштаба всего мегаполиса. Используя грубое определение классов для кластеризации мест проживания пациентов в Санкт-Петербурге, территория была разделена на 129 кластеров, а для кластеризации мест работы пациентов территория была разделена на 69 кластеров.

Таким образом, методом кластеризации возможно установить территории потенциального риска развития заболевания как для мест проживания пациентов,



так и для мест их работы, что позволяет направить усилия противотуберкулезной службы на самые отягощенные территории.

## Обсуждение

Результаты исследования показывают, что в настоящее время имеется техническая возможность определения границ территорий повышенного риска возникновения случаев заболевания с использованием геоинформационных систем. В рамках исследования был введен новый термин — «территория потенциального риска распространения заболевания», характеризующий пространство, на котором отмечается длительное устойчивое воспроизведение возникновения случаев заболевания.

Методы геоинформационного анализа многолетней регистрации случаев заболевания демонстрирует неравномерность их территориального распределения. Применение методов геоинформационного картографирования доказывает, что традиционный анализ интенсивных показателей распространения заболеваний, проводящийся в соответствии с административно-территориальными границами (районов, муниципалитетов, поликлиник, медицинских участков), часто не соответствует зонам эпидемического неблагополучия, что диктует необходимость проведения эпидемиологического надзора с использованием современных методов геоинформатики. К аналогичным выводам ранее пришли как отечественные, так и зарубежные исследователи при работе по смежным тематикам [15].

При этом, в отличие от приведенного ранее зарубежного опыта геоинформационного картографирования в медицине, в данном исследовании в большей степени используется персонифицированная информация о пациенте. Агрегирование пространственных сведений отдельных районов или муниципалитетов при исследовании заболеваний в крупных мегаполисах было впервые дополнено агрегированием данных о пациентах вплоть до отдельного жилого дома и даже квартиры! И результаты исследования подтвердили, что только при этом условии возможно объективно и наглядно определять территории потенциального риска в мегаполисах.

Очевидно, что определение территорий потенциального риска развития заболевания, особенно в крупных мегаполисах, где число контактов с носителями воздушно-капельных инфекций максимально высоко, должно стать приоритетной задачей эпидемиологов и фтизиатров. Тогда как проведение эпидемиологического анализа без учета географического фактора распределения заболевания представляется чрезвычайно затратным, не предполагает перехода к рискоориентированной технологии противотуберкулезной работы, что существенно снижает эффективность всей проводимой работы.

Метод агрегирования пространственных данных по отдельным адресам показывает, что территорией потенциального риска выступает как отдельный жилой дом, так и группа домов в пределах одного микрорайона. Являясь самым доступным для реализации, данный метод может быть использован без применения геоинформационных систем, поскольку в конечном результате сводится к вычислению повторяющихся значений (случаев заболевания) для одного адреса. Инструменты для его реализации встроены в программный продукт Microsoft Excel, а также в системы управления базами данных на базе языка SQL. Применение метода регулярных сеток обусловлено его простотой и доступностью, что делает данный метод удобным при первичном определении территорий потенциального риска.

С другой стороны, кластеризация медицинских пространственных данных свидетельствует о настоятельной необходимости перехода на трансграничный уровень противотуберкулезной работы, исключающий ограничения, задаваемые установленными административно-территориальными границами районов, муниципальных округов и т.д. При этом дополнительные характеристики кластера с повышенными показателями заболеваемости должны способствовать разработке уникальных и эффективных программ профилактических и противотуберкулезных мероприятий на территории конкретного кластера, что может и должно стать важнейшим приоритетом противотуберкулезной службы в ближайшие годы.

Метод ядерной оценки плотности (KDE) может служить дополнительным инструментом при проведении эпидемиологического исследования территории, поскольку наилучшим образом отражает интенсивность протекающего процесса. С его помощью возможно зафиксировать как негативные, так и позитивные тенденции. Например, выбрать территории, на которых напряженность эпидемической обстановки снижалась, изучить практику борьбы с заболеванием на данной территории и применить данные методы на других, более пораженных территориях.

При определении территорий повышенного риска распространения заболеваний важную роль будет играть расчет вероятности возникновения новых случаев заболевания. Ее расчет, а также пространственная автокорреляция туберкулеза с другими, социально значимыми заболеваниями, является на текущий момент наиболее перспективной задачей. С помощью данного инструмента станет возможно дать не только пространственную оценку развития заболевания, но и предоставить качественные сведения о корреляции туберкулеза с другими заболеваниями на данных территориях. Кроме того, дальнейшие исследования в области медико-географической оценки данных

территорий предполагают и обязательную оценку социально-демографических характеристик населения, проживающего на данных территориях, что позволит выстроить планомерную работу с населением.

## Заключение

Таким образом, использование геоинформационных технологий для визуализации и агрегирования данных, построения карты плотности, реляционных сетей и кластеризации позволяет определять территории, в которых на протяжении ряда лет регистрируется наибольшее количество случаев туберкулеза, что дает возможность выделять их в территории потенциального риска дальнейшего распространения заболевания.

Впервые визуализированные многолетние данные в двух мегаполисах России позволяют по-иному взглянуть на проблему распределения зарегистрированных случаев туберкулеза при разных масштабах: от города в целом до отдельного дома, что в свою очередь дает новое знание для выработки доказательных противоэпидемических и профилактических мероприятий.

В работе приведены методы определения подобных территорий на примере распространения туберкулеза в мегаполисах. Однако эти же методы применимы в любых территориально-административных образованиях и при других социально значимых инфекциях, включая ВИЧ-инфекцию и гепатиты, инфекциях, распространяющихся воздушно-капельным путем, и других заболеваниях.

Все это свидетельствует о создании объективного механизма визуализации, позволяющего перестроить и усовершенствовать противотуберкулезную и профилактическую работу в территориях, а в перспективе пересмотреть и принципы диспансеризации и профилактических осмотров населения, в зависимости от напряженности эпидпроцесса в конкретном кластере, сделать работу противотуберкулезной службы адресной и более эффективной как с медицинской, так и с экономической точки зрения.

Авторский коллектив выражает глубокую благодарность студентам кафедры картографии и геоинформатики ФГБОУ ВО «СПбГУ» В.В. Ивановой, Н.С. Полицинскому и А.Е. Пикулевой за активное участие в подготовке материалов для исследования.

## Список литературы

1. Васильева И.А., Стерликов С.А., Тестов В.В., Михайлова Ю.В., Обухова О.В., Пономарев С.Б., Кудлай Д.А. Отраслевые и экономические показатели противотуберкулезной работы в 2020–2021 гг. Статистические материалы. М.: ФГБУ «НМИЦ ФПИ» Минздрава России, 2022: 64. [Vasil'eva I.A., Sterlikov S.A., Testov V.V., Mihajlova Ju.V., Obuhova O.V., Ponomarjev S.B., Kudlaj D.A. Otrasleyve i jekonomicheskie pokazateli protivotuberkuljoznoj raboty v 2020–2021 gg. Statisticheskie materialy. Moscow: FGBU «NMIC FPI» Minzdrava Rossii, 2022: 64 (In Russ.)]. ISBN 978-5-6048392-2-5.
2. Галкин В.Б., Ариэль Б.М., Чужов А.Л. Сравнительная оценка динамики заболеваемости легочным и внелегочным туберкулезом в Санкт-Петербурге за полвека наблюдения. Медицинский альянс 2020; 8 (3): 6–14. [Galkin V., Ariel' B., Chuzhov A. Comparative assessment of the dynamics of the incidence of pulmonary and extrapulmonary tuberculosis in St. Petersburg for half a century of observation. Medicinskij al'yans 2020; 8 (3): 6–14 (In Russ.)]. doi: 10.36422/23076348-2020-8-3-6-14.
3. The END TB strategy: WHO/HTM/TB/2015.19 (available at: [https://www.who.int/tb/strategy/End\\_TB\\_Strategy.pdf](https://www.who.int/tb/strategy/End_TB_Strategy.pdf)).
4. Галкин В.Б., Яблонский П.К., Пантелеев А.М., Туркина Е.Н., Бармина Н.А., Михайловская Т.Л., Карпов А.В., Беляева Е.Н., Тоина С.В. Перспективы ликвидации туберкулеза в Северо-Западном федеральном округе России. Медицинский альянс 2022; 10 (1): 16–26. [Galkin V., Yablonskiy P., Panteleev A., Turkina E., Barmina N., Mikhailovskaya T., Karpov A., Belyaeva E., Toinova S. Prospects for the elimination of tuberculosis in the Northwestern Federal District of Russia. Medicinskij al'yans 2022; 10 (1): 16–26 (In Russ.)]. doi: 10.36422/23076348-2022-10-1-16-26.
5. Коровка В.Г., Галкин В.Б., Паниди Е.А., Кузнецов И.С., Бельтюков М.В., Соколович Е.Г., Пантелеева О.В., Воронов Д.В., Козлов В.В., Федоров С.В., Яблонский П.К. Возможности геоинформационных технологий для улучшения качества мониторинга очагов социально значимых инфекций. Профилактическая медицина 2021; 24 (10): 7–13. [Korovka V.G., Galkin V.B., Panidi E.A., Kuznetsov I.S., Belyukov M.V., Sokolovich E.G., Panteleeva O.V., Voronov D.V., Kozlov V.V., Fedorov S.V., Yablonskiy P.K. Potential of geoinformation technologies to improve the monitoring of socially significant infections outbreaks. The Russian Journal of Preventive Medicine 2021; 24 (10): 7–13 (In Russ.)]. doi: 10.17116/profmed2021241017.
6. Ефимов Е.И., Вьюшков М.В., Побединский Г.Г., Сарсков С.А. Современное состояние и направления развития ГИС «Эпидемиологический атлас России». ИнтерКарто. ИнтерГИС: Материалы Междунар. конф. М.: Географический факультет МГУ 2022; 28 (2): 276–297. [Efimov E.I., Vyushkov M.V., Pobedinskiy G.G., Sarskov S.A. The current state and directions of development of GIS «Epidemiological Atlas of Russia» InterCarto. InterGIS. GI support of sustainable development of territories: Proceedings of the International conference. Moscow: MSU, Faculty of Geography 2022; 28 (2): 276–297 (In Russ.)]. doi: 10.35595/2414-9179-2022-2-28-276-297.
7. Куликова И.Б., Паролина Л.Е., Тестов В.В., Стерликов С.А., Бурыхин В.С. Мониторинг эпидемической ситуации как основа повышения качества оказания медицинской помощи при социально значимых инфекционных заболеваниях в Российской Федерации. Туберкулез и болезни легких 2022; 100 (12): 6–14. [Kulikova I.B., Parolina L.E., Testov V.V., Sterlikov S.A., Burykhin V.S. Epidemic situation monitoring as a basis for improving the quality of medical care for socially important infectious diseases in the Russian Federation. Tuberculosis and Lung Diseases 2022; 100 (12): 6–14 (In Russ.)]. doi: 10.21292/2075-1230-2022-100-12-6-14.
8. Nazia N., Butt Z.A., Bedard M.L., Tang W.-C., Sehar H., Law J. Methods Used in the Spatial and Spatiotemporal Analysis of COVID-19 Epidemiology: A Systematic Review. Int. J. Environ. Res. Public Health 2022; 19: 8267. doi: 10.3390/ijerph19148267.

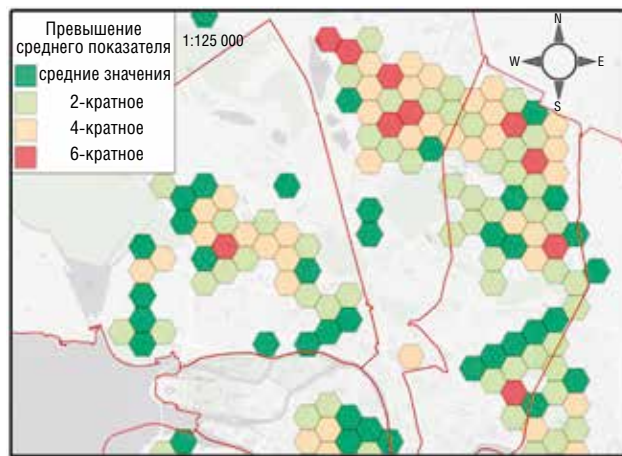
9. Alene K.A., Python A., Weiss D.J., Elagali A., Wagaw Z.A., Kumsa A., Gething P.W., Clements A.C.A. Mapping tuberculosis prevalence in Ethiopia using geospatial meta-analysis. *Int. J. Epidemiol.* 2023 Aug 2; 52 (4): 1124–1136. doi: 10.1093/ije/dyad052.
10. Wong E.B., Olivier S., Gunda R. et al. Convergence of infectious and non-communicable disease epidemics in rural South Africa: a cross-sectional, population-based multimorbidity study. *The Lancet Global Health* 2021; 9 (7): e967-e976. doi: 10.1016/S2214-109X(21)00176-5.
11. Kuznetsov I., Panidi E., Korovka V., Galkin V., Voronov D. Web-based representation and management of infectious disease data on a city scale, case study of St. Petersburg, Russia. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, XLIV-3/W1-2020: 87–91. doi: 10.5194/isprs-archives-XLIV-3-W1-2020-87-2020.
12. Kuznetsov I., Panidi E., Korovka V., Yakovlenko A. Application of the Gis-Based 3D Modeling of Multiflat Buildings to Assess the Prevalence of Tuberculosis on a City Scale. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.* XLVI-4/W3-2021: 171–175. doi: 10.5194/isprs-archives-XLVI-4-W3-2021-171-2022.
13. Приказ Минздрава России от 21.03.2003 № 109 «О совершенствовании противотуберкулезных мероприятий в Российской Федерации». [Приказ Минздрава России от 21.03.2003 № 109 «О совершенствовании противотуберкулезных мероприятий в Российской Федерации» (In Russ.)]. <http://docs.cntd.ru/document/901868614>.
14. Birch C.P.D., Oom S.P., Beecham J.A. Rectangular and hexagonal grids used for observation, experiment, and simulation in ecology. *Ecological Modelling* 2007; 206 (3–4): 347–359. doi: 10.1016/j.ecolmodel.2007.03.041.
15. Darques R., Trottier J., Gaudin R., Ait-Mouheb N. Clustering and mapping the first COVID-19 outbreak in France. *BMC Public Health* 2022; 22: 1279. doi: 10.1186/s12889-022-13537-7.

Поступила в редакцию: 13.10.2023 г.

**Приложение 1. Фрагменты гексагональных сеток для территорий повышенного риска развития туберкулеза в Санкт-Петербурге. Длина ребра 3700 (а), 1400 (б), 530 (в) метров**



а

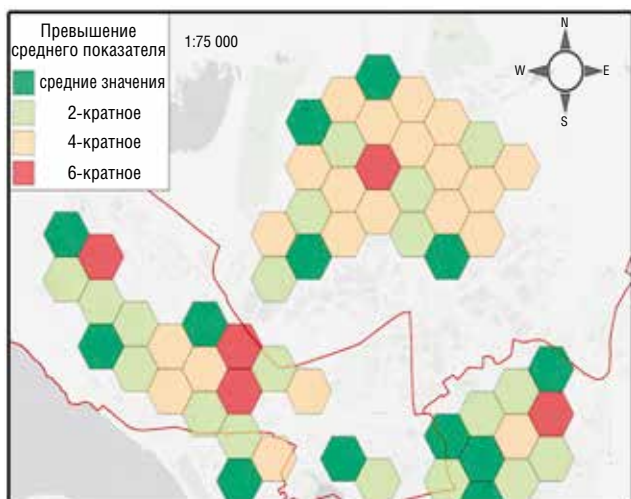


б



в

**Приложение 2. Фрагменты гексагональных сеток для территорий повышенного риска развития туберкулеза в Екатеринбурге. Длина ребра 1400 (а), 530 (б) метров**

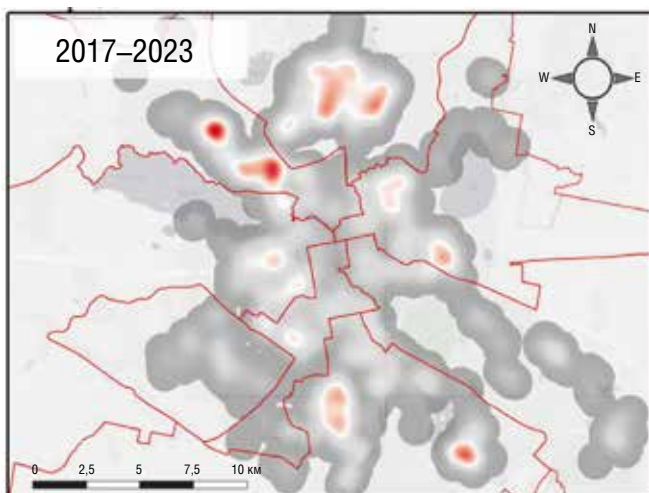


а

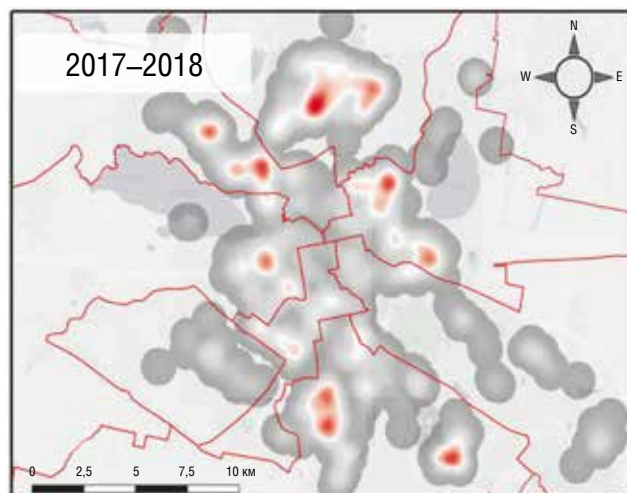


б

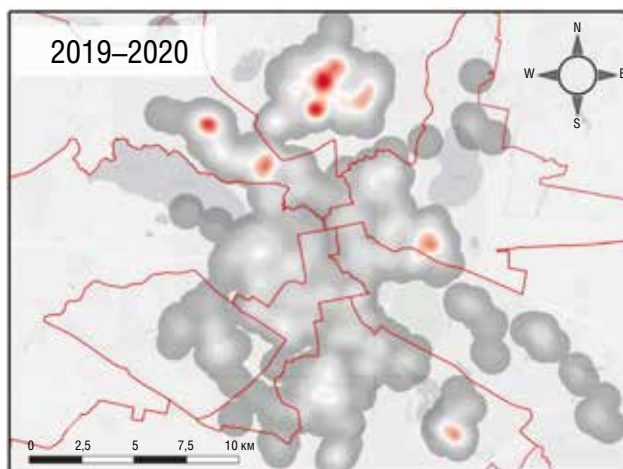
**Приложение 3. Результат алгоритма KDE для данных за различные временные промежутки**



а

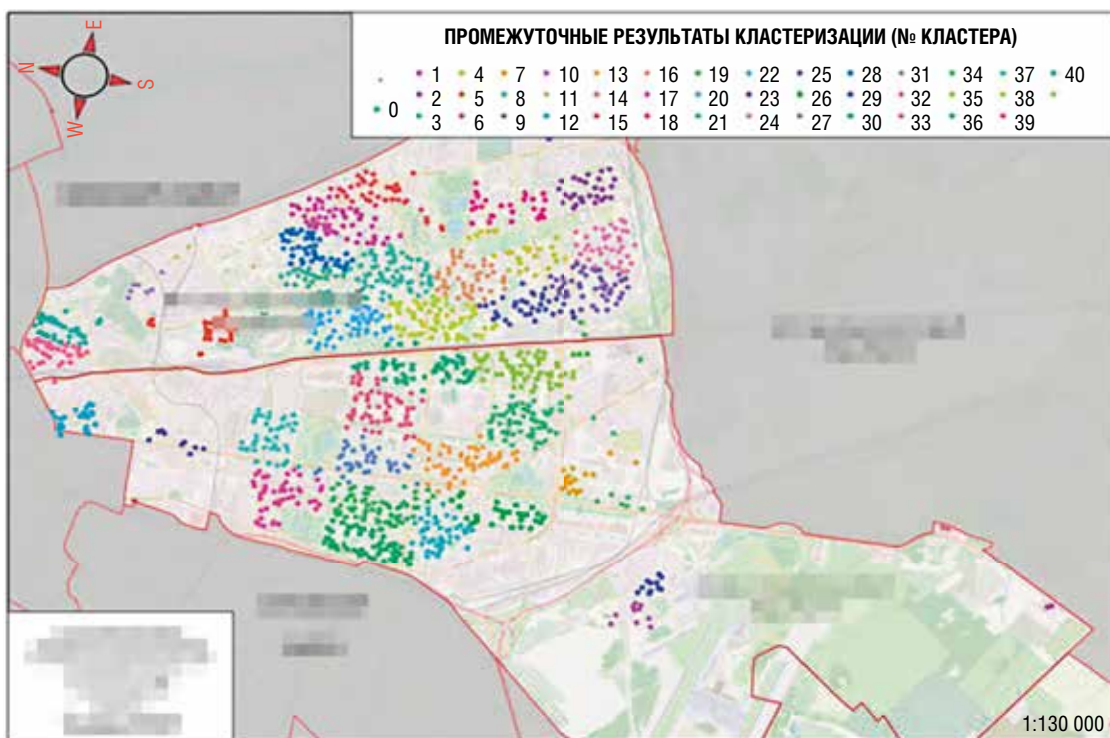


б

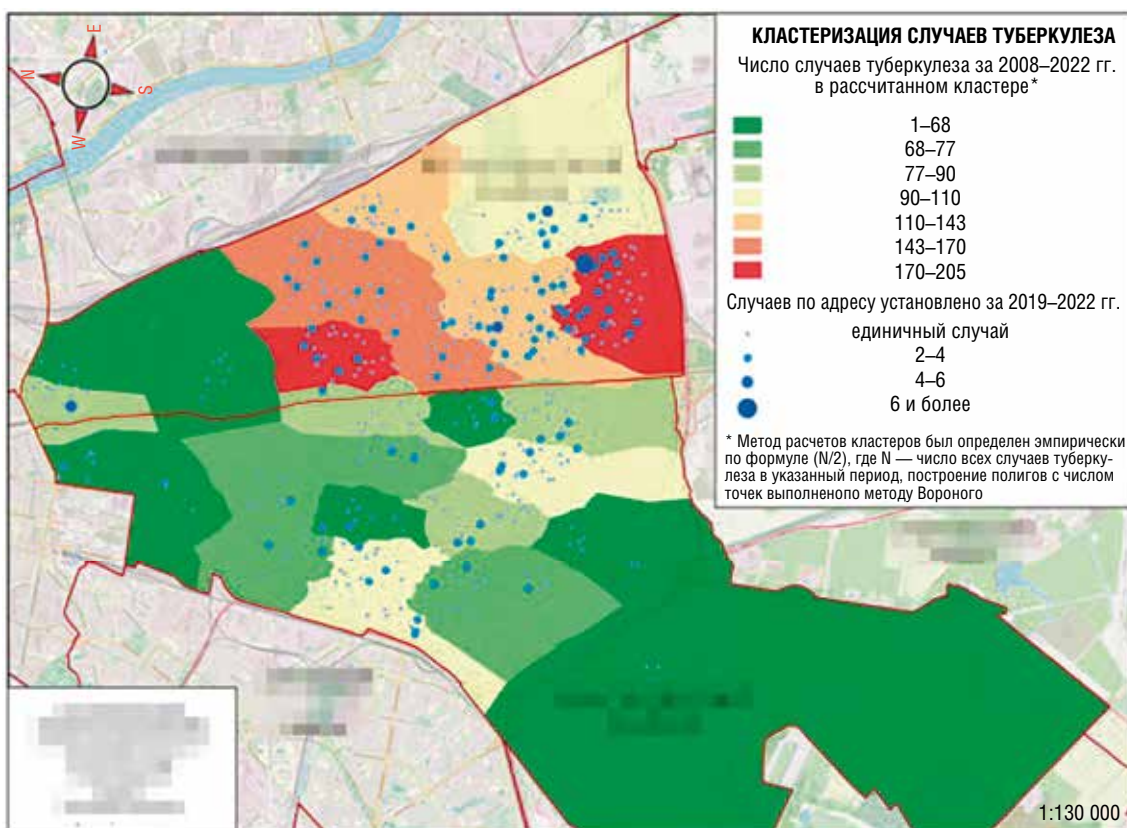


в

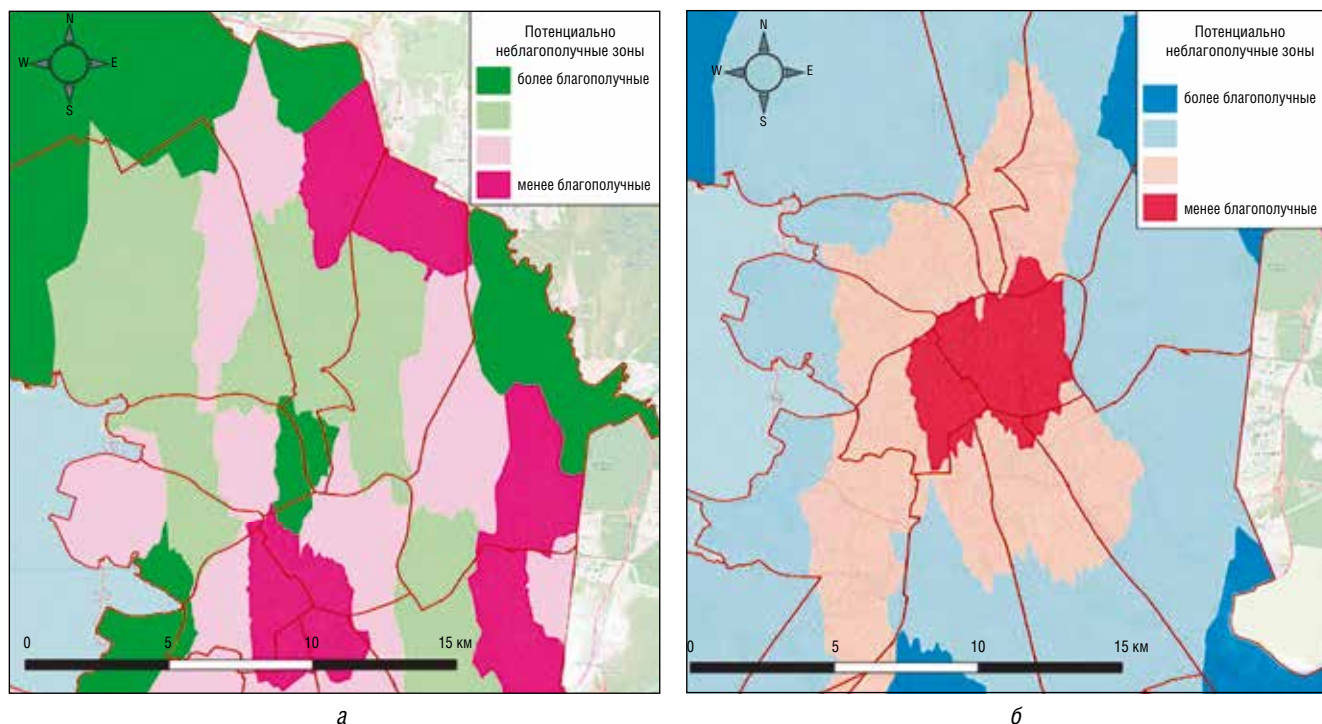
**Приложение 4. Результаты кластеризации случаев туберкулеза**



**Приложение 5. Потенциально опасные территории распространения туберкулеза и эндемичные очаги на территории двух районов Санкт-Петербурга**



## Приложение 6. Фрагменты кластерного анализа случаев туберкулеза по местам проживания пациентов (а) и по местам работы (б) в Санкт-Петербурге



### Сведения об авторах:

*Куликова Инна Борисовна* — директор Департамента организации экстренной медицинской помощи и управления рисками здоровью Министерства здравоохранения Российской Федерации; 127994, Москва, Рахмановский пер., д. 3; e-mail: info@minzdrav.gov.ru; ORCID 0000-0002-3547-5474;

*Кузнецов Илья Сергеевич* — аспирант Санкт-Петербургского государственного университета; 199034, Санкт-Петербург, Университетская наб., д. 7/9; лаборант-исследователь Санкт-Петербургского научно-исследовательского института фтизиопульмонологии; 191036, Санкт-Петербург, Лиговский пр., д. 2–4; e-mail: ilya.kuznetsov.ilya@gmail.com; ORCID 0000-0002-8820-1637;

*Коровка Владислав Григорьевич* — врач по коммунальной гигиене Санкт-Петербургского научно-исследовательского института фтизиопульмонологии; 191036, Санкт-Петербург, Лиговский пр., д. 2–4; e-mail: vl.korovka@gmail.com; ORCID 0000-0001-6782-3915;

*Бельтюков Михаил Витальевич* — кандидат технических наук, доцент, старший научный сотрудник Санкт-Петербургского научно-исследовательского института фтизиопульмонологии; 191036, Санкт-Петербург, Лиговский пр., д. 2–4; доцент кафедры биотехнических систем Санкт-Петербургского государственного электротехнического университета; 197376, Санкт-Петербург, ул. Профессора Попова, д. 5; e-mail: mikebat@mail.ru; ORCID 0000-0002-3364-561X;

*Галкин Владимир Борисович* — кандидат медицинских наук, ведущий научный сотрудник Санкт-Петербургского научно-исследовательского института фтизиопульмонологии; 191036, Санкт-Петербург, Лиговский пр., д. 2–4; e-mail: vbgalkin@gmail.com; ORCID 0000-0003-0672-2816; SPIN-код: 9601-5362;

*Соколова Ольга Павловна* — кандидат медицинских наук, главный врач Санкт-Петербургского научно-исследовательского института фтизиопульмонологии; 191036, Санкт-Петербург, Лиговский проспект, д. 2-4; e-mail: op.sokolova@spbniif.ru; ORCID 0000-0003-0608-4826;

*Паниди Евгений Александрович* — кандидат технических наук, доцент, заведующий кафедрой картографии и геоинформатики Санкт-Петербургского государственного университета; 199034, Санкт-Петербург, Университетская наб., д. 7/9; e-mail: panidi@ya.ru, e.panidi@spbu.ru; ORCID 0000-0002-1492-4218;

*Яблонский Петр Казимирович* — доктор медицинских наук, профессор, директор Санкт-Петербургского научно-исследовательского института фтизиопульмонологии; 191036, Санкт-Петербург, Лиговский пр., д. 2–4; проректор по медицинской деятельности, заведующий кафедрой госпитальной хирургии Санкт-Петербургского государственного университета; 199034, Санкт-Петербург, Университетская наб., д. 7/9; e-mail: piotr\_yablonskii@mail.ru; ORCID 0000-0003-4385-9643.