



Munich Personal RePEc Archive

Problems for spatial dimensions at different levels of data aggregation

Timiryanova, Venera and Krasnoselskaya, Dina

Bashkir State University, Ufa State Petroleum Technological
University

15 September 2021

Online at <https://mpra.ub.uni-muenchen.de/125024/>
MPRA Paper No. 125024, posted 28 Jun 2025 11:30 UTC

Venera M. Timiryanova, Senior Researcher, Deputy Head of the Laboratory for the Study of Socio-Economic Problems of Regions, Bashkir State University
Dina K. Krasnoselskaya, Associate Professor, Department of Economic Security, Ufa State Petroleum Technological University

PROBLEMS FOR SPATIAL DIMENSIONS AT DIFFERENT LEVELS OF DATA AGGREGATION

Annotation. The paper discusses the problems associated with the formation of data necessary to assess the spatial dependence of territories at different levels of aggregation. The analysis is carried out on the basis of data from settlements, municipal districts and urban districts of the Russian Federation. The problems associated with both the formation of spatial matrices and the structure, features of the publicly available statistical information are noted.

Key words: spatial analysis, administrative division, MAUP, geographic coordinates, spatial matrix

ПРОБЛЕМЫ ДАННЫХ ДЛЯ ПРОСТРАНСТВЕННЫХ ИЗМЕРЕНИЙ НА РАЗНЫХ УРОВНЯХ АГРЕГАЦИИ

Венера Маратовна Тимирьянова, старший научный сотрудник, зам.зав. лаборатории исследования социально-экономических проблем регионов ФГБОУ ВО «Башкирский государственный университет»

Дина Хадимовна Красносельская, доцент кафедры экономической безопасности ФГБОУ ВО «Уфимский государственный нефтяной технический университет»

Аннотация. В работе рассматриваются проблемы связанные с формированием данных необходимых для оценки пространственной зависимости территорий на разных уровнях агрегации. Анализ проводится на уровне поселений, муниципальных районов и городских округов. Отмечаются проблемы связанные как с формированием пространственных матриц, так и структурой, особенностями находящейся в открытом доступе статистической информации.

Ключевые слова: пространственный анализ, административно-территориальное деление, MAUP, географические координаты, пространственная матрица

Административно-территориальное устройство - одна из важнейших на сегодняшний день составляющих организации общества в территориальном отношении, предусматривающая разделение территории на части (административно-территориальные единицы). На основе такого устройства выстраивается вся система территориального управления, предусматривающая разделение управленческих функций на макро и микро уровни.

Согласно Конституции Российской Федерации, Россия является федеративным государством и состоит из 85 равноправных субъектов Российской Федерации (иногда называемых просто «регионами»). При этом применяется два вида территориального деления. Для упорядоченного осуществления функций государственного управления это административно-территориальное устройство (ОКАТО), а для организации местного самоуправления - муниципальное устройство (ОКТМО). При этом Министерство Финансов для целей бюджетного управления периодически выпускает таблицу соответствия кодов ОКАТО кодам ОКТМО муниципальных образований и входящих в их состав населенных

пунктов и межселенных территорий по состоянию на 1 января 2021 года. Начиная с этого пункта возникает первая проблема: соответствие делений и наименований территорий, представленных в списках. Она усложняется наличием нормативно-правовых актов субъектов Российской Федерации, определяющих наименование муниципального образования, населенного пункта. В первом случае мы сталкиваемся с различным порядком слов (например: муниципальный район "Абыйский улус (район)" согласно ОКАТО, в ОКТМО обозначен как Абыйский муниципальный район) или даже различным написанием ключевого слова (например, Муниципальное образование "Кэнтикский наслег" Верхневилуйского улуса в ОКАТО и Кентикский наслег, Верхневилуйский муниципальный район в ОКТМО). Во втором случае мы сталкиваемся с тем, что даже внутри одного документа встречаются два варианта написания наименования. Например, в законе Республики Саха (Якутия) от 30 ноября 2004 года N 173-З N 353-III «Об установлении границ и о наделении статусом городского и сельского поселений муниципальных образований Республики Саха (Якутия)» присутствует два варианта: Бёкчегинский и Бекчегинский. Отдельно отметим, что Федеральная служба государственной статистики ведет свой справочник территориальных единиц разработки сведений о населении (ТЕРСОН) для фиксации территориальных, административных, муниципальных изменений в регионах Российской Федерации и ряда других целей. В нем в частности обозначенные выше территории представляются как Кэнтикский и Бекчегинский наслеги. Казалось бы это несущественные нюансы, но они имеют важное последствие: осложнение обработки данных. Одно дело когда речь идет о 85 субъектах Российской Федерации и совсем другое когда речь о почти 19 тыс. муниципальных образований или даже 155 тыс. поселений. Такие различия приходится отслеживать вручную, т.к. при объединении данных, полученных из разных источников возникают ошибки, вызванные несовпадением символов в наименованиях.

Другой проблемой, возникающей в ходе работы со статистическими массивами данных является присутствие территорий с одинаковыми наименованиями в рамках одной территориальной единицы более высокого уровня. Такие особенности отмечаются и в таблице соответствия кодов ОКАТО кодам ОКТМО муниципальных образований и входящих в их состав населенных пунктов и межселенных территорий по состоянию на 1 января 2021 года включением пояснений типа: д Поповка Пояснение: Ведерковского сельсовета, 8 км от центра сельсовета - п Бушуиха; д Поповка Пояснение: Ведерковского сельсовета, 9 км от центра сельсовета - п Бушуиха.

Соответственно при объединении двух баз вы также должны отследить, что сведения перенесены верны по каждой из деревень. При этом следует отметить, что согласно реестру ОКТМО они обе находятся в Комьянском сельском поселении.

Обозначенные особенности осложняют не только процесс объединения статистических данных, полученных из разных источников, но и последующий процесс определения координат объектов необходимый для пространственных измерений. И мы плавно переходим к следующей проблеме: на текущий момент в открытом доступе нет хорошей базы с точными координатами всех территориальных единиц Российской Федерации, позволяющей автоматизировать процесс объединения пространственных данных со статистическими. Наиболее известными источниками получения координат являются: Росреестр, сервисы Яндексa и Google, Openstreetmap. У каждого источника есть свои недостатки. Наибольшее доверие вызывают карты Росреестра, так именно данная организация осуществляют кадастровый учет земель. Следует отметить, что на сайте Росреестра отражаются данные с координатами не только на текущий момент, но и за предыдущие периоды, что удобно при проведении исследований по разным периодам. Однако формат предоставления ими данных

не позволяет автоматизировать процесс слияния их с данными из других источников. В частности обозначенная выше д. Поповка в них никак не идентифицирована (табл.1), более того, учитывая отсутствие в описании принадлежности к тому или иному сельсовету, автоматически возникает проблема с появлением третьей д.Поповка в Грязовецком районе.

Таблица 1. Формат предоставления сведений из государственного каталога географических названий Росреестра

Наименование объекта	Тип объекта	Субъект Район РФ	Район	Широта	Долгота
Поповка	деревня	Вологодская область	Грязовецкий район	58,86667	40,9
Поповка	деревня	Вологодская область	Грязовецкий район	59,01667	40,4
Поповка	деревня	Вологодская область	Грязовецкий район	59,06667	40,48333

Сервисы Яндексa, Google, Openstreetmap для автоматизированного предоставления координат используют API. Но согласно положений использования API Яндексa в рамках открытой лицензии есть ограничения на количество запросов и право сохранения данных. Последнее ограничение фактически не позволяет их использовать для последующих расчетов. При этом минимальная стоимость коммерческой лицензии, допускающей сохранение данных составляет 620 тыс. руб. У сервиса Google так же есть ограничения в рамках бесплатного лимита, но они касаются только количества отправляемых запросов. Проблемой API сервиса Google является ограниченная возможность оформления запросов на русском. Основным недостатком Openstreetmap является строгое соблюдения порядка символов в наименованиях. Если в запросе вместо «е» будет отражено «ё», то данные не будут выгружены. В то же время именно для автоматизированной работы с данными этот сервис является достаточно удобным. В частности используя библиотеки R «dplyr» и «tidygeocoder» можно автоматически сформировать столбцы широты и долготы у уже существующего набора статистических данных, при соблюдении определенного порядка запроса (табл. 2). Однако если не все уровни территориального деления будут представлены, при совпадении символов в наименованиях данные будут найдены. Полнота данных определяет точность нахождения именно требуемого объекта. Здесь фактически работает правило списка, в случае неполных данных будут предоставлены данные первого найденного объекта соответствующего критериям. Так например отсутствие слова «городское»/«сельское» в случае Фировского городского поселения/Фировского сельского поселения приведет к появлению у населенных пунктов одинаковых координат принадлежащих одному из вариантов.

Таблица 2. Формат предоставления сведений Openstreetmap

Наименование	Широта	Долгота
Поповка, Сидоровское сельское поселение, Грязовецкий район, Вологодская область, Россия	58,8664461	40,9039922
Поповка, Комьянское сельское поселение, Грязовецкий район, Вологодская область, Россия	59,0185890	40,4035760
Поповка, Комьянское сельское поселение, Грязовецкий район, Вологодская область, Россия	59,0661130	40,4910240

Следующая проблема которая выявляется при сопоставлении данных таблиц 1 и 2, связана с тем, что Росреестр, сервисы Яндексa и Google, Openstreetmap показывают близкие

но разные координаты. И если речь идет о деревнях, то различие не такое большое. Когда же мы пытаемся получить данные по более крупным территориальным единицам (например, сельсоветы и районы), то различие может стать существенным для пространственных измерений. Основная причина кроется в детализации границ объектов и способе расчета точки. Точка может являться центром тяжести (геометрическим центром) полигона или внутренней точкой полигона со значениями координат, полученными, например, осреднением координат всех точек, образующих полигон. В случае невыпуклого полигона или составного полигона, включающего внутренние полигоны "острова", анклавов, их положение может не совпадать. Сильнее всего это проявляется на территории Дагестана, где есть прикутанные селения (земли отгонного животноводства). Например, если посмотреть на Ботлихский район, то он имеет несколько таких земель расположенных фактически в других районах республики: Алибекотар, Бутуш, Бюру-кутан, Каплановка, Комар-хутор, Полтавка, Старое Каратюбе, Старый Артезиан (рис. 1).

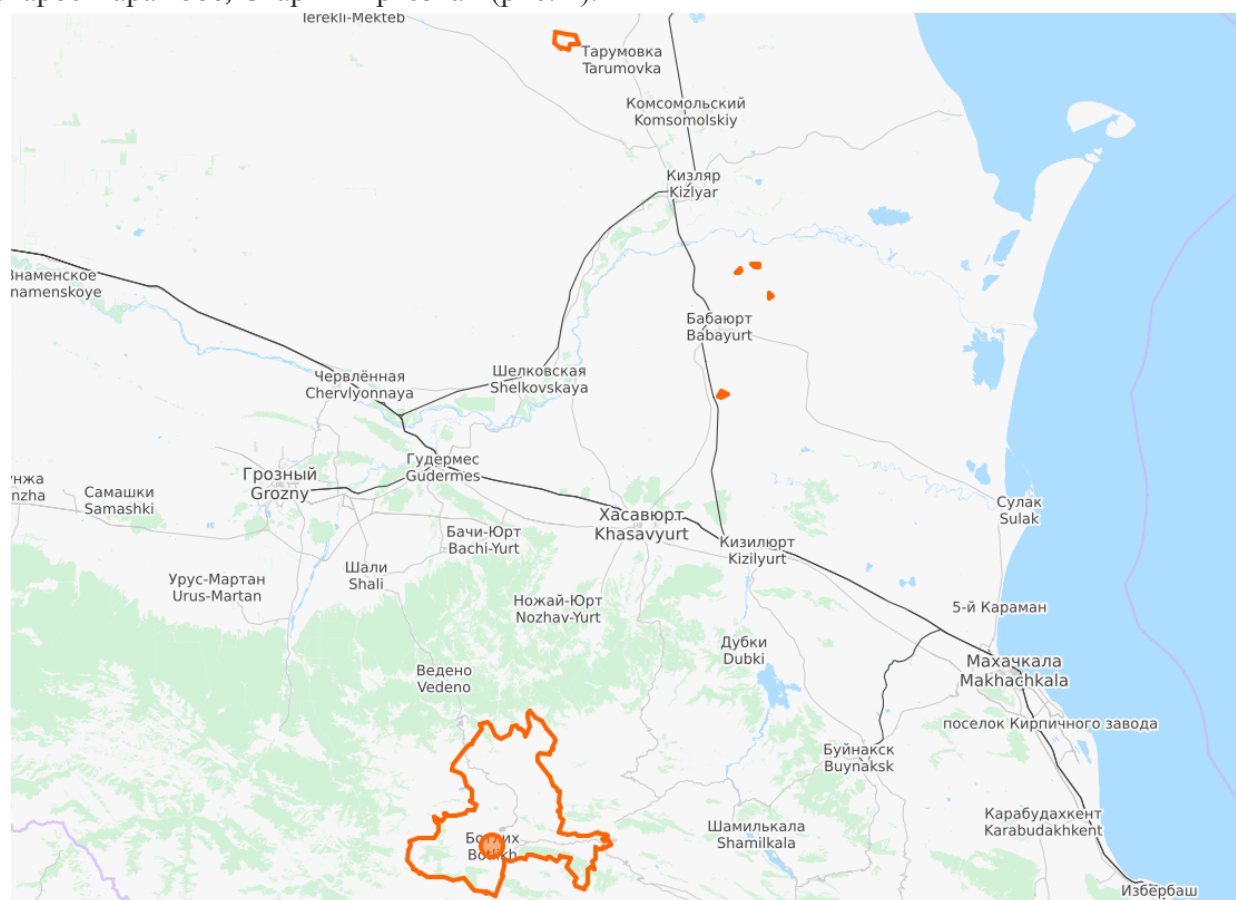


Рис. 1. Ботлихский район на openstreet map

При этом следует учитывать, что данные Openstreetmap формируются сообществом энтузиастов картографов, которые добавляют и поддерживают данные об объектах по всему миру на добровольной основе. В связи с этим многие объекты имеют комментарий «Граница нарисована из общей схемы. Требуется уточнения из Генплана поселения». Так как значение точек для полигона в большинстве случаев рассчитывается автоматически, результат может быть иным для карт с различной детализацией границ. Интересным является и то, что Росреестр в отличие от автоматизированных систем центром района определяет его административный центр. Что в случае отдельных экономических тем исследования является более предпочтительным для анализа, так как в расчетах учитывается фактически центр социально-экономической активности территории (табл. 3).

Отдельно отметим и то, что в случае охвата всей территории РФ на уровне

муниципальных образований и более низких уровней выявляется другая проблема, связанная с особенностями географического ландшафта и расселения населения. Она напрямую влияет на формируемые матрицы расстояний. Фактически становится нецелесообразным создание матрицы расстояний по центрам полигонов по двум причинам. Первая связана с тем, что даже при расчете с учетом кривизны земного шара не учитывается особенность местности: горы, болота, реки. Например из административного центра Каргасокского района в административный центр Александровского района Томской области, являющегося соседом первого порядка, вместо 222 км по прямой фактически потребуются проехать 2800 км. Вторая причина связана с тем, что расстояния между объектами и плотность заселения на юге и севере страны разная. В ряде случаев охватывая радиус в 200 км вы получите одного соседа, в то время как в другой части страны включите в соседи более 100 населенных пунктов. В первом случае соседний населенный пункт очень значим, во втором случае деревня на расстоянии в 200 км от другой деревни скорее всего не имеет существенных экономических связей для их анализа.

Таблица 3. Данные географических координат Ботлихского района

Наименование	Openstreetmap		Яндекс		Росреестр	
	Широта	Долгота	Широта	Долгота	Широта	Долгота
Село Ботлих	42.6688	46.2099	42.667179	46.211045	42,6651	46,2174
Ботлихский район	42.7490	46.2524	42.761464	46.264980	42,6627	46,2181

Все эти нюансы отмеченные выше кажутся не такими значимыми когда речь в целом о небольшом массиве данных: 85 субъектов РФ или 2340 муниципальных районов и городов. Переход на уровень поселений и в последующем на уровень конкретных адресов потребует автоматизации процессов сбора и обработки данных. В целом пространственный анализ данных относится к категории задач Data Mining, а в совокупности с категорией «время» автоматически переводит его в категорию «Big Data», т.к. для анализа требуется применение специальных инструментов обработки данных [1]. Нельзя не согласиться с тем, что есть и другие факторы указывающие на то, что «существующие способы социально-экономического районирования не готовы к обработке больших объёмов постоянно обновляемой эмпирической информации» [2]. И если обозначить первый этап пространственного анализа, то для начала каких либо пространственных вычислений нужно иметь качественную пространственную матрицу, которую можно легко объединить со статистическим массивом, полученным из другого источника. И если на уровне муниципальных образований все более менее понятно: есть коды ОКТМО (к слову которые не упоминаются в данных Росреестра), то при переходе на адреса возникают естественные проблемы с наличием одинаковых улиц в рамках одного города (например, возникшее в результате исторического поглощения одних территорий другими), отслеживание которых вручную видится весьма непродуктивным.

Благодарность: Статья подготовлена в соответствии с государственным заданием Министерства науки и высшего образования Российской Федерации (код научной темы FZWU-2020-0027)

Литература:

1. Гараева А.Р., Минниханов Р.Н., Дагаева М.В., Кильдеева С.С., Аникин И.В. Интеллектуальный анализ больших пространственно-временных данных для служб экстренного реагирования // Современные информационные технологии и ИТ-образование. 2018. Т. 14. № 3. С. 679-685.
2. Блануца В. И. Социально-экономическое районирование в эпоху больших данных. М.: Инфра-М, 2017. 194 с.