

УДК 327

DOI 10.63115/6063.2025.90.80.006

ИНТЕРВЬЮ

Современные методы пространственного анализа: Интервью с Тимирьяновой Венерой Маратовной



Тимирьянова Венера Маратовна

доктор экономических наук, доцент; главный научный сотрудник лаборатории исследования социально-экономических проблем регионов, Уфимский университет науки и технологий, Уфа, Россия

79174073127@mail.ru, <https://orcid.org/0000-0002-1004-0722>

Venera Timiryanova

Doctor of Economics, Associate Professor; Chief Researcher, Laboratory for the Study of Socio-Economic Problems of Regions, Ufa University of Science and Technology, Ufa, Russia

79174073127@mail.ru, <https://orcid.org/0000-0002-1004-0722>

ИНТЕРВЬЮЕР **НЕСМАШНЫЙ АЛЕКСАНДР ДМИТРИЕВИЧ**, НАУЧНЫЙ СОТРУДНИК ИНСТИТУТА МЕЖДУНАРОДНЫХ ИССЛЕДОВАНИЙ МГИМО МИД РОССИИ:

— Многие из ваших работ связаны с анализом пространственной закономерности в России на муниципальном и на региональном уровне. На основе тех выводов, к которым вы приходили, как бы вы обобщённо описали пространственную структуру Российской Федерации? Можем ли мы опираться на одну из существующих концепций, или любые генерализации здесь вредны?

ТИМИРЬЯНОВА ВЕНЕРА МАРАТОВНА:

— Если обобщённо описывать пространственную структуру Российской Федерации, то можно сказать, что она очень неоднородная. Это отмечается многими. Вопрос встаёт, как правило, с тем, как оценивать, насколько подходит нам эта неоднородность. На самом деле с неоднородностью в большинстве своём люди предпочитают не бороться, а корректировать её чрезмерные проявления. То есть, в принципе, неоднородность — это нормальное состояние экономических систем. И здесь просто нужно отслеживать, то, каким образом она сказывается на отдельных территориальных образованиях, и отслеживать, чтобы не было излишне сильной неоднородности, то есть крайних её проявлений. Например, мы наблюдаем подобную

неоднородность в реакции отдельных регионов на изменение валютных курсов, которые выражаются в отношении приобретения техники, например. То есть в каких-то регионах люди будут просто её массово скупать, а в других регионах спокойно реагировать. Где-то цены могут очень сильно повышаться на этом фоне, где-то могут не повышаться. И вот именно этот вариант неоднородности, его, конечно, необходимо регулировать. Он в большей степени связан с устойчивостью этих территориальных систем. То есть если где-то идёт, скажем так, не совсем адекватная реакция населения на изменения... Она может говорить о том, что данная система не совсем устойчива к тем всплескам, которые наблюдаются. В других системах возможно больше запас продуктов был или какая-то более развитая логистическая система, которая смогла восполнить недостатки и не было вот этих скачков и так далее.

— *А вот эти вот пространственные кластеры, которые такие паттерны поведения формируют, они для разных вопросов разные или они воспроизводятся?*

— У нас на самом деле очень разнообразная территория. И в целом можно говорить о том, что поведенческие кластеры они разные. То есть юго-западная часть России она, в принципе, ведёт себя иначе, нежели северная часть и центральная. И на самом деле тут очень много вопросов встаёт именно в пространственной организации. К сожалению, мы точно можем сказать, что нельзя объективно использовать одни и те же инструменты управления ко всем этим группам территорий, и необходимо учитывать именно региональные особенности.

— *Мы записываем это интервью в стенах Московского государственного института международных отношений, и поэтому хотелось бы задать вопрос относительно применения методов пространственного анализа уже на международных данных. Как вы считаете, насколько те исследовательские подходы, методы, которые себя зарекомендовали в региональном анализе, могут и должны применяться в международных отношениях или в сравнительной политологии? Более оправдан именно сравнительный кросс-странный подход в таких исследованиях, или мы можем в каких-то случаях брать государство за единицу анализа и пытаться проверить те гипотезы, которые работали на субнациональном уровне?*

— Ну первое тут необходимо отметить, что пространственные зависимости можно оценить только в случае, если рассматривается несколько терри-

торий. Соответственно, оценка пространственных зависимостей, когда мы рассматриваем только одну страну, она, в принципе, невозможна. Соответственно, это в любом случае кросс-секционное исследование, может быть, на панельных данных, но так или иначе мы затрагиваем разные территории. Что касается того, что инструменты региональных исследований применяются для анализа стран, например, связей стран, то, в принципе, могу сказать, что, наоборот, ряд инструментов пришли из исследований межстрановых отношений в региональные исследования. То есть многие из инструментов, которые сейчас используются в региональной экономике пришли из исследований, связанных с межстрановым сопоставлением, анализом взаимодействия стран. Те же самые межрегиональные взаимодействия они во многом похожи на межстрановые взаимодействия, и пространственные зависимости также могут оцениваться теми же инструментами.

Единственное, на межстрановом уровне более ярко проявляются факторы территории, то есть их особенности. На региональном уровне очень много общих условий, например, сложившаяся система управления. Командная, административная или какая-то другая. На уровне стран, естественно, мы видим и унитарные государства, и федерации; могут выделяться монархии, республики, и другие варианты. А на уровне региона мы фактически видим единообразие системы управления. Вот в этом контексте получается, что мы на региональном уровне просто должны учитывать чаще несколько другие наборы факторов, в особенности что касается системы управления.

На региональном уровне многие компоненты управления являются одинаковыми. Можно только попытаться оценить, насколько то или иное управление действительно эффективно, но сама сложившаяся система административно-территориального деления и территориального управления — она является единой.

— Но в международных исследованиях ещё большое внимание уделяется связям между странами, вот этим потокам, которые могут иметь как коммуникативную природу, так и транзакционную. И есть ли какие-то вот направления исследований, подходы, может быть, в региональных исследованиях, где именно связи между регионами, связи ставятся в центр анализа, и те или иные гипотезы проверяются вокруг них?

— В целом, да, есть миграционные потоки, есть товарные потоки, и те, и другие часто, скажем так, включаются в региональные исследования, в том числе и в рамках пространственного анализа. В частности, в простран-

ственной эконометрики можно оценивать расстояние между территориями не только по длине дорог или по времени перемещения, можно оценивать и по объёму товарооборота. То есть чем больше товарооборот между двумя регионами, тем эти регионы, получается, с точки зрения экономики, они ближе. Это не так часто наблюдается, но, в целом, это технически возможно. Достаточно редко, но такие исследования встречаются, в том числе и на территории Российской Федерации. Например, когда мы рассматриваем «ядра притяжения» и анализируем методами пространственного анализа связи территорий. Естественно, на межстрановом уровне такое тоже достаточно часто встречается, то есть анализ потоков миграционных, товарных, и т.д. Экспорт-импорт в данном случае очень активно используется.

Да, на уровне регионов очень сложно уловить экспорт-импорт. В частности, одно из моих исследований делало такую попытку — выделить, скажем так, между федеральными округами объём перемещаемых товаров, и он, в частности, показал, завышенную роль Москвы в этих потоках. Это, кстати, скажем так, не секрет: спросить любого водителя грузовой машины, и он Вам подтвердит, что, в принципе, в тот же самый Благовещенск, который находится на границе с Китаем, товар поступает следующим образом: он проезжает мимо Благовещенска, едет до Москвы, там перекомплектовывается и едет обратно в Благовещенск. Это Вам покажется фантастичным, но это реальность. Это связано с тем, что в Москве расположены штаб-квартиры и основные логистические центры компаний. Соответственно, вот этот вот объём товарооборота, который излишне перемещается через Москву, он занимает, наверное, процентов тридцать всего товарооборота страны.

Тем не менее, многие — и «Озон», и «Вайлдберриз» — строят большие логистические центры в регионах, потому что они понимают, что им распределительный центр содержать в Москве и везти товар через всю страну для того, чтобы потом его привезти обратно в Сибирь, чрезмерно затратно. Поэтому да, у нас начали строить распределительные центры, по крайней мере в Башкирии это видно.

— Вы затронули такой вопрос: откуда получать информацию для исследований? Можете рассказать, как Вы обычно получаете данные для своих исследований регионов, муниципальных образований Российской Федерации? Это в основном какая-то официальная статистика, государственные данные или есть другие источники сведений?

— На самом деле, в последние пять лет я наблюдаю большой прорыв в рас-

крытии данных. Хотя сейчас последние три года отмечается, что многие официальные источники закрываются, но появляется на самом деле большое количество данных, которые лежат на поверхности, и нужно просто уметь их взять и обработать.

Я здесь хотела бы отметить Ивана Бегтина¹². Это человек, который активно занимается сбором данных. Он архивирует какие-то, возможно, базы, которые закрываются. Если он узнает о том, что они закрываются, он их архивирует, сохраняет. И на самом деле он максимально погружён в наборы данных. Когда я листаю его ленту, я думаю: «Боже мой, откуда такие объёмы данных и как их вообще можно охватить». И зачастую, наборов весьма интересных. Здесь их очень много. Их самое главное — это научиться их обрабатывать. И, к сожалению, многие эти данные очень сырые. Там могут содержаться ошибки, и тут очень важно именно научиться исключать выбросы, восстанавливать возможные пропуски, и т.д.

А что касается того, откуда я беру данные, всё зависит от того, какое исследовательское направление мы разрабатываем. Например, когда мы анализировали пространственную зависимость Коронавируса, то мы просто, скажем так, делали то же самое, что Яндекс.DataLens¹³. Например, Яндекс собирал сведения со всех информационных площадок. Каждая республика/область/край вывешивала информацию о том, сколько было заражённых коронавирусом. Собирая данные Яндекс формирует различные наборы/базы. Есть открытые базы, в том числе по коронавирусу, которые мы использовали, делая соответствующие ссылки. Но по тем данным, которые он, например, в последний год перестал собирать, мы самостоятельно настроили сборщик, который обходит соответствующие страницы и ежедневно собирает сведения с публикуемой информацией о количестве заболевших.

Если же мы говорим о, например, таких элементах, которые иногда нужно для пространственного анализа знать: сколько торговых объектов находится в этом поселении или как далеко находится транспортный узел. То есть разные варианты. Скажем так, можно приобретать эти данные у Яндекса, хотя это очень дорого. Раньше была возможность бесплатно загружать через API Google Maps API, но она также закрылась. Сейчас мы используем в основном OpenStreetMap, хотя это данные, скажем так, они часто имеют ошибки, они не обновляются так хорошо, как данные Яндекса. Но так или иначе транспортные узлы, например, не могут быстро пе-

12 Иван Бегтин — соучредитель АНО «Информационная культура», основатель платформы поиска данных Dateno.io, один из ведущих экспертов России по открытым данным и открытому государству (<https://t.me/begtin>) — прим. редакции.

13 Яндекс.DataLens — инструмент для анализа и визуализации данных, с использованием которого Яндекс распространял агрегированную информацию по пандемии Ковид-19 — прим. редакции.

ремещаться и открываться в новых местах. Соответственно, мы используем эти наборы данных, в том числе по количеству зданий, по количеству объектов инфраструктуры. Мы можем эти данные собирать. Медицинские данные мы чаще всего всё-таки берём непосредственно у системы здравоохранения. Там очень важные нюансы, связанные с деперсонализацией. Существует отдельный протокол по тому, какие данные могут быть проанализированы, и важно понимать, что здесь даже сами научные журналы требуют от нас, чтобы мы предоставили сведения о том, что данные собраны с согласия пациентов. Есть данные по погоде — они в принципе открыты.

Данные социальных сетей в целом можно частично тоже использовать, хотя API ВКонтакте запрещает хранить и обрабатывать эти данные. Мы взаимодействуем с Университетским консорциумом исследователей больших данных на базе Томского государственного университета, который помогает получить доступ и работать с такими данными. Все данные социальных сетей с определённой погрешностью (сразу скажу) тоже имеют географическую привязку. Все пользователи так или иначе указывают местоположение, либо место жительства. Мы понимаем, что такая привязка может быть неправильной, недостоверной. Также хотелось бы отметить, что есть, например, такие наборы данных, как госзакупки. Есть данные HeadHunter'a, там тоже всё структурировано, есть географическая привязка.

Единственная проблема, которая есть у всех этих наборов, — это то, что их очень сложно объединять. То есть мэтчинг этих данных он очень проблематичен. Проблематичен из-за того, что каждый называет территорию так, как он её видит. Есть полное название, есть сокращённое. А при этом ещё нужно понимать, что многие муниципальные районы и поселения периодически меняют свои названия, и это очень сильно усложняет работу, когда речь идет о двух тысячах четырехстах муниципальных образованиях или о девятнадцати тысяч поселений. И то, что сегодня, скажем, называется городом на следующий год становится муниципальным округом, и просто техническими методами это очень сложно объединять. Это большая и трудоёмкая проблема, но решаемая. А данных, которые можно собрать, на самом деле очень много, не обязательно из официальной статистики.

— *Продолжая разговор о данных, что, я думаю, для всех количественников достаточно болезненная тема. Один из основателей критических теорий международных отношений, неограммшианец Роберт Кокс, в одной из своих работ сформулировал ключевой тезис этого направления о том, что «все теории создаются для кого-то и с какой-то целью»¹⁴. Насколько мы можем*

¹⁴ Cox R. (1981), Social Forces, States and World Orders: Beyond International Relations Theory, Millennium Journal of International Studies, vol. 10, no. 2, pp.126–154.

тоже самое сказать про данные? Сталкивались ли вы с предвзятостью при фиксации данных, как преднамеренной, так и нет, которая могла бы повлиять на результаты анализа?

— Я в принципе согласна с этим мнением, что данные собираются и собираются для кого-то, хотя я часто собираю данные, просто складываю их на компьютер. У меня уже терабайты данных лежат, не все даже я анализирую. А по поводу непредвзятости, тут хотела бы сказать, что да, к сожалению, с этим периодически мы сталкиваемся. И наиболее яркий пример — это, наверное, данные по коронавирусной инфекции. Отмечалось во многих публикациях, в том числе российскими авторами то, что данные были занижены, особенно в первый период. Людей, видимо, боялись испугать, и официально регистрировали меньший объём смертей именно по причине коронавируса, списывая их на другие виды и причины болезней. То есть это наиболее яркое проявление предвзятости данных за последний период, скажем так.

— Переходя от данных обратно к методам, вы можете побольше рассказать про иерархические подходы, которые вы применяли в своих исследованиях? На основе каких методов, концепций, какого рода данных, в какой структуре можно осуществлять такой анализ? Как он стыкуется с пространственным анализом? К каким выводам он Вас привёл в исследованиях российских муниципальных образований и регионов?

— С методами иерархического анализа я, наверное, познакомилась примерно в 2018 году. И здесь необходимо отметить, что на территории Российской Федерации, именно в региональных исследованиях, методы иерархического анализа практически не развиты. Я могу назвать ну максимум две-три научные школы, которые их используют именно в региональных исследованиях. В основном это инструмент, который используется социологами. К слову, у меня есть второе высшее образование в социологии, поэтому мне этот инструмент попался. Я встретила публикации, в которых я видела, что за рубежом этот инструмент применяется для региональных исследований. Идея метода состоит исходно с точки зрения социологии в том, что есть несколько классов в школе, и необходимо понять, почему в некоторых классах один и тот же ученик может получать более высокие оценки, нежели в другом. И соответственно, это анализ межгрупповых и внутригрупповых различий, фактически. По аналогии в региональной экономике делается попытка выяснить почему одни муниципалитеты, расположенные в одних регионах, показывают лучшие результаты, чем другие,

расположенные в других регионах, но имеющие близкие характеристики.

В иерархическом анализе очень важна вложенность. То есть дети находятся в каком-то классе, этот класс находится в определённой школе. Применительно к регионам: есть поселения, поселения находятся в определённом муниципальном районе, муниципальный район относится к определённому региону, регион относится к определённому федеральному округу. Вот эта чистая вложенность она очень важна для инструментов иерархического анализа.

Что касается методов, есть несколько способов анализа. Самый простой, с которого, наверное, в 2018 году началась работа — это просто разложение данных на три уровня. Он был предложен примерно семьдесят лет назад Х. Моельрингом и В. Тоблером. Затем появились публикации, в которых рассматривалось иерархическое разложение, а именно выделение вклада поселений, районов и стран в сложившийся уровень социального неравенства через разложения индекса Тейла¹⁵. Примерно в конце прошлого века появились так называемые многоуровневые иерархические модели. И на самом деле они как бы появились одновременно в нескольких, ну, скажем так, в двух интерпретациях. Из-за этого вы иногда можете встретить учебник, где будет написано «многоуровневая модель», а не «иерархическая», и там, казалось бы, коэффициент рассчитывается точно так же, но он будет называться не Intraclass Correlation Coefficient, а Variance Partition Coefficient. Вот и, соответственно, хотя вроде как инструменты одни и те же, но они развивались параллельно и имеют, скажем так, некоторые нюансы, связанные оформлением, с тем, как описывать вот эти многоуровневые модели или как обозначать коэффициенты.

Вы спросили по поводу того, каким образом они соединились с пространственным анализом. На самом деле даже Люк Анселин¹⁶ отмечал, что любые территории они имеют свойства иерархии. Это естественно, то есть мы видим что административно территориальное деление оно иерархически организовано соответственно.

— *Правильно ли я вас понимаю, что основная идея иерархического анализа заключается в том, что объединённые наличием единого верхнего уровня элементы анализа, как например несколько регионов, которые входят в один федеральный округ, более тесно между собой связаны, чем несколько*

¹⁵ Индекс Тейла — мера измерения социального неравенства, предложенный в 1967 году нидерландским экономистом Анри Тейлом [H. Theil, Economics and Information Theory, North-Holland, 1967.] — *прим. редакции.*

¹⁶ Люк Анселин — американский исследователь и разработчик программного обеспечения, один из основателей современного пространственного анализа — *прим. редакции.*

регионов, пусть и расположенных по соседству, но входящих в разные федеральные округа?

— На примере Российской Федерации было четко показано что регион оказывает влияние на результативность тех муниципальных районов, которые в него вложены. Ярким примером является Татарстан. Если мы берем муниципальные районы на границе Татарстана и сравниваем их с муниципальными районами соседних республик, той же Чувашии или Марий Эл, даже Республики Башкортостан, мы видим, что эффективность муниципальных районов, которые подчинены Татарстану выше, и это во многом определяется именно работой регионального правительства. То есть, если бы этот муниципальный район на границе между Татарстаном и Чувашией находился не в Татарстане, а в Чувашии, то скорее всего, его показатели были бы хуже.

Что касается связи пространственного анализа и иерархического анализа, на текущий момент нужно отметить, что развитие инструментов идёт настолько быстро, и оно такое разноплановое, что одновременно две научные школы пришли примерно к одному и тому же. Первая группа учёных — это те, кто развивал иерархический анализ заметили, что можно учесть соседство регионов и пространственную матрицу вставить в верхний уровень анализа. И появились так называемые иерархические модели с пространственной зависимостью.

Одновременно те, кто развивал пространственные модели, они обратили внимание на то, что группа оказывает влияние, и тоже вышли на эти инструменты. И они предложили вначале учесть пространственную зависимость, то есть включить матрицу соседства на нижнем уровне, а затем наложить матрицу групповых связей.

И на текущий момент уже существуют такие модели, которые могут одновременно учитывать и иерархию, и пространственные связи. То есть, это такая сетка получается... Это я считаю, один из наиболее правильных подходов, потому что и иерархия, и пространственное соседство оказывается очень важно. В том же примере с районом на границе Чувашии с Татарстаном важно то, что район является соседом по отношению к Татарстану. Люди из Чувашской Республики могут ездить в Татарстан на заработки и привозить эти деньги в Чувашию.

Соответственно, пространственно-иерархические модели появились. Они не так активно, может быть, используются, потому что они достаточно сложные. Но мне хотелось бы ещё добавить, что усложнение моделей идёт очень быстро, сейчас уже дополнительно можно учесть и эффекты времени. То есть одновременно можно учесть и эффекты времени, и пространства, и

иерархии. И единственная проблема в том, чтобы эти инструменты активно применялись. Дело в том, что очень сложно сразу сходу разобраться в том, каким образом вычислять эти эффекты, как оценивать. HSAR-модель — достаточно сложная, ну, хотя, в принципе, всё возможно...

— *Вот не могу отказать себе в удовольствии задать вопрос: а всё-таки, есть какие-то попытки уже интегрировать и страновой уровень, вот этот иерархический анализ? То есть, когда мы смотрим на государство как следующий уровень иерархии, там может быть даже дальше — макрорегион как следующий уровень иерархии? Или это пока вот эти модели пока применяются исключительно для субнационального анализа?*

— Нет, есть такие исследования, которые вот эти иерархические модели применили на данных Евросоюза. А Евросоюз — это всё-таки разные страны с различным устройством. То есть, в принципе, проводятся исследования, которые учитывают вот эти два уровня, NUTS¹⁷ второго и третьего уровня. Они опубликованы. Я не могу сказать, что их тоже очень много, но, наверное, штук пять я могла бы выделить таких исследований. Достаточно интересные выводы там были.

— *Переходя к смежному подходу, к которому Вы активно обращаетесь в Ваших исследованиях, анализу полицентричности. Каких данных требует данный подход? Насколько мы можем заимствовать этот аппарат измерения полицентричности и применять его уже в международных исследованиях, где вопросы там структуры мирового порядка — полицентрический, биполярный, однополярный — достаточно тоже широко дискутируются и в основном всё-таки базируются на достаточно субъективных оценках при том, что количественных метрик немного.*

— Вот смотрите, полицентричность на самом деле она начала исходно рассматриваться именно на межстрановом уровне. Если мы возьмём исследования середины прошлого века, Ципфа¹⁸ — это один из первых исследователей, который обратил внимание, что численность населения и ранги выстраиваются в определённую последовательность. И вот в этой книге, которую он опубликовал в середине прошлого века, там как раз сопоставляются некоторые европейские страны, США, Индия именно по распре-

17 Номенклатура территориальных единиц для целей статистики Евросоюза (фр. nomenclature des unités territoriales statistiques, NUTS) — прим. автора.

18 Zipf G.K. (1949), Human behavior and the principle of least effort, Cambridge, MA: Addison-Wesley Press, 573 p.

делению населения в привязке к определённым экономическим видам деятельности. То есть, не только демографические, но и экономические отдельные показатели рассматриваются. И как раз показывается вот эта зависимость. Начиная с этого момента, можно говорить о том, что исследование полицентричности начало активно проводиться с применением различных методов.

На текущий момент сложились два основных подхода к анализу полицентричности: морфологический и функциональный. Морфологический подход опирается на анализ данных о численности населения, о распределении поселений, ну городов в основном — от больших к малым, с тем, чтобы выявить, насколько велика доля больших регионов, городов, и насколько велика доля маленьких городов. Может использоваться различный показатель, в том числе тот вариант, который предложил Ципф. Используется регрессионное уравнение, вычисляется коэффициент регрессии, на основе которого делается вывод относительно порядка распределения. Может использоваться даже индекс Херфиндала-Хиршмана, который, как правило, изначально использовался для анализа монополии, степени монополизации рынка.

На самом деле, в рамках этого направления разные показатели применяются, которые анализируют размеры территориальных единиц в рамках определённой территории и делают вывод относительно того, насколько велика доля одних территорий, насколько равномерно распределено население между территориальными образованиями. И далее, как правило, ключевые направления — это понять, насколько полицентризм или моноцентризм оказывают влияние на общее развитие территории. И консенсуса в этом вопросе до сих пор нет. То есть кто-то считает, что полицентризм — это зло, кто-то что моноцентризм — это зло, кто-то говорит, что он оказывает влияние на то, что территория развивается более эффективно и так далее.

Второй подход — функциональный. Он учитывает потоки. Здесь исследуются миграционные потоки. Очень активно используются современные данные GPS-трекеров, которые анализируют перемещение между территориями. Чаще всего функциональный подход используется в анализе городских территорий и агломераций, чтобы выявить определённые агломерационные эффекты. Но в Российской Федерации на текущий момент таких исследований, по крайней мере, я не встречала, возможно, они есть. Я исследования полицентризма провожу в коллаборации с к.э.н. Красносельской Д.Х., которая формирует основные задачи. К сожалению, мы пока тоже не применяем функциональный подход, потому что у нас нет доступа к необходимым данным, именно о перемещениях, концентрациях людей,

маршрутах перемещения и так далее. Но такие исследования активно проводятся в Китае, например.

— *Яндекс.Карты пока не отдаёт данные по движению?*

— Нет, он отдаёт разные данные, но очень дорого. Вопрос всегда в цене. У Яндекса очень дорогие данные, на мой взгляд. По крайней мере, когда раньше был доступен Google API, там определённый объём запросов можно было делать бесплатно. У Яндекса такого практически нет. Есть Единый коммерческий тариф. Есть возможность бесплатного получения данных, но там разные условия, в том числе то, что все данные, полученные средствами API, должны быть отражены на карте Яндекса. Но главное запрещается сохранять или изменять данные, без чего невозможно проведение дальнейшего анализа. Мы решением этого вопроса начали заниматься, но данные Яндекса по картам пока не используем.

— *Говоря снова про данные, вы немало писали об эффектах от использования пространственных данных в бизнесе, в том числе в малом и среднем бизнесе. Расскажите, насколько, на ваш взгляд, сегодня в России использование пространственных данных и результатов пространственного анализа востребовано в бизнесе, так может быть, и в государственном, муниципальном управлении? Что такое высокочастотные данные (англ. high-frequency data)?*

— На самом деле, все данные, которые существуют, они географически структурированы, вплоть до дома или только до населённого пункта. Просто кто-то отбрасывает этот географический компонент. Соответственно, исследование и построение каких-то моделей и проведение какого-то анализа в привязке к географии делает результаты оценок более качественными и менее смещёнными. Поэтому в перспективе объём таких исследований будет только расти.

Интерес к такого рода низкоуровневым исследованиям действительно есть. Предприниматели, например, это касается фискальных данных, исследований на рынке, часто к нам приходят и просят провести именно анализ с позиции пространства, особенно торговых объектов. Безусловно, изменение цен в одной торговой точке оказывает определённое влияние на цены соседней торговой точки.

Но есть то, что ограничивает наши возможности проведения такого анализа. Как правило, мы можем сделать анализ только если данные существуют. Есть определённые проблемы: у владельцев торговых объектов,

которые к нам приходят, нет тех данных, которые принадлежат соседям — если какой-то мониторинг и осуществляется, то не на постоянной основе.

Еще одна существенная проблема связана с тем, что пока мало опубликованных исследований на таком уровне. При этом когда мы переходим от уровня годовых данных до дневных или минутных (высокочастотных) данных, возникает большое количество сложностей. Пространственное взаимодействие проявляется иначе, когда данные становятся менее агрегированными.

Если говорить про розничную торговлю, на уровне минут анализировать тренды бессмысленно. На уровне часа уже можно попытаться что-то спрогнозировать, особенно в моменты всплеска, когда начинается ажиотаж, можно понять, когда волна нарастает, когда она уже начинает спадать. Мы не только можем отследить, насколько сильно происходят эти всплески в отдельных объектах, но и то, как они при этом ещё перетекают через пространство. То есть если какой-то товар заканчивается в одном объекте, естественно, люди перебегают в другой объект и скупают его там. То есть это явление, которое имеет чёткую зависимость и в пространстве, и во времени.

При этом значение пространственной автокорреляции, сделанное на месячных данных для той же самой территории, будет выше, чем расчёт, полученный на дневных данных. Тем более будет ниже автокорреляция на данных, полученных по ежечасным замерам.

— Значит ли это, что здесь вообще пространственный анализ уже не нужен? И тот же анализ временных рядов, например, даст лучший результат?

— Анализ таких данных в кросс-секции, не даст оптимальный результат, так же как и просто анализ временных рядов. Нужен анализ одновременно и в пространстве, и во времени. Мы такой анализ проводим, а именно для того, чтобы понять, как на каждом отдельном промежутке времени себя ведёт вот это пространственная связь. И тут мы, кстати, нашли зависимости во времени! То есть пространственные зависимости в выходные дни проявляются сильнее, чем в будние дни. А есть ещё сезонные колебания определённых групп товаров, и мы выявили, что по отдельным группам товаров пространственные зависимости с учётом изменения цен могут проявляться по-разному в течение года. То есть оценки, полученные осенью, будут однозначно отличаться от тех оценок, которые получены весной. И соответственно, сопоставлять оценку, полученную там в ноябре определённого года с данными другого исследования, полученного на оценках,

сделанных весной, некорректно.

Вот это тоже такой нюанс, который раньше не выделялся, потому что в основном исследовались годовые данные. И что касается высокочастотных данных, вот если задаваться вопросом, насколько вы вообще видите перспективы использования высокочастотных данных, то я однозначно считаю, что мы придём к тому, что все модели будут строиться в режиме, близком к реальному времени, для того, чтобы максимально быстро и эффективно реагировать на те импульсы, которые происходят на рынках, в частности.

Но на текущий момент очень мало исследований, вот прямо совсем мало исследований, которые проводятся и публикуются. Мы знаем точно, что на основе высокочастотных фискальных данных Федеральная налоговая служба делает расчёты, но она до сих пор их не применяет активно, потому что есть очень много нюансов. Центральный банк Российской Федерации также проводил исследования на этих данных для того, чтобы понять, можно ли их на текущий момент использовать для прогнозирования потребительского поведения.

И я могу сказать, что такие же исследования проводятся и за рубежом. То есть США и Европейский Союз, они используют эти данные. То есть у них тоже данные собираются фискальными операторами и анализируются. При этом у них, помимо вот этих фискальных данных, была намного раньше создана инфраструктура, в частности, в Америке, по сбору данных от обычных покупателей. Я не знаю, когда-нибудь вы видели, например, в России такое приложение, как «Едадил», где вы фотографируете и загружаете чек, и он выводит информацию о том, сколько вы товаров купили. В США существует похожая система сбора данных с обычных потребителей. Потребителям предлагают инструменты для планирования их бюджетов, и люди действительно выходят из магазинов, просто фотографируют свой чек. Он соединяет эту информацию для того, чтобы в приложении потом делать какие-то рекомендации пользователю этого приложения. Но все понимают, что эти данные одновременно уходят в базу данных о покупках людей, и, соответственно, собирается информация, в том числе и о том, что приобретается, по каким ценам и так далее. То есть у них есть такие наборы данных.

У нас такая инфраструктура развернулась полноценно только с запуском операторов фискальных данных. И на сегодняшний день эти данные предоставляются на коммерческой основе. Мы, к слову, — единственный университет, которому удалось заключить договор с оператором фискальных данных, это я бы сказала просто случайность и удача. Когда только открывались операторы фискальных данных, я уже знала о том, что такие исследования проводятся в США. Это ещё было в 2018 году, и я обратилась

к одному из операторов фискальных данных, и он сказал: «Ну давай посмотрим, что ты сделаешь»...

И на текущий момент мы по-прежнему с ними взаимодействуем, и они, кстати, тоже нам ставят, скажем так, некоторые амбициозные задачи, на которые мы иногда говорим, что, к сожалению, этого сейчас мы решить не можем. Часто это действительно глобальные задачи, для решения которых пока нет инструментов. У них, естественно, есть свои собственные команды. Мы, скажем так, наблюдаем за ними, они наблюдают за нами. В рамках договора о сотрудничестве они нам для исследовательских целей передавали обезличенные данные, есть доступ к платформе Продажи.рф. При этом с практической точки зрения есть много очень вопросов к фискальным данным. И за рубежом, и в России эти данные исследуются, но очень мало открытых публикаций. В этих публикациях, во всех, во всех абсолютно, отмечается, что это очень сложные данные с позиции их обработки и анализа. Как пример, вы можете получить чек с полной детализацией, если вы приобретаете товары где-то в хорошем торговом магазине. Если вы покупаете на рынке, в чеке может быть забито всё, что угодно. Индивидуальный предприниматель, например, вообще не ограничен в формировании чеков. Он может просто писать «овощи», без детализации. Когда мы анализировали, именно черновые данные, необработанные, мы видели много проблем в них. Например, мы запросили сведения о хлебе, как о наиболее социально значимом товаре. Под наименованием очень похожим на хлеб могла быть водка. То есть, образно говоря, в строке просто «Хлеб-в» было написано, и число «сорок». Всё. И мы просто понимаем, что это водка. То есть, скажем так, что забивают люди в кассовые аппараты — это на их личное усмотрение. Поэтому это очень сложный набор данных. Несколько лет назад государство в том числе для решения этой проблемы ввело систему цифровой маркировки и прослеживания товаров. Сейчас для молока, например, она введена: на каждую единицу товара выдаётся QR-код. Он стоит одну копейку, но на самом деле копейки — это рынок целого миллиарда в день, наверное, если так разобраться. Но всё идёт к тому, что на каждый товар будет свой QR-код. И тогда можно будет действительно говорить о том, что мы точно знаем, что продано, когда продано и зачем.

— В последнее время довольно популярной стала тема географического искусственного интеллекта, так называемого Geo-AI. Насколько вы за этим следите, насколько, на Ваш взгляд, это — такое желание следовать модным течениям, а насколько — какие-то действительно новые методы, которые могут расширить исследовательский инструментарий?

— Что касается искусственного интеллекта и его применения по отношению к географически структурированным данным, на самом деле, это очень интересная тема. Мы в неё активно погружаемся... Я даже могу сказать, что мы в принципе подходим к тому, что соответствующие методы начнём использовать в ближайшее время. Здесь единственное необходимо разделить: есть более традиционные инструменты машинного обучения, включая Random Forest, Градиентный бустинг и нейронные сети.

Нейронные сети, на самом деле, достаточно специфичны. Их основная проблема в том, что они слабо интерпретируемы, скажем так. А большинство, заказчиков исследований, в том числе и организации, и госорганы, хотят понимать и интерпретировать, и понимать, как им воздействовать. А нейронные сети — это больше такой чёрный ящик, где мы вложили данные и получили определённую картинку. Как это будет без объяснения, почему это так, как это на это можно оказать влияние и так далее. Вот в этом смысле классическое машинное обучение может быть более интерпретируемо.

Инструменты машинного обучения тоже развиваются, и появляются все новые модели. Они сравниваются, мы рассматриваем эти работы, смотрим, какие инструменты дают наиболее эффективные результаты.

Могу сказать, что в рамках нашего университета мы активно взаимодействуем именно с географическим факультетом. У нас есть там достаточно интересный молодой учёный, который проходил стажировку в Европе, как раз у него геологические данные, но так или иначе, их распределение также подвержено пространственному закону. И он рассказывал о тех инструментах, которые сейчас активно применяются в Европе, в том числе и машинное обучение, и какие результаты могут быть получены и как их потом визуализировать. Поэтому мы, в принципе, уже готовы к тому, что мы будем применять эти инструменты в ближайшей перспективе.

Что из этого получится? Я надеюсь, что это будет что-то интересное. И, естественно, мы будем сравнивать те результаты, которые получают у нас по обычным пространственным моделям, в том числе географически взвешенной регрессии, и по моделям, полученным на основе машинного обучения... Хотя тут есть важная особенность, состоящая в том, что регрессия представляется как тип задач машинного обучения, но она проще в исполнении и интерпретации. Фактически же основной переход от традиционных количественных методов к машинному обучению состоит в оценке результатов исследования — в одном случае мы смотрим на некие статистические коэффициенты, а в другом — разделяем наши выборки на обучающую и тестовую, и уже на основе оценки последней делаем выводы.

— У вас, как у исследователя с очень широким спектром интересов, от эко-

номики и демографии до медицины, хотел бы спросить: насколько можно говорить о существовании общих пространственных закономерностей? Насколько Вы в своих работах видели сначала закономерность в одном домене, а потом видели тоже самое, возникающее в другом домене? Или разные социальные процессы можно объединить только на уровне очень широких общих закономерностей, как там законы Тоблера, например?

— На самом деле, закон Тоблера однозначно работает! Что касается всего спектра тех исследований, которые я провожу, тут необходимо в первую очередь отметить, что действительно он достаточно широкий. Не потому, что я разбираюсь во всех этих исследовательских направлениях, там в медицине, например, или в технических системах, например, в нефтяной промышленности (у неё тоже есть исследования, связанные с пространственной зависимостью добычи нефти), а в том, что исследователи из разных областей исходно сами понимают, что они видят определённые пространственные зависимости, и уже обращаются к нам с тем, чтобы мы помогли им их оценить. То есть мы в данном случае во многих исследованиях выступаем просто как консультанты, как люди, которые выстраивают эту модель и взаимодействуют на этапе интерпретации. То есть исходные исследовательские задачи часто ставятся не нами, и вопрос о том, существует ли там пространственные зависимости, исходно не нами выявляется. Делается гипотеза другими научными коллективами, которые уже к нам обращаются, чётко понимают, что да, здесь есть пространственная зависимость, и далее мы тестируем, есть ли она действительно или нет.

Соответственно, что касается вопроса о том, можно ли говорить о том, что одна и та же пространственная зависимость может проявляться для всех регионов, для всех, скажем так, территорий одинаково, нет. На самом деле, это происходит не так. Можно однозначно сказать, что даже на территории Российской Федерации существуют разные зоны, и это неоднократно отмечалось, например, в исследованиях Коломак Е.А., Демидовой О.А., что пространственная зависимость, то есть связь между территориями на Юго-Западе и Северо-Востоке, отличается. И это, в том числе, связано с тем, что там по-разному учитывается расстояние. То есть расстояние между населёнными пунктами в Сибири может быть сто километров, и люди привыкают перемещаться на сто километров в соседний населённый пункт, для них других вариантов нет, и для них это самое близкое поселение. Для Кавказа, где более плотное размещение населённых пунктов, они более мелкие, всё несколько иначе происходит.

Кроме того, есть определённая социальная особенность. Например, если мы поднимаем вопрос цен, то изменение цен в одном объекте в цен-

тральной части России может спровоцировать отток покупателей с одного торгового объекта в тот объект, где цена ниже. В отдельных территориях, где очень важны межличностные отношения, в том числе на уровне «покупатель-продавец», покупатель никогда не бросит того продавца, к которому он ходил, даже если продукты у него теперь дороже, просто потому что ну это не принято на этой территории. А соответственно, пространственная зависимость там будет работать иначе.

Что касается заболеваний, по заболеваниям мы тоже можем сказать, что даже внутри инфекционных заболеваний пространственные зависимости могут проявляться по-разному на разных территориях, и передача инфекции, например, коронавируса, она происходила быстрее, а такая болезнь, как туберкулёз, она передаётся иначе, хотя она также инфекционной является. Но скорость размножения туберкулёзной палочки очень низкая, в результате чего те модели пространственные, которые мы строим для быстро развивающихся инфекций, они не работают для медленных инфекций или для других болезней. Очень много нюансов, различий и в разрезе отдельных социально-экономических, медицинских, технологических процессов, и в разрезе отдельных территорий.

— *Вы сейчас в основном на территориальные особенности обратили внимание. А сталкивались ли вы с тем, чтобы в какой-то территории одни и те же факторы влияли на паттерны, к примеру, распространение инфекционных заболеваний и экономического поведения, или Вы такого не встречали?*

— Особенность территории, безусловно, оказывает влияние. Например, если это неблагоприятная территория, то такие зависимости, как потребление алкоголя и распространение туберкулёзной палочки, будут иметь схожие рисунки. Опять-таки, те же самые территории, где люди живут большими семьями, они достаточно закрыты. Там и болезни будут распространяться, возможно, медленнее, и пространственная передача информации будет медленнее.

— *Как вы считаете, нужно ли владеть языками программирования исследователю, который интересуется пространственными методами? Просматривая Ваши работы, я обратил внимание, что Вы двигались к всё более сложным методам, и в какой-то момент в основном перешли на уже написание кода, а не на использование каких-то программ.*

— На мой взгляд, всё-таки необходимо. Я отмечу следующее: если вы ра-

ботаете в большом коллективе, и у вас достаточно узкая задача, например, анализировать только кросс-секционные зависимости на каком-то определенном наборе данных, то, да, скорее всего, вам функционала отдельных программ будет достаточно.

Если у вас достаточно сложная исследовательская задача и вам необходимо сделать добор данных, загрузить какие-то сведения из стороннего ресурса, которые, как правило, не всегда бывают чистыми, то скорее всего потребуется умение работать с кодом. Нужно понимать, как собирать, объединять наборы данных, как делать их очистку, как учесть достаточно большое количество зависимостей. В простых программах, как правило, учитываются либо временные, либо пространственные факторы, либо какое-то их сочетание. Например, панельно-пространственную модель можно построить. Но когда вы захотите туда ещё добавить иерархическую плоскость, то столкнётесь с тем, что на текущий момент разработанного инструмента такого не будет.

Инструменты развиваются очень быстро. Вначале появляется метод, и уже позже если он пользуется спросом появляется ПО с дружественным интерфейсом, но на это уходит время. И есть такой важный нюанс: разработчики инструментов часто не заинтересованы делать их так, чтобы они вам были удобны и вы могли их быстро перенастроить. Те, кто разрабатывают эти инструменты и алгоритмы, они находятся уже на определенном уровне развития. И им кажется, что Вы должны в этом прекрасно разбираться и для Вас не будет никакой сложности этим пользоваться. И зачем для вас ещё что-то упрощать, делать дружественный интерфейс и так далее? То есть у них, как правило, таких даже мыслей не возникает, потому что они уже на совсем другом уровне полёта. И для того, чтобы начинать использовать более сложные инструменты, просто не обойдётся без языков программирования.

Я могу сказать так, что мне пришлось, хотя я доктор экономических наук, овладеть инструментами сбора данных, средствами API. Я умею парсить сайты, настраивать сборщики данных с этих сайтов, в том числе с помощью CSS-селекторов. Я владею SQL, R, Python. Я проходила переподготовку в МФТИ, я знаю Hadoop, Spark. Но на текущий момент я сталкиваюсь с тем, что теперь возникла проблема, связанная с тем, что я слишком далеко ушла.

И когда я обращаюсь к нашим программистам со словами: «Настройте мне среду для того, чтобы я это сделала», мне говорят: «У нас есть ограничения». При этом при работе с высокочастотными данными мы уже столкнулись с насущной необходимостью задействовать довольно серьезные вычислительные мощности и соответствующие инструменты по работе с

большими данными на вычислительных кластерах университета.

Общаясь с администраторами сетей, программистами вы должны уметь доносить до них то, что вам необходимо. Поэтому нужно осваивать эти инструменты, хотя бы понимать их функционал. Не везде нужно, скажем так, всё самим охватить. Это практически невозможно. Нужно собирать команду. У нас в лаборатории два экономиста и два кандидата технических наук, потому что мы без них не можем. Дело в том, что люди, которые получали математическое/техническое образование, они немножко иначе мыслят, они иначе видят те же самые инструменты и модели. И мы, часто получая определенные результаты с ними, безусловно, их обсуждаем. У нас есть супер-кластер, где есть отдельная команда, которая помогает нам в распараллеливании каких-то расчётов, потому что, когда мы строим модель на больших данных, на одном ядре её не посчитать. Вот и в этом контексте, безусловно, нужно развиваться не только самому, нужно ещё, чтобы с вами была определённая команда, которая готова это поддерживать и развивать.