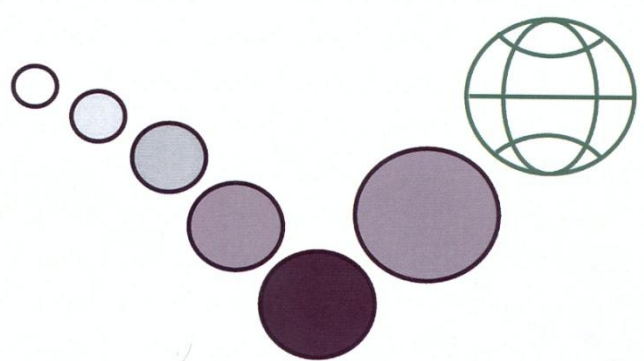


InterCarto
18 InterGIS

**Устойчивое развитие территорий:
теория ГИС и практический опыт**

Смоленск (Россия), 26-28 июня 2012
Сен-Дье-де-Вож (Франция), 2-4 июля 2012



Smolensk (Russia), June 26-28, 2012
St. Die des Vosges (France), July 2-4, 2012

**Sustainable development of territories:
GIS theory and practice**

18 InterGIS

ББК Д8с512
УДК 911.2/3:528.9/519.8

Редакционная коллегия:

С.П. Евдокимов (отв. ред.), В.С. Тикунов, Т.В. Ватлина

ИнтерКарто-ИнтерГИС-18: Устойчивое развитие территорий: теория ГИС и практический опыт. Материалы международной конференции / Редкол.: С.П. Евдокимов (отв. ред.) [и др.]. Смоленск, 26–28 июня, 2012 г. Смоленск, 2012. 532 с.

Сборник включает материалы, представленные на Международную конференцию ИнтерКарто-ИнтерГИС-18. Рассматриваются теоретические и методические аспекты геоинформационного обеспечения задач устойчивого развития, в том числе вопросы разработки и создания различных геоинформационных систем и новых геоинформационных технологий и виртуальных сред, создания инфраструктур пространственных данных, баз данных, их интеграции и метаданных, геосервисов, обработки дистанционных изображений и исследования Земли из космоса. Предлагаются пути решения управленческих, экономических, градостроительных, планировочных, экологических, медицинских и рекреационных задач.

ИнтерКарто-1 – Москва, 23–25 мая 1994 г.
ИнтерКарто-2 – Иркутск, 26–29 июня 1996 г.
ИнтерКарто-3 – Новосибирск, 27–31 января 1997 г.
ИнтерКарто-4 – Барнаул, 1–4 июля 1998 г.
ИнтерКарто-5 – Якутск, 17–19 июня 1999 г.
ИнтерКарто-6 – Апатиты, Мурманской обл., 22–24 августа 2000 г.
ИнтерКарто-7 – Петропавловск-Камчатский, 30 июля – 1 августа 2001 г.
ИнтерКарто-8 – Хельсинки (Финляндия), Санкт-Петербург, 28 мая – 1 июня 2002 г.
ИнтерКарто-9 – Новороссийск, Севастополь (Украина), 25–29 июня 2003 г.
ИнтерКарто-ИнтерГИС-10 – Владивосток, Чанчунь (Китай), 12–19 июля 2004 г.
ИнтерКарто-ИнтерГИС-11 – Ставрополь-Домбай, Будапешт (Венгрия), 25 сентября – 3 октября 2005 г.
ИнтерКарто-ИнтерГИС-12 – Калининград, Берлин (Германия), 25 – 1 августа 2006 г.
ИнтерКарто-ИнтерГИС-13 – Ханты-Мансийск, Йеллоунайф (Канада), 12–24 августа 2007 г.
ИнтерКарто-ИнтерГИС-14 – Саратов, Урумчи (Китай), 24 июня – 1 июля 2008 г.
ИнтерКарто-ИнтерГИС-15 – Пермь, Гент (Бельгия), 29 июня – 5 июля 2009 г.
ИнтерКарто-ИнтерГИС-16 – Ростов-на-Дону, Зальцбург (Австрия), 3–8 июля 2010 г.
ИнтерКарто-ИнтерГИС-17 – Барнаул, Денпасар (Индонезия), 14–19 декабря 2011 г.
ИнтерКарто-ИнтерГИС-18 – Смоленск, Сен-Дье-де-Вож (Франция), 26 июня – 4 июля 2012 г.

*Издание осуществляется при финансовой поддержке Российского фонда
фундаментальных исследований (проект 12-05-06036-г)
и Администрации Смоленской области*

© Международная Картографическая Ассоциация
© Коллектив авторов

ПРОСТРАНСТВЕННО-СТАТИСТИЧЕСКИЙ АНАЛИЗ
ДЕМОГРАФИЧЕСКИХ ПРОЦЕССОВ²

С.А. Тимонин

Географический факультет МГУ имени М.В. Ломоносова
Москва, Россия, ser-timonin@yandex.ru

SPATIAL STATISTICAL ANALYSIS OF DEMOGRAPHIC PROCESSES

Sergey Timonin

Faculty of Geography, Lomonosov Moscow state university
Moscow, Russia, ser-timonin@yandex.ru

Abstract. Different aspects of spatial statistic analysis of demographic processes are presented in the article. Among the main are geovisualisation, analysis of geographical distribution and patterns, spatial autocorrelation. As an example, analysis of spatial patterns of ethnicities in Russia on the basis of spatial autocorrelation measures is carried out.

Традиционно статистические методы являются ключевыми при анализе демографических процессов. Однако для изучения геодемографической ситуации и выявления пространственных закономерностей в развитии демографических процессов требуется привлечение дополнительного пространственно-статистического аппарата³.

Большинство специалистов, занимающихся демографическими вопросами, не всегда уделяют достаточное внимание пространственным аспектам демографического развития. Это, в первую очередь, объясняется тем, что основная группа специалистов в области демографических исследований, как правило, фокусируется на изучении определенной группы населения и происходящими в ней демографическими процессами больше с временной, нежели пространственной точек зрения. Даже очень немногие специалисты-географы в области народонаселения используют пространственные методы в своих исследованиях, за исключением измерения миграционной подвижности населения и сравнения демографических трендов между регионами.

Еще одна причина недостаточного использования пространственно-статистического анализа в демографических исследованиях – отсутствие подходящего программного обеспечения (программных решений) для анализа статистических (в том числе геодемографических) процессов в пространстве. Тем не менее в последнее время наблюдается тенденция к внедрению модулей пространственно-статистического анализа и моделирования в геоинформационное программное обеспечение, что существенно расширяет возможности статистического анализа пространственных данных. Ведется разработка и отдельных программных продуктов для проведения пространственно-статистического анализа. Исключительная роль в этом направлении принадлежит специалистам Центра пространственно-интегрированных социальных наук, США.

В конце 1990-х годов Люк Анселин сформулировал понятие «исследовательский пространственный анализ данных» (exploratory spatial data analysis), который определил как «набор методов для описания и визуализации пространственных распределений, установления (распознавания) атипичных расположений или пространственных выбросов, определения пространственных закономерностей, кластеров или «горячих точек», и предположение пространственных режимов или иных форм пространственной неоднородности (гетерогенности)» [Maguire, 2005]⁴.

В соответствии с данным определением, а также рядом других работ по пространственной статистике [Edon, 1977; Handbook..., 2009], могут быть выделены следующие варианты исследовательского пространственного анализа данных:

- геовизуализация данных на основе группировки одного из значений атрибутов объектов и построения статистических графиков;
- выявление и математическая оценка распределение/размещения объектов или явлений в пространстве;

- изучение пространственной автокорреляции и пространственной регрессии.

Также отметим, что для выполнения полноценного пространственно-статистического анализа, следующие функции анализа должны быть реализованы в программном обеспечении:

- построение традиционных статистических графиков (гистограммы, «ящик с усами» (box plot), график рассеяния и прочие);
- визуализация данных посредством создания карт;
- динамическая связь между объектами на карте, графике и в таблице;
- выполнение анализа – оценка распределения объектов, выявление пространственных закономерностей и взаимосвязей.

Геовизуализация данных на основе группировки одного из значений атрибутов объектов.

² Исследование выполнено при поддержке гранта РФФИ № 12-06-00310-а

³ В качестве исключения можно назвать элементарные пространственно-статистические манипуляции с точечными объектами, которые, как правило, могут быть выполнены в стандартном статистическом программном обеспечении. К примеру, расчет описательных статистических показателей местоположения, таких как медианный или средний центр. Однако анализ пространственных закономерностей в распределении точечных объектов или изучение объектов, представленных в виде полигонов, не может быть выполнен в стандартном статистическом ПО.

⁴ «a collection of techniques to describe and visualize spatial distributions, identify atypical locations or spatial outliers, discover patterns of spatial association, clusters or hot spots, and suggest spatial regimes or other forms of spatial heterogeneity»

СФЕРЫ ПРАКТИЧЕСКОГО ПРИМЕНЕНИЯ ГИС И КАРТОГРАФИИ В ЭКОНОМИКЕ

Существует несколько традиционных способов картографической передачи информации (картографические способы изображения), которые используются для создания тематических карт, в том числе и картодемографических характеристик населения [Тимонин, 2010]. В тоже время, для анализа больших массивов статистических данных зачастую достаточно применения способа картограмм. Широкое внедрение именно этого способа картографического изображения в изучении геомемографических процессов связано как с особенностью сбора демографических данных по единицам административно-территориального деления, так и с разработкой статистических методов классификации единиц картографирования по их атрибутивным показателям. Выбор того или иного метода группировки, как правило, зависит от характера распределения исходных данных и целей проводимого анализа. Остановимся на основных из них:

- *метод естественных интервалов (метод оптимизации Йенка)*. Границы интервалов определяются таким образом, чтобы сгруппировать схожие значения в пределах класса (минимизация среднеквадратичных отклонений в каждом классе) и максимально увеличить различия между классами (максимизация среднеквадратических отклонений между классами). Объекты делятся на группы, границы которых устанавливаются там, где встречаются относительно большие скачки значений.

- *метод равных интервалов*. В рамках данной схемы классификации весь диапазон анализируемых значений делится на равные промежутки в соответствии с заданным пользователем количеством интервалов. После этого объекты распределяются по группам в соответствии с определенными границами классов.

- *метод равных классов (метод квантилей)*. Каждая группа содержит одинаковое число единиц картографирования. Такая классификация хорошо подходит для линейно распределенных данных. Поскольку объекты сгруппированы по принципу их одинакового количества в классе, полученная карта может ввести в заблуждение. Похожие объекты могут попасть в разные классы, а объекты с существенно разными значениями могут оказаться в одном классе. Минимизировать искажения возможно путем увеличения числа классов.

- *метод стандартных отклонений*. Эта схема классификации показывает, насколько значения атрибутов объектов отличаются от среднего значения. Вычисляется среднее арифметическое и стандартные отклонения от среднего по формуле:

$$\bar{x} = \frac{\sum_{i=1}^N x_i}{N}, \quad \sigma = \sqrt{\frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})^2}$$

Границы классов устанавливаются с шагом, кратным 0,25, 0,5, 1 стандартному отклонению в зависимости от характера распределения данных. На рисунке 1 приведен график распределения показателей ожидаемой продолжительности жизни при рождении для субъектов ЦФО (2009 год). Анализ графика позволяет установить, что в ЦФО нет регионов, в которых ОПЖ более чем на 1 стандартное отклонение хуже среднего. Однако имеется 2 региона, в которых ОПЖ существенно лучше среднего значения (более чем на 2-2,5 стандартных отклонения).

При визуализации данных методом стандартных отклонений следует учитывать, что геоинформационные продукты автоматически рассчитывают среднее арифметическое значение из представленной выборки и соответственно стандартное отклонение. Однако полученная таким образом средняя не всегда является реальной средней величиной. Так, простая средняя арифметическая величина ОПЖ для регионов ЦФО составляет 69,35 лет, в то время как ОПЖ в среднем по ЦФО составляет 68,06 лет. В связи с этим, следует учитывать этот нюанс, и принимать решение относительно какой средней необходимо рассчитывать стандартные отклонения и визуализировать данные.

На рисунке 2 приведена серия карт, отражающая ОПЖ при рождении в субъектах ЦФО, построенная разными методами группировки значений атрибута.

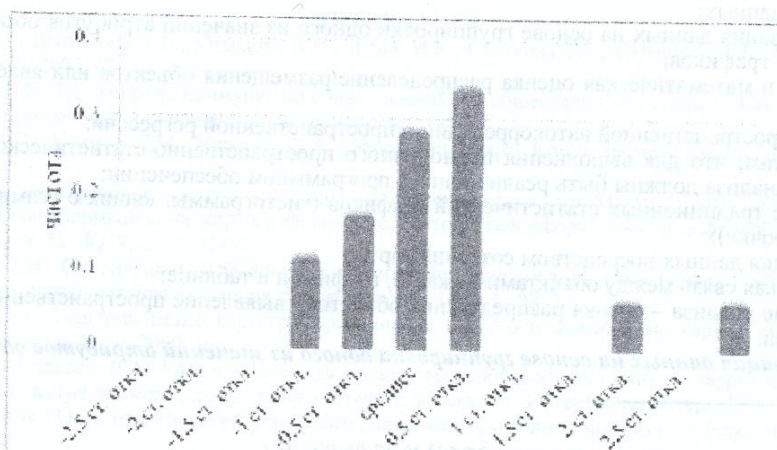


Рис. 1. График распределения субъектов ЦФО по ОПЖ при рождении, 2009 год

Чаще всего рекомендуется использовать различные методы классификации для поиска наилучшего метода группировки и визуализации данных, а также же для выявления особенностей распространения изучаемого явления. Помимо использования стандартных статистических методов классификации, пользователь может самостоятельно вводить значения заранее определенных интервалов для дальнейшей визуализации явления.

СФЕРЫ ПРАКТИЧЕСКОГО ПРИМЕНЕНИЯ ГИС И КАРТОГРАФИИ В ЭКОНОМИКЕ

Выявление и математическая оценка распределение/размещения объектов и явлений в пространстве.

Существует несколько методов оценки размещения объектов в пространстве, которые могут быть очень востребованы в геодемографических исследованиях. Первая группа методов характеризуется показателями центра распределения, которые вычисляются следующим образом:

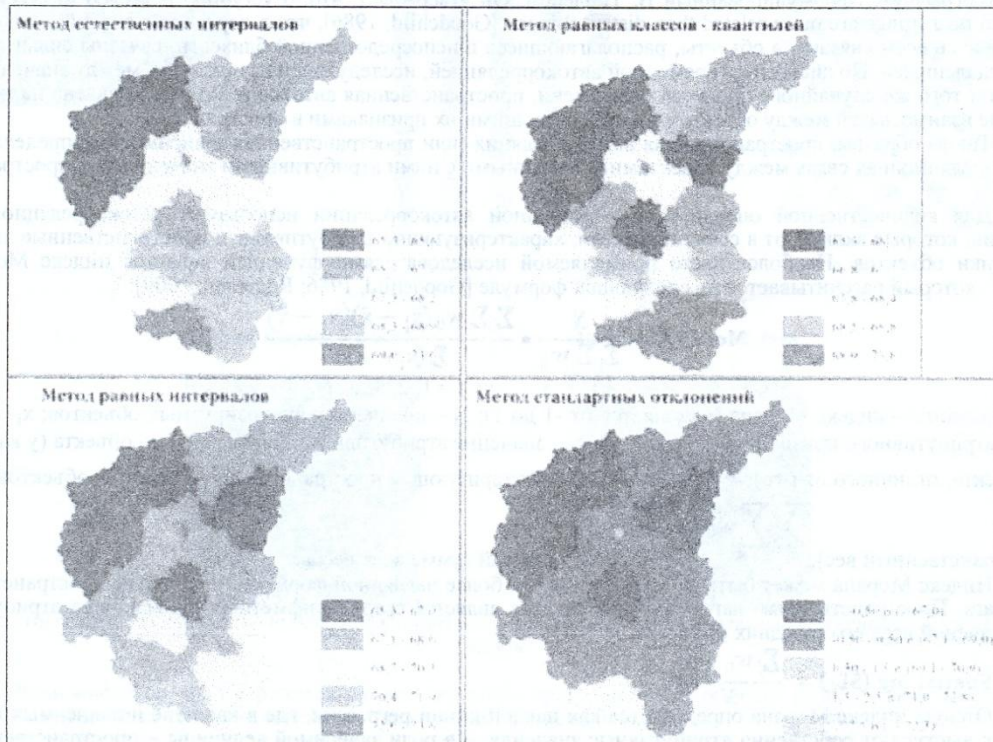


Рис. 2. Варианты визуализации ОПЖ в регионах ЦФО, основанные на различных методах классификации значений атрибутивного признака

Центральный объект. Из набора анализируемых пространственных объектов выбирается наиболее центрально расположенный объект (для точек, линий, полигонов). Для этого рассчитывается сумма расстояний от каждого объекта до всех других. Минимальная величина характеризует центральный объект. Возможно также использование системы весовых коэффициентов. К примеру, учет численности населения при поиске центрального объекта.

Средний центр. Создается новый точечный объект, чьи координаты соответствуют средней или средней взвешенной величине из координат анализируемого набора объектов (для полигональных объектов – центроидов).

$$\bar{x} = \frac{\sum_{i=1}^n w_i x_i}{\sum_{i=1}^n w_i}, \quad \bar{y} = \frac{\sum_{i=1}^n w_i y_i}{\sum_{i=1}^n w_i}$$

Медианный центр (центр минимального пути). Создается новый точечный объект, который находится на минимальном расстоянии от всех остальных анализируемых объектов. Процесс поиска этого объекта итерационный.

Вторая группа методов представлена показателями, характеризующими дисперсию в пространственном размещении объектов:

Стандартное расстояние. Характеризует компактность размещения объектов в пространстве путем вычисления стандартного расстояния относительно среднего центра:

$$\sigma = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2 + \sum_{i=1}^n (y_i - \bar{y})^2}{n}}$$

При желании рассчитать среднее взвешенное стандартное расстояние пользуются следующей формулой:

$$\sigma_w = \sqrt{\frac{\sum_{i=1}^n w_i (x_i - \bar{x}_w)^2}{\sum_{i=1}^n w_i} + \frac{\sum_{i=1}^n w_i (y_i - \bar{y}_w)^2}{\sum_{i=1}^n w_i}}$$

По результатам расчетов строится окружность в точке с координатами $(\bar{x}; \bar{y})$ и радиусом в n стандартных расстояний.

Эллипс стандартных отклонений. Принимается во внимание тот факт, что распределение объектов относительно среднего центра может быть различно по различным направлениям. Для этого вычисляется длина малой оси, длина большой оси и ориентация эллипса. По результатам расчетов строится эллипс стандартных отклонений.

СФЕРЫ ПРАКТИЧЕСКОГО ПРИМЕНЕНИЯ ГИС И КАРТОГРАФИИ В ЭКОНОМИКЕ

Анализ закономерностей и взаимосвязей в пространственном распределении явлений. Изучение пространственной автокорреляции

Исследование пространственной автокорреляции, кластеризации и «выбросов» является одним из наиболее востребованных и часто используемых методов анализа закономерностей в пространственном распределении явлений. Гипотезой в изучении пространственной взаимосвязи служит так называемый «первый закон географии», сформулированный В. Тоблером. Он утверждает, что «Everything is related to everything else, but near things are more related than distant things» [Goodchild, 1986], что в переводе с английского означает «всё со всем связано, а объекты, располагающиеся в непосредственной близости, связаны сильнее, нежели удаленные». По аналогии с временной автокорреляцией, исследующей взаимосвязь между значениями одного и того же случайного процесса во времени, пространственная автокорреляция направлена на исследование взаимосвязей между объектами и описывающими их признаками в пространстве.

Таким образом, пространственная автокорреляция (или пространственная зависимость) определяется как корреляционная связь между объектами и связанными с ними атрибутивными значениями в пространстве.

Для количественной оценки пространственной автокорреляции используют автокорреляционные функции, которые включают в себя показатели, характеризующие атрибутивные и пространственные характеристики объектов. Наиболее часто применяемой исследователями функцией является индекс Морана (1948)⁵, который рассчитывается по следующей формуле [Goodchild, 1986; Rogerson, 2009]:

$$\text{Moran } I = \frac{N \sum_i \sum_j w_{ij} (x_i - \bar{x})(x_j - \bar{x})}{\sum_i \sum_j w_{ij} \cdot \sum_i (x_i - \bar{x})^2}$$

$i = 1, 2, 3, \dots, N; j = 1, 2, 3, \dots, N.$

Moran I – индекс Морана (изменяется от -1 до 1); N – количество анализируемых объектов; x_i – значение атрибутивного признака у i -го объекта; x_j – значение атрибутивного признака у j -го объекта (у каждого объекта, отличного от i -го); w_{ij} – величина, характеризующая пространственную близость объектов i и j

(пространственный вес); $\sum_i \sum_j w_{ij}$ – показатель, равный сумме всех весов.

Индекс Морана может быть также определен в более наглядной форме через понятие пространственного лага. Пространственным лагом для i -го объекта является среднеарифметическая величина атрибутивных значений всех его соседних объектов.

$$\text{Spatial lag } (SL_i) = \frac{\sum_j w_{ij} \cdot x_j}{N}$$

Отсюда индекс Морана определяется как наклон линии регрессии, где в качестве независимых переменных выступают собственно атрибутивные значения, а в роли зависимой величины – пространственный лаг этих значений. Такой подход позволяет легко рассчитать индекс Морана, а также представить промежуточные и итоговые результаты с помощью графика рассеяния.

Глобальный индекс Морана изменяется от -1 до 1. Значение «-1» соответствует идеальному дисперсному размещению объектов в пространстве (объекты с низкими атрибутивными значениями располагаются в пространстве рядом с объектами с высокими показателями). Значение «1» характерно для близко расположенных в пространстве объектов со схожими атрибутивными значениями (объекты с высокими значениями атрибутов располагаются рядом с друг другом, и с низкими – тоже рядом с друг другом). Значение «0» свидетельствует о случайном (рандомном) характере распределения объектов и их атрибутивных признаков в пространстве. Чем ближе значение глобального индекса Морана к нулю, тем сложнее говорить о каких-либо закономерностях размещения объектов в пространстве. В качестве примера на рисунке 3 приведены несколько решеток размером 10*10 ячеек. Каждой ячейке присвоено одно из двух возможных значений: 1 или 0, и для нескольких вариантов размещения атрибутов по ячейкам рассчитан индекс Морана.

Исключительно важную роль в расчете пространственных лагов, а вместе с тем и индекса Морана, является определение соседства (взаиморасположение объектов в пространстве), что, в свою очередь, описывается так называемой матрицей пространственных весов. Матрица пространственных весов представляет собой квадратную матрицу, состоящая из N^2 элементов (N – количество пространственных объектов). Если 2 объекта являются соседями, то на пересечении столбца и строки, описывающих эти объекты, фиксируется значение 1, в противном случае – 0. Такой подход (бинарный) позволяет описывать только наличие или отсутствие связи между объектами в пространстве. При необходимости более точного описания степени связи объектов в пространстве, могут указываться веса, характеризующие степень удаленности объектов друг от

⁵ Альтернативным показателем оценки степени пространственной автокорреляции является индекс Гери (Geary's Ratio C, 1954), рассчитываемый по формуле:

$$\text{Geary's } C = \frac{(n-1) \sum_i \sum_j w_{ij} (x_i - x_j)^2}{2 \left(\sum_i \sum_j w_{ij} \right) \cdot \sum_i (x_i - \bar{x})^2}$$

Значения индекса изменяются от 0 до 2. Чем ближе величина к 2, тем выше пространственная взаимосвязь. Значение равное 1 свидетельствует о случайном характере распределения объектов и связанных с ними атрибутивных признаков. Основное отличие индекса Гери от индекса Морана состоит в том, что при его расчете сравниваются атрибутивные значения двух соседних объектов друг с другом, а не со средним значением, как при вычислении индекса Морана.

СФЕРЫ ПРАКТИЧЕСКОГО ПРИМЕНЕНИЯ ГИС И КАРТОГРАФИИ В ЭКОНОМИКЕ

друга. В некоторых случаях проводят стандартизацию сток матрицы пространственных весов (row standardization) – каждый вес делится на сумму весов в строке или, иными словами, сумму весов всех соседей. В результате, все веса принимают значения от 0 до 1.

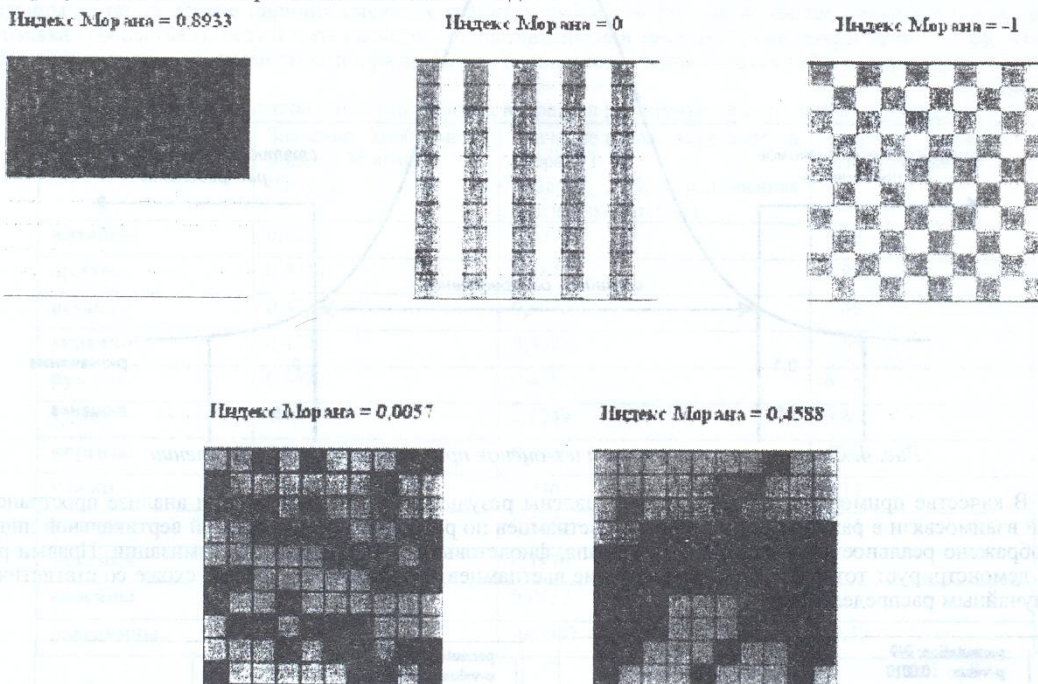


Рис. 3. Визуализация расчетов индекса Морана для нескольких вариантов размещения атрибутивных признаков

Возможно несколько вариантов расчета матрицы пространственных весов (определения соседства), которые зависят от целей исследования, а также характера исходных данных:

- на основании наличия у анализируемых объектов смежных ребер или узлов;
- путем указания расстояния между объектами в пространстве;
- путем установления заданного числа ближайших соседей;
- на предположение, что с увеличением расстояния от конкретного объекта в пространстве, уменьшается вес других объектов (относительно каждого анализируемого).

В первом случае соседями являются объекты, имеющие общую границу или узел. При втором подходе вокруг каждого объекта (центроида объекта) строится окружность заданного радиуса, и все единицы, попадающие в пределы окружности, являются соседями. Такой прием следует использовать в том случае, когда анализируемые полигональные объекты сильно варьируют по своему размеру. В некоторых случаях необходимо использовать не только «воздушные» расстояния, а реальные расстояния по дорожной сети или ценовые расстояния. В третьем случае количество соседей у каждого объекта постоянно и определяется заданным числом k . В последнем – объектам присваивается вес, обратно пропорциональный расстоянию от анализируемого объекта. В этом случае также имеет смысл устанавливать пороговые значения, превышение которых указывает на отсутствие связи («соседства») между объектами. Пороговые значения также задаются для того, чтобы гарантировать, что у каждого объекта имеется хотя бы один сосед. Как в случае проведения любого статистического анализа, при оценке пространственной автокорреляции необходимо оценить статистическую значимость полученных результатов, т.е. выявить насколько анализируемое распределение объектов и их атрибутивных значений отличается от случайной модели размещения. При этом используют нулевую гипотезу, в рамках которой предполагается полная пространственная случайность (complete spatial randomness) как в размещении объектов, так и в описывающих их атрибутивных признаках. Для того, чтобы признать тест на пространственную автокорреляцию статистически значимым, а результаты исследования корректными, необходимо опровергнуть нулевую гипотезу. Для этого рассчитывают ожидаемое значение индекса Морана (E_i), стандартное отклонение для данной величины, p -значение (вероятность того, что наблюдаемое распределение объектов и ассоциированных с ними атрибутивных признаков, является результатом случайных процессов) и z -оценку (стандартное отклонение). Взаимосвязь p -значений и z -оценок является однозначной и может быть проиллюстрирована на графике нормального распределения (рис. 4).

Существует несколько вариантов проверки статистической значимости результатов. Один из наиболее часто используемых подходов, реализованных в ряде ГИС-пакетов, – рандомизация. Суть рандомизации заключается в следующем:

- исходные атрибутивные значения анализируемых пространственных объектов случайным образом перераспределяются среди объектов заданное число раз;
- для каждого нового распределения рассчитывается глобальный индекс пространственной автокорреляции (например, индекс Морана);
- результаты расчетов индекса Морана для случайно перераспределенных наборов значений (их средняя величина и стандартное отклонение) сравниваются с индексом Морана, полученным для реально наблюдаемых объектов с их атрибутивными признаками.

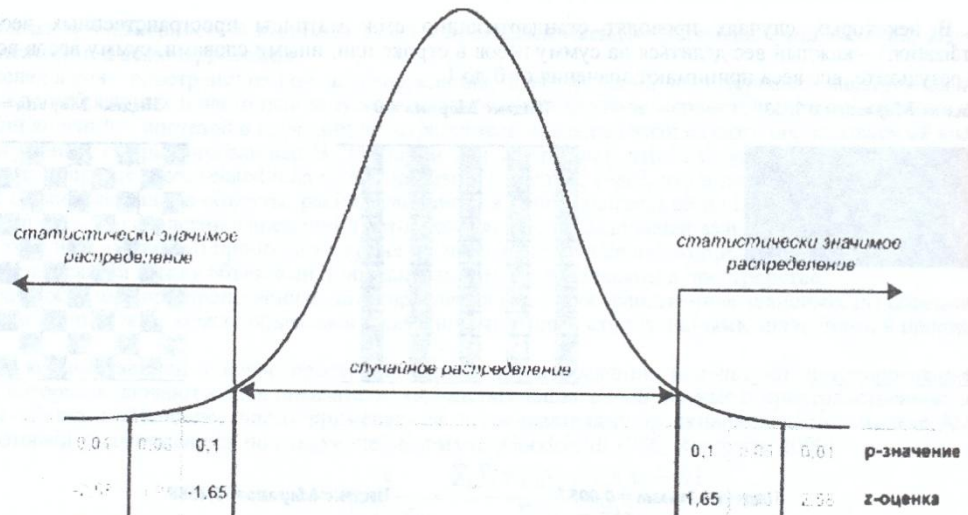


Рис. 4. Соотношение p-значений и z-оценок при нормальном распределении

В качестве примера на рисунке 5 представлены результаты рандомизации при анализе пространственной взаимосвязи в размещении китайцев и вьетнамцев по регионам России. Желтой вертикальной линией отображено реальное значение индекса Морана, фиолетовыми – результаты рандомизации. Правый рисунок демонстрирует тот факт, что распределение вьетнамцев по территории России, схоже со статистически случайным распределением.

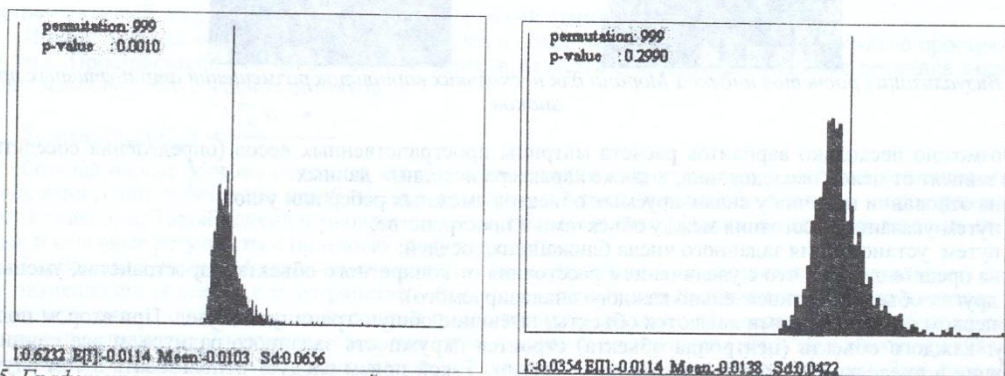


Рис. 5. График проверки статистической значимости расчетов индекса Морана для двух национальностей: китайцы (слева) и вьетнамцы (справа)

Ожидаемое значение Индекса Морана определяется по следующей формуле:

$$E_i = - \frac{1}{N-1}$$

Z-оценка рассчитывается с учетом реально рассчитанного значения индекса Морана, ожидаемой величины и стандартного отклонения:

$$z = - (I - E_i) / \sqrt{Var(E_i)}$$

Глобальные показатели автокорреляции позволяют охарактеризовать ситуацию в целом (наличие и степень пространственной взаимосвязи). Помимо этого, важным является вычисление локальных индикаторов пространственной автокорреляции для обнаружения кластеров (групп близко расположенных в пространстве объектов с высокими или низкими значениями атрибутов) и выбросов (объектов в пространстве, атрибутивные признаки которых значительно отличаются от атрибутивных значений соседних объектов). Для этих целей используют локальный индикатор пространственной ассоциативности (с англ. – local indicators of spatial association - LISA), предложенный Анселином и являющийся модификацией глобального индекса Морана [Maguire, 2005]:

$$LISA = \frac{x_i - \bar{x}}{\sigma} * \sum_j w_{ij} \left(\frac{x_j - \bar{x}}{\sigma} \right)$$

Данный показатель рассчитывается для каждого из анализируемых пространственных объектов, что позволяет картографировать результаты вычислений и выявлять типы пространственных взаимосвязей – кластеры и выбросы. Кроме того, для каждого объекта рассчитывается и показатель статистической значимости, что дает возможность определять лишь статистически значимые взаимосвязи в пространстве:

$$z = - \frac{LISA - E_{LISA}}{Var_{E_i}} \quad E_{LISA} = - \frac{w_i}{N-1}$$

В качестве примера практического внедрения представленного выше математического аппарата, проведем анализ пространственных закономерностей в распределении национальностей по субъектам Российской Федерации.

СФЕРЫ ПРАКТИЧЕСКОГО ПРИМЕНЕНИЯ ГИС И КАРТОГРАФИИ В ЭКОНОМИКЕ

ской Федерации и регионами бывшего РСФСР на основе данных переписей населения 1970-2002 гг. Вначале были проведены расчеты индекса Морана для каждой из 12 анализируемых национальностей. Вычисления выполнялись в двух программных продуктах – GeoDa и ArcGIS Desktop (использовались одинаковые матрицы весов на основе наличия смежных границ у субъектов РФ как наиболее очевидного и понятного признака «соседства»). Результаты расчетов, представленные в таблице 1, свидетельствуют о том, что имеются определенные закономерности в размещении большинства национальностей по территории России.

Таблица 2

Таблица результатов расчетов индекса Морана в различных программных продуктах

Название национальности	Значение глобального индекса Морана I (GeoDa)	Значение глобального индекса Морана I (модуль Пространственная статистика, ArcGIS)	z-оценка (стандартные отклонения)
китайцы	0,6232	0,6375	9,93
армяне	0,5714	0,5845	8,98
немцы	0,4962	0,5076	7,49
украинцы	0,4524	0,4628	6,55
русские	0,4477	0,458	6,25
турки	0,4153	0,4249	6,6
киргизы	0,3936	0,4046	6,98
узбеки	0,3385	0,3463	5,11
поляки	0,2201	0,2252	3,44
таджики	0,1934	0,1978	3
корейцы	0,0446	0,0457	2,11
вьетнамцы	-0,0354	-0,0362	-0,52

В таблице также приведены z-оценки для каждого из расчетов. Для 10 национальностей z-оценка выше величины 2,58. Из этого следует, что вероятность того, что распределение этнических групп по территории Российской Федерации носит случайный характер, менее 0,01 и нулевая гипотеза может быть отклонена. Расчеты индекса Морана для двух других этнических групп являются не достаточно достоверными, поскольку их размещения по территории России не сильно отличаются от случайного распределения.

На следующем этапе было проведено сравнения расчетов индекса Морана при использовании различных подходов к определению «соседства». Два метода представляются наиболее целесообразными к использованию при анализе пространственной взаимосвязи в регионах России – метод наличия общих границ и метод на основе обратных расстояний. Результаты расчетов представлены в табл. 2 и на рис. 6.

Таблица 3

Результаты расчетов индекса Морана с применением различных матриц весов

Название национальности	Индекс Морана (весовая матрица на основе наличия общих границ)	z	Индекс Морана (весовая матрица на основе обратных расстояний)	z	Индекс Морана (весовая матрица на основе обратных расстояний, стандартизованная)	z
китайцы	0,6375	9,93	0,0474	1,35	0,3517	9,32
армяне	0,5845	8,98	0,2366	5,61	0,2423	6,49
немцы	0,5076	7,49	0,2623	5,96	0,4962	10,57
украинцы	0,4628	6,55	0,1762	3,93	0,4222	8,64
русские	0,458	6,25	0,4173	8,7	0,3909	7,73
турки	0,4249	6,6	0,3133	7,36	0,2749	7,26
киргизы	0,4046	6,98	0,1415	3,81	0,2366	7,02
узбеки	0,3463	5,11	0,0851	2,08	0,1717	4,44
поляки	0,2252	3,44	0,1326	3,16	0,093	2,55
таджики	0,1978	3	0,0273	0,84	0,0971	2,61
корейцы	0,0457	2,11	0,0705	1,95	0,1596	10,73
вьетнамцы	-0,0362	-0,52	-0,1609	-2,82	-0,0192	-0,276

Сравнение результатов, представленных в таблице и на графике, свидетельствует о существенных различиях в расчетах при использовании различных матриц весов. При использовании метода обратных расстояний, влияние одного объекта на другой уменьшается с увеличением расстояния. Учитывая тот факт, что регионы России существенно различаются по площади, то при использовании данного подхода (с учетом задания порогового значения), регионы центральной России имеют существенно большее количество соседей, нежели, к примеру, восточносибирские и дальневосточные. Стандартизация матрицы весов заметно корректирует расчеты, уменьшая весовые коэффициенты для регионов, имеющих много соседей (маленькие по площади субъекты РФ), и увеличивая значения крупных по площади регионов России.

СФЕРЫ ПРАКТИЧЕСКОГО ПРИМЕНЕНИЯ ГИС И КАРТОГРАФИИ В ЭКОНОМИКЕ

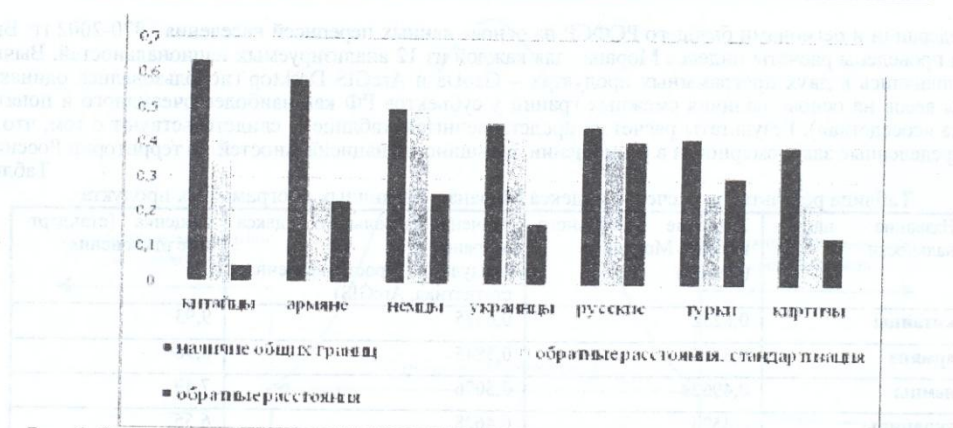


Рис. 6. Сравнение результатов расчета индекса Морана с применением различных матриц весов

На следующем этапе была предпринята попытка оценки степени пространственной автокорреляции с использованием анаморфированных изображений. Анаморфирование регионов Российской Федерации по общей численности населения позволяет представить более реальную картину расселения населения, и тем самым точнее оценить пространственную взаимосвязь в размещении этнических групп (рис. 7).

На заключительном этапе предпринималась попытка выделить кластеры и выбросы в размещении национальностей по территории Российской Федерации. Расчеты локального индикатора пространственной ассоциативности проводились для трех национальностей (китайцы, немцы, русские) за несколько переписных лет. Некоторые результаты представлены на рис. 8. В размещении каждой из национальностей наблюдаются кластеры высокой и низкой концентрации этноса.



Рис. 7. Сравнение расчетов индекса Морана при традиционном представлении регионов России и при анаморфировании субъектов России по численности населения

ЛИТЕРАТУРА

- Тимонин С.А. Методы математико-картографического и геоинформационного моделирования для изучения демографических процессов в регионах Российской Федерации. // Вестн. Моск. ун-та, Серия 5. География. М.: Издательство МГУ, 2010. № 5. С.11–18.
- Ebdon D. Statistics in geography: a practical approach. Oxford, Great Britain: Basil Blackwell, 1977. 195 с.
- Goodchild, Michael F. Spatial autocorrelation. Ontario, Canada: University of Western Ontario, 1986. 57 с. URL: <http://qmgr.org.uk/files/2008/11/47-spatial-autocorrelation.pdf>. Дата обращения: 04.07.2011.
- Handbook of applied spatial analysis: software tools, methods and applications. Edited by Fischer M., Getis A. Springer, 2009. 811 с.
- Maguire J. David, et al. GIS, spatial analysis and modeling. ESRI Press, 2005 480 с.
- Rogerson Peter and Yamada Ikuho. Statistical Detection and Surveillance of Geographic Clusters. New York, USA: Chapman & Hall / CRC, 2009. 322 с.

СФЕРЫ ПРАКТИЧЕСКОГО ПРИМЕНЕНИЯ ГИС И КАРТОГРАФИИ В ЭКОНОМИКЕ

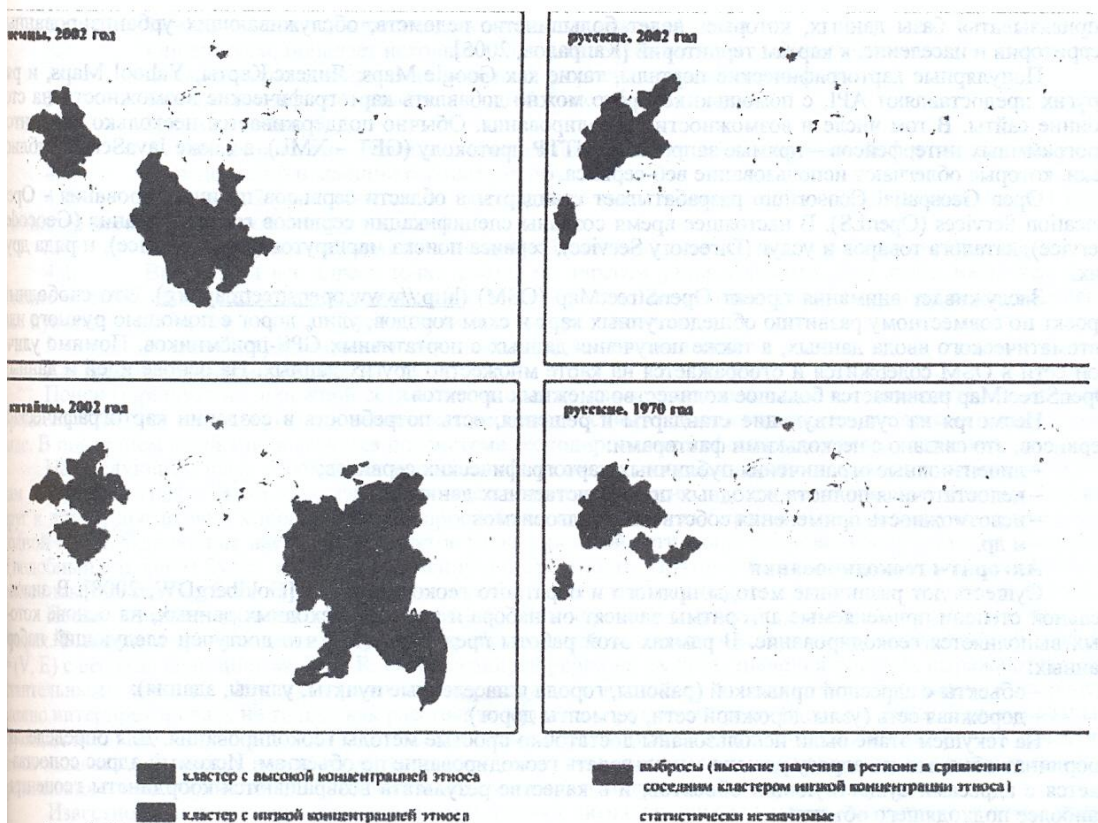


Рис. 8. Кластеры и выбросы в размещении некоторых этнических групп по территории России

ПОСТРОЕНИЕ ГЕОПРОСТРАНСТВЕННЫХ ВЕБ-СЕРВИСОВ ДЛЯ ЗАДАЧ ОПЕРАТИВНОГО МОНИТОРИНГА ТРАНСПОРТНЫХ СРЕДСТВ

А.В. Токарев
Институт вычислительного моделирования СО РАН
г. Красноярск, Россия

BUILDING GEOSPATIAL WEB-SERVICES FOR ON-LINE MONITORING OF TRANSPORT VEHICLES

A.V. Tokarev
Institute of Computational Modeling SB RAS
Krasnoyarsk, Russia Federation

Abstract. The application of service-oriented architecture in GPS / GLONASS vehicle monitoring systems is considered. The algorithms of forward and reverse geocoding based on address full text search and the road network data is proposed. Developed geospatial web services provide geocoding and shortest path problem solving and based on the open source software. The solution actively used in satellite transport monitoring system "Regnass" on the Krasnoyarsk Territory.

Введение

Системы мониторинга на основе технологий спутникового позиционирования GPS/ГЛОНАСС обеспечивают эффективный контроль за движением транспорта и управление транспортным комплексом города или автопарком предприятия. Одним из существенных модулей системы мониторинга является блок работы с картографической информацией. Он обеспечивает не только наглядное отображение местоположений транспортных средств, но и решение ряда вспомогательных задач связанных с адресным поиском и дорожной сетью. Применение сервис-ориентированной архитектуры позволяет выделить и разнести отдельные компоненты программного обеспечения. Такое решение повышает масштабируемость системы и позволяет развивать отдельные модули независимо друг от друга.

В настоящей работе рассматривается реализация нескольких гео-сервисов – сервиса геокодирования и сервиса поиска маршрута. **Геокодирование** (geocoding) – процесс преобразования адресов пространственных объектов в их географические координаты. Обратный процесс, преобразование точки на карте в читаемый для пользователя адрес, известен как **обратное геокодирование**. Функции геокодирования позволяют