

УДК 004.9:519.2

МЕТОДЫ АНАЛИЗА ДАННЫХ В ГЕОИНФОРМАЦИОННЫХ СИСТЕМАХ¹

О. В. Пашковская*, Д. В. Бренинг

Сибирский государственный университет науки и технологий имени академика М. Ф. Решетнева
Российская Федерация, 660037, г. Красноярск, просп. им. газеты «Красноярский рабочий», 31

*E-mail: pashkovskaya@sibsau.ru

Рассматриваются методы анализа данных в геоинформационных системах (ГИС). Предлагается использовать статистические и эконометрические методы для анализа и прогноза в различных «слоях» ГИС.

Ключевые слова: геоинформационные системы (ГИС), анализ данных, пространственные данные.

DATA ANALYSIS METHODS IN GEOGRAPHIC INFORMATION SYSTEMS

O. V. Pashkovskaya*, D. V. Brening

Reshetnev Siberian State University of Science and Technology
31, Krasnoyarskii rabochii prospekt, Krasnoyarsk, 660037, Russian Federation

*E-mail: pashkovskaya@sibsau.ru

The methods of data analysis in geographic information systems (GIS) are considered. It is proposed to use statistical and econometric methods for analysis and forecasting in various "layers" of GIS.

Keywords: geographic information systems (GIS), data analysis, cross-section data.

Геоинформационная система (географическая информационная система, ГИС) — система сбора, хранения, анализа и графической визуализации пространственных (географических) данных и связанной с ними информации о необходимых объектах. Пространственные данные (географические данные, геоданные) — данные о пространственных объектах и их наборах. Пространственные данные составляют основу информационного обеспечения геоинформационных систем

Любая современная ГИС содержит в себе набор средств для анализа пространственно-атрибутивной информации.

Пространственные данные обычно состоят из двух взаимосвязанных частей: координатных и атрибутивных данных.

Координатные данные определяют позиционные характеристики пространственного объекта. Они описывают его местоположение в установленной системе координат.

Атрибутивные данные представляют собой совокупность непозиционных характеристик (атрибутов) пространственного объекта. Атрибутивные данные определяют смысловое содержание (семантику) объекта и могут содержать качественные или количественные значения.

Данные в геоинформационных системах описывают, как правило, реальные объекты, такие как дороги, здания, водоемы, лесные массивы. Реальные объекты можно разделить на две абстрактные категории: дискретные (дома, территориальные зоны) и непрерывные

¹ Работа выполнена с использованием гранта президента Российской Федерации на развитие гражданского общества, предоставленного фондом президентских грантов (договор № 18-2-017584)

(рельеф, уровень осадков, среднегодовая температура). Для представления этих двух категорий объектов используются векторные и растровые данные.

Растровые данные хранятся в виде наборов величин, упорядоченных в форме прямоугольной сетки. Ячейки этой сетки называются пикселями. Наиболее распространенным способом получения растровых данных о поверхности Земли является дистанционное зондирование, проводимое при помощи спутников и БПЛА. Хранение растровых данных может осуществляться в графических форматах, например TIFF или JPEG.

Векторные данные обычно имеют намного меньший размер, чем растровые. Их легко трансформировать и проводить над ними бинарные операции. Векторные данные позволяют проводить различные типы пространственного анализа, к примеру поиск кратчайшего пути в дорожной сети. Наиболее распространёнными типами векторных объектов являются точки, полилинии (ломаные), полигоны (многоугольники).

Для анализа данных в ГИС используют различные методы.

При обработке цифровых векторных слоев с учетом атрибутов геообъектов использует векторный анализ: элементарный пространственный анализ; пространственная статистика; расширенный пространственный анализ; сетевой анализ.

При обработке цифровых растровых слоев с учетом их атрибутов, используется растровый анализ: интерполяция раstra; анализ поверхностей; картирование плотности; картирование расстояний; использование функции картографической растровой алгебры.

Среди операций пространственной статистики в ГИС выделяют: статистическую обработку атрибутов, описательную статистику выборки, работу с базами атрибутивной информации, а также расширенные операции пространственной статистики. Статистическая обработка атрибутов позволяет выполнять расчет значений новых атрибутивных полей на основе существующих. Функции описательной статистики выборки дают возможность рассчитывать основные статистические показатели (максимум, минимум, среднее и т. д.) массива атрибутивных данных векторных объектов.

Целью геоинформационного анализа и прогнозирования является снижение уровня неопределённости при принятии решений. Области применения такого анализа чрезвычайно разнообразны: от лесоведения до медицины и энергетики [1-5]. Несомненно, в подавляющем большинстве случаев технологии географических информационных систем применяются для создания карт. Но, наряду с геоинформационными методами, ГИС позволяет использовать и традиционные методы регрессионного анализа [6] при наличии пространственных факторов [7-8]. В ГИС разработаны специальные методы регрессионного анализа, которые акцентированы именно на особенности пространственных данных. Они предназначены для улучшения моделирования взаимоотношений именно таких данных. Одни пространственные регрессионные методы наиболее эффективно учитывают пространственную автокорреляцию, другие – географическое непостоянство явлений.

При построении пространственной регрессионной модели необходимо учитывать особенности пространственных данных. Во-первых, географические данные часто автокоррелированы, что приводит к искажению оценки влияния фактора. Во-вторых, для пространственных данных в первую очередь важна именно их география.

Часто самыми важными для модели являются непостоянные процессы; эти процессы протекают по-разному на различных участках. Эти особенности могут быть описаны как территориальные изменения или пространственный дрейф. Поэтому в ГИС разработаны специальные методы регрессионного анализа, которые акцентированы именно на эти две особенности пространственных данных [9]. Они предназначены для улучшения моделирования взаимоотношений именно таких данных. Одни пространственные регрессионные методы наиболее эффективно учитывают пространственную автокорреляцию, другие – географическое непостоянство явлений.

В настоящее время не существует пространственных регрессионных методов, которые были бы эффективны сразу для всех особенностей пространственных данных. Поэтому при

адаптации методов регрессионного анализа данных для ГИС необходима дополнительное теоретическое обоснование этих методов. Возможно, решение этой проблемы будет получено при дополнительном использовании кластерного или факторного анализа [10].

Библиографические ссылки

1. Барановский Н.В. Моделирование и прогноз лесной пожарной опасности // *Аэрокосмические методы и геоинформационные технологии в лесоведении, лесном хозяйстве и экологии* : материалу VII Всероссийской конференции (Москва, 22-24 апреля 2019 г.) – Москва: ЦЭПЛ РАН, 2019 г. С. 4-5.
2. ВIKANова А.А. Методы прогнозирования в геоинформационных системах // *Научное сообщество студентов XXI столетия. Технические науки*: сб. ст. по мат. IV междунар. студ. науч.-практ. конф. № 4. URL: <http://sibac.info/archive/technic/4.pdf> (дата обращения 22.04.20).
3. Зависимость частоты пожаров в Ильменском заповеднике от типа леса и доступности территории / Пустовалова Л.А., Чашина О.Е., Куянцева Н.Б. и др. // *Аэрокосмические методы и геоинформационные технологии в лесоведении, лесном хозяйстве и экологии: доклады VII всероссийской конференции* (Москва, 22-24 апреля 2019 г.) – Москва: ЦЭПЛ РАН, 2019 г. С. 155-157. URL: http://www.cepl.rssi.ru/confs/ASGIS2019/files/ASGIS2019_Abstract_book.pdf (дата обращения 22.04.20).
4. Самохина А.В., Овчаренко А.Н. ГИС на службе энергетиков // *ГИС для управления инфраструктурой ArcReview* | № 1 (92) за 2020 год. URL: https://www.esri-cis.ru/news/arcreview/detail.php?ID=28172&SECTION_ID=1125 (дата обращения 23.04.20).
5. Струков Д.Р., Мельник М.А. Геостатистические методы при выявлении причинно-следственных связей между факторами и откликами на территории города // *ГИС в здравоохранении и медицине ArcReview* | № 1 (60) за 2012 год. URL: https://www.esri-cis.ru/news/arcreview/detail.php?ID=6659&SECTION_ID=221 (дата обращения 23.04.20).
6. Математическое моделирование стоимости вторичного жилья в городе Красноярске в 2017 году / Д.В. Бренинг, О.В. Пашковская, С.И. Сенашов, И.Л. Савостьянова // *Сибирский журнал науки и технологий*. 2018. №4 (19). С.581–588.
7. Серебрянная О.Л. Регрессионный анализ пространственных данных помогает ответить на вопрос «ПОЧЕМУ?» // *Профессиональные ГИС-услуги: Проектная деятельность компании DATA+*. ArcGIS 10.1 | № 4 (63) за 2012 г. URL: https://www.esri-cis.ru/news/arcreview/detail.php?ID=8329&SECTION_ID=265. (дата обращения 23.04.20).
8. Сравнение возможностей интерполяционных модулей QGIS для морских климатических исследований при работе с массивом данных малой обеспеченности / Новикова А.М., Полонский А.Б., Новиков А.А. // *Геоинформационное обеспечение устойчивого развития территорий в условиях глобальных изменений климата* : материалы междунар. науч. конф. ИнтерКарто-ИнтерГИС. Т.1, 2016. Москва: НАУЧНАЯ БИБЛИОТЕКА. С. 76-88. URL: <https://istina.msu.ru/> (дата обращения 24.04.20).
9. Митчелл Энди. Руководство по ГИС анализу. Часть 1: Пространственные модели и взаимосвязи / ESRI, 2000. URL: https://www.studmed.ru/view/mitchell-endi-rukovodstvo-po-gis-analizu-chast-1-prostranstvennyye-modeli-i-vzaimosvyazi_c0632c90fd5.html (дата обращения 24.04.20).
10. Пашковская О.В., Коршакевич И.С. Использование факторного анализа в информационных медицинских системах // *Актуальные проблемы авиации и космонавтики [Электронный ресурс]* : сб. материалов XIII Междунар. науч.-практ. конф., посвящ. Дню космонавтики (10-14 апреля 2017 г., Красноярск) : в 3 т. Т. 2. С. 363-364. URL: <https://disk.sibsau.ru/index.php/s/OCqTVFnwoa2ENp8> (дата обращения 24.04.20).