DOI: 10.17117/no.2015.04.020

Поступила (Received): 16.06.2015

<http://ucom.ru/doc/no.2015.04.020.pdf>

Kafedra_pi@rshu.ru

**Истомин Е.П., Колбина О.Н.,
Степанов С.Ю., Сидоренко А.Ю.
Модели обработки гетерогенных данных
с помощью ГИС программ для решения
задач оценки вероятностных событий**

**Istomin E.P., Kolbina O.N., Stepanov S.Yu., Sidorenko A.Yu.
Model the processing of heterogeneous data using GIS
software to solve problems evaluation of probabilistic events**

В статье рассматриваются некоторые аналитические модели и системы обработки гетерогенных данных. Приводится описание предложенных технологий и программных решений, направленных на совместное использование гетерогенной информации из различных источников, в целях совместного анализа разнородных данных и решения прикладных задач. Рассматривается частный случай решения задачи оценки риска наводнения
Ключевые слова: модель, гетерогенные данные, обработка данных, разноформатные данные, геоинформационная система, пространственная информация, риск, вероятность, событие

This article discusses some of the analytical models and system of processing heterogeneous data. The description of the proposed technologies and software solutions aimed at the sharing of heterogeneous information from various sources for joint analysis of heterogeneous data and decisions applied problems. The particular case of the solution of the flood risk assessment

Key words: model, heterogeneous data, data processing, multi-format data, geographic information system, spatial information, risk, probability, events

Истомин Евгений Петрович

Доктор технических наук, профессор, декан
Российский государственный
гидрометеорологический университет
г. Санкт-Петербург, Малоохтинский проспект, 98

Istomin Eugeny Petrovich

Doctor of Engineering Sciences, Professor, Dean
Russian state hydrometeorological university
Saint-Petersburg, Malookhtinsky ave., 98

Колбина Ольга Николаевна

Старший преподаватель
Российский государственный
гидрометеорологический университет
г. Санкт-Петербург, Малоохтинский проспект, 98

Kolbina Olga Nikolaevna

Senior Lecturer
Russian state hydrometeorological university
Saint-Petersburg, Malookhtinsky ave., 98

Степанов Сергей Юрьевич

Старший преподаватель
Российский государственный
гидрометеорологический университет
г. Санкт-Петербург, Малоохтинский проспект, 98

Stepanov Sergey Yurievich

Senior Lecturer
Russian state hydrometeorological university
Saint-Petersburg, Malookhtinsky ave., 98

Сидоренко Артем Юсупович

Старший преподаватель

Российский государственный

гидрометеорологический университет

г. Санкт-Петербург, Малоохтинский проспект, 98

Sidorenko Artem Yusupovich

Senior Lecturer

Russian state hydrometeorological university

Saint-Petersburg, Malookhtinsky ave., 98

Введение (Introduction)

На сегодняшний день количество опасных всевозможных стихийных бедствий постоянно возрастает. Важной задачей для своевременного принятия решений по урегулированию возникших чрезвычайных ситуаций является предоставление геопространственных данных в рамках всего цикла развития стихийного бедствия. Основные данные включают в себя данные моделирования, мониторинга и прогнозирования соответствующих рисков, построение карт, демонстрирующих вероятность наступления события и ущерб, который может быть им причинен [1].

Наиболее качественные модели оценки рисков могут быть построены при использовании различных и всевозможных геопространственных данных, имея в своем наборе обилие статистической информации, такая ГИС система обеспечит своевременным актуальным анализом все государственные службы и органы власти.

Анализ геопространственных данных зависит в первую очередь от консолидации данных. Основной задачей технологии слияния данных является объединение данных из разных источников в интересах принятия управленческих решений, анализе состояния объектов, прогнозирования ситуаций. Интеграция гетерогенных данных в единую информационную среду обеспечивает возможность их комплексного анализа и позволяет получить качественно новые знания об объекте исследования. Такая информационная среда может объединять, следующие типы информационных ресурсов: геоинформационные ресурсы (картографические материалы, спутниковые снимки, данные наблюдений и т. п.); электронные каталоги; цифровые изображения, аудио и видео материалы; другие ресурсы [2,3].

В работе рассматривается модель обработки гетерогенных данных для решения частной прикладной задачи:

– Модель оценки риска наводнений на основе разнородных геопространственных данных. Модель оценки риска затопления в Намибии на основе разнородной геопространственной информации.

Материалы и методы (Materials and Methods)

Для достижения цели исследования вышеописанной модели оценки риска наводнений на основе разнородных геопространственных данных и модели оценки риска затопления в Намибии на основе разнородной геопространственной информации, в работе были использованы общенаучные и специальные методы.

Стадии и этапы исследования:

1. Проведение анализа исследуемой модели.
2. Математическая модель оценки рисков.

3. Исследование метода и технологии слияния данных.
4. Определение используемых источников разнородных данных.
5. Анализ внутренней инфраструктуры гео модели. Уровни обработки данных.
6. Подведение результатов исследования.

Результаты (Results)

Создание системы оценки рисков на основе разнородной геопространственной информации выполнены в рамках международного пилотного проекта «Sensor Web Project for Flood Monitoring in Namibia», выполняемого по инициативе программы UN-SPIDER ведущими космическими агентствами: NASA, NOAA, DLR при участии Института космических исследований НАНУ- НККУ, представлены в отчете [7].

Основные задачи системы:

- Интегрированное использование спутниковых данных, наземных измерений и данных моделирования для мониторинга и прогнозирования наводнений.
- Моделирование и прогнозирование распространения заболеваний, связанных с водой.
- Оценка рисков на основе разнородной геопространственной информации и метод ее решения.

Для оценки вероятности риска стихийного бедствия необходимо анализировать информацию, поступающую из различных источников с различным временным и пространственным разрешением. Для совместного анализа такой информации применяются методы и технологии слияния данных data fusion [4].

Рис. 1. Схема оценивания плотности распределения вероятности стихийного бедствия p_{xy} на основе разнородной информации

Необходимо подчеркнуть, что плотность вероятности стихийного бедствия должна оцениваться на основе слияния преимущественно не первичных данных, а информации более высоких уровней обработки. Для оценивания плотности вероятности стихийного бедствия предлагается следующая обобщенная схема (рис. 1).

На рис. 1 блоки $DS_i, i=1,..., n$ представляют различные источники данных – спутниковые данные, данные наземных сенсоров и данные моделирования. Блоки $C_i, i=1,..., m$ обеспечивают обработку данных и получение информации более высокого уровня. В общем случае число блоков обработки не равно количеству источников данных, поскольку данные из одного источника могут обрабатываться разными обработчиками и, в то же время, один обработчик может анализировать данные из разных источников одновременно. То есть слияние данных происходит уже на стадии их преобразования обработчиками. Вид преобразования в каждом из блоков $C_i, i=1,..., m$ определяется типом входных данных. Например, спутниковые данные проходят несколько уровней предварительной обработки (перепроектирование, геопривязка, коррекция и т.д.), а затем подлежат так называемой «тематической» обработке [5].

Рис. 2. Схема определения вероятности затопления для территории Намибии

Для оценки риска затоплений для Намибии в рамках совместного проекта платформы ООН UN-SPIDER, NASA, DLR, NOAA и Института космических исследований НАНУ-НКАУ используются следующие разнородные источники данных (рис. 2):

- спутниковые данные: радиолокационные: Envisat/ASAR; оптические: EO-1, MODIS (Terra и Aqua); TRMM;
- данные моделирования: метеорологические данные (численная модель прогнозирования погоды); гидрологические данные (водосборный бассейн);
- наземные данные: уровень осадков и сток;
- статистические данные: статистические данные о наводнениях за прошлые годы.

Рассмотрим детальнее эти наборы данных и приведем их примеры (рис. 2). Картографирование наводнений с использованием спутниковых данных. Для создания карт затоплений используются как радиолокационные, так и оптические спутниковые данные. Преимущество использования данных, полученных в микроволновом диапазоне электро-магнитного излучения, состоит в невазависимости от погодных условий, в частности облачного покрова [6]. Для создания маски затоплений применяются интеллектуальные вычисления.

На сегодняшний день доступными для обработки являются следующие данные:

- Envisat/ASAR (в рамках гранта Европейского космического агентства): среднее пространственное разрешение (150 м), продукты обработки поставляются в течение 24 ч. после получения данных; высокое пространственное разрешение (30 м): продукты поставляются по запросу

- RADARSAT-2 (по запросу в рамках Международной Хартии “Space and Major Disasters” или Рабочей группы по чрезвычайным ситуациям комитета GEO): высокое пространственное разрешение (от 3 до 30 м).

- В случае отсутствия облачного дополнительно используются следующие оптические данные:

- Envisat/MERIS: среднее пространственное разрешение (300 м);
- Terra и Aqua/MODIS: среднее пространственное разрешение (250 м – 1 км);
- NASA EO-1: высокое пространственное разрешение (30 м).

Продукты обработки поставляются в форматах KML (для визуализации в Google Earth), GeoTiff, WMS и др. Для мониторинга уровня осадков используются измерения, получаемые в рамках совместной миссии NASA и JAXA Tropical Rainfall Measuring Mission (TRMM). Для метеорологического прогнозирования используется численная модель прогнозирования погоды WRF. Специалисты ИКИ НАНУ-НКАУ имеют значительный опыт настройки, адаптации и использования данной модели для территории Украины. Для территории Намибии данная модель была запущена в тестовом режиме. Пространственное разрешение составляет 30 км (может быть улучшено до 1 км с использованием технологии вложенных сеток). Для начальных условий используются данные глобальной модели GFS. Прогнозы предоставляются каждые 3 часа. Информация о водосборном бассейне рек предоставляется подразделениями Министерства по сельскому хозяйству, водным ресурсам и лесному хозяйству Намибии. Кроме

того, для каждого бассейна предоставляются текущие и архивные значения уровня осадков и стока. Эти архивные данные, наряду со спутниковыми измерениями используются для построения статистических моделей для прогнозирования наводнений и оценки рисков. Данные полученные прибором MODIS, который установлен на спутниках Aqua и Terra, используется для статистического анализа карт затоплений за разные года. Данные карты доступны с 1999 года и предоставляются центром Dartmouth Flood Observatory. Для глобального детектирования наводнений используются система, разрабатываемая в центре JRC (Joint Research Center) Европейской комиссии. Для этого используются данные, получаемые микроволновым сканирующим радиометром AMSR-E, который установлен на спутнике Aqua [7].

Обсуждение и заключение (Discussion and Conclusion)

В работе предложен единый подход к оцениванию вероятностных событий, связанных с чрезвычайными ситуациями природного характера на основе разнородной геопространственной информации. В основе подхода лежит применение статистической модели и предполагает использование классификаторов для определения плотности вероятности стихийного бедствия. Несомненным преимуществом данного подхода является более высокая точность определения риска при использовании оптимальной сложности модели [7,8]. Такой подход можно применять и в других категориях рисков: финансового, социального, экономического и т.д.

Во многих задачах устойчивого развития территорий, необходимо одновременно обрабатывать аэрокосмические данные, данные наземных измерений, разнообразную текстовую информацию и т.д. Это накладывает существенные особенности на все процессы обработки данных и применения прикладных инструментов, используемых для этого [10,11].

Подводя итоги по вышеописанной модели, можно сделать вывод, что качественный анализ требуют привлечения большого количества разнообразной информации, получаемой из различных источников, в наибольшей степени используются геопространственные данные, необходимы также и другие данные, без которых решение задач невозможно или недостоверно. Комплексное использование такой информации связано с большими проблемами, обусловленными, прежде всего, гетерогенностью данных различной природы. Использование геопространственных данных связано с необходимостью постоянно интегрировать эти данные с геопространственными данными другой природы. При этом большая часть проблем связана с реализацией механизмов интегрирования и преодоления гетерогенности данных [9,12].

В данной работе были рассмотрены общие структуры моделей слияния гетерогенных данных для решения прикладных задач, в частном случае оценка вероятностных событий, разработанные данные могут быть взяты за основу при создании подобных моделей в других областях применения пространственных данных.

Список используемых источников:

1. Колбина О.Н., Зоринова Е.М., Истомин Е.П., Степанов С.Ю. Обзор технологических методов доступа к базам данных распределенных ГИС // Информационные технологии и системы: управление, экономика, транспорт, право. Вып. 3(14). СПб: Андреевский издательский дом, 2014. 116 с.
2. Степанов С.Ю., Гришин Н.М., Колбина О.Н., Соловьева А.Д. ГИС. Обзор программного обеспечения // Информационные технологии и системы: управление, экономика, транспорт, право. Вып. 3(14). СПб: Андреевский издательский дом, 2014. 116 с.
3. Управление федеральной службы государственной регистрации, кадастра и картографии – информационный портал.
URL: http://www.to33.rosreestr.ru/upload/to33/files/Концепция_ситуационного_центра_ABТ
4. Жижимов О.Л., Молородов Ю.И., Пестунов И.А., Смирнов В.В., Федотов А.М. Интеграция разнородных данных в задачах исследования природных экосистем // Вестник НГУ. 2011. Т. 9. Вып. 1.
5. Кашников А.В. Интеграция гетерогенных источников данных на основе рекурсивной декомпозиции // International Journal "Information Technologies & Knowledge". Vol. 5. № 3. 2011.
6. Ильин Е.В. Технологическая платформа для распределенных гетерогенных систем сбора и обработки данных // Студент и научно-технический прогресс: Информационные технологии. Новосибирск, 2012.
7. Скакун С.В. Оценка риска наводнений на основе разнородных геопространственных данных. // Наукові праці ДонНТУ. Вип. 12(165). 2010.
8. Степанов С.Ю. Сравнительный анализ открытых геоинформационных систем // Информационные технологии и системы: управление, экономика, транспорт, право. Вып. 10. СПб: Андреевский издательский дом. 2013. 156 с.
9. Котиков П.Е. Варианты построения темпоральных баз данных в геоинформационных системах // Научный аспект. № 4. 2014. Самара: Аспект, 2014. №4. С. 118-120.
10. Котиков П.Е., Муратова О.Н. Существующие сложности перехода к применению пространств данных в геоинформационных системах // Актуальные проблемы технических наук. Уфа: Аэтерна, 2015. С. 42-44.
11. Котиков П.Е. Некоторые решения технологических проблем хранения пространственных данных ГИС // Научный аспект № 2-2015. Аспект, 2014. №2. Т. 2. С. 212-214.
12. Котиков П.Е., Муратова О.Н. Обзор вариантов архитектурных решений построения распределенных и параллельных баз данных для геоинформационных систем // Закономерности и тенденции развития науки. Уфа: Аэтерна, 2015. С. 52-54.

References:

1. Kolbin O.N., Zharinova E.M., Istomin E.P., Stepanov S.Yu. Review of technological methods of access to databases distributed GIS. Information technology and systems: management, economy, transport, right. Vol. 3(14). SPb: Andriivskiy, 2014. 116 p.
2. Stepanov S.Y., Grishin N.M., Kolbina O.N., Soloveva A.D GIS. Software review. Information technology and systems: management, economy, transport, right. Vol. 3(14). SPb: Andriivskiy, 2014. 116 p.
3. The Federal service of state registration, cadastre and cartography and information portal.
URL: http://www.to33.rosreestr.ru/upload/to33/files/Концепция_ситуационного_центра_ABТ
4. Zhizhimov O.L., Molorodov I., Pestunov I.A., Smirnov V.V., Fedotov A.M. Integration of heterogeneous data in study of natural ecosystems. Vestnik NSU. 2011. Vol. 9. Issue 1.
5. Kashnikov V.A. Integration of heterogeneous data sources based on a recursive decomposition. Information Technologies & Knowledge. Vol. 5. № 3. 2011.
6. Ilyin E.V. Technology platform for distributed heterogeneous systems of data collection and processing. Student and scientific-technical progress: Information technologies. Novosibirsk. 2012.
7. Skakun S. Flood risk Assessment based on heterogeneous geospatial data. Naukow Pratsi DonNTU. Issue 12(165). 2010.
8. Stepanov S.Yu. Comparative analysis of open geographic information systems. Information technology systems: management, economy, transport, right. Vol. 10. St. Petersburg: Andrew Izdatelsky house. 2013. 156 p.
9. Kotikov P.E. Options for the design of temporal databases in geoinformacionnyh systems. Scientific aspect №4. 2014. Samara: Aspect, 2014. №4. P. 118-120.
10. Kotikov P.E., Muratova O.N. The Existing difficulties of the transition to the use of data spaces in geographic information systems. Actual problems of technical Sciences : materials of International scientific-practical conference. Ufa: Aeterna, 2015. P. 42-44.

11. Kotikov P.E. *Some of the technological problems of spatial data storage GIS. Scientific aspect.* №2. 2015. Samara: Aspect, 2014. №2. Vol. 2. P. 212-214.
12. Kotikov P.E., Muratova O.N. *Overview of architectural solutions for building distributed and parallel data bases for geographic information systems. Regularities and tendencies of development of science.* Ufa: Aeterna, 2015. P. 52-54.

© 2015, Истомин Е.П., Колбина О.Н., Степанов С.Ю., Сидоренко А.Ю.

Модели обработки гетерогенных данных с помощью ГИС программ для решения задач оценки вероятностных событий

© 2015, Istomin E.P., Kolbina O.N., Stepanov S.Yu., Sidorenko A.Yu.

Model the processing of heterogeneous data using GIS software to solve problems evaluation of probabilistic events