

УДК 681.512:00.467:519.246.8

Я. И. Зельк, Н. Н. Куусуль, С. В. Скакун, А. Ю. Шелестов

Институт космічних досліджень Національної академії наук України
та Національного космічного агентства України, Київ

ОЦЕНКА РИСКОВ СТИХИЙНЫХ БЕДСТВИЙ НА ОСНОВЕ АНСАМБЛЕВОЙ ОБРАБОТКИ И ТЕХНОЛОГИИ СЛИЯНИЯ РАЗНОРОДНЫХ ГЕОПРОСТРАНСТВЕННЫХ ДАННЫХ

Поставлена задача оцінки ризиків, пов'язаних з наслідками стихійних лих на основі різномірної геопросторової інформації (даних супутникових, наземних спостережень і даних моделювання) і запропоновано метод її розв'язання. Основною методу є ансамблеве оброблення і технологія злиття гетерогенних даних (data fusion) стосовно оцінювання за вибіркою невідомої густини ймовірності стихійного лиха, що залежить від скінченної кількості параметрів. Проаналізовано джерела різномірних геопросторових даних, що використовуються у створюваному операційному сервісі картографування ризиків затоплень для території Намібії. Розглянуто принципову схему системи оцінювання густини ймовірності для визначення ризику затоплень для території Намібії, побудовану у відповідності до запропонованого методу. На продовження робіт за міжнародним пілотним проектом «Sensor Web Project for Flood Monitoring in Namibia» співробітниками ІКД НАНУ — НККАУ буде створений операційний сервіс картографування ризиків затоплень з використанням сучасних Internet та ГІС технологій, який буде відповідати міжнародним стандартам консорціуму OGC (Open Geospatial Consortium) надання геопросторової інформації.

При прийнятті рішень, касаючихся зашити от воздействия стихийных бедствий на население, социально-экономические объекты и среду, возникает проблема оценивания рисков неблагоприятных последствий от воздействия различных факторов окружающей среды на значительных территориях. Исходная информация для решения проблемы оценивания такого рода рисков содержится в разнородных геопространственных данных (спутниковых, наземных данных моделирования), касающихся конкретного региона, различных характеристик территории, объектов и факторов внешней среды. Мера риска пропорциональна ожидаемым потерям, которые могут быть причинены рисковым событием, и вероятности этого события. В настоящее время в мире есть системы оценки рисков в операционном режиме. Однако используемые в таких операционных системах методы оценки рисков зачастую являются слишком упрощенными. Они не опираются на достаточно развитый матема-

тический аппарат оценки среднего риска по эмпирическим данным, разработанный для задач оценивания качества восстановления функциональных зависимостей на основе эмпирических данных, который применяется в статистической теории обучения [1, 2, 7].

В работе [4] содержится детальный обзор более 25 известных метрик для определения рисков отдельного индивидуума, а также социальных, экономических, рисков для окружающей среды и других, связанных со стихийными бедствиями. В общем случае риск описывается в виде функции вероятности наступления неблагоприятного события и приносимого этим событием ожидаемого ущерба. Во многих странах мира такие метрики оценки рисков стали государственными и отраслевыми стандартами. Они используются при создании информационных продуктов и сервисов для заинтересованных государственных органов предотвращения чрезвычайных ситуаций, органов планирования развития урбанизированных территорий, имущественного и коммунального сектора, страховых компаний. В Украине проблема оценки рисков стихийных

бедствий еще далека от ее постановки на государственном уровне.

В настоящей работе поставлена задача оценки рисков, связанных со стихийными бедствиями на основе разнородной геопространственной информации, предложен и обоснован метод ее решения и приведена принципиальная схема системы оценивания плотности вероятности для определения риска затоплений для территории Намибии, построенная в соответствии с предложенным методом.

Совокупным ожидаемым риском последствий стихийного бедствия (совокупными ожидаемыми потерями) в области A назовем величину

$$R_A = \iint_A r(x, y) dx dy,$$

где $r(x, y)$ — локальный ожидаемый риск последствий стихийного бедствия z (локальные ожидаемые потери) в точке (x, y) , вычисляемый как математическое ожидание функции ущерба $h_{x,y}(z)$ последствий в позиции (x, y) , т. е. как значение функционала среднего риска [1, 2, 7]

$$r(x, y) = \int_0^{\infty} h_{x,y}(z) p_{x,y}(z) dz,$$

где $p_{x,y}(z)$ — плотность распределения вероятности стихийного бедствия z в точке (x, y) , оцениваемая на основе совместного анализа разнородных геопространственных данных. Незвестная плотность вероятностей стихийного бедствия в точке зависит от различных факторов окружающей среды, которые могут быть прямо или косвенно измерены с помощью наземных средств и дистанционных методов, либо получены с помощью моделирования.

Задача восстановления плотности вероятности в классе непрерывных функций сводится к некорректной задаче численного дифференцирования функции распределения вероятностей [7]. Ее можно решать непараметрическими методами (например, метод Парзена, метод упорядоченной минимизации риска с использованием ковариационной матрицы коррелированных ошибок измерения), которые учитывают некорректность задачи и опираются на статистическую теорию регуляризации. Однако в случаях, когда имеется априорная информация об иско-

мой плотности вероятности, можно избежать некорректной постановки этой задачи. В частности, если восстанавливаемая плотность вероятности известна с точностью до конечного числа параметров (что имеет место в нашем случае), задача ее восстановления по эмпирическим данным корректна, и для ее решения можно использовать эффективные методы параметрической статистики. Так, в классе задач минимизации среднего риска, связанном с классификацией (обучением распознаванию образов), восстановление неизвестных параметров плотности распределения вероятности $p_{x,y}(z)$ возможно различными методами параметрической статистики (в зависимости от контекста задачи): методом максимального правдоподобия, методом наилучших несмещенных приближений, методом Байесовых приближений. Кроме того, можно восстанавливать параметризованную плотность вероятности стихийного бедствия методами статистической теории обучения — нейронные двухслойные сети; ядерные методы — радиальных базисных функций (RBF), метод машин опорных векторов (SVM) и др. [1, 2, 7].

Для оценки плотности вероятности стихийного бедствия необходимо анализировать (классифицировать) информацию, поступающую из различных источников с различным временным и пространственным разрешением. Совместный анализ такой информации выполняется с применением ансамблевого подхода [3] и реализацией методов и технологий слияния данных (data fusion) [5, 6].

Предлагаемая система для оценивания плотности вероятности стихийного бедствия $p_{x,y}(z)$ состоит из ансамбля отдельных («слабых») классификаторов (решающих правил), каждый из которых обеспечивает результат анализа данных одного или нескольких источников. Агрегированные в систему ансамблевой обработки данных, отдельные классификаторы составляют ансамбль экспертов («сильный» классификатор), мнения которых с соответствующими весами учитываются в блоке слияния, на выходе которого в результате реализации машинного обучения получается с высокой точностью оценка плотности вероятности стихийного бедствия.

Преимуществом ансамблевого метода классификации является возможность повышения точности классификации за счет приема «усиления» (boosting), который сводится к оценке функции потерь (ошибки классификации) и минимизации этой функции путем добавления новых компонентных классификаторов до тех пор, пока очередное добавление не перестанет снижать значение функции потерь.

Работы по созданию системы оценки рисков на основе разнородной геопространственной информации проводятся в рамках международного пилотного проекта «Sensor Web Project for Flood Monitoring in Namibia», выполняемого по инициативе платформы ООН UN-SPIDER ведущими космическими агентствами: NASA, NOAA, DLR при участии Института космических исследований НАН Украины и НКА Украины (ИКИ НАНУ-НКАУ). Для оценки риска затоплений для территории Намибии используются следующие разнородные источники данных.

1. Спутниковые данные:

- радиолокационные: Envisat/ASAR — среднего пространственного разрешения (150 м) (продукты обработки поставляются в течение 24 ч после получения данных), высокого пространственного разрешения (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) — высокого пространственного разрешения (от 3 до 30 м);

- оптические: Envisat/MERIS — среднего пространственного разрешения (300 м), Terra и Aqua/MODIS — среднего пространственного разрешения (250 м — 1 км), EO-1 (NASA) — высокого пространственного разрешения (30 м);

- данные мониторинга уровня осадков, получаемые со спутников в рамках совместной миссии NASA и JAXA Tropical Rainfall Measuring Mission (TRMM), оснащенных радаром, приемниками микроволнового диапазона, видимого и инфракрасного диапазона.

2. Данные математического моделирования:

- метеорологические, полученные с использованием численной мезомасштабной модели прогнозирования погоды WRF — с простран-

ственным разрешением 30 км (может быть улучшено до 1 км); для задания начальных условий используется глобальная модель (система) прогнозирования GFS;

- гидрологические — о водосборном бассейне рек, предоставляемые Министерством сельского хозяйства, водных ресурсов и сельского хозяйства Намибии.

3. Данные наземных наблюдений о текущих и архивных значениях уровня осадков и уровня стока для каждого водосборного бассейна.

4. Статистические данные об уровне осадков, уровне стока, данные спутников Terra и Aqua/MODIS для статистического анализа карт затоплений за предыдущие годы, начиная с 1999 г., предоставляемые Dartmouth Flood Observatory.

Продукты обработки поставляются в форматах KML (для визуализации в Google Earth), GeoTiff, WMS и др.

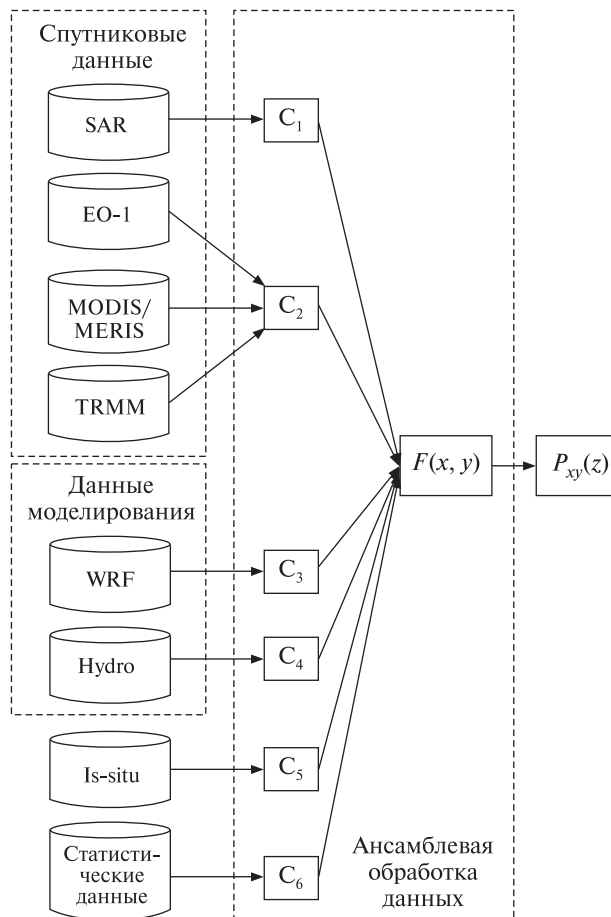
Принципиальная схема системы оценивания плотности вероятности для определения риска затоплений для территории Намибии представлена на рисунке. Для ансамблевой обработки спутниковых данных используются нейросетевой (C_1) и пороговый (C_2) классификаторы. Для обработки данных математического моделирования — предсказатель C_3 временных рядов количества осадков, пороговый классификатор C_4 водосборного бассейна. Для классификации данных, полученных в месте нахождения (*in-situ*), и разнообразных статистических данных используются пороговый классификатор C_5 количества осадков и уровня стока и классификатор C_6 по методу максимума правдоподобия.

В настоящее время в рамках международного пилотного проекта «Sensor Web Project for Flood Monitoring in Namibia» сотрудниками ИКИ НАНУ-НКАУ осуществляется обработка спутниковых снимков и создание карт потенциальных затоплений на основании слияния разнородных геопространственных данных. Был создан соответствующий Web-интерфейс пилотной версии разрабатываемой системы мониторинга наводнений для территории Намибии. В будущем создаваемая система мониторинга паводков в Намибии будет снабжена операционными сервисом картографирования рисков затоплений,

который будет функционировать на основе предложенного метода.

Работы по созданию системы раннего предупреждения наводнений на базе технологии SensorWeb (Flood SensorWeb Early Warning System) в рамках пилотного проекта, выполняемого при международном партнерстве NASA, платформы ООН UN-SPIDER, Намибийского департамента гидрологии, Канадского космического агентства, Института космических исследований НАН Украины и НКА Украины, Немецкого космического агентства (DRL) и др. демонстрируются сайтом <http://sensorweb.nasa.gov/NamibiaFlood.html>. В продолжение этих работ сотрудниками ИКИ НАНУ-НКАУ будет создан операционный сервис картографирования рисков затоплений с использованием современных интернет- и ГИС-технологий, который будет соответствовать международным стандартам консорциума OGC (Open Geospatial Consortium) предоставления геопространственной информации.

Таким образом, предложен интегрированный подход к оцениванию различных видов риска, связанных с чрезвычайными ситуациями природного характера на основе разнородной геопространственной информации. Подход базируется на применении статистической теории обучения и предполагает использование ансамбля классификаторов для определения плотности вероятности стихийного бедствия. Преимуществом предложенного подхода является более высокая точность определения риска при использовании оптимальной сложности модели ансамблевой обработки данных, чем при использовании каждого из классификаторов определенного типа геопространственных данных в отдельности. Проанализированы источники разнородных геопространственных данных, используемые в создаваемом операционном сервисе картографирования рисков затоплений для территории Намибии. Рассмотрена принципиальная схема системы оценивания плотности вероятности для определения риска затоплений для территории Намибии, построенная в соответствии с предложенным методом. В продолжение работ по международному пилотному проекту «Sensor Web Project for Flood Monitoring



Принципиальная схема системы оценивания плотности вероятности для определения риска затоплений для территории Намибии (см. текст)

in Namibia» сотрудниками ИКИ НАНУ-НКАУ будет создан операционный сервис картографирования рисков затоплений с использованием современных Internet и ГИС технологий, который будет соответствовать международным стандартам консорциума OGC (Open Geospatial Consortium) предоставления геопространственной информации.

1. Bishop C. M. Pattern recognition and machine learning. — New York: Springer, 2006. — 738 p.
2. Haykin S. Neural networks. A comprehensive foundation. — New Jersey: Prentice Hall, 1, 1994. — 768 p.
3. Jaakkola T. Course materials for 6.867 machine learning, Fall 2006. MIT OpenCourseWare, Massachusetts Institute of Technology. — 10 p. (<http://ocw.mit.edu/>).

4. Jonkman S. N., van Gelder P. H. A. J. M., Vrijling J. K. An overview of quantitative risk measures for loss of life and economic damage // J. Hazardous Materials. — 2003. — A99. — P. 1–30.
5. Kussul N., Shelestov A., Skakun S. Grid and sensor web technologies for environmental monitoring // Earth Sci. Informatics. — 2009. — 2, N 1-2. — P. 37–51.
6. Mitchell H. B. Multi-sensor data fusion — An introduction — Berlin, Springer-Verlag, 2007. — 282 p.
7. Vapnik V. Statistical learning theory. — New York: Wiley, 1998. — 740 p.

Надійшла до редакції 17.12.10

Ya. I. Zelyk, N. M. Kussul', S. V. Skakun, A. Yu. Shelestov

NATURAL DISASTER RISK ASSESSMENT BASED ON THE ENSEMBLE PROCESSING AND TECHNOLOGY OF HETEROGENEOUS GEOSPATIAL DATA FUSION

The natural disaster risk assessment problem is stated which is based on heterogeneous geospatial data, namely, satel-

lite data, ground-based observations and simulation data. A method for solving the problem is proposed. The heart of the method is the ensemble data processing and technology of the heterogeneous data fusion with respect to the unknown disaster probability density estimation based on a sample of data. This probability density depends on the finite number of parameters. The sources of heterogeneous geospatial data are analyzed which are used in the developed operational flooding risk mapping service for the territory of Namibia. We consider a conceptual sketch of the probability density estimation system to determine the flooding risk for the territory of Namibia. It is constructed in accordance with the method proposed. To continue the investigation according to the international pilot project «Sensor Web Project for Flood Monitoring in Namibia», the staff of SRI NASU-NSAU will elaborate an operational flood risk mapping service with the use of modern Internet and GIS technologies. The operational service will satisfy the international standards of Open Geospatial Consortium (OGC) to provide geospatial information.