

# Геопространственный анализ рисков на основе слияния данных

Н.Н. Куссуль, Я.И. Зельк, С.В. Скаун, А.Ю. Шелестов

Институт космических исследований НАНУ-НКАУ.  
Проспект Академика Глушкова, 40, корп. 4/1, Киев, Украина,  
03680 [inform@ikd.kiev.ua](mailto:inform@ikd.kiev.ua)

Рассматривается задача оценки рисков, связанных с чрезвычайными ситуациями природного характера на основе анализа разнородной геопространственной информации (спутниковых и наземных данных, а также данных моделирования). Предлагается формализованная постановка задачи, обосновывается метод ее решения и рассматривается пример его практического применения для оценки риска затоплений в Намибии. Основой метода является ансамблевый подход к анализу разнородных данных с использованием технологий слияния данных (data fusion) и оценивание с его помощью плотности распределения вероятности стихийного бедствия.

**Ключевые слова:** Оценка рисков, стихийные бедствия, геопространственные данные, минимизация среднего риска, функция плотности вероятности, статистическая теория обучения, принцип максимального правдоподобия, слияние данных, ДЗЗ, ансамблевая обработка данных, нейросетевой классификатор.

## Введение

Изменения климата становятся причиной многочисленных стихийных бедствий: паводков, засух, обильных снегопадов, лесных пожаров и т.п., приносящих огромный ущерб экономике отдельных стран и целых регионов. В последние годы для мониторинга стихийных бедствий все чаще используют геопространственные данные разной природы: аэрокосмические снимки и продукты (например, цифровую модель рельефа, карты землепользования), а также данные двумерного или трехмерного моделирования (в частности, метеорологические или гидрологические модели). Результатом мониторинга при использовании такого рода информации становятся цифровые карты или многослойные геопространственные данные, существенно облегчающие процесс принятия решения соответствующими органами. Подобная информация может использоваться не только для картографирования районов стихийных бедствий во время или после самого события, но и на других этапах цикла развития стихийного бедствия (disaster cycle) – в том числе для построения карт рисков, иллюстрирующих вероятность наступления события и ущерб, который может быть им причинен.

Проблеме создания операционных сервисов оценки рисков стихийных бедствий в Европе посвящен проект SAFER ([http://www.emergencyresponse.eu/site/FO/scripts/myFO\\_accueil.php?lang=EN](http://www.emergencyresponse.eu/site/FO/scripts/myFO_accueil.php?lang=EN)) программы GMES (Global Monitoring for Environment and Security. В отношении создания информационных сервисов для мониторинга, картографирования, моделирования рисков затоплений и оценке вызванных ими потерь аналогом создаваемой системы можно указать разработанные в международных стандартах оценки рисков информационные продукты и сервисы компании Ambiental (Великобритания) (<http://www.ambiental.co.uk/>). Ambiental является мировым лидером в области разработки детализированного картографирования рисков затоплений и их моделирования в интеграции с социо-экономической информацией. Компания предоставляет информационные продукты и сервисы заинтересованным государственным органам предотвращения чрезвычайных ситуаций, органам планирования развития урбанизированных территорий, имущественному и коммунальному сектору, страховым

компаниям. Заслуживает внимания опыт оценивания рисков затоплений, предоставление через портал целой системы информационных услуг для поддержки принятия решений по этому поводу организациям и отдельным гражданам Великобритании Агентством по окружающей среде (Environment-agency) (Великобритания) (<http://www.environment-agency.gov.uk/homeandleisure/floods/default.aspx>) Французское космическое агентство CNES активно развивает подход к оценке рисков инфекционных заболеваний, вызываемых распространением насекомых – переносчиков инфекций в результате наводнений в странах Африки (<http://www.redgems.org/spip.php?rubrique4>). Однако используемые сегодня в операционных системах методы оценки рисков зачастую являются слишком упрощенными, и не опираются на достаточно развитый математический аппарат оценки среднего риска по эмпирическим данным, разработанный для задач оценивания качества восстановления функциональных зависимостей на основе эмпирических данных, применяемый в статистической теории обучения [1-3].

В данной работе предлагается постановка задачи оценки рисков, связанных со стихийными бедствиями на основе разнородной информации (спутниковых и наземных данных, а также данных моделирования), обосновывается применяемый для решения этой задачи метод и рассматривается пример его практического применения для оценки риска затоплений в Намибии.

### **Существующие подходы к оценке рисков стихийных бедствий на основе геопространственной информации**

В самых различных предметных областях (экономике, здравоохранении, финансовой сфере) общее понятие риска определяется примерно одинаково. Под риском  $R$  понимают функцию двух переменных: вероятности наступления неблагоприятного события  $p$  и наносимого этим событием ущерба  $l$  ([http://www.wired.com/science/planetearth/magazine/17-01/ff\\_dutch\\_delta?currentPage=3](http://www.wired.com/science/planetearth/magazine/17-01/ff_dutch_delta?currentPage=3)):

$$R = f(p, l). \quad (1)$$

Зачастую зависимость (1) представляют в виде матрицы риска (risk matrix) – двумерного массива, элементами которого являются значения риска для фиксированных дискретных градаций значений  $p$  и  $l$  [4], определяемые методом экспертного оценивания. Такой подход применим только в простейших случаях и требует строгой математической формализации для обеспечения его применимости в системах оценки рисков стихийных бедствий.

В статистической теории принятия решений функция риска для оценки  $\delta(x)$  параметра  $\theta$  (с помощью классификатора или решающего правила), вычисленная на основе наблюдений  $x$  параметра  $\theta$ , определяется как ожидаемое значение функции потерь  $L$  [5]

$$R(\theta, \delta(x)) = \int L(\theta, \delta(x)) f(x/\theta) dx \quad (2)$$

В работе [6] содержится детальный обзор метрик для определения рисков отдельного индивидуума, а также социальных, экономических и других рисков, связанных со стихийными бедствиями. При этом, в общем случае, риск описывается в виде функции вероятности нанесенного ущерба. Например, простой мерой социального риска считается ожидаемое число жертв за год, вычисляемое по формуле

$$E(N) = \int_0^\infty xf_N(x)dx, \quad (3)$$

где  $f_N(x)$  – функция плотности вероятности числа жертв  $x$  за год.

Другим примером функции риска [7] является функция совокупного взвешенного риска (AWR – Aggregated weighted risk), определяемая соотношением

$$AWR = \iint_A IR(x, y)h(x, y)dxdy, \quad (4)$$

где  $IR(x, y)$  – это риск стихийного бедствия (так называемый, индивидуальный риск) в позиции с координатами  $(x, y)$ ,  $h(x, y)$  – число домов в позиции  $(x, y)$ ,  $A$  – область, для которой определяется совокупный риск AWR.

В последующих разделах будет formalизовано понятие риска последствий стихийного бедствия на основе разнородной геопространственной информации, а также определены источники данных и изложен метод оценивания риска.

### **Постановка задачи оценки рисков на основе геопространственной информации и метод ее решения**

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

$$R_A = \iint_A r(x, y)dxdy \quad (5)$$

где  $r(x, y)$  – индивидуальный ожидаемый риск последствий стихийного бедствия  $z$  (индивидуальные ожидаемые потери) в точке  $(x, y)$ , вычисляемый как математическое ожидание функции ущерба  $h_{xy}(z)$  последствий в позиции  $(x, y)$

$$r(x, y) = \int_0^\infty h_{xy}(z)p_{xy}(z)dz, \quad (6)$$

где  $p_{xy}(z)$  – плотность распределения вероятности стихийного бедствия  $z$  в точке  $(x, y)$ , оцениваемая на основе совместного анализа разнородных геопространственных данных. Один из методов оценивания плотности вероятности  $p_{xy}(z)$  и функции ущерба определяется типом стихийного бедствия и будет описан ниже.

Плотность распределения вероятности стихийного бедствия  $p_{xy}(z)$  определяется различными факторами окружающей среды и метеорологическими условиями, которые

могут быть прямо или косвенно измерены с помощью наземных средств и дистанционных методов, либо получены с помощью моделирования. Например, вероятность весеннего паводка определяется влагозапасом снега, интенсивностью снеготаяния, температурой воздуха и количеством осадков в период снеготаяния на данной территории и выше по течению, а также структурой почвы, ее степенью промерзания и другими факторами. Ожидаемый объем стока и уровень воды в реке можно оценить с помощью гидрологических моделей при ассимиляции в нее спутниковых данных и данных наземных измерений.

Для восстановления плотности распределения вероятности  $p_{xy}(z)$  можно использовать хорошо разработанные в статистической теории обучения [1-3] метод минимизации эмпирического функционала в проблеме минимизации среднего риска. Конструирование эмпирического функционала основано на аппроксимации неизвестной плотности вероятности, входящей в функционал среднего риска типа (6) (в нашем случае), некоторой эмпирической плотностью, восстанавливаемой на основе данных измерений, и использованием ее вместо неизвестной плотности в получившемся эмпирическом функционале. Далее нам не нужно решать задачу минимизации этого эмпирического функционала среднего риска, как это делается в классической теории восстановления зависимостей по эмпирическим данным. В нашем случае интерес представляет лишь классическая в математической статистике задача восстановления плотности вероятности  $p_{xy}(z)$  по выборочным данным. Восстановив плотность вероятности, мы сможем оценить индивидуальный риск согласно (6) как ожидаемую величину ущерба от стихийного бедствия, совокупный ожидаемый риск последствий стихийного бедствия (совокупные ожидаемые потери) в области  $A$  согласно (5) и использовать полученную информацию о риске для принятия решения о мероприятиях, снижающих ущерб от стихийного бедствия.

Задача восстановления плотности распределения вероятностей в классе непрерывных функций сводится к некорректной задаче численного дифференцирования функции распределения вероятностей [1]. 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 [1]. Однако в случаях, когда имеется априорная информация об искомой плотности вероятности, можно избежать некорректной постановки этой задачи. Например, если восстанавливаемая плотность вероятности известна с точностью до конечного числа параметров, то задача ее восстановления по эмпирическим данным корректна, и для ее решения можно использовать эффективные методы параметрической статистики [1]. Так в классе задач минимизации среднего риска, связанном с классификацией (обучением распознаванию образов), восстановление неизвестных параметров плотности распределения вероятности  $p_{xy}(z)$  возможно, в частности [1], различными методами параметрической статистики (в зависимости от контекста задачи): методом Байесовых приближений; методом наилучших несмещенных приближений; методом максимального правдоподобия. Кроме того, в качестве моделей для оценки плотности вероятности стихийного бедствий можно использовать различные регрессионные модели или модели в виде черного ящика (например, нейросетевые, ядерные методы, другие методы статистической теории обучения и др.) [1-3].

Таким образом, для оценки вероятности риска стихийного бедствия необходимо анализировать (классифицировать) информацию, поступающую из различных источников с различным временным и пространственным разрешением. Для совместного анализа

такой информации применяются методы и технологии слияния данных (data fusion) [8]. Необходимо подчеркнуть, что плотность вероятности стихийного бедствия должна оцениваться на основе слияния преимущественно не сырых данных, а информации более высоких уровней обработки [9]. Для оценивания плотности вероятности стихийного бедствия  $p_{xy}(z)$  предлагается следующая обобщенная схема (рис. 1).

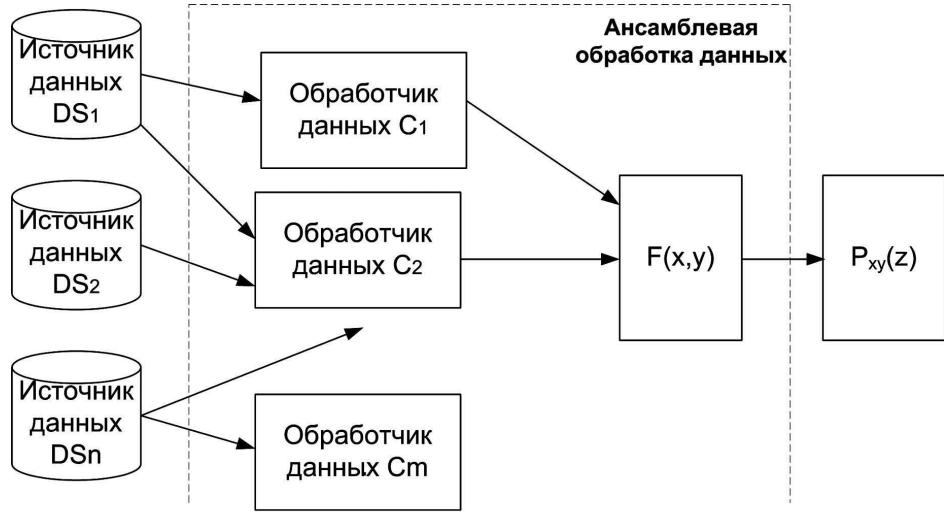


Рис. 1. Схема оценивания плотности распределения вероятности стихийного бедствия  $p(z) x y$  на основе разнородной информации

На рис. 1 блоки  $DS_i, i=1, \dots, n$  представляют различные источники данных – спутниковые данные, данные наземных сенсоров и данные моделирования. Блоки  $C_i, i=1, \dots, m$  обеспечивают обработку данных и получение информации более высокого уровня. В общем случае число блоков обработки не равно количеству источников данных  $n$  ( $m \neq n$ ), 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  $C_i, i=1, \dots, m$  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. Будем иметь в виду, что каждый из обработчиков  $C_i, i=1, \dots, m$  может обеспечивать несколько уровней преобразования входных данных, но для ясности при дальнейшем изложении явно не будем выделять эти уровни.

Без ограничения общности можно считать, что каждый из обработчиков  $C_i, i=1, \dots, m$  представляет собой отдельное решающее правило или классификатор (так называемый «слабый» или компонентный классификатор), обеспечивающий результат анализа данных одного или нескольких источников. Таким образом, классификаторы  $C_i, i=1, \dots, m$  составляют ансамбль экспертов (или «сильный» классификатор), «мнения» которых с соответствующими весами  $\alpha_i$  учитываются в блоке слияния  $F$ .

$$F(x, y) = \sum_{i=1}^m \alpha_i C_i(x, y). \quad (7)$$

Представленная на рис. 1 схема обработки данных реализует ансамблевый метод анализа данных [10], обеспечивающий более высокую точность оценивания разнородной информации, чем любой отдельно взятый классификатор. Отдельные классификаторы выступают в такой схеме в качестве «ядер», обеспечивающих переход от анализа отдельных признаков к оценке ситуации в целом. Преимуществом ансамблевого метода классификации является возможность повышения точности классификации за счет приема «усиления» (boosting) [11], который сводится к оценке функции потерь (ошибки классификации) и минимизации этой функции путем добавления новых компонентных классификаторов. Если добавление нового «слабого» классификатора не снижает значение функции потерь, этот классификатор не привносит новых информативных признаков для оценки ситуации и его можно не учитывать. Таким образом, обеспечивается оптимальная сложность модели.

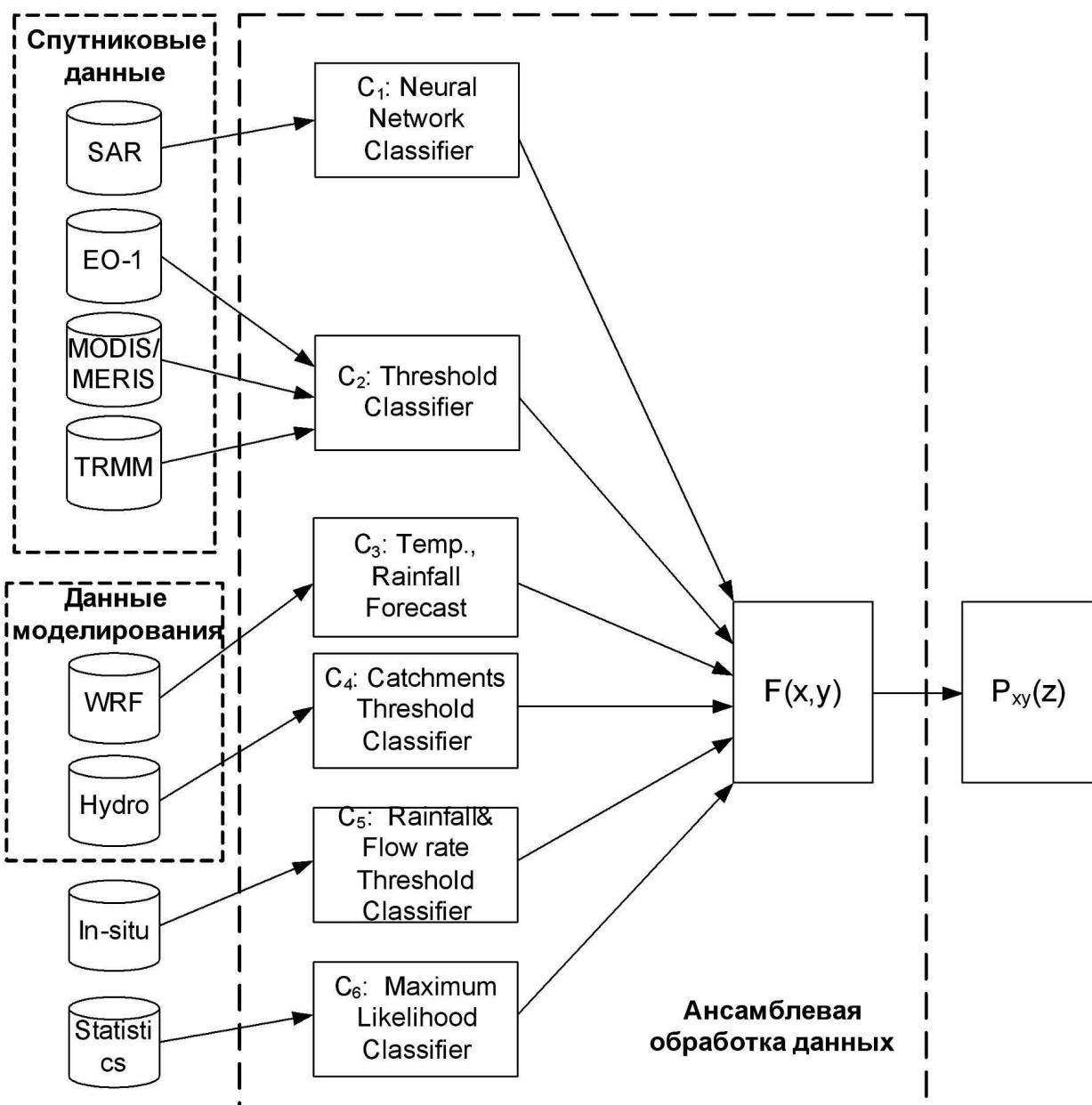


Рис. 2. Схема определения вероятности затопления для территории Намибии

## Пример: оценка риска затопления на основе разнородной геопространственной информации

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 источники данных:

- 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 продукты их обработки с приведением примеров визуализации.

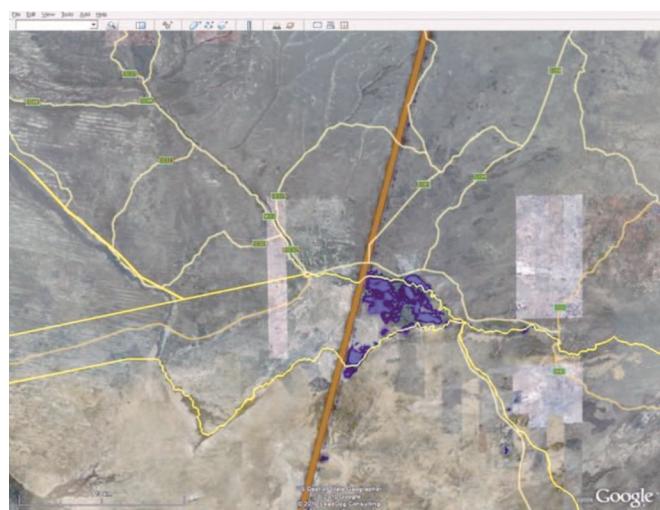


Рис. 3. Карта затоплений для региона Катима-Мулило (Намибия), полученная на основе данных Envisat/ASAR, 03.03.2010

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 электромагнитного излучения, состоит в невязкимости от погодных условий, в частности облачного покрова. Для создания маски затоплений применяются интеллектуальные вычисления [Ошибка! Источник ссылки не найден.2-14]. 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).



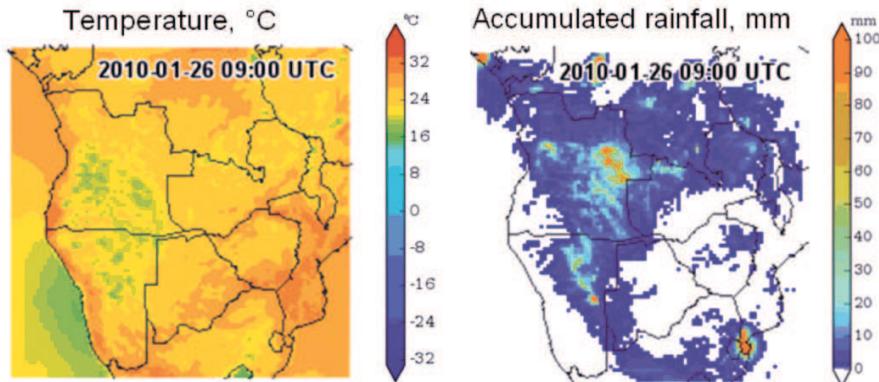
*Рис. 4. Измерения TRMM*

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. Примеры представлены на рис. 3.

Для мониторинга уровня осадков используются измерения, получаемые в рамках совместной миссии NASA и JAXA Tropical Rainfall Measuring Mission (TRMM). Пример предоставляемых данных показано на рис. 4.



*Рис. 5. Результаты работы модели WRF:  
(а) прогноз температуры; (б) прогноз осадков*

Метеорологические данные. Для метеорологического прогнозирования используется численная модель прогнозирования погоды WRF (см. рис. 5). 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 речек предоставляемая подразделениями Министерства по сельскому хозяйству, водным ресурсам и лесному хозяйству Намибии (рис. 6). Кроме того, для каждого бассейна предоставляются текущие и архивные значения уровня осадков (рис. 7) и стока (рис. 8). Эти архивные данные, наряду со спутниковыми измерениями используются для построения статистических моделей для прогнозирования наводнений и оценки рисков.

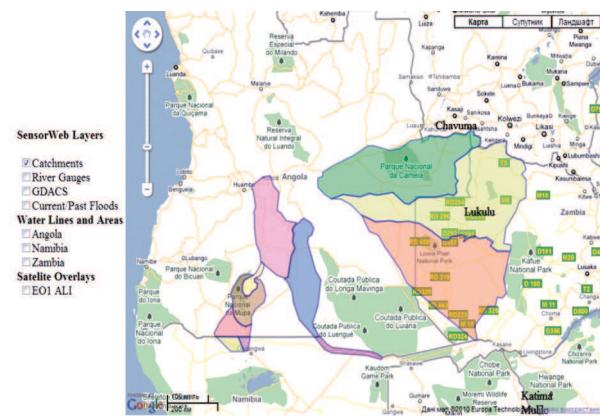


Рис. 6. Водосборный бассейн

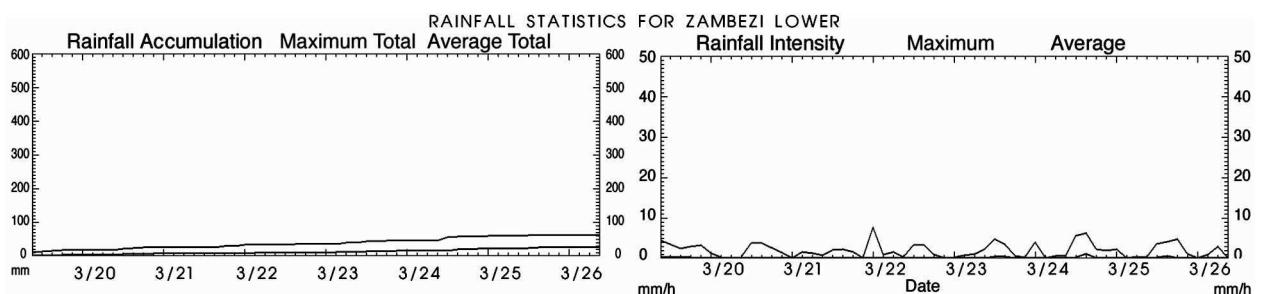
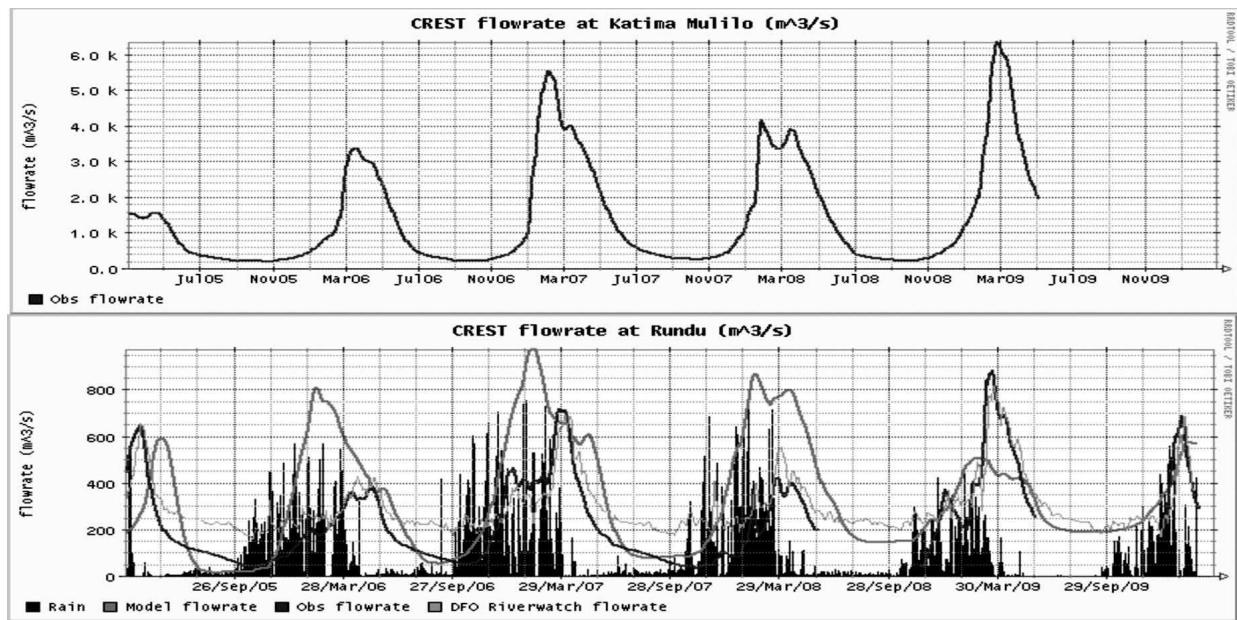


Рис. 7. Измерения уровня осадков

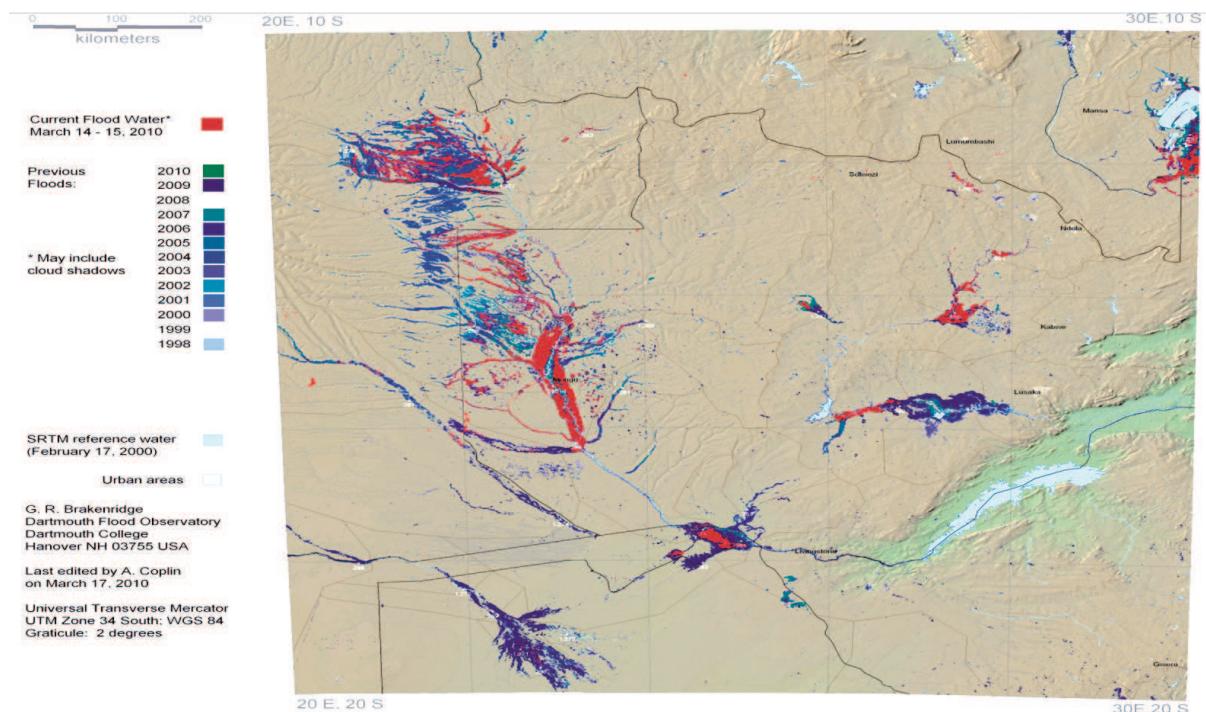
Статистические данные. Данные, полученные прибором MODIS, который установлен на спутниках Aqua и Terra, используется для статистического анализа карт затоплений за разные годы. Данные карты доступны, начиная с 1999 года и предоставляются центром Dartmouth Flood Observatory. Пример продукта обработки данных MODIS представлен на рис. 9.

Для глобального детектирования наводнений используются система [15], разрабатываемая в центре JRC (Joint Research Center) Европейской комиссии. Для этого используются данные, получаемые микроволновым сканирующим радиометром AMSR-E, который установлен на спутнике Aqua. Примеры продуктов представлены на рис. 10.



*Рис. 8. Наземные измерения стока*

Работы по созданию системы оценки рисков на основе разнородной геопространственной информации выполняются 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-НКАУ. Web-интерфейс пилотной версии разрабатываемой системы представлен на рис. 11.



*Рис. 9. Изменение маски затоплений в разные годы*

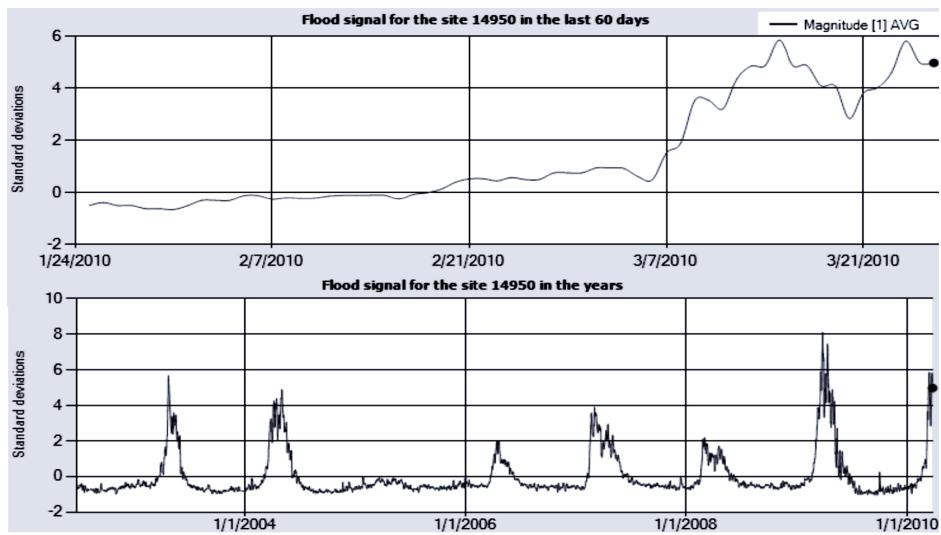


Рис. 10. Изменение уровня затоплений на разных временных интервалах

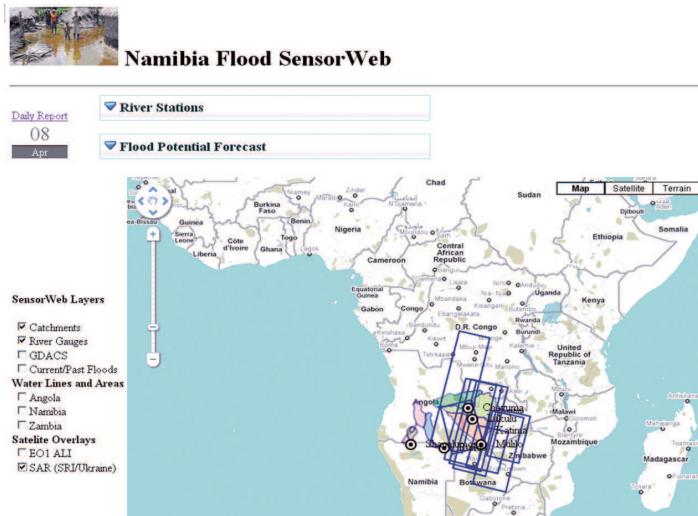


Рис. 11. Web-интерфейс pilotной версии системы мониторинга паводков для Намибии

## Заключение

В работе предложен единый подход к оцениванию рисков, 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. В дальнейшем этот подход планируется развивать для оценки конкретных категорий риска: финансового, социального, экономического и т.д.

## Благодарности

Работа частично поддерживается грантом УНТЦ #4928 а также проектом МОН Украины № М/72-2008 «Разработка системы комплексной обработки данных ДЗЗ с использованием Grid-технологий».

## **Литература**

1. Vapnik, V. – Statistical Learning Theory. New York: Wiley, 1998.
2. Haykin, S. Neural Networks. A comprehensive Foundation. – New Jersey: Prentice Hall, 1
3. Bishop C.M. Pattern Recognition and Machine Learning. - New York: Springer Science+Business Media, 2006. – 738 p.
4. Flage R., Aven T., Expressing and communicating uncertainty in relation to quantitative risk analysis// R&RATA, # 2(13), part 1 (Vol.2), 2009, PP. 9-18.
5. Robert, Christian (2007). The Bayesian Choice (2nd ed.). New York: Springer. doi:10.1007/0-387-71599-1. MR1835885. ISBN 0-387-95231-4.
6. S.N. Jonkman et al. An overview of quantitative risk measures for loss of life and economic damage // Journal of Hazardous Materials A99 (2003). P. 1–30
7. M. Piers, Methods and models for the assessment of third party risk due to aircraft accidents in the vicinity of airports and their implications for societal risk// In: R/E/ Jorissen, P.J.M. Stallen (Eds.), Quantified Societal Risk and Policy Making, Kluwer Academic Publishers, Dordrecht, 1998.
8. H. B. Mitchell, Multi-sensor Data Fusion – An Introduction (2007) Springer-Verlag, Berlin, ISBN 9783540714637
9. S. Das, High-Level Data Fusion (2008), Artech House Publishers, Norwood, MA, ISBN 9781596932814 and 1596932813
10. Tommi Jaakkola, course materials for 6.867 Machine Learning, Fall 2006. MIT OpenCourse-Ware(<http://ocw.mit.edu/>), Massachusetts Institute of Technology.
11. S. Kotsiantis, P. Pintelas, Combining Bagging and Boosting, International Journal of Computational Intelligence, Vol. 1, No. 4 (324-333), 2004.
12. Скакун С.В. Нейросетевой метод картографирования паводков на основе спутниковых изображений // Научные труды ДонНТУ серия "Информатика, кибернетика и вычислительная техника". 2009. – Вып. 10(153). – С. 115-120.
13. Kussul N., Shelestov A., Skakun S. Intelligent Computations for Flood Monitoring// International Book Series "Advanced Research in Artificial Intelligence" (ed. Markov K., Ivanova K., Mitov I.), 2008, number 2, pp.48-54.
14. Kussul N., Shelestov A., Skakun S. Grid System for Flood Extent Extraction from Satellite Images // Earth Science Informatics. – 2008. – 1(3–4). – P. 105-117.
15. De Groot, T., P. Riva, 2009. Global real-time detection of major floods using passive microwave remote sensing. Proceedings of the 33<sup>rd</sup> International Symposium on Remote Sensing of Environment Stresa, Italy, May 2009.

## **Geospatial risk analysis based on data fusion**

**N.N. Kussul, Ya.I. Zyelyk, S.V. Skakun, A.Yu. Shelestov**

*Space Research Institute of NAS and NSA of Ukraine.  
40 Academician Glushkov Avenue, build. 4/1, Kyiv, Ukraine, 03680  
inform@ikd.kiev.ua*

The problem statement of disaster risk assessment, based on heterogeneous information (from satellites and in-situ data, and modeling data) is proposed, the problem solving method is grounded and considers its practical use for risk assessment of flooding in Namibia. The basis of the method is the ensemble approach to the heterogeneous data analysis with the use of the data fusion techniques and evaluation the probability density function of a natural disaster using this method.

**Keywords:** risk assessment, natural disasters, geospatial data, remote sensing, data fusion, ensemble data processing, average risk minimizing, probability density function, maximum likelihood principle, neural network classifier.