

УДК 519.254, 519.257, 004.932, 004.42

С.В. Скакун
Институт космических исследований НАНУ-НКАУ
serhiy.skakun@ikd.kiev.ua

Оценка плотности вероятности наводнений на основе анализа спутниковых и наземных данных

В статье рассмотрены вопросы оценки плотности вероятности наводнений на основе анализа разнородных геопространственных данных. Проводится статистический анализ наземных гидрологических данных для оценки периода повторяемости наводнений. Предлагается новый метод оценки плотности вероятности наводнений на основе анализа временных рядов спутниковых данных. Для оценки плотности вероятности для отдельно взятого события предлагается использовать интеллектуальные методы, в частности, нейронные сети.

Ключевые слова: наводнения, геопространственные данные, ДЗЗ, ансамблевая обработка данных

Введение

В последние годы проблема мониторинга наводнений существенно сместилась в сторону управления рисками. Об этом отмечается в директиве Европейской комиссии "Flood risk directive", принятой в октябре 2007 года [1]. В ней говорится о необходимости анализа рисков, связанных с наводнениями, путем разработки соответствующих карт рисков наводнений и планов действий для уменьшения последствий наводнений. Поэтому на сегодняшний день задача разработки методов оценки рисков наводнений на основе разнородной геопространственной информации является чрезвычайно актуальной.

Согласно директиве "Flood risk directive" [1], риск определяется как функция (комбинация) двух аргументов: вероятность наступления наводнения и соответствующий ущерб (потери). Наиболее популярным является подход, при котором риск оценивается как математическое ожидание функции ущерба [2, 3, 4, 5, 6]. При этом возникает задача оценки плотности распределения вероятности наводнения. Для этого обычно применяются гидравлические модели с целью анализа разных сценариев затоплений. Однако использование таких моделей связано с неопределенностями, которые возникают в результате частичного отсутствия данных, необходимых для запуска, калибровки и валидации моделей, ненадежности данных и разных способов интерпретации данных [7, 8].

Другим подходом, который дополняет существующие модели, является ансамблевый подход, при котором для восстановления плотности распределения вероятности наводнения анализируется информация из различных источников. В работе [9] прогнозирование уровня воды осуществляется путем объединения выходов

следующих моделей: нейросетевой, на основе нечеткой логики, статистической и гидрологической модели TOPMODEL. Показано, что точность прогнозирования ансамблевой архитектурой выше, чем каждой модели в отдельности. В работах [4, 10] предложен метод оценки риска для отдельно взятого события (наводнения) путем создания ансамбля классификаторов при обработке радиолокационных спутниковых данных. Таким образом, для каждого пикселя изображения вычисляется вероятность затопления, которая умножается на значение функции ущерба для получения риска. Стоит отметить, что в этой работе для оценки вероятности затопления существенным является использование двух спутниковых изображений, сделанных в одно и то же время, что не всегда выполнимо. В работах [5, 6] ансамблевый подход используется для оценивания плотности распределения вероятности стихийного бедствия. Результат анализа данных одного или нескольких источников обеспечиваются отдельными классификаторами, «мнения» которых интегрируются в блоке слияния.

В данной статье развит подход, предложенный в работах [5, 6], путем разработки новых классификаторов для анализа данных. В частности, проводится статистический анализ наземных гидрологических данных для оценки периода повторяемости наводнений. Предлагается новый метод оценки плотности вероятности наводнений на основе анализа временных рядов спутниковых данных. Для оценки плотности вероятности для отдельно взятого события предлагается использовать интеллектуальные методы, в частности, нейронные сети. В отличие от подхода, который применяется в работах [4, 10], достаточно использовать одно спутниковое изображение. Кроме того, рассматриваются вопросы разработки

информационной системы поддержки принятия решений для оценки рисков чрезвычайных ситуаций. Предложенные в данной статье методы верифицированы на реальных данных для оценки риска затоплений в Намибии.

Общая постановка задачи оценки рисков на основе разнородной геопространственной информации

Совокупным ожидаемым риском стихийного бедствия (совокупными ожидаемыми потерями) в области A назовем величину [5, 6]

$$R_A = \iint_A r(x, y) dx dy \quad (1)$$

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

$$r(x, y) = \int_0^{\infty} h_{xy}(z) p_{xy}(z) dz, \quad (2)$$

где $p_{xy}(z)$ — плотность распределения вероятности стихийного бедствия z в точке (x, y) , оцениваемая на основе совместного анализа разнородных геопространственных данных. Плотность распределения вероятности стихийного бедствия $p_{xy}(z)$ определяется различными факторами окружающей среды и метеорологическими условиями.

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) [11]. Необходимо подчеркнуть, что плотность вероятности стихийного бедствия должна оцениваться на основе слияния преимущественно не сырых данных, а информации более высоких уровней обработки [12]. Для оценивания плотности вероятности стихийного бедствия $p_{xy}(z)$ предлагается следующая обобщенная схема (рис. 1).

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.

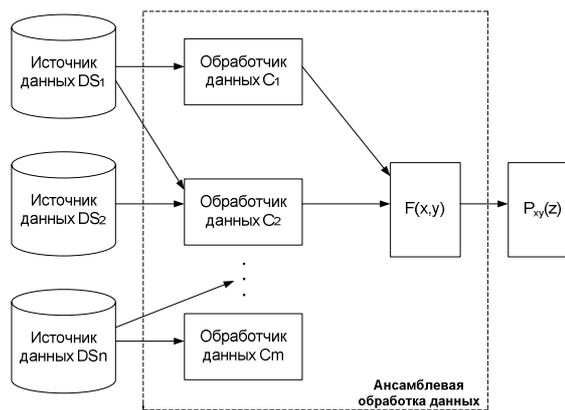


Рисунок 1 – Схема оценивания плотности распределения вероятности стихийного бедствия $p_{xy}(z)$ на основе разнородной информации

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 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. Каждый из обработчиков C_i 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 составляют ансамбль экспертов (или «сильный» классификатор), «мнения» которых с соответствующими весами α_i учитываются в блоке слияния F . Для интеграции «мнений» можно использовать смеси экспертов (mixture of experts) [13].

Описание данных

В рамках данной работы используются разнородные геопространственные данные для региона Катима Мулило (Намибия).

- Наземные гидрологические измерения стока и уровня воды за 1965-2010 гг.
- Временные ряды спутниковых данных Landsat-5/TM и Landsat-7/ETM+ (2000-2010 гг.).
- Данные об уровне осадков, измеренные в рамках миссии Tropical Rainfall Measuring Mission (TRMM).
- Социо-экономические данные: плотность застройки жилыми зданиями.

Оценка периода повторяемости наводнений на основе наземных данных

Для оценки периода повторяемости наводнений использованы методы теории экстремальных значений [14]. Для определения наводнения n -летней повторяемости для каждого года определено максимальное значение стока. Эти значения затем ранжируются, и году с наибольшим значением стока присваивается ранг $m=1$, следующему году – $m=2$, и т.д. На основе ранга и общего количества данных ($n=46$) вычисляется период повторяемости (так называемое уравнение Вэйбулла):

$$R = (n+1)/m. \quad (3)$$

В таблице 1 приведены соответствующие значения для первых десяти лет с наибольшим значением стока.

Таблица 1. Максимальное значение стока

Ранг, m	Год	Сток, м ³ /сек	R
1	1969	6817	47,00
2	2009	6365	23,50
3	1978	6251	15,67
4	2010	5704	11,75
5	1979	5675	9,40
6	1976	5568	7,83
7	2007	5564	6,71
8	1975	5409	5,88
9	1968	5312	5,22
10	1966	5276	4,70

Затем строится зависимость значений стока от периода повторяемости (в логарифмическом масштабе, рис. 2). В результате аппроксимации полиномом 3-го порядка получена следующая зависимость:

$$y = 2969,8x^3 - 9567,7x^2 + 11163x + 1181,$$

где y – значение стока; $x = \log_{10}(R)$.

Значение достоверности аппроксимации составило $R^2 = 0,9907$. Значения стока и периода повторяемости, полученная зависимость с 95% интервалом показаны на рис. 2. Полученные значения n -летней повторяемости наводнений для значений $n = 10, 50$ и 100 представлены в таблице 2.

Таблица 2. Значения n -летней повторяемости наводнений

n , лет	Сток, м ³ /сек	Интервал достоверности
10	5746	[5419; 6073]
50	7093	[6654; 7532]
100	8993	[8131; 9855]

Таким образом, вероятность наступления наводнения со значением стока 8993 м³/сек и более для некоторого года составляет 1% (или 0,01), со значением стока 5746 м³/сек и более – 10% (или 0,1). Для прогнозирования стока

используются значения осадков, полученные в рамках миссии TRMM (рис. 3). Дневные значения получены с использованием системы NASA Giovanni (GES-DISC Interactive Online Visualization ANd aNalysis Infrastructure)¹.

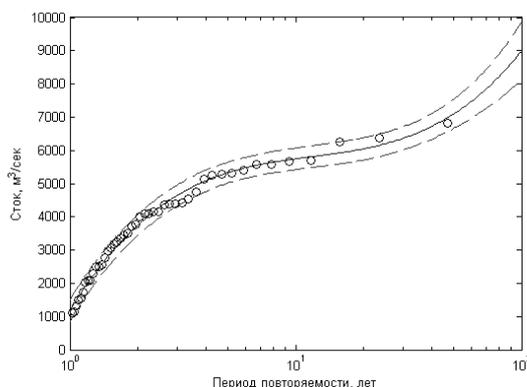


Рисунок 2 – Зависимость максимального стока от периода повторяемости в логарифмической шкале, пунктирной линией показан 95% доверительный интервал (регион Катима Мулило, Намибия, 1965-2010 гг.)

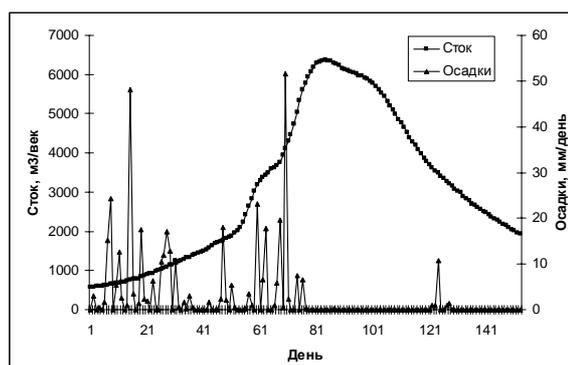


Рисунок 3 – Изменение стока и уровня осадков за 2009 год (регион Катима Мулило, Намибия)

Значения стока и вероятности используются для разработки различных сценариев наводнений. Для определения зон затоплений при разных значениях стока предлагается использовать временные ряды спутниковых данных.

Оценка плотности вероятности наводнений с использованием спутниковых данных

Оценка плотности вероятности на основе анализа временных рядов спутниковых данных. Данные, полученные со спутников Landsat-5/TM и Landsat-7/ETM+ в период 2000-2010 гг. используются для оценки плотности вероятности затоплений. Процесс обработки каждого изображения состоит из следующих этапов:

¹ Giovanni. - <http://disc.sci.gsfc.nasa.gov/giovanni>

1. Идентификация облачности и теней на изображении, согласно алгоритму [15], а также артефактов на изображениях Landsat-7/ETM+ вследствие технических неисправностей. Эти значения помечаются как «No Data» и исключаются из последующего анализа.

2. Идентификация водной поверхности, используя метод пороговой сегментации [16].

Таким образом, в результате обработки и корегистрации всех изображений каждый пиксель может принимать следующие значения: 0 - «Суша», 1 - «Вода», 2 - «No Data». Для вычисления вероятности затопления при конкретном значении стока для различных событий воспользуемся следующим выражением:

$$PI_{pixel} = \frac{1}{N} \sum_{i=1}^N I_{pixel=i} \quad (4)$$

где N – количество доступных снимков.

Оценка плотности вероятности для отдельно взятого события. В случае отсутствия статистических временных рядов спутниковых данных можно оценивать плотности вероятности для отдельно взятого события (т.е. для значения стока за дату, когда получено спутниковое изображение). В работах [17, 18, 19] использовались нейронные сети (НС) Кохонена для классификации изображения на два класса: «Вода» и «Суша». Для определения, какому классу соответствует нейрон, применялась следующая процедура калибровки НС: нейрон относится к классу «Вода», если он активируется большинством векторов из обучающей выборки, которые маркированы как класс «Вода». Т.е. любой пиксель обработанного изображения мог принимать только два значения (1 для класса «Вода» и 0 – «Суша»). Для получения «вероятностного» результирующего изображения предлагается следующий подход: каждому нейрону НС ставится в соответствие не бинарное значение, а значение из промежутка [0; 1], которое вычисляется как относительное число векторов из обучающей выборки, активирующих данный нейрон и маркированных как «Вода». Таким образом, каждый пиксель изображения будет принимать значение из промежутка [0; 1], которое и будет составлять искомое значение вероятности.

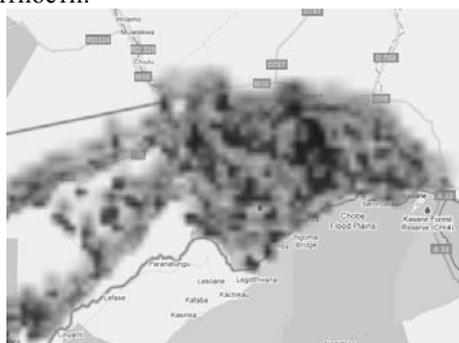


Рисунок 4 – Плотность застройки жилыми зданиями (регион Катима Мулило, Намибия)

Полученные в результате обработки одним из вышеизложенных методов изображения интегрируются, например с социо-экономической информацией, для оценки риска (выражение (2)). Пример плотности застройки жилыми зданиями представлен на рис. 4.

Информационная система поддержки принятия решений при чрезвычайных ситуациях

Архитектура системы. Основная цель состоит в создании системы, которая должна предоставлять информационные продукты мониторинга чрезвычайных ситуаций на регулярной основе с привлечением данных дистанционного зондирования Земли (ДЗЗ). Основными преимуществами системы являются: динамическое формирование контента системы; удобные средства визуализации геопространственной информации и работы с системой; использование международных стандартов обмена и предоставления информации. Архитектуру разработанной системы представлено на рис. 5. Система состоит из следующих уровней: уровень работы пользователя; уровень бизнес-логики, который обеспечивает основную функциональность по обработке и предоставлению данных и геоинформационных продуктов; уровень данных.

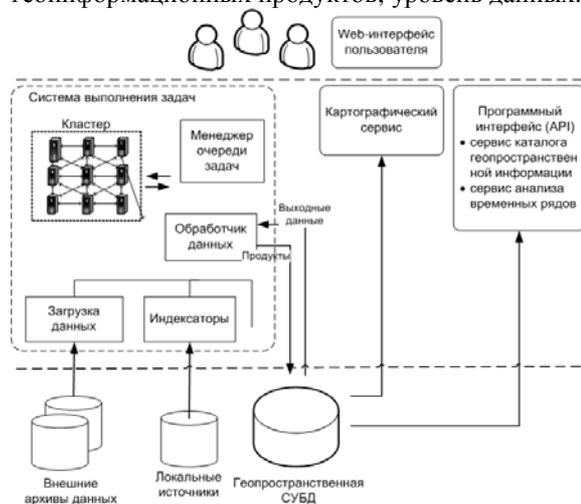


Рисунок 5 – Архитектура системы

Компоненты системы. На основе функциональной декомпозиции системы выделены следующие ее компоненты [20]:

- Подсистема взаимодействия с пользователями (Web-портал);
- Серверная подсистема, обеспечивающая обработку запросов и отвечающая за выполнение заданий;
- Подсистема управления содержимым (административная часть);
- Подсистема поддержки баз данных (БД) с соответствующими программными интерфейсами

для доступа к данным.

При разработке клиентской Web-системы и административной подсистемы использовались JavaScript-библиотеки OpenLayers и Prototype. Использование технологии AJAX позволяет динамически изменять части Web-страницы, при этом сама страница не перезагружается. Система работает в браузерах Internet Explorer версии 6 и выше, Firefox, Google Chrome, Safari и Opera. В панели управления пользовательского интерфейса размещены следующие разделы: «Спутник» и «Тип ЧС» для выбора имеющихся спутниковых данных и продуктов обработки и различных видов ЧС; «Дата» и «Период» для поиска соответственно по дате или за период; «Area search» для поиска по географическому региону (рис. 6).



Рисунок 6 – Web-интерфейс системы

В качестве Web-сервера системы используется бесплатный, масштабируемый, неблокирующий Web-сервер Tornado для поддержки режима реального времени. Программный интерфейс API (Application Programming Interface) взаимодействия с сервером построен на основе стандартного протокола HTTP. Для вызова методов интерфейса используется GET-запрос к серверу, а для передачи результатов вызова метода (т.е. обработки запросов) - формат JSON. Программный интерфейс реализован на языке программирования Python в пределах программного каркаса Django. Преимуществами данного каркаса является его высокая надежность, способность работать в режиме высокой нагрузки, удобство и простота создания сервисов.

Для хранения геопространственных данных, в частности спутниковых снимков и продуктов их обработки, используется геопространственная реляционная СУБД (система управления БД). В результате анализа существующих решений выбрана такая модель хранения данных, при которой в реляционной БД хранятся только метаданные и ссылки на данные. При этом сами спутниковые данные хранятся в отдельном сетевом хранилище данных. По сравнению с моделью хранения всех данных под управлением СУБД, выбранная модель имеет недостатки, связанные с обеспечением целостности БД. К сожалению, на данный момент

надежные решения, предусматривающие сохранение самих геопространственных данных в БД с поддержкой необходимых операций над данными (выборка подмножества данных, изменение пространственного разрешения, перепроектирование и т.д.) находятся на стадии разработки. Исходя из этого, реальных альтернатив отдельного хранения метаданных и данных, на наш взгляд, не существует. Для хранения данных выбрана реляционная СУБД PostgreSQL, поскольку, кроме стандартной функциональности, позволяет эффективно хранить геопространственные векторные данные и имеет расширение для хранения слабоструктурированных данных. Для хранения географической области покрытия снимков и продуктов, а также координат наземных объектов используется расширение PostGIS² для реляционной СУБД PostgreSQL. Расширение PostGIS включает поддержку пространственных индексов R-Tree/GiST и функции обработки геопространственных данных.

Геопространственные объекты, поддерживаемые PostGIS, является надмножеством, так называемых, простых объектов (Simple Features), определенных консорциумом OGC (Open Geospatial Consortium). В рамках данной работы используются объекты PostGIS Point для хранения координат наземных станций и Polygon для представления области, которую занимает каждый слой данных, а также для хранения границ наземных объектов (например, областей и районов). Для реализации программных интерфейсов для доступа к БД использованы CGI-приложения на языке программирования Python. Эти приложения обеспечивают функциональность по добавлению/удалению слоев, поиску слоев по имени, дате съемки, спутнику и другим параметрам.

Визуализация геопространственных данных в системе мониторинга осуществляется по международным стандартам OGC, в частности WMS (Web Map Service) и KML (Keyhole Markup Language). WMS – это стандартный протокол для отображения географически привязанных изображений через Интернет, которые генерируются map-сервером, используя данные из геопространственной БД. Запрос передается по протоколу HTTP и определяет географические слои и области, представляющие интерес для обработки, целевую систему координат (проекцию). В ответ на запрос выдается изображение в форматах JPEG, PNG и прочее, которые могут отображаться в браузере. Интерфейс также поддерживает возможность наложения картографических слоев с нескольких серверов. Разработанный сервис WMS

² PostGIS. - <http://postgis.refractor.net>

функционирует на основе программного обеспечения (ПО) MapServer³, которое является одной из самых популярных сред для создания картографических Web-сервисов с открытым кодом. Среди преимуществ ПО MapServer можно выделить мощные инструменты конфигурации, работа с геопространственными СУБД, поддержка большого количества растровых и векторных форматов данных.

Заключение

В данной статье рассмотрены вопросы оценки плотности вероятности наводнений на основе анализа разнородных геопространственных данных для региона Катима Мулило (Намибия). Проведен статистический анализ наземных гидрологических данных для оценки периода повторяемости наводнений. Определено, что вероятность наступления наводнения со значением стока 8993 м³/сек и более для некоторого года составляет 1% (или 0,01), со значением стока 5746 м³/сек и более – 10% (или 0,1). Предложен новый метод оценки плотности вероятности наводнений на основе анализа временных рядов спутниковых данных. Для оценки плотности вероятности для отдельно взятого события используются нейронные сети. В отличие от существующих ансамблевых подходов, достаточно использовать один спутниковый снимок для получения «вероятностного» изображения. Рассмотрены вопросы разработки информационной системы поддержки принятия решений для оценки рисков чрезвычайных ситуаций.

Литература

1. Mostert E., Junier S. J. The European flood risk directive: challenges for research // *Hydrology and Earth System Sciences Discussions*, Vol. 6, N 4, 2009. - p. 4961-4988.
2. 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 // *Journal of Hazardous Materials*, A99, 2003. - p. 1-30.
3. Hoes O., Schuurmans W. Flood standards or risk analyses for polder management in the Netherlands // *Irrig. Drain.*, 55, 2006. - p. 113-119.
4. Schumann G., Di Baldassarre G. The direct use of radar satellites for event-specific flood risk mapping // *Remote Sensing Letters*, Vol. 1, N 2, 2010. - p. 75-84.
5. Оценка рисков стихийных бедствий на основе разнородной геопространственной информации / Н.Н. Кукуль, Б.В. Соколов, Я.И. Зельк и др. // *Проблемы управления и информатики*. – 2010. - № 6. - С. 97-110.

6. Скакун С.В. Оценка риска наводнений на основе разнородных геопространственных данных / С.В. Скакун // *Наукові праці ДонНТУ Серія "Інформатика, кібернетика та обчислювальна техніка"*. – 2010. – Вип. 12(165). - С. 94-98.
7. Horritt M.S. A methodology for the validation of uncertain flood inundation models // *Journal of Hydrology*, 326, 2006. - p. 153-165.
8. Faulkner H., Parker D., Green C., Beven K. Developing a translational discourse to communicate uncertainty in flood risk between science and the practitioner // *AMBIO: A Journal of the Human Environment*, 36(8), 2007. - p. 692-704.
9. See L., Abrahart R.J. Multi-model data fusion for hydrological forecasting // *Computers & Geosciences*, Vol. 27, N 8, 2001. – p. 987-994.
10. Schumann G., Di Baldassarre G., Bates P.D. The Utility of Spaceborne Radar to Render Flood Inundation Maps Based on Multialgorithm Ensembles // *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 47, N 8, 2009. – p. 2801-2807.
11. Mitchell H.B. Multi-sensor Data Fusion – An Introduction. – Springer-Verlag, Berlin, 2007.
12. Das S. High-Level Data Fusion. – Artech House Publishers, Norwood, MA, 2008.
13. Jacobs R.A., Jordan M.I., Nowlan S.J., Hinton G.E. Adaptive mixtures of local experts // *Neural Computation*, Vol. 3, 1991. - p. 79-87.
14. Holmes R.R., Dinicola K. 100-Year flood—it's all about chance: U.S. Geological Survey General Information Product 106, 2010. - 1 p.
15. Irish R., Barker J., Goward S., Arvidson T. Characterization of the Landsat-7 ETM+ Automated Cloud-Cover Assessment (ACCA) Algorithm // *Photogrammetric Engineering & Remote Sensing*, Vol. 72, N 10. - p. 1179-1188.
16. Frazier P.S., Page K.J. Water Body Detection and Delineation with Landsat TM Data // *Photogrammetric Engineering & Remote Sensing*, Vol. 66, N 12, 2000. - p. 1461-1467.
17. 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.
18. Скакун С.В. Нейросетевой метод картографирования паводков на основе спутниковых изображений / С.В. Скакун // *Научные труды ДонНТУ серия "Информатика, кибернетика и вычислительная техника"*. 2009. - Вып. 10(153). - С. 115-120.
19. Skakun S. A Neural Network Approach to Flood Mapping Using Satellite Imagery // *Computing and Informatics*, Vol. 29, N 6, 2010. - p. 1013-1024.
20. Кукуль Н.Н. Grid-системы для задач исследования Земли. Архитектура, модели и технологии / Н.Н. Кукуль, А.Ю. Шелестов. – К.: "Наукова думка", 2008. - 452 с.

³ MapServer. - www.mapserver.org