THE USE OF TIME-SERIES OF SATELLITE DATA TO FLOOD RISK MAPPING

Sergii Skakun

***Abstract**: In this paper we propose a novel approach for flood hazard mapping by processing and analyzing a time-series of satellite data and derived flood extent maps. This approach is advantageous in cases when the use of hydrological models is complicated by the lack of data, in particular high-resolution DEM. We applied this approach to the time-series of Landsat-5/7 data acquired 2000 to 2010 for the Katima Mulilo region in Namibia. We further integrated flood hazard map with dwelling units database to derive flood risk map.*

***Keywords**: flood hazard, flood risk assessment, Earth remote sensing, Earth observation, satellite data processing, UN-SPIDER.*

***ACM Classification Keywords**: H.1.1 [Models and Principles] Systems and Information Theory; I.4.8 [Image Processing and Computer Vision] Scene Analysis - Sensor Fusion.*

Introduction

Over last decades we have witnessed the upward global trend in natural disaster occurrence. Hydrological and meteorological disasters are the main contributors to this pattern [Knight, 2006; Rodriguez et al, 2009]. In 2007, hydrological disasters, such as floods and wet mass movements, represented 55% of the overall disasters reported.

It should be noted that in recent years flood management has shifted from protection against floods to managing the risks of floods (European Flood risk directive) [Mostert and Junier, 2009]. To enable flood risk assessment, corresponding flood hazard and flood risk maps should be developed. Flood risk is a function of two arguments: hazard probability and vulnerability [Mostert and Junier, 2009; Schumann and Di Baldassarre, 2010]. In other words, risk is a mathematical expectation of vulnerability (consequences) function [Jonkmana et al, 2003; Hoes and Schuurmans, 2006; Kussul et al, 2010]. Flood probability density is to be estimated in order to produce flood hazard maps. Usually, this is done through hydraulic modeling of a peak flow []. But running such models faces many uncertainties [Horritt, 2006] due to the lack of hydrological and other required data, their incompleteness and imperfection [Mostert and Junier, 2009]. The use of space-borne remote sensing data to flood risk mapping is a complement approach to the existing flood modeling techniques [Schumann and Di Baldassarre, 2010; Bates et al, 1997; Bates 2004; Horritt 2006; Lecca et al, 2011].

In [Schumann and Di Baldassarre, 2010], a novel approach for rapid flood risk mapping is proposed based on the use of radar satellite data. An event-specific weighted hazard map was generated based on plausible flood area observations from an aggregation of widely applied image-processing techniques. The map is further augmented to an event-specific fuzzy flood risk map by fusing the multialgorithm ensemble map with vulnerability-weighted land cover vector data. In [See and Abrahart, 2001], an ensemble approach to hydrological forecasting is exploited. River level is predicted by fusing outputs from different models, in particular fuzzy neural network, statistical model, and hydrological model TOPMODEL. It was shown that accuracy of the ensemble model is higher than for separate models.

In this paper we propose a novel approach for flood hazard mapping by analyzing a time-series of satellite data. In particular, satellite images are processed in order to derive flood extent maps and the latter are used to estimate flood probability density. In particular, each pixel of the flood extent maps can be one of the following values: 0 - «No water», 1 - «Water», 2 - «NoData». The specific value «NoData» is used to mark pixel that contain no valuable information due to the cloud cover and shadows, or specifics of the satellite instrument (for example, Landsat-7/ETM+ SLC-off pixels). Therefore, if «NoData» pixels are excluded from considerations, each pixel is binary with Bernoulli distribution. Maximum likelihood method is used to derive a parameter of Bernoulli distribution, a *success probability*, from sampling set. This parameter shows probability of inundation, and can be viewed as flood probability density function.

Study area and available data description

The study area is the Katima Mulilo region in Namibia (Fig. 1). Since 2009, Namibia has experienced a surge of flooding in the Northern portion of the country. It was estimated that during 2009, 700,000 of the approximately 2 million people in Namibia were impacted by the floods of 2009, furthermore around 50,000 people were displaced and 102 people lost their lives. During the 2000-2011 periods, each year, except 2005, was characterized by floods that usually occurred from the month of February through May. Three floods from this period were in top 10 water level records historically, and 8 floods were in top 20.

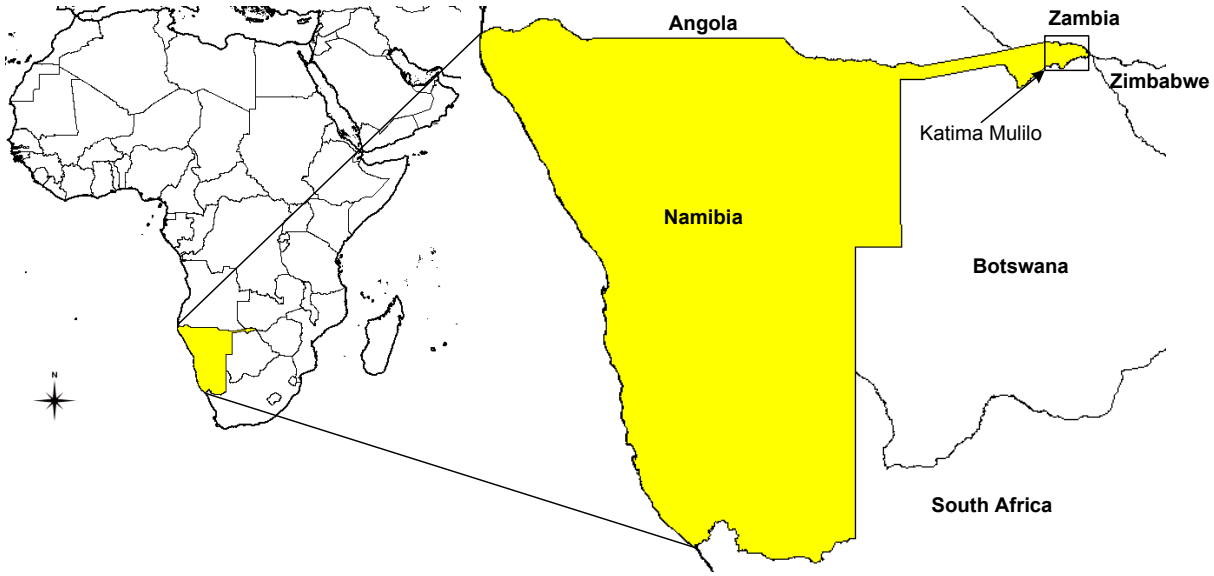

Figure1. Location of the study area

As a follow-up to the assistance provided by the United Nations Platform for Space-based Information for Disaster Management and Emergency Response (UN-SPIDER), National Aeronautics and Space Agency (NASA), German Space Agency (DLR), Ukraine Space Research Institute (USRI), and other space agencies in 2009, an international multi-disciplinary initiative, titled Namibia SensorWeb Pilot Project, was established. The Pilot aims at developing an operational trans-boundary flood management decision support system for the Southern African region to provide useful flood information and water-borne disease forecasting tools for local decision makers.

One of main tasks within the Pilot is flood risk assessment. In order to provide comprehensive hydrological modeling, a high-resolution (approximately 1 m) digital elevation model (DEM) is required since the topography is very flat in the study area. At present, such DEM is not available, and a 90 m resolution SRTM DEM is not enough to accomplish that task.

To tackle this problem we exploit different data sets in order to obtain flood hazard map. In particular, we benefit from large number of freely available images that were acquired by Landsat-5/7 satellite over 2000-2010 years.

The following are characteristics of the geospatial data that were used in the study:

-- 44 images acquired by Landsat-5/TM and Landsat-7/ETM+ from 2000 to 2010 (Fig. 2).

-- River gauge data: water level and discharge from 1965 to 2010 (provided by Hydrological Services Namibia, (Fig. 3).

-- The Tropical Rainfall Measuring Mission (TRMM) rainfall estimates and global flood potential forecast [Yilmaz et al, 2010]. Rainfall estimates are 3-, 24-, 72- and 169-hour rainfall accumulation. Flood potential forecasts are provided for 24-, 72- and 128-hour in advance. Real-time global estimation of flood areas using satellite-based rainfall and a hydrological model are run globally, every three hours at 0.25° resolution. Real-time product are produced within 6 h after observations made by TRMM.

-- Namibia dwelling unit database.

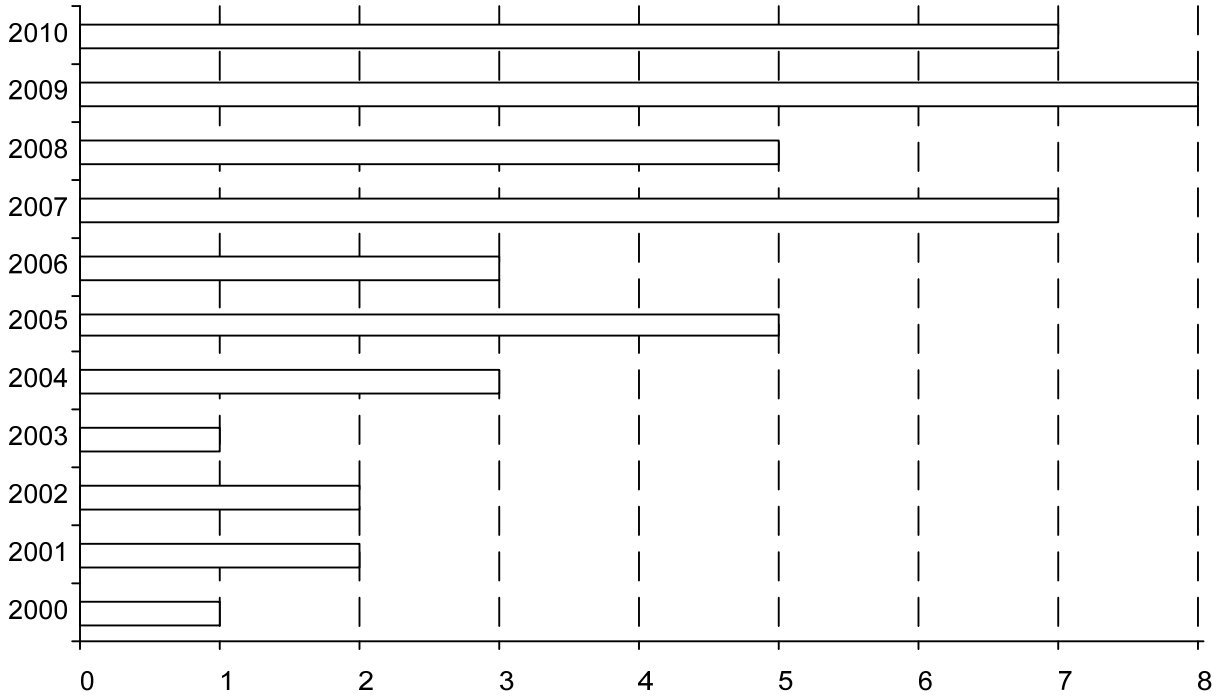

Figure 2. Distribution of the number of Landsat-5/7 images over 2000-2010 years

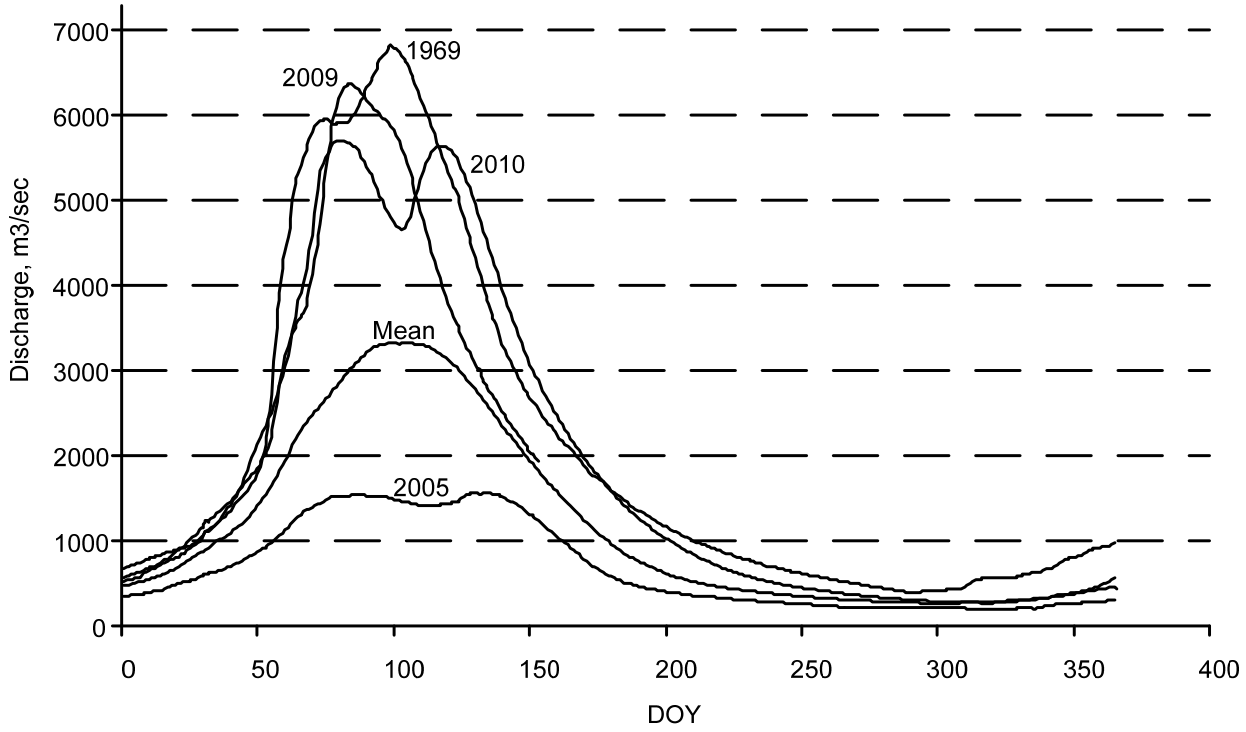

Figure 3. Hydrograph for 1969, 2009, 2010 and 2005 years along with the mean averaged for 1965 to 2010

Recurrence interval estimation

Based on river gauge data we estimated recurrence interval. For each year a maximum discharge was calculated, and these maximum values were sorted in descending order. The year with the maximum discharge was given m=1 magnitude, the year with the second maximum discharge was given m=2 magnitude and so on. These values along with total number records, n=46, were used to calculate recurrence interval (Weibull equation):

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

Table 1 shows recurrence interval for the top 10 years with maximum discharge.

A polynomial fit was constructed to predict discharge from recurrence interval. The following 3-order polynomial dependence was obtained:

$$y = 2969.8x^3 - 9567.7x^2 + 11163x + 1181, \quad (2)$$

where y is discharge, and $x = \log_{10}(R)$.

The obtained coefficient of determination was 0.99. Figure 4 shows the plot along with 95% confidence interval.

Table 1. Recurrence interval of floods for the Katima Mulilo region in Namibia

Magnitude, *m*	Year	Discharge, m^3/sec	*R*
1	1969	6817	47.0
2	2009	6365	23.5
3	1978	6251	15.7
4	2010	5704	11.8
5	1979	5675	9.4
6	1976	5568	7.8
7	2007	5564	6.7
8	1975	5409	5.9
9	1968	5312	5.2
10	1966	5276	4.7

Table 2 gives 10-, 50- and 100-year floods values.

Table 2. 10-, 50- and 100-year floods

n, year	Discharge, m^3/sec	Confidence interval
10	5746	[5419; 6073]
50	7093	[6654; 7532]
100	8993	[8131; 9855]

That is, probability of the flood with discharge exceeding 8993 m^3/sec in any given year is equal to 0.01.

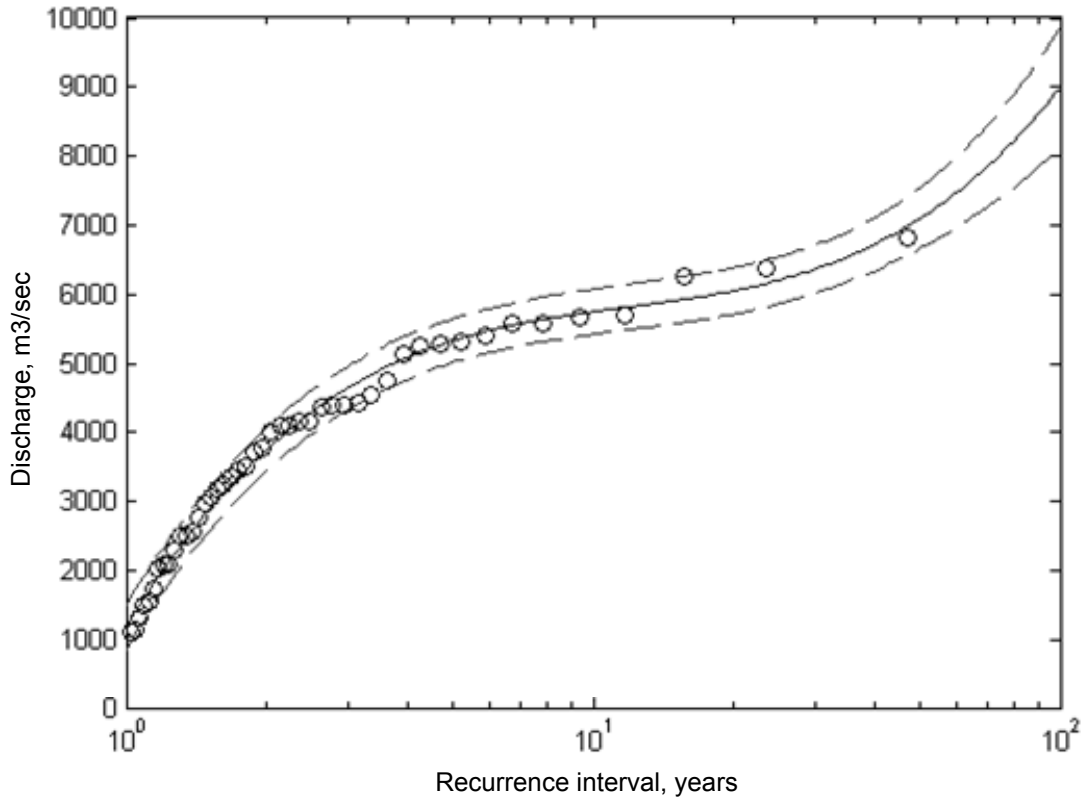

Figure 4. Dependency between maximum discharge values from recurrence interval

Flood hazard mapping using satellite data

We used a time-series of 44 Landsat TM and ETM+ images at 30 m spatial resolution to estimate flood probability density. Firstly, clouds and shadows were identified on Landsat images using the Automated Cloud-Cover Assessment (ACCA) Algorithm [Irish et al, 2006]. Also SLC-off pixels on ETM+ images were marked and removed from consideration. All these pixels were assigned with the "NoData" value, and were removed from the further analysis. At second, water bodies were detected using a density sliding method [Frazier and Page, 2000]. Therefore, each pixel in the image was assigned one of the following values: 0 - «No water» class, 1 - «Water» class, 2 - «No Data». Figure 5 shows the original and processed Landsat-5 image.

Two approaches were used to estimate probability density function. They were different in how images were integrated within the single year. Let A be the set of all satellite images, i.e. $A=\bigcup\{a\}$, where the image a is characterized by the following tuple:

$$a = \{y, doy, (i, j)\}, \tag{3}$$

where $a.y$ and $a.doy$ are the year and day of the year the satellite image was acquired, and $a(i,j) \in \{0, 1, 2\}$ is the value of image pixel with coordinates i and j.

Within the first approach, for each year we selected an image that was closest to the DOY with maximum discharge, and then aggregated these images into a probability of inundation map.

$$A_1 = \left\{ \forall y \in \{2000,\ldots,2010\} : a^* = \min_{a \in A, a.y = y} \left| a.doy - \text{doy\_min\_discharge(y)} \right| \right\}, \tag{4}$$

$$PI_1(i,j) = \frac{1}{\left| \{ a \in A_1 : a(i,j) \neq 2 \} \right|} \sum_{\{a \in A_1 : a(i,j) \neq 2\}} a(i,j). \tag{5}$$

Within the second approach, for each year we aggregated all available images into the single image by assigning a pixel the "Water" class, if at least on one of the images it was identified as the "Water" class. The same was for the "No Water" class. Therefore, the pixel was assigned the "NoData" value only if on all images it had the "NoData" value. These yearly images were then into a probability of inundation map:

$$A_2 = \left\{ \forall y \in \{2000,\ldots,2010\} : a^*(i,j) = 1 \text{ if } \exists a \in \bigcup \{ a \in A : a.y = y \} : a(i,j) = 1 \right\}, \tag{6}$$

$$PI_2(i,j) = \frac{1}{\left| \{ a \in A_2 : a(i,j) \neq 2 \} \right|} \sum_{\{a \in A_2 : a(i,j) \neq 2\}} a(i,j). \tag{7}$$

The main difference in these two approaches is the following. The image generated using the former approach will show flooded areas on the day of a peak discharge (or the closest day). Since the satellite images were not acquired on the day of maximum discharge we, obviously, will miss some flooded areas. Moreover, in the case of Katima Mulilo region there is latency between river flow at the gauge and flow coming to Liambezi Lake. In contrast, the latter approach allows us to identify all the pixels that were flooded during the flood season. Figure 6 and Figure 7 show two maps that were generated using both approaches. We can see that main differences between images are in the south-west part of the area. First approach does not allow us to capture latency in

flood wave that is coming through the region to the Liambezi Lake and to the south, while the second approach does.

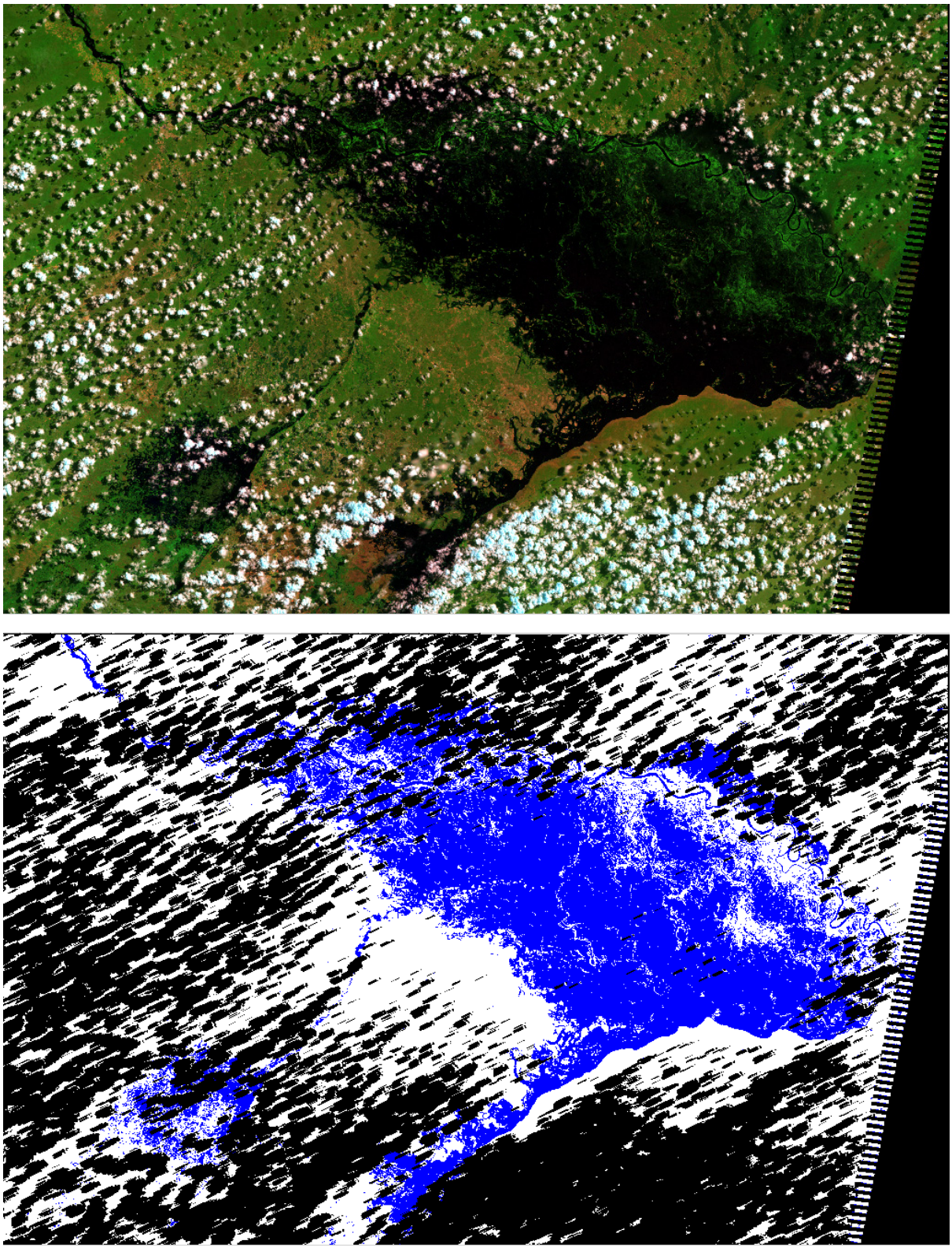

Figure 5. Original (top) and processed (bottom) Landsat-5 image acquired in 2010, DOY=81

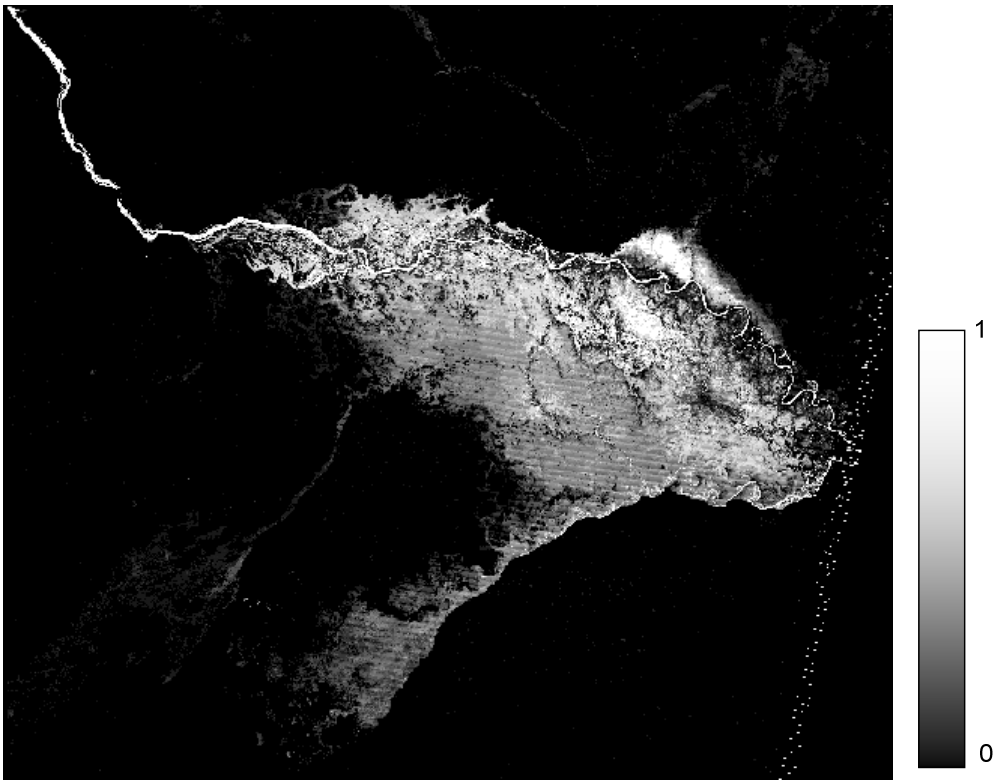

Figure 6. PI for the Katima Mulilo region, Namibia, obtained using Eq. (4)-(5)

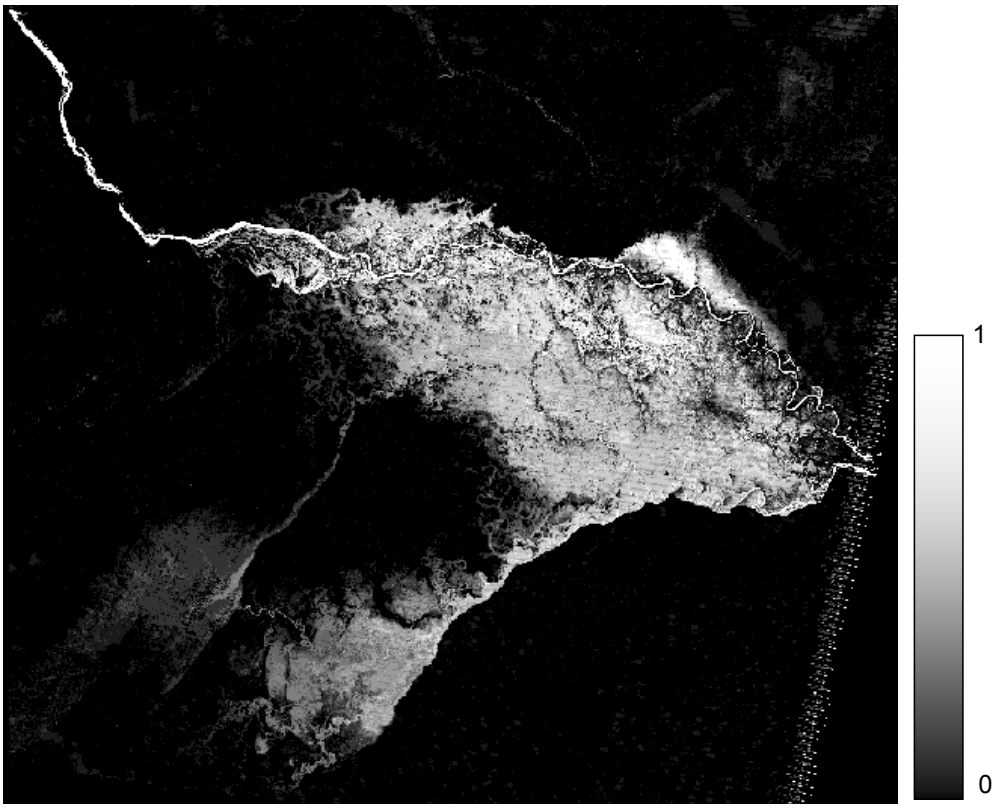

Figure 7. PI for the Katima Mulilo region, Namibia, obtained using Eq. (6)-(7)

Flooded area estimation

Based on the yearly flood extent maps obtained from Eq. (6) we calculated the area of flooded territories. In order to predict the expected flooded area we built a regression that predicts the flooded area in dependence of discharge (Fig. 8). The figure shows a good correspondence between the flooded area and discharge with coefficient of determination of 0.83. The only outlier is the values for the year of 2000. It could be probably explained that for this year a single satellite image was only available, and no maximum discharge was recorded on the date of image acquisition (there was a 1 day difference). In contrast, in 2003, when a single image was also available, the satellite image was acquired on the day when maximum discharge was recorded.

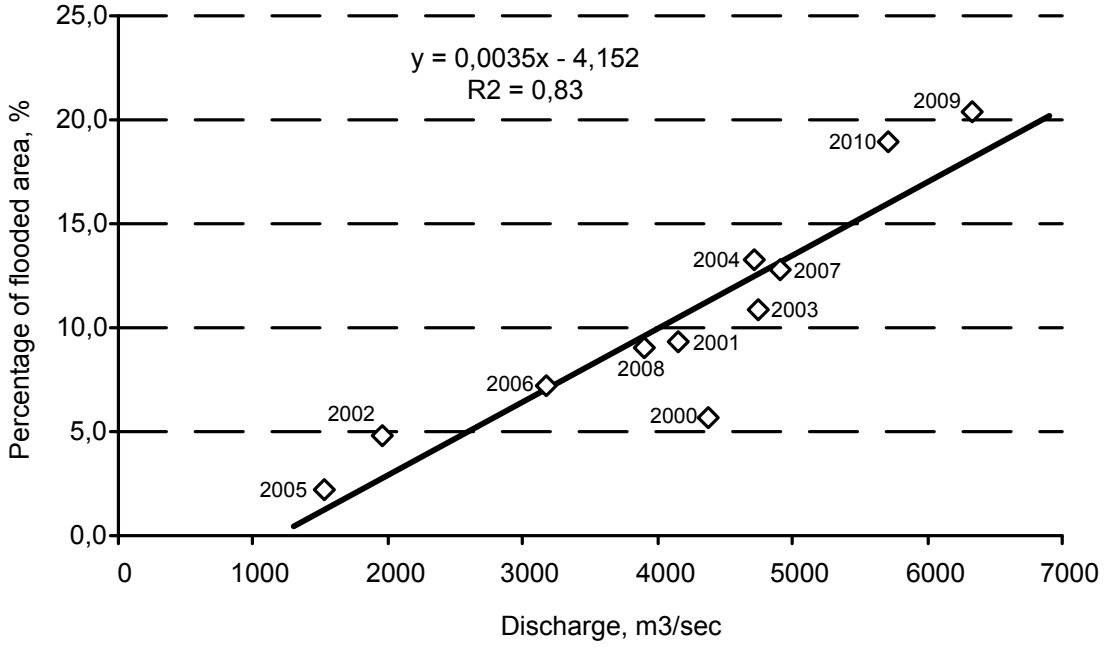

Figure 8. Dependence of the area of flooded territories from discharge

Another issue to be addressed is that we used a pixel counting method for area estimation. Since our classification algorithm does not provide 100% accuracy, the pixel counting method is usually downward biased [Gallego, 2004]. Counting pixels and multiplying by the area of each pixel will result in biased area estimates and should be considered raw numbers needing bias correction. One way to tackle this problem is to provide area frame sampling (AFS) data and then use a regression estimator to improve estimates [Carfagna and Gallego, 2005]. Since it is impractical to provide AFS using ground observations, AFS could be done by photointerpretation. Flooded waters can be reliably identified by visual inspection.

Flood risk mapping

The obtained flood hazard map was integrated with dwelling unit database to provide flood risk mapping. Such analysis allows us to identify the dwellings that are more likely to be inundated during the flood season, and specify probability of being flooded under different scenarios. Figure 9 shows integration flood hazard map with dwellings.

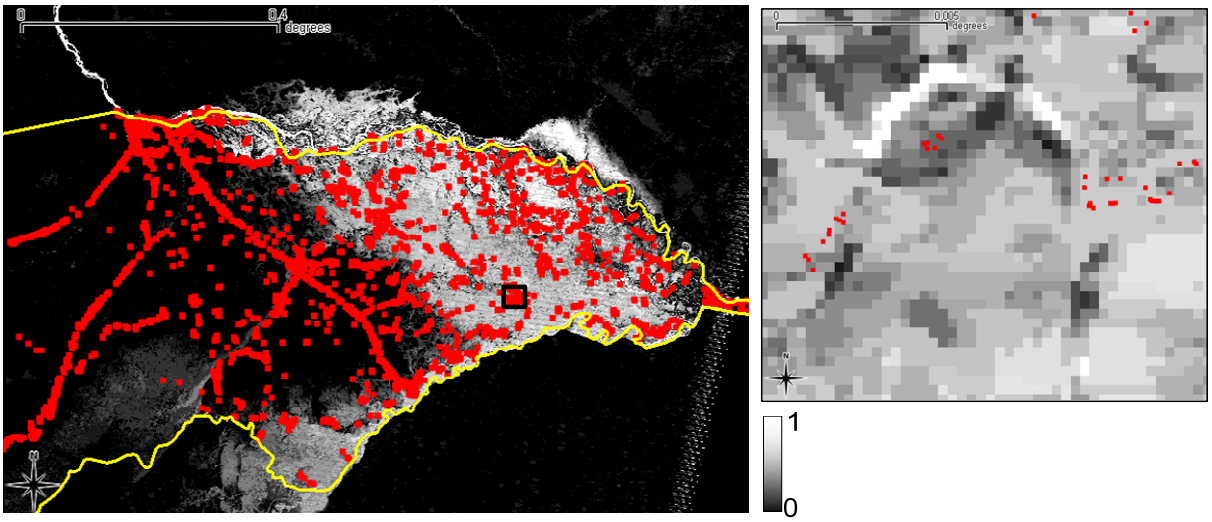

Figure 9. Flood risk map. Red squares show dwelling units

Conclusion, discussion and future works

In this paper we proposed a novel approach for flood hazard mapping by processing and analyzing a time-series of satellite data and derived flood extent maps. This approach is advantageous in cases when the use of hydrological models is complicated by the lack of data, in particular high-resolution DEM. Two approaches were investigated for generating flood extent maps for each year: by selecting an image with date of acquisition closest to the day when the maximum discharge was recorded, and integrating all flood extent maps available for the year. Due to the cloud cover and shadows the former method tends to miss areas that were flooded during the flood season, while the latter accounts for all areas that were flooded. Each pixel of the yearly flood extent map is viewed as Bernoulli distribution value, and maximum likelihood method was applied to estimate a *success probability* from sampling set. This parameter shows probability of inundation, and can be viewed as flood probability density function. Also, we believe that the derived flood extent maps will be very valuable in validating hydrological models once high-resolution DEM is available.

The future works should be directed: (1) to account for uncertainties in pixels with the “NoData” value (which can be either “Water” or “No water”); (2) to build a model that based on flood extent maps, gauge records and low-resolution DEM will predict flooded areas; (3) to provide flood risk mapping based on infrastructure facilities (e.g. roads, enterprises, etc); (4) to integrate radar satellite images with optical ones to reduce the effect of cloud cover.

Acknowledgement

The paper is published with financial support by the project ITHEA XXI of the Institute of Information Theories and Applications FOI ITHEA (www.ithea.org) and the Association of Developers and Users of Intelligent Systems ADUIS Ukraine (www.aduis.com.ua).

Bibliography

[Bates, 2004] P.D. Bates. Invited commentary: remote sensing and flood inundation modelling. Hydrological Processes, 2004, 18, pp. 2593–2597.

[Bates et al, 1997] P.D. Bates, M.S. Horritt, C.N. Smith, D.C. Mason. Integrating remote sensing observations of flood hydrology and hydraulic modelling. Hydrological Processes, 1997, 11, pp. 1777–1795.

[Carfagna and Gallego, 2005] E. Carfagna, F.J. Gallego. The use of remote sensing in agricultural statistics, International Statistical Review, 2005, 73, 3, pp. 389-404.

[Frazier and Page, 2000] P.S. Frazier, K.J. Page. Water Body Detection and Delineation with Landsat TM Data. Photogrammetric Engineering & Remote Sensing, 2000, Vol. 66, N 12, pp. 1461-1467.

[Gallego, 2004] F.J. Gallego. Remote sensing and land cover area estimation. International Journal of Remote Sensing. 2004, Vol. 25, n. 15, pp. 3019-3047.

[Hoes and Schuurmans, 2006] O. Hoes, W. Schuurmans. Flood standards or risk analyses for polder management in the Netherlands. Irrig. Drain., 2006, 55, pp. 113-119.

[Horritt, 2006] M.S. Horritt. A methodology for the validation of uncertain flood inundation models. Journal of Hydrology, 2006, 326, pp. 153-165.

[Irish et al, 2006] R. Irish, J. Barker, S. Goward, T. Arvidson. Characterization of the Landsat-7 ETM+ Automated Cloud-Cover Assessment (ACCA) Algorithm. Photogrammetric Engineering & Remote Sensing, 2006, Vol. 72, N 10, pp. 1179–1188.

[Jonkmana et al, 2003] S.N. Jonkmana, P.H.A.J.M. van Gelder, J.K. Vrijling. An overview of quantitative risk measures for loss of life and economic damage. Journal of Hazardous Materials, A99, 2003, pp. 1-30.

[Knight, 2006] D.W. Knight. Introduction to flooding and river basin modelling. In: River Basin Modelling for Flood Risk Mitigation, D. Knight and A. Shamseldin (Eds), 2006, pp. 1–20.

[Kussul et al, 2010] N.N. Kussul, B.V. Sokolov, Y.I. Zyelyk, V.A. Zelentsov, S.V. Skakun, A.Yu. Shelestov. Disaster Risk Assessment Based on Heterogeneous Geospatial Information. Journal of Automation and Information Sciences, 2010, Volume 42, Issue 12, pp. 32-45.

[Kussul et al, 2012] N. Kussul, D. Mandl, K. Moe, J.-P. Mund, J. Post, A. Shelestov, S. Skakun, J. Szarzynski, G. Van Langenhove, M. Handy. Interoperable Infrastructure for Flood Monitoring: Sensor Web, Grid & Cloud. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing (J-STARS), 2012, doi: 10.1109/JSTARS.2012.2192417.

[Lecca et al, 2011] G. Lecca, M. Petitdidier, L. Hluchy, M. Ivanovic, N. Kussul, N. Ray, V. Thieron. Grid computing technology for hydrological applications. Journal of Hydrology, 2011, Volume 403, Issues 1-2, pp. 186-199.

[Mostert and Junier, 2009] E. Mostert, S.J. Junier. The European flood risk directive: challenges for research. Hydrology and Earth System Sciences Discussions, 2009, Vol. 6, N 4, pp. 4961-4988.

[Rodriguez et al, 2009] J. Rodriguez, F. Vos, R. Below, D. Guha-Sapir. Annual Disaster Statistical Review 2008: The numbers and trends. Centre for Research on the Epidemiology of Disasters, Jacoffset Printers, Melin (Belgium), 2009, 33 p.

[Schumann and Di Baldassarre, 2010] G. Schumann, G. Di Baldassarre. The direct use of radar satellites for event-specific flood risk mapping. Remote Sensing Letters, 2010, Vol. 1, N 2, pp. 75-84.

[See and Abrahart, 2001] L. See, R.J. Abrahart. Multi-model data fusion for hydrological forecasting. Computers & Geosciences, 2001, Vol. 27, N 8, pp. 987-994.

[Yilmaz et al, 2010] K.K. Yilmaz, R.F. Adler, Y. Tian, Y. Hong, H.F. Pierce. Evaluation of a satellite-based global flood monitoring system. International Journal of Remote Sensing, 2010, vol. 31, no. 14, pp. 3763–3782.

Authors' Information

Sergii Skakun *– Senior Scientist, Space Research Institute NASU-NSAU, Glushkov Prospekt 40, build. 4/1, Kyiv 03680, Ukraine; e-mail: serhiy.skakun@ikd.kiev.ua*
Major Fields of Scientific Research: Grid computing, Sensor Web, Earth observation, satellite data processing, risk analysis.