

A PRELIMINARY APPROACH TO FLOOD RISK MAPPING AND FLOOD FORECASTING SYSTEM FOR THE LDCs

A. Albanese^b, F. Disabato^a, O. Terzo^c, R. Vigna^a, M. Giardino^b, L. Perotti^b

^aITHACA (Information Technology for Humanitarian Assistance, Cooperation and Action), Via P.C. Boggio 61 Torino, Italy – (franca.disabato, rossella.vigna)@polito.it

^bDept. of Earth Science, University of Turin, Via Valperga Caluso 35 Torino, Italy – (adriana.albanese, marco.giardino, luigi.perotti)@unito.it

^cISMB (Istituto Superiore Mario Boella), Via P.C. Boggio 61, Torino, Italy – terzo@ismb.it

KEY WORDS: Floods, Precipitation, Satellite, Database, Geomorphology, Hydrology, Mapping, Real-time

ABSTRACT:

The aim of the project is the development of an Early Warning system for flood events based on precipitation analysis and related to historical flooded area detection. The recent change of Humanitarian International Agencies in the way of approaching the emergency preparedness has led the UN (United Nations) system to create a tightened collaboration with different research centers, where ITHACA (Information Technology for Humanitarian Assistance, Cooperation and Action) is one of these, with the aim to find new suited technologies to support the interventions. It could be extremely useful mapping the risk scenarios and establishing a flood forecasting system in the so called LDCs (Less Developed Countries), where an extreme meteorological event usually involves crisis on the local socio-economic order, in terms of huge damages and a very large number of displaced and dead people. The methodology is based on the individuation of critical precipitation events in the last ten years on worldwide extent and the evaluation of related fields effects in terms of flooded areas; these will be extracted from satellite images using suitable classification procedures and analyzing geomorphological aspects to identify the potential floodable areas. Pluviometric thresholds, determined by different analysis (hydrological and statistical), executed on historical rainfall data, will be stored in a database structure where they will be compared with real-time rainfall values. This Early Warning System allows to monitor near real time rainfall values, creating an alert for critical values and producing a map with a flood scenario referred to the past event with similar rainfall field.

RESUME:

L'objectif du projet est centré sur la détection préventive des inondations à partir des événements de précipitations. Les analyses sont conduites à partir de l'étude des données historiques de pluies qui indiquent une détection des régions inondées. Récemment les Agences Humanitaires Internationales ont changé leur méthodes de gestion des situations d'urgences créant une étroite collaboration entre les Nations Unies et plusieurs centres de recherches, ITHACA (Information Technology for Humanitarian Assistance, Cooperation and Action) est l'un d'entre eux. Un des objectifs est de mettre en place de produits technologiques innovants à support des interventions. Se doter d'un système de prévention, basé sur la cartographie des scénarios de risques d'inondations, pourrait être extrêmement utile pour les pays en voie de développement, où généralement des phénomènes météorologiques d'extrême intensité provoquent une crise locale au niveau socio-économique. Les conséquences généralement sont graves, tant en pertes de vies humaines que pour les énormes dommages provoqués aux structures. La méthode est basée sur la détermination d'événements critiques de précipitations durant ces dix dernières années, à une échelle mondiale. L'étude prend aussi en considération l'évaluation des conséquences des inondations grâce à une analyse géomorphologique à partir d'images satellitaires, l'objectif étant d'identifier des zones à risque d'inondations. En se basant sur les données historiques, des seuils de pluviométrie sont déterminés à partir de plusieurs analyses (hydrologiques et statistiques) et sont ensuite mémorisés dans la base de données pour pouvoir ensuite procéder à une comparaison avec les valeurs de pluie en temps réels. Le système de pré alerte permet de contrôler, dans une mesure proche du temps réel, les valeurs de pluies et de lancer une alerte pour les valeurs de pluie critiques en fournissant une carte des différents scénarios d'inondations basée sur des événements du passé similaires.

1. INTRODUCTION

1.1 Workplace and general purposes

The Early Warning project is developed within the ITHACA organization (Information Technology for Humanitarian Assistance, Cooperation and Action), founded by Politecnico di Torino and SiTI (Istituto Superiore sui Sistemi Territoriali per l'Innovazione) in cooperation with WFP (World Food Programme), the largest UN operational Agency. ITHACA conducts research activities in the fields of Geomatics, in particular it is devoted to monitor, analyze and forecast natural

disasters, with special attention to less developing countries (LDCs).

The collaboration between ITHACA and WFP has increased efficacy in approaching emergency preparedness through different activities such as early warning, early impact, spatial data infrastructure, snow cover, web applications (www.ithacaweb.org). In particular, the Early Warning phase represents one of the most important purposes, considering that WFP is comprised inside the IATF/DR (Inter-Agency Task Force for Disaster Reduction) and recognizes the importance of

Early Warning systems to save lives, to protect properties and to secure livelihood and sustainable development.

In this paper the approach used by ITHACA to create the Early Warning System for Flood Events will be explained, which was specifically defined by WFP as a necessary instrument for the evaluation of potentially floodable areas and to understand the gravity of the event.

1.2 Data used for the project

This Early Warning system is based on precipitation analysis and it is related to historical flooded areas detection.

The choice of the dataset used in this project is conditioned by specific agreements with WFP, which prefers data of public domain with global coverage.

Different precipitation products have been used, coming from different sources.

Data from the GPCC (Global Precipitation Climatology Centre) are:

- The Full Data Reanalysis Product, that varies from less than 10,000 to more than 43,000 stations with a geometric resolution of 1°x1° latitude/longitude, temporal resolution of a month and temporal coverage from 1951 to 2004.
- The Monitoring Product has continuously been done every month starting from 1986. It derives from 7500 ground stations with a geometric resolution of 1°x 1° lat/lon and temporal resolution of a month.

From TRMM (Tropical Rainfall Measuring Mission) satellite the 3B42 product has been used, whose algorithm uses, as source data, High Quality microwave - infrared (IR) and root-mean-square (RMS) precipitation-error estimates. These gridded estimates are on a 3 hour temporal resolution and a 0.25 degree by 0.25 degree spatial resolution in a global belt extending from 50 degrees South to 50 degrees North latitude, its temporal coverage is from 1998 to present (Huffman et al., 1995; Huffman, 1997).

For the real time the product 3B42RT has been used. This product is a combination of the TRMM real-time merged passive microwave (HQ; 3B40RT) and microwave-calibrated IR (VAR; 3B41RT). The 3B42RT combinations are computed every 3 hours but are posted to the web about 6 hours after observation time. They have a grid resolution of 0.25 degrees by 0.25 degrees lat-lon over the latitude band 50 degrees North Sud. The observation are available from 2002-12-29 till today (Huffman, 2007). These estimates have experimental nature and were developed to apply new concepts in merging quasi-global precipitation estimates and to take advantage on the increasing of availability of input data sets in near real time.

The analysis has been lead on basin scale since flood effects depend on the characteristics of river basin. A river basin is considered as an extent of land where water drains downhill into a body of water, on the basis of the terrain elevation.

It has been used a GIS layer of hydrographical basin (HYDRO1k from GTOPO30) which is constituted of a subdivision in different levels, increasing details from level 1 to level 6. It be comes necessary to resample the precipitation data on the basis of this layer: the grid of the TRMM data (0.25 x

0.25 degrees) have been overlapped upon the level 6 basin, obtaining a territorial entity named *geodb_id* and rainfall values have been associated to this entity.

The analysis of the effects on the field is carried out using MODIS Land Imagery, from NASA EOS which have a geometric resolution of 250 meters, daily resolution and world coverage.

Geomorphological analysis is executed on Landsat Imagery, from NASA EOS with a geometric resolution of 30 meters and using DTM from Shuttle Radar Topographic Mission (SRTM), a NASA product with a geometric resolution of 90 meters and coverage from 180 E-W and 60.005 North – 56.005 South.

2. METHODOLOGY

2.1 Definition of extreme rainfall periods

As a first step of this project, a rough definition of the different extreme rainfall periods have been set out, on the basis of the pluviometric regime derived from the analysis of 50 years worldwide precipitation data, that belong to GPCC (paragraph 1.2).

Starting from these data, it has been calculated with an automatic procedure the mean monthly value over the whole period (1951-2007), for each grid cell (1°x1° lat/lon). In this way the most rainy months have been found and the results have been plotted in a map where different colours show when the maximum occurred; in this map isolines of average annual rain are also drawn (figure 1).

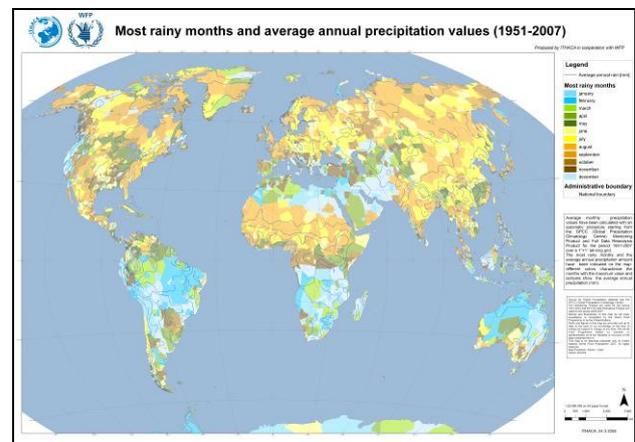


Figure 1. Most rainy months and average annual precipitation values on the period 1951-2007.

2.2 Early Warning System architecture

The second step of this project is the individuation of the different extreme meteorological events that occurred in the last ten years and the related calculation of cumulated rainfall values during these events. The use of only ten years of data is due to the necessity of compare satellite images (from MODIS) and satellite rainfall data.

To offer a service that monitor precipitation all over the world in near real time involves the use of data with high temporal resolution (near hourly), that imply in its turn a huge amount of data. Only a database can store and extract information from this enormous data set.

In consideration of this aspect it has been decided to use Oracle 10g. For the architecture it has been chosen a solution that takes in consideration both the historical and real-time data loading (figure 2).

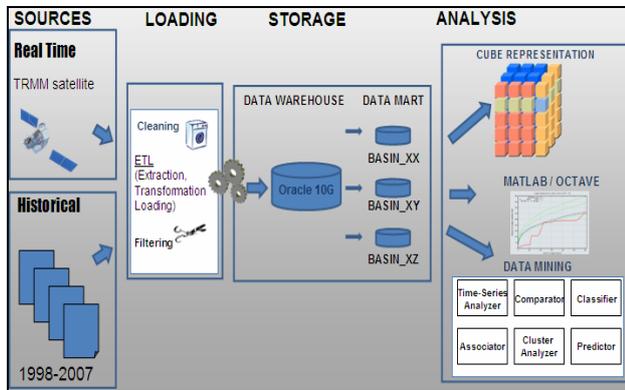


Figure 2. Early Warning Architecture

3. DEVELOPMENT

3.1 Pre-elaboration

The gridded data of the TRMM 3B42 product need to be transformed from the raster format, where the rainfall estimates are associated to lat/lon values, into the Ascii format required by the database architecture, where the rainfall estimates values are associated to *geodb_id* index and their temporal collocation. In the general oracle architecture, each *geodb_id* is associated to a latitude, a longitude and to an area value, which indicates the area covered by the single 0.25°x0.25° TRMM product pixel referenced on a level 6 basin. It has been assumed that the precipitation value remains constant in the whole area of the pixel.

According to these considerations, the processing steps for the transformation are:

- identification of the temporal collocation of the product (year, month, day, hour);
- reading of the rainfall estimate value corresponding to each *geodb_id*;
- output of the requested output Ascii file, which provides *geodb_id*, temporal collocation and associated rainfall estimates.
-

This transformation is performed by an ITHACA developed routine, elaborated in a suitable environment for raster elaboration, the IDL (Interactive Data Language).

3.2 Data Loading

Two different approaches have been considered; the first one is for the historical rain data and the second one is for the real time rain data.

The ETL process is the first cut of filtering information by deleting the sea regions and the error values. A data quality check is executed before the storage processes. There are only three information inside the table:

- the spatial georeferenced indications;
- the period information;
- the rain value.

The load process for historical data is executed by using *Oracle sqlloader*. The time expected for loading one year of rain values related at a basin of level 1 is 100 minutes. One year is split in 2920 files and each file have 311999 records.

The load process for real time data is executed by using an ftp for getting the 3 hours file and *Oracle sqlloader* for loading data in the database.

It is being used an HP workstation xw4600 (intel core 2 duo 2.66 Ghz and with 4 Gb of RAM memory).

3.3 Data Modeling

It has been used a mixed approach: one, called top-down, which offers a general view of the problem and another one, called bottom-up, where the individual parts of the system are developed in detail. In particular the data warehouse, that is a repository of an organization electronically stored data to facilitate reporting and analysis (Golfarelli & Rizzi, 2002), has been planned in bottom-up. The different data-mart have been defined inside the data warehouse and the single one has been designed in top-down method, from conceptual to fact schema. The data model has been implemented with the star schema concept that is composed of two important domains: the dimension table and the fact table.

The *dimension table* is defined with a set of DT_1, \dots, DT_n relations. Every DT_i is characterised with a primary key d_i in association of a set of attributes. The advantage is about the different levels of aggregation for the data analysis (for example year, month, day).

The *fact table* is an *FT* relation and it is defined with the set of external DT_1, \dots, DT_n dimension table key. The primary key of *FT* is composed by the association of all external DT_i key. Fact tables have several attributes for the measures considered. In our cases the measure is the rain value.

Two levels of dimensional aggregation have been defined: the time and spatial references. Time reference is defined with the year, month, day and 3 hours fields. A *time_ref* table has been created for filtering data in a specific period.

The spatial reference is defined with the georeferenced information from level 1 to level 6 basin.

The *basin_ref* table has been created, where there are the geodatabase identifications for all world regions and where it is easy to select a specific region by selecting the continent of interest and the basin level needed. For each database identifier there are:

- the continent;
- the single information about level 1 to 6 of the basin;
- the longitude and latitude;
- the area covered by the single pixel referenced on a level 6 basin;
- the unique identifier for each lat/lon pixel;
- the referenced table with the rain values data for a specific level 1 basin.

The early warning project takes into consideration 47 basins of level 1, corresponding to 47 fact tables. For performance reasons the fact table contains only one year of rain data. Each table has an order of 80 millions of records.

For increasing the level of performances during the database interrogation, it has been decided to elaborate two types of data aggregation: the 3 hours and day information of rain.

The schema (figure 3) reports the historical rain storage Data Model.

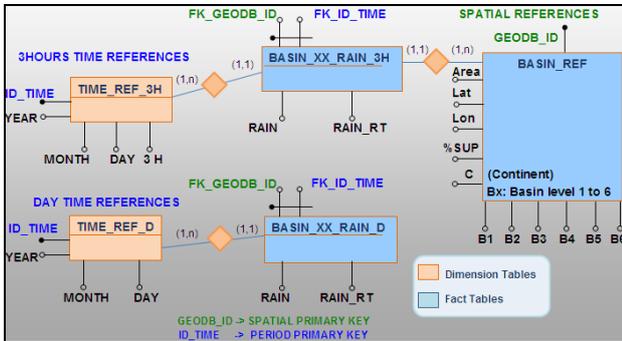


Figure 3. Early Warning Data model (historical data)

3.4 Analysis inside historical database

It has been decided to test the whole method in a specific area, constituted by a basin of level 6 in North Korea, that contains the capital city Pyongyang. This area has been chosen on the basis of the results obtained from precipitation study (paragraph 2.1), which shows critical events in North Korea in July and August. Furthermore, it has to be said that North Korea represents one of the WFP special countries of interest; in fact the Democratic People’s Republic of Korea (DPRK) continues to suffer from widespread food shortages, due to economic problems, limited arable land, lack of agricultural machinery and energy shortages, it means that cereals production remains well below the minimum required (source: www.wfp.org/english/).

3.4.1 Spatial and temporal graphical representation

The database architecture allows to create specific selection of data with the both temporal and spatial parameters of the area of interest, using a specific query in sql language (table 1). For the selection and projection data we have fixed values for the two dimensional references. In figure 4 it is represented a common slice-and dice operation in a Data Warehouse by selecting all rain values referred to the year 2000 and all the georeferenced areas corresponding to the level 6 basin (North Korea); this kind of graphics, generated by Octave, gives a good indication of what rainfall periods are and what their severity is.

SELECT GEODB_ID, FK_TIME, RAIN FROM basin_29_3h_dim WHERE fk_time in (select id_time FROM time_ref WHERE y=2003 and m=1 and d=1) and geodb_id in (select geodb_id FROM basin_ref WHERE b6=2911960)	FIELDS SELECTED
	TEMPORAL FILTER
	SPATIAL FILTER

Table 1. Query in sql language to select data in temporal and spatial dimension.

3.4.2 Hydrological analysis

Floods are usually caused by great rainfall amounts during discrete periods of time. Consequently, it becomes necessary to estimate the rainfall height of an event, for several durations, and to confront it with the historical maxima records for the correspondents durations.

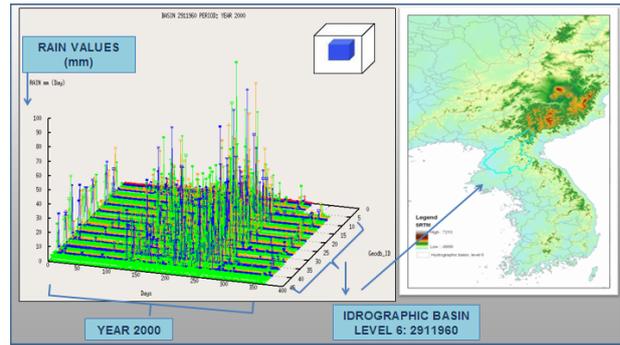


Figure 4. 3D graphical view (spatial and temporal).

A well known procedure includes the extraction of annual rainfall maxima for several durations to obtain the average of the annual rainfall maxima for each durations (figure 5).

The empirical curve is generally well interpolated by an analytic curve, modulating its two coefficients *a* and *n*.

$$h = a \cdot d^n \quad (1)$$

where *h* = height of rainfall (mm)
d = duration of rainfall (hour)

It is now possible to determine the entity of any rain event (in terms of frequency) comparing the cumulated rain of the event with the historical values. In fact, for any temporal instant in the historical series of the rainfall data, it is possible to create a cumulated curve of rainfall heights for several durations, by calculating the total rainfall during *n* previous hours (where *n* varies between 0 and a sufficiently number of hours, in function of the dimensions of the river basin). In figure 6 some critical curves are plotted in black, while the red line refers to a moment characterized by less cumulated precipitations, calculated on a period that extends until approximately 10 days before the considered moment (29/2/2008, hours 21).

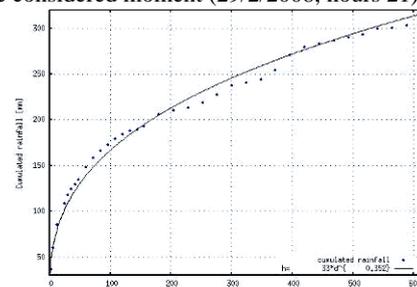


Figure 5. Empirical curve showing the average of the annual maxima of rainfalls for different durations and related analytic curve.

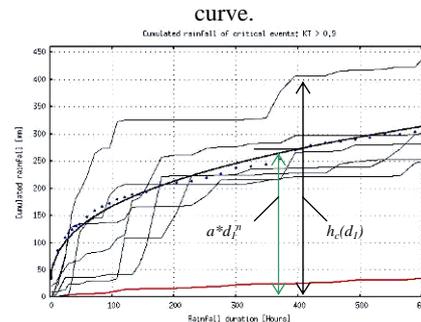


Figure 6. Comparison between different events and reference calculated curves.

The comparison allows to evaluate the criticality of the event through the calculation of a *KT* factor.

$$KT = h_c(d_j) / (a * d_j^n) \quad (2)$$

where h_c = cumulated rainfall at d_j duration
 d_j = duration

The procedure of individuation of the extreme rain events has been created with OCTAVE (version 2.9), a MATLAB porting open source software. It has been created a weighted average of all rainfall values of *geodb_ids* contained by a basin of level 6 processing the whole historical precipitation data set (ten years). The critical events have been individuated extracting the higher *KT* related to the durations considered critical for the basin in consideration. These critical durations depend on the dimensions, soil and hydrological characteristics and other parameters of the basin; in fact the procedure takes in consideration the time lag t_L of the basin, calculated with the formula (Chow et al., 1988):

$$t_L = 0.6 * t_c \quad (3)$$

where t_c = time of concentration (the time required for a drop of water to travel from the most hydrologically remote point in the subcatchment to the point of collection)

This procedure can be used for monitoring in real time, creating each time the curve of the rain cumulated for the last moment in which rain data are available.

3.4.3 Geomorphological analysis

The geomorphological diagnosis method aims to identify areas potentially at risk of flooding along watercourses by evaluating the present and the past geomorphological settings, both at the scale of drainage basins and of single river segments.

The geomorphological analysis evidences diagnostic erosional and depositional landforms along the fluvial environments: scarps and terraces risers, abandoned streamcourses, cutoff meanders, and so on; they are interpreted in terms of linear features of natural instability and areas prone to erosion, floods and fluvial deposition (Richards, 1982; Carling & Petts, 1992). These elements are checked in particular to highlight the flood prone areas; to give an example, alluvial plains where the low slopes can cause the reconnection of abandoned landforms as relict channels to the current rivercourse (figure 7).

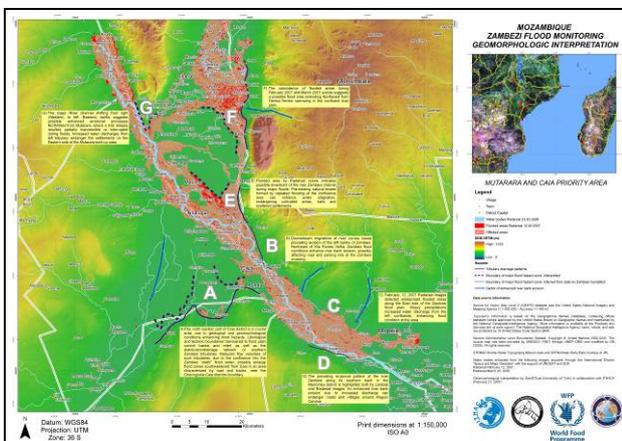


Figure 7. Geomorphological analysis for the Zambesi floodplain near Mutarara (Mozambique).

The analysis is executed mainly on satellite images and DSM (Landsat TM/ETM and SRTM). The multispectral characteristic allows to combine different bands obtaining opportune synthesis that highlight particular landforms; indexes like NDVI and the thermal Band 6 have been used deriving vegetation and surface temperature information. Different morphometric parameters such as elevation and slope are also used to define the flood prone areas (figure 8).

The results of geomorphological analysis from Landsat imageries merged with digital elevation model will be stored in vector format in the database.

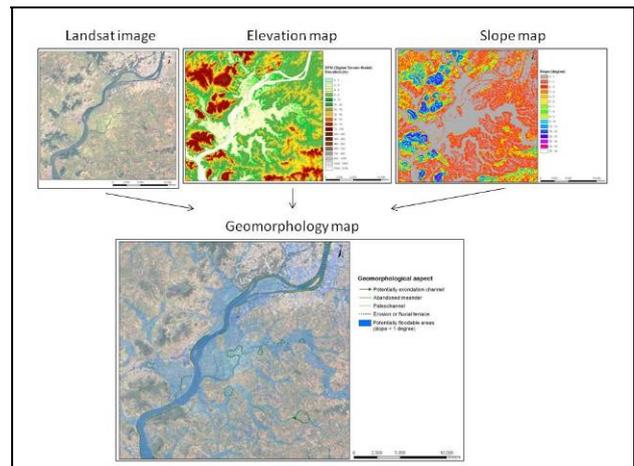


Figure 8. Geomorphological analysis from Landsat image and DSM processing for the Taedong floodplain near Pyongyang (North Korea).

3.4.4 Classification techniques and extraction of flooded areas

Flood events will be divided into different classes of severity according to the geographic region and the period of the year. An historical satellite-detected flooded area will be associated to each class of the event at the scale of a basin.

Flooded areas are detected from the classification of the MODIS/Terra Surface Reflectance Daily product (MOD09GQK and MOD09GQ), which has a spatial resolution of 250 meters and provide a daily estimate of the surface spectral reflectance for two bands, the red one (620-670 nm) and the IR one (841-876 nm). The advantage of this product is that it contains the surface spectral reflectance as it would be measured at ground level in the absence of atmospheric scattering or absorption.

In the test phase of the methodology these data are classified with the supervised Maximum Likelihood Method only for the test area (the North Korea).

Finally the results of the geomorphological analysis will be integrated to the result of classifications, merging areas classified as water and floodable areas. This data will be stored as shapefiles into the geodatabase used by Ithaca.

3.4.5 Identification of pluviometric thresholds

Starting from the calculation of the KT factor it has been extracted a list of critical events to be correlated with classified images; different scenarios have been associated to cumulated rainfall for each sub-basin, creating different associations between rainfall and field effects (in terms of severity of flooding). For every sub-basin the amount of rain that can cause flooding has been stored in the DB as a pluviometric threshold.

4. CONCLUSION AND FUTURE PLANS

Through the historical analysis of TRMM rainfall data (3B42) merged with analysis of satellite images (MODIS) it is possible to create a pluviometric thresholds database, associated to related possible flood scenarios.

This dataset is necessary to build an operative Early Warning System in near real time.

The system is connected to the TRMM ftp site every *dt* to allow the 3B42RT data downloading, the data are loaded in the procedure, which is constituted by data pre-processing, data loading and data analysis (figure 2). The rainfall data go on loading since the thresholds achievement, when the alert starts; in this case the system is logged to the ITHACA geodatabase, where shapes features of flooded areas have been stored; this allows the creation of maps that give to WFP useful indications about where and how a pluviometric event takes place, and what its consequences could be in terms of flooded areas.

The whole system is near real time because the rainfall data are available with a lag of 6-7 hours every 3 hours.

As further development, it is in progress the elaboration of an algorithm for an automated detection of the flooded areas with a global coverage. This algorithm has to receive as input the date and the area of the event and has to provide as output the flooded areas. Since the most important problem for the classification in the presence of a cloud coverage, a cloud and cloud-shadow masking process need to be performed. Furthermore, a compositing algorithm is necessary for the presence of no-data area due to cloud cover.

According to these considerations, the intermediate steps between input and output are:

- automated detection of satellite imagery according to date and area of interest;
- data masking from cloud effects;
- data compositing;
- data classification;
- data extraction of flooded areas.

Furthermore a comparison between the two different types of rainfall data, TRMM 3B42 and 3B42RT is needed to find a correction factor and find any existing biases (Liu et al., 2003).

REFERENCES

Alfieri L., Laio F. & Claps P., 2008. *A simulation experiment for optimal design hyetograph selection*. Hydrol. Process., 22, pp. 813-820.

Applications of GTOPO30 and HYDRO1K:

<http://edc.usgs.gov/products/elevation/gtopo30/hydro/>

Carling P.A. & Petts G.E. (eds), 1992. *Lowland floodplains rivers: geomorphological perspectives*. Wiley, Chichester.

Chow V.T., Maidment D.R. and Mays L.W., 1988. *Applied Hydrology*. McGraw-Hill, New York, USA.

Golfarelli M., Rizzi S., 2002. *Data Warehouse, Teoria e pratica della progettazione*. Mc Graw-Hill.

Hamid Minoui, Practical Space Management in data Warehouse Environments
www.dbspecialists.com/presentations/dw_space.ppt

Huffman G.J., 1997. Estimates of Root-Mean-Square Random Error for Finite Samples of Estimated Precipitation. *J. Appl. Meteor.*, pp. 1191-1201.

Huffmann G. J., 2007. Readme for accessing experimental real-time TRMM Multi-satellite Precipitation analysis (TMPA-RT) Data Sets
ftp://mesoa.gsfc.nasa.gov/trmmrt/docs/3B4XRT_README.pdf

Huffman G.J., Adler R.F., Rudolph B., Schneider U. and Keehn P., 1995. Global Precipitation Estimates Based on a Technique for Combining Satellite-Based Estimates, Rain Gauge Analysis, and NWP Model Precipitation Information. *J. Clim.*, 8, pp. 1284-1295.

ITHACA:
www.ithaca.polito.it

Liu Z., Riu H., Teng W. & Chiu L., 2003. Online intercomparison of TRMM and global ridge precipitation products. *Proc. 17th Conf. Hydrology, 2003 AMS Annual Meeting, Long Beach, CA*.

MODIS Imagery:
<ftp://e4ftl01u.ecs.nasa.gov/MOLT/MOD09GQ.005>
<http://deleenn.gsfc.nasa.gov/~imswww/pub/imswelcome/index.html>

GPCC dataset:
<ftp://ftp-anon.dwd.de/pub/data/>

Richards K., 1982. *Rivers, Forms and Processes in alluvial Channels*. Methuen, London and New York.

Tufféry S., 2005. *Data Mining et statistique décisionnelle. L'intelligence dans les bases de données*. Editions TECHNIP.

TRMM 3B42:
http://disc.gsfc.nasa.gov/data/datapool/TRMM_DP/01_Data_Products/02_Gridded/06_3-hour_Gpi_Cal_3B_42/

TRMM 3B42RT:
<ftp://trmmopen.gsfc.nasa.gov/pub/merged/mergeIRMicro>

World Food Programme:
www.wfp.org/english/