Integration of remote sensing data and GIS for accurate mapping of flooded areas

P. A. BRIVIO†, R. COLOMBO, M. MAGGI and R. TOMASONI
Remote Sensing Department, IRRS, National Research Council, Via Bassini 15, 20133-Milano, Italy

(Received 8 April 1999; in final form 11 July 2000)

Abstract. This paper describes a synergetic use of satellite radar images and ancillary information to detect flooded areas at their peak and evaluates its potential with mapping. The procedure was tested on the catastrophic flood that occurred in Regione Piemonte in Italy in November 1994. Two ERS-1 synthetic aperture radar (SAR) images were processed, one acquired one month before the flood and the other acquired three days after the event. Visual interpretation and two different thresholding techniques were performed. The flood map derived shows only a small fraction (20%) of the actually flooded lands because of the time delay between the flood peak and the satellite overpass. To overcome this limitation, the authors developed a new procedure to estimate the flooded area at the peak time by integrating the flooded area from SAR imagery with digital topographic data from a GIS technique. This method allowed inundated areas covering 96.7% of the flooded area officially recorded by the local government to be mapped. The proposed procedure is suitable for mapping flooded areas even when satellite data are acquired some days after the event, thus overcoming the constraint of temporal resolution in the application of SAR imagery in hydrology.

1. Introduction

In densely populated areas, both in tropical and temperate regions, floods are one of the most important natural disasters as they inundate large areas causing damage to agricultural crops and property, road and rail links, and the loss of human lives. Appropriate land use planning, which requires accurate knowledge of flood extent for locating flood prone areas, is a key tool to improve flood management and to mitigate its catastrophic effects. The knowledge of the spatial extent of inundated areas is essential both during the emergency phase, when it is necessary to have an overall view of the phenomenon in order to plan relief efforts, and after the event for the assessment of damaged areas by authorities and insurance companies.

Traditional methods of flood mapping are based on ground surveys and aerial observations, but when the phenomenon is widespread, such methods are time consuming and expensive; furthermore timely aerial observations can be impossible

†e-mail: brivio@irrs.mi.cnr.it
due to prohibitive weather conditions. An alternative option is offered by satellite remote sensing (RS) technology. In recent decades optical data acquired by sensors onboard spacecraft have been used in many studies to map inundated areas over regions characterized by very different conditions in climate, morphology and land use. These approaches for flood extent evaluation have involved digital or visual analyses of Système Probatoire pour l’Observation de la Terre Multispectral (SPOT XS) scenes of Europe and Asia (Robin and Le Rhun 1989, Blasco et al. 1992), Landsat Thematic Mapper (TM) data in North America (Lougeay et al. 1994) and Landsat Multispectral Scanner (MSS) images of West Africa (Brivio et al. 1984, Zilioli et al. 1993). Thermal data collected during the night-time by NOAA-9 AVHRR (National Oceanic and Atmospheric Administration Advanced Very High Resolution Radiometer) were also used by Barton and Bathols (1989) to map the extent of flooding in central Australia. An estimate of the extent of flooded areas was conducted by Berg et al. (1981) using the NOAA satellite thermal infrared sensor. Despite their coarse spatial resolution, AVHRR images have the advantage of high temporal resolution, and sometimes are the only available source of information.

Bad weather conditions during and after flood events can represent a strong constraint to the utilization of optical remotely sensed data. For this reason, optical sensors are generally used to assess inundated fields only some days after the event, either by recognition of fluvial sediments left on the land (Rosso 1995) or by the detection of vegetation stress (Michener and Houhoulis 1997). On the contrary, spaceborne radar systems, because of their exclusive cloud penetration capacity, offer a primary tool for real-time assessment of flooded areas. Different methodologies have been developed using special sensor microwave imager SSM/I (Tanaka et al. 2000) or single-band (Sharma et al. 1996) and multi-frequency (Ormsby et al. 1985, Hess et al. 1995) synthetic aperture radar (SAR) data. The disadvantages of the use of radar sensors lie in the difficult classification of the acquired signal because of the influence of complex ground and system variables. This is particularly true when single-band and single-polarization data are available, as is the case in the ERS-1/2 satellites. To improve this situation, multi-temporal techniques based on detection of changes between radar images acquired before and after the inundation event are usually recommended (Badji et al. 1994, Wang et al. 1995, Profeti and Macintosh 1997, Takeuchi et al. 1999). Moreover if a high temporal resolution dataset is available, change detection analysis can be used to produce a flood evolution map (Noyelle et al. 1995, Oberstadler et al. 1995, Laugier et al. 1997).

If radar images are acquired some days after the event, when only a few areas are still submerged by water, it is advisable to use a multi-sensor approach. In this case flooded areas as derived from radar data are complemented with information extracted from optical images, such as the areal extent of fine material left during the flood (Imhoff et al. 1987, Bonansea 1995). Another possibility is the integration in a Geographical Information System (GIS) of radar imagery acquired during the flood event with information derived from digital topography (Giacomelli and Mancini 1996, Brakenridge et al. 1998).

This paper proposes a new methodology which, integrating results obtained from ERS SAR images and topography in an appropriate GIS methodology, allows a satisfactory mapping of inundated areas, even when satellite data are acquired some days after the flooding event. The catastrophic flood event that occurred in Regione Piemonte (northern Italy) on November 1994 is used to illustrate the methodology.
2. Mapping flooded areas through an integrated RS and GIS approach

The magnitude of the flood and the extent of flood-affected areas depend on the intensity of rainfall, its duration, the watershed topography and its conditions at the time of the heavy rainfall. Observation from a RS satellite provides a time-instantaneous portrait of the flood stage over a wide area. However, the period of submergence can vary from hours to weeks and the time delay between the flood peak phase and the satellite observation may severely limit the capabilities of mapping the widest spread of flooding. Figure 1 illustrates the relationships between flooded area extent, flood peak phase and SAR image acquisition.

2.1. Flood mapping on SAR imagery

The fundamental characteristic recorded on a radar image is the backscattering coefficient, which may vary from surface to surface. The strength of the returned signal from the surface is influenced by combined system and ground parameters, including the average surface roughness and soil dielectric properties. Horizontal smooth surfaces, such as water bodies, reflect nearly all incident radiation away from the sensor and the weak return signal is represented by dark tonality on radar images with the result that standing water areas are easily recognizable. This specular reflection can be decreased by bad weather conditions and/or the presence of vegetation, roughening the surface and making the detection of flooded areas more difficult (Laugier et al. 1997). Moreover, SAR data are subject to speckle, a multiplicative random noise that considerably reduces the interpretability of the images and limits classification techniques, and SAR images have to be filtered in order to increase the signal-to-noise ratio. The enhanced Lee adaptive filter is an efficient tool for reducing SAR image speckle by removing high-frequency noise while preserving edges or sharp features in the image (Zhengao and Fung 1994). Shadows at the

![Figure 1](image-url)

Figure 1. Temporal relationships between flood event, flooded area and SAR observation highlighting the difference between flooded area at the peak ($T_p$) and flooded area after the event ($T_p + \Delta T$) seen on an SAR image.
foothills, paved roads and the hydrological network itself can be confused with inundated areas since they also appear in a dark grey tone. The determination of paved roads and river streams can be established with the help of auxiliary information derived from digitized topographic maps.

Different kinds of approaches can be adopted to map the extent of flooded areas. The first one is based on segmentation techniques applied to SAR imagery acquired after the flood. A second approach consists of comparing two SAR images, one taken before and the other after the flood, through visual interpretation or change detection analysis. A third approach, which is not considered in this study, involves the use of coherence information deduced from an interferometric pair of SAR images (Nico et al. 2000).

When a segmentation approach is adopted to detect flooded areas on the after-flood SAR image, optimal threshold definition is required. The threshold value can be determined on a frequency histogram of the areas corresponding to dark grey tones in the SAR image and to flooded areas in the aerial photographs. A possible alternative is to use an adaptive thresholding algorithm. Ober et al. (1998) developed an iterative process which, assuming a Gaussian distribution of the data, calculates the mean value (μ) and the standard deviation (σ) and assigns to the Digital Number (DN) values out of the interval \([μ−σ; μ+σ]\), the extreme value of the same interval. This iterative process continues until all pixels are one of only two values (resulting in a binary image (flooded and non-flooded areas)) or when all the image pixels have a value belonging to the interval \([μ−σ; μ+σ]\); in a such case μ is the searched threshold value.

When using a multi-temporal approach, principal component analysis (PCA) of SAR images, taken at different times, offers great opportunities for flooded area detection (Calabresi 1995, Henebry 1997). In this approach a flood map is drawn up directly on a computer screen by means of visual interpretation of the false colour composite represented in RGB by images taken after and before the flood event.

These approaches allow satisfactory determination of flooded zones where water is still remaining at the time of space observation. However, this situation does not always represent the widest spread of the flooding. Often, flash flooding occurs and then the water level decreases very quickly; in this case observation from remote sensing satellites cannot meet the peak flooding phase.

2.2. An integrated RS and GIS approach

In order to overcome these limitations, a new procedure was developed in a GIS environment with the purpose of estimating the flooded area at the peak using standing water area observed after the flood event seen on SAR imagery and ancillary information extracted from the digital elevation topography. The main steps of the overall methodology are presented in figure 2. The basic assumption of this approach is that water has to flow out from the main river channel up to reach at least the areas recognized as flooded in the SAR data. The technique, based on the computation of least accumulative cost–distance surface (Douglas 1994), fills up the flooded place by tracing backwards from the main river stream towards still remaining water observed on SAR image.

From the digital topography of the study area a ‘least accumulative cost–distance’ matrix is derived. In such a raster image, each cell is characterized by a value equal to the least accumulative cost that water has to spend to get to the cell departing from main stream of the river. Cost corresponds to the work necessary for the water
1. Flooded Cells from SAR data

![Image of Flooded Cells from SAR data]

2. Least Accumulative Cost Distance Matrix

<table>
<thead>
<tr>
<th></th>
<th>2.8</th>
<th>3.0</th>
<th>2.6</th>
<th>2.7</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1.9</td>
<td>2.3</td>
<td>2.4</td>
<td>2.2</td>
</tr>
<tr>
<td></td>
<td>1.2</td>
<td>1.5</td>
<td>1.9</td>
<td>2.1</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>1.1</td>
<td>1.5</td>
<td>1.7</td>
</tr>
</tbody>
</table>

3. Estimated results

![Image of Estimated results]

Legend:
- River (origin cell)
- Flooded
- Non Flooded

Figure 2. Schematic diagram of the combined methodology: 1, detect flooded areas from SAR data; 2, generate least accumulative cost–distance matrix from digital topography of the study area, where the origin is the river; 3, a cell in the study area is flooded if its cost is lower than the cost of the cell detected from SAR data.

to overcome resistance formed by natural and anthropic features such as topography, surface roughness, vegetation cover, land use—all factors that affect water flow over the surface.

The least accumulative cost–distance matrix is modelled as a continuous, cumulative function from the main river stream up to the limits of the study area using a recursive procedure common to many GIS environments. An automatic procedure implemented in ARC/INFO (ESRI 1992) was used for this study.

First the cost–distance, i.e. the cost required for water to move from the source cell to each of the neighbour cells, is calculated. Initially the source cells correspond to the cells of the main stream. This process is then repeated considering the least costly neighbouring cell (i.e. the cell that presents the least resistance to the water flow) as the new source cell. The iterative process stops when all the cells have been assigned an accumulative cost.

In this procedure the cost–distance between two any adjacent cells A and B (cost–distA_B) is calculated according to the following formula:

$$\text{cost} - \text{distA}_B = \frac{(\text{cost}_A + \text{cost}_B)\text{distA}_B}{2} \times (VF)(HF)$$

(1)

where costA and costB are values associated with each cell A and B, distA_B is the
distance between cells A and B, and VF and HF are vertical and horizontal factors accounting for topography and surface typology respectively.

The cost value (costA and costB) associated with each cell is a weight factor proportional to the resistance incurred by water when flowing over the cell. It is a per-unit distance measure and in this case is represented by the elevation of the digital topography. The Pythagorean theorem, using a representation in which the centres of two adjacent cells are connected by links, calculates the distance between two cells (distA_B).

The vertical factor (VF) takes into account the cost necessary to overcome the height difference between two cells. Its computation is based on the slope map and values are assigned using a linear function relating the slope to the interval 0 to 2, as required by the used ARC/INFO procedure.

The horizontal factor (HF) accounts for any friction encountered by water flowing over the surface and is linked to the surface characteristics—roughness, land use, vegetation cover, etc. Because this factor is very complex to model, the typology of the surface was assumed constant over the entire area and set equal to one everywhere.

The least accumulative cost–distance matrix is then integrated with a base layer consisting of the flooded area map derived from the previous analysis of SAR image. Cost matrix cells corresponding to flooded points detected on the SAR image are considered to be border limits of the inundation event: the cost value assumed by the cell at the largest distance from main river stream is selected as a threshold.

Application of these criteria allows selection of all those cells of the cost matrix where water could be expected. In fact, because the cost value is a function of topography and distance, the energy necessary for water to reach higher cost cells must also be sufficient to get water to all those cells with lower cost value. The final result is a map of all zones affected by the widest spread of the inundation event within the study area.

3. A case study: Tanaro flood in 1994

Our study is concerned with the flood that occurred in Regione Piemonte (northern Italy) on the days of 4, 5 and 6 November 1994. The persistence of a wide atmospheric depression over north-west Europe caused heavy rains all over the Po river plain. The Tanaro basin, located in southern Piemonte, was the most affected by this extreme rainfall. Over 200 mm of rain fell on 4 and 5 November in the upper and middle part of the Tanaro catchment and in the upper portions of its tributaries, the Belbo, Bormida and Orba rivers. Precipitation reached a maximum hourly intensity of 55 mm, with maximum cumulative value of 264.4 mm over 24 h, causing the saturation of both the drainage capacity of the slopes and the impoundment of the hydrographic network. Rainfall produced exceptional discharge of the rivers and hydrometric levels never recorded before (unit discharge 0.6–7.3 m³ s⁻¹ km⁻²). Floods caused widespread damage to 38 urban areas, with the creation of 44 victims and 2000 homeless (Luino 1999). A comparable case in the Tanaro basin has not been recorded since 1879.

3.1. Study area and data used

The study area is a reach of the Tanaro basin included between the towns of Asti and Felizzano. This part of the catchment is an almost flat area with an average slope of 1% and is mainly devoted to agricultural activities. The land cover as
derived from supervised classification of Landsat 5 TM data taken in July 1994 is as follows: 79% cropland (wheat, sugar beet, grass), 8% poplar plantations, 6% urban areas and 7% water bodies.

The satellite data used were two ERS-1 SAR images, the first one acquired on 4 October 1994, one month before the flood and the second one acquired on 9 November 1994, three days after the event (figure 3). The SAR mounted onboard ERS-1 is a C-band system (5.6 cm wavelength) with a nominal incidence angle of 23° at mid-swath, which transmits and receives vertically polarized microwave electromagnetic energy. ERS-1 Precision Image (PRI) products used are three-look digital images, corrected for antenna elevation gain pattern and range spreading loss, with pixel dimensions of 12.5 m × 12.5 m.

In addition to the satellite images, ancillary data were used. Contour lines were extracted from a digital topographic map in raster format (1:10000) and used to derive the digital elevation model (DEM); the main streams of the river network and paved roads were extracted from the same map. Colour aerial photographs taken on 9 and 10 November 1994 over some areas were also available and used to define ground truth areas. Moreover, maps (1:10000 scale) of the actual inundated areas, acquired from ground surveys by the Settore Prevenzione del Rischio Geologico, Meteorologico e Sismico of Regione Piemonte, were used as reference data for accuracy evaluation. The areal extension of the actual inundated areas was estimated to be 54 km² from this reference map.

Both radar images and ancillary data were pre-processed and arranged in a GIS environment. A set of ground control points (GCPs) taken from the 1:10000 topographic map was used to rectify the 9 November ERS-1 image to the Universal Transverse Mercator (UTM) map projection, producing a root means square error (RMSE) of ±11.6 m. The pre-flood image of 4 October was then registered to the after-flood scene (RMSE ±8.75 m) through a first-degree polynomial transformation, to form a perfectly overlapping multi-temporal dataset. A nearest-neighbour resampling technique was adopted for both images in order to retain radiometric integrity.

Figure 3. Tanaro river basin study area where a catastrophic flood occurred in November 1994. ERS-1 SAR image acquired on 9 November, three days after the flood event.
The digital topography was generated from contour lines having a vertical interval of 5 m over the mountain areas and 10 m along the floodplain. The interpolation process was carried out using an iterative finite difference technique coupled with a drainage enforcement algorithm (Hutchinson 1989) specifically developed to generate hydrologically corrected digital topography. The grid size was defined at the same resolution of the SAR data (12.5 m × 12.5 m). Areas below an elevation threshold value were extracted and used to mask (on the radar images) the area (151 km²) on which to apply the mapping procedures.

3.2. Analysis of the results

Processing of a single radar image acquired some days after the flood peak and multi-temporal analysis of radar data taken before and after the inundation event allowed only a limited assessment of the flooded areas. The two thresholding procedures and visual interpretation were able to map only a small fraction of the total inundated area, amounting to 54 km².

Visual interpretation of a false colour composite obtained from multi-temporal SAR images and Second Principal Component (figure 4) gave larger extension of flooded areas (10.23 km²) compared to maps determined by thresholding approaches applied to a single-date SAR imagery. An iterative threshold algorithm applied to a single-date SAR image allowed us to map an area of 7.08 km²; using a simple threshold value derived from the cumulative frequency curve mapped only 3.96 km² as flooded zones.

Figure 5 shows a map of flooded areas derived from visual interpretation of the false colour composite of the multi-temporal dataset. In the same figure, the study area extracted from digital topography of the Tanaro watershed is also presented. Areas classified as inundated are patches spread around the main river stream. The water surfaces in this map cover only 18.9% of the flooded area recorded by Regione Piemonte. These large differences between the flooded zone detected on SAR imagery

![Figure 4. False colour composite of multi-temporal ERS-1 radar images (RGB: 9 November 1994; 4 October 1994; Second Principal Component). Areas still submerged by water three days after the flood event are easily recognizable due to their intense blue colour.]
Integration of SAR data and GIS for flood mapping

Figure 5. Flood map derived from visual interpretation of false colour composite of multi-temporal ERS-1 radar images (RGB: 9 November 1994; 4 October 1994; Second Principal Component), compared with the reference flood map compiled by the Settore Prevenzione del Rischio Geologico, Meteorologico e Sismico of Regione Piemonte.

and reference data can be explained by considering the time delay between the ERS-1 passage and the flood peak.

On 9 November, when the satellite image was acquired, most of the water had already flown back into the river or infiltrated into the soil. Aerial photographs (figure 6) taken on 9 and 10 November 1994 over some sites confirmed that, at the time of the ERS-1 overpass, most of the water was not actually still standing on land, and areas estimated as inundated corresponded to few zones still under water three days after the flood event. Moreover, the presence of a layer of fluvial wet fine sediments made the recognition of areas affected by the flood event difficult. In fact, in radar images, wet sediments produce intermediate backscattering values between low-intensity signals of water and the higher intensity signals of agricultural areas. In some cases poplar plantations close to the Tanaro river were also a constraint to ERS-1 in monitoring standing water, because of the rather low penetration capacity of the C microwave band at which it operates.

In contrast to the previous results, which reflect the situation three days after the flood peak, the flood map of figure 7 obtained by application of the combined methodology describes the maximum flooded area (60 km$^2$). The flooded zones are seen to extend both sides of the river and form a continuous belt around the main stream.

This map was obtained using the flooded area map given by the simple threshold approach as a base layer to be integrated with the least accumulative cost–distance matrix. The Tanaro river basin was subdivided into different sections reflecting
different morphological situations, each one corresponding to an area mapped as flooded in ERS data.

The map of flooded areas at the peak estimated by the new integrated approach was compared pixel-to-pixel with the digitized reference map of inundated areas compiled by Regione Piemonte. From this analysis a coincidence of 96.7% (52.2 km$^2$) was found between the flooded areas estimated and those officially recorded by the local government office.

The results obtained appear very promising, especially because the methodology allows mapping of inundated areas accurately over a large area and without the need for a complex simulations approach. Even though hydraulic models may provide a more correct description of the flooding process, they are also characterized by a more critical parameterization.
4. Conclusions

Techniques based on visual interpretation of multi-temporal satellite data and thresholding algorithms applied to single-date satellite data show serious limitations in determination of inundated areas at the peak of the flood. This is because of the inappropriate time of acquisition of satellite radar data with respect to the flood peak event. In fact, in the case study presented here, most of the water had already flown back into the river or had been infiltrated into the soil at the time of the satellite overpass.

To overcome these limitations, a new approach was proposed which integrates results obtained from SAR data with ancillary information using a 'least accumulative cost–distance' matrix. The catastrophic flood that occurred in northern Italy in November 1994 was used to test the obtained results through detailed comparison with a reference map of the flooded area compiled by Regione Piemonte. When compared with the reference map showing the area actually flooded, the flooded area map obtained from the new integrated RS and GIS approach shows an accuracy of 96%. This result indicates that images collected by ERS-1 SAR some days after an event can be useful in generating flood maps of the peak if used in conjunction with an appropriate methodology, whereas without such an approach SAR data do not allow such an accurate delimitation of flooded areas.

Acknowledgments

This study was supported by the National Group for Prevention from Hydrogeological Hazards (GNDCI) of the National Council of Research (CNR) of
Italy. We are also grateful to Giovanna Ober for software implementation of the iterative threshold technique. Suggestions of two anonymous referees greatly improved the quality of the manuscript. The 4 October 1994 and the 9 November 1994 ERS-1 PRI images were kindly provided by ESA-ESRIN, Frascati, Italy.

References


