HYDROLOGICAL PROCESSES
Hydrol. Process. **19**, 3699–3716 (2005)
Published online 1 August 2005 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/hyp.5852

Remote sensing and GIS-based flood vulnerability assessment of human settlements: a case study of Gangetic West Bengal, India

Joy Sanyal* and X. X. Lu
Department of Geography, National University of Singapore, Singapore 117570, Singapore

Abstract:

Flooding due to excessive rainfall in a short period of time is a frequent hazard in the flood plains of monsoon Asia. In late September 2000, a devastating flood stuck Gangetic West Bengal, India. This particular event has been selected for this study. Instead of following the conventional approach of flooded area delineation and overall damage estimation, this paper seeks to identify the rural settlements that are vulnerable to floods of a given magnitude. Vulnerability of a rural settlement is perceived as a function of two factors: the presence of deep flood water in and around the settlement and its proximity to an elevated area for temporary shelter during an extreme hydrological event. Landsat ETM$^+$ images acquired on 30 September 2000 have been used to identify the non-flooded areas within the flooded zone. Particular effort has been made to differentiate land from water under cloud shadow. ASTER digital elevation data have been used to assess accuracy and rectify the classified image. The presence of large numbers of trees around rural settlements made it particularly difficult to extract the flooded areas from their spectral signatures in the visible and infrared bands. ERS-1 synthetic aperture radar data are found particularly useful for extracting the settlement areas surrounded by trees. Finally, all information extracted from satellite imageries are imported into ArcGIS, and spatial analysis is carried out to identify the settlements vulnerable to river inundation. Copyright © 2005 John Wiley & Sons, Ltd.

KEY WORDS flood; remote sensing; settlement vulnerability; GIS

INTRODUCTION

Flood is a perpetual natural hazard in the flood plains of monsoon Asia, where over 80% of annual precipitation is received in the four wet months from June to September. The problem of river flooding is of great concern in the Indian state of West Bengal. The Irrigation and Waterways Department of West Bengal Government reports that, since its independence in 1947, there has only been 5 years in which it has been spared from the effects of monsoon inundation. One of the most devastating of its kind was experienced in September–October 2000. A total of 23 756 km^2 land area was inundated and 22·1 million people were affected (Rudra, 2001). This particular flood event caught everybody unaware. The unprepared condition of the administration to cope with this kind of natural calamity was thoroughly exposed. In recent years, a number of studies have recognized the importance of estimating people's vulnerability to natural hazards, rather than retaining a narrow focus on the physical processes of the hazard itself (Hewitt, 1997; Varley, 1994; Mitchell, 1999). Cannon (2000) argued that natural disaster is a function of both natural hazard and vulnerable people. He emphasized the need to understand the interaction between hazard and people's vulnerability.

Although most of the developed countries are well equipped with detailed flood hazard maps, up-to-date flood insurance maps and post-disaster hazard mitigation technical support (FEMA, 2003), there is hardly any

* Correspondence to: Joy Sanyal, Department of Geography, 1 Arts Link, National University of Singapore, Singapore 117570, Singapore. E-mail: g0202381@nus.edu.sg

Received 23 January 2004
Accepted 24 September 2004

detailed spatial database for flood prevention and mitigation in the developing countries. In recent years, efforts have been made to use remote sensing and geographic information systems (GISs) for creating national-level flood hazard maps for Bangladesh (Islam and Sado, 2000a). Population density and other socio-economic data have been integrated with hydrologic information to identify priority zones for implementing anti-flood measures (Islam and Sado, 2002). These studies were undertaken on a regional scale using coarse-resolution AVHRR imageries from NOAA satellites. The results of such investigations would only be useful for national-level macro planning.

The focus of this paper is on individual settlements. Its purpose is to analyse how the location of an individual settlement vis-à-vis the flood-prone zone and its socio-economic characteristics make it vulnerable to monsoon floods. The theoretical framework of this study is based upon the hypothesis that settlements are vulnerable to flood from three aspects: (1) whether the people have access to relatively higher ground to take shelter during an extreme hydrological event; (2) whether a settlement falls in a zone that is expected to experience a high flood discharge causing extraordinary damage of life and property; (3) whether the population density of the area is high enough to result in huge loss of property even in a moderate flood.

The example used for this study is Gangetic West Bengal, which is a natural flood-prone area whose fluvial characteristics have made it very suitable for rice cultivation. The population density in this region is one of the highest in the world. Although human settlement is abundant in this region, the local people traditionally settled only on relatively higher ground, known locally as *danga*. Rapidly increasing population density, due to both natural growth and the influx of millions of refugees in the post-independence era from the then East Pakistan, left very little choice for the people to settle on higher ground selectively. Severe shortage of land has forced people to settle indiscriminately over the highly flood-prone zone.

Landsat ETM^+ and ERS synthetic aperture radar (SAR) imageries are used in this study to classify non-flooded areas and flood depth within flooded zones, and to delineate human settlements at village level. The high spatial resolution of satellite imageries enables us to obtain detailed classification results that are suitable for formulating planning measures on a small scale. An added advantage is that the high resolution hydrologic information can be conveniently integrated with demographic data collected from smaller administrative units. This would greatly enhance the capability of the spatial database to estimate vulnerability of individual settlements to an extreme flood event.

STUDY AREA

The study area extends over three major river basins of southern West Bengal, namely Bhagirathi-Hoogly, Jalangi and Churni. All these three rivers are distributaries of the main branch of the River Ganga. Although we have tried to cover the natural region of the three river basins, the extent of the investigation is marginally compromised due to limited availability of digital terrain data. The eastern parts of the districts of Bardhaman, Murshidabad, and most of Nadia form the administrative entity for the area (Figure 1). The three river basins are overwhelmingly rural, with agriculture as the main source of livelihood. According to Bagchi's (1945) subregional classification of the Bengal Delta, the study area is identified as a moribund delta. In this section of the delta, the rivers are decaying and the land-building process has entirely ceased. Owing to its comparatively higher elevation and high levees, this area is traditionally less flood prone than the area that lies further south. The area falling between the Rivers Bhagirathi and Jalangi is an elongated depression, and the Churni basin area is almost entirely low lying in comparison with rest of Gangetic West Bengal. Therefore, this zone is liable to flooding. The study states that, in the Nadia and Hoogly districts, this belt is bounded by 10 m contour lines. Interfluves of the numerous distributaries are ill drained (Spate, 1965) and very often cause waterlogging during the monsoon season. This situation ultimately led to stagnation of water and development of cut-off channels known as *bills*. An abundance of oxbow lakes and misfit river channels also characterize this part of Gangetic West Bengal. There is a marked distinction in the channel pattern of the streams lying east and west of the River Hoogly. Sinuosity indices of the rivers on the eastern side of the

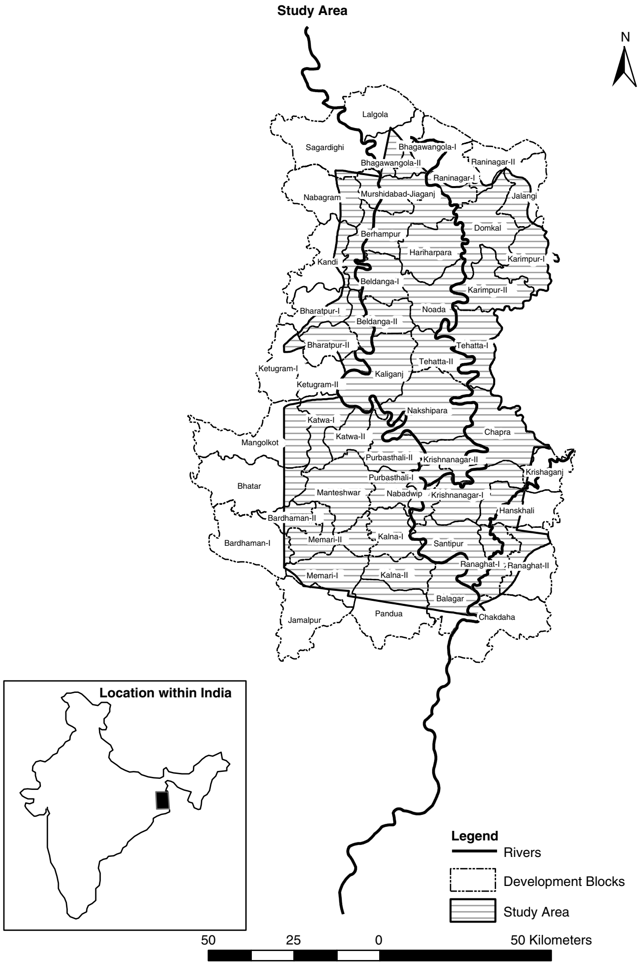

Figure 1. Administrative boundary of the study area

River Hoogly are very high compared with the western side (Goswami, 1983). The overall geomorphology of the study area depicts a degenerating fluvial system.

Owing to its geographical location, i.e. at the tail end of the extensive Ganga basin, West Bengal has a very limited capacity to control extreme hydrological events ensuing from the upper catchment of the River Ganga and its tributaries. Very high precipitation over a short period of time is cited as the most important factor responsible for triggering devastating floods in Gangetic West Bengal (*The Times of India*, 2000). After the Independence of India, 1956, 1959, 1978, 1995, 1999 and 2000 are identified as years that received abnormally high precipitation and, hence, severe floods (Basu, 2001). The 2000 flood in September–October was the worst in terms of its scale and damage caused. The West Bengal Government estimated that a total of 171 blocks of the state (23 756 km^2) area was affected. The total loss was estimated to be 56 600 million rupees (US$1132 million) (Ganashakti, 2000). Abnormally high rainfall for 4 days in the upper catchment areas of the western tributaries of the River Bhagirathi were responsible for this natural calamity. The severity of the event was so high that many low-lying areas of the Nadia district remained waterlogged for over 3 weeks, with the depth of water estimated as high as 3 m (Rudra, 2001).

DATA AND METHODS

Work flow of the current investigation has four components: (1) delineating extent of non-flooded surface in and around the major flood zone; (2) extracting the high flood-depth zone from the flooded areas; (3) demarcating the human settlements of the area; (4) importing the above three layers into a vector GIS environment and performing spatial analysis to obtain relevant results.

Data sources for the first three components are satellite images. Semi-automatic digital image processing techniques and manual digitization have been employed to extract the relevant information from remotely sensed data. The fourth component involves incorporation of demographic parameters with the information extracted from satellite imageries to identify effectively the settlements that are highly vulnerable to flood hazard. In this section, each component of the total work flow has been dealt with separately. Special attention has been given to illustrating how each of these components contributes to the ultimate objective of this study.

Delineating non-flooded areas from flooded areas

The current study is concerned more with dry/land area than flooded area. Delineation of the non-flooded area is particularly important because these areas can serve as a temporary shelter for the nearby settlements. This information is necessary for identifying the settlements that are highly vulnerable to flooding. Settlements having no immediate access to dry areas would be considered highly vulnerable to flooding.

From the early era of passive remote sensing, special attention has been given to distinguishing water from dry surface. MSS band 7 ($0{\cdot}8$–$1{\cdot}1$ μm) has been found to be particularly suitable for distinguishing water or moist soil from dry surface due to its strong absorption of water in the near-infrared (NIR) range of the spectrum (Smith, 1997). MSS data were used to deal with the flood-affected areas in Iowa (Rango and Solomonson, 1974), Arizona (Morrison and Cooley, 1973), and the River Mississippi basin (Deutsch *et al.*, 1973; Deutsch and Ruggles, 1974; Rango and Anderson, 1974; McGinnis and Rango, 1975).

From the early 1980s, Landsat TM data with an improved spatial resolution of 30 m have become one of the major sources of remotely sensed data for flood management research. Landsat TM band 4 is spectrally a near equivalent of MSS band 7. Water yields very low reflectance in the NIR region of the spectrum and, therefore, can be effectively used to discriminate water from land surface. This property of Landsat band 4 has been extensively used to delineate flooded areas in West Africa (Berg and Gregiore, 1983), India (Bhavsar, 1984) and Thailand (Raungsiri *et al.*, 1984).

For the current research, two Landsat ETM$^+$ scenes of the study area acquired on 30 September 2000 have been obtained. The imageries were acquired nearly at the peak of the flood. The scenes were geometrically and radiometrically corrected (level 1G product from US Geological Survey (USGS)). The two scenes have been accurately georeferenced using GPS control points collected during a field visit to the study area. All bands of the two scenes have been mosaicked. TM bands 4, 3 and 2 have been projected in RGB to generate a standard false colour composition (FCC) of the study area. Although TM band 4 is useful in delineating land and water boundaries, asphalt road surfaces and rooftops also yield very low reflectance in this band. Reflectance from water varies sufficiently from roads and dark rooftops in Landsat band 7 ($2{\cdot}28$–$2{\cdot}35$ μm, mid infrared). Therefore, water and non-water can be effectively discriminated by adding TM band 4 and band 7 (Wang *et al.*, 2002).

The predominance of cloud cover during the flooding season is one of the major obstacles to using optical remote sensing in flood management. Although SAR can penetrate cloud cover, scientific communities in the developing countries still prefer optical data over radar because of two reasons. Purchasing radar data is beyond the means of most of the government agencies in the developing countries because of its high price. Radar satellite needs prior programming to capture data for a particular area. Owing to its very high rate of data capture it requires a real-time data transfer to a receiving station near the coverage area. Lack of ground receiving stations and proper coordination has resulted in limited coverage of SAR data for developing countries. Major satellites that operate in the optical portion of the spectrum maintain a year-round archive of global coverage and, therefore, are still more acceptable in the field of remote sensing.

After adding TM bands 4 and 7, an effort was made to mask out the cloud-contaminated pixels, as their existence can interfere with any classification effort. There are a number of algorithms for screening clouds in optical imageries (England and Hunt, 1985; Saunders, 1986). Cloud appears very bright in a standard FCC image. After a detailed comparison of the band $(4 + 7)$ image with the FCC, it has been found empirically that any pixel having a digital number (DN) of over 124 in band $(4 + 7)$ is cloud. The pixels with a (DN) range >124 have been masked. It has been observed that it is very easy to extract cloud-covered pixels over a dark background (e.g. flooded area in band $(4 + 7)$), but over a bright background, like settlements or healthy vegetation, pixels at the periphery of a cloud cluster have a very similar reflectance to the background. Therefore, the above-mentioned threshold is not able to remove small numbers of cloud-covered pixels over bright and dry surfaces. Since this study focuses on flooded areas, the existence of small numbers of cloud-contaminated pixels in the predominantly dry areas is not of great concern.

Band $(4 + 7)$ is quite efficient in discriminating flood water from dry land. Major boundaries of flood water can be delineated without much effort. However, there are still a few 'water pixels' that are not classified as flooded within the rural settlements (Figure 2). These pixels can be readily identified in a standard FCC. The reflectance of these pixels is very close to the nearby wet soil surface. During floods, the albedo from the water body increases significantly because of a high concentration of debris and silt in the water. Thus, the reflectance peak moves toward the red band. On the other hand, increasing soil moisture decreases soil albedo, making reflectance from some non-flooded pixels very similar to flood pixels (Sheng *et al.*, 1998). From a close comparison of band $(4 + 7)$ with the FCC, a binary classification has been done as follows for band $(4 + 7)$:

$$\text{Pixel value} > 78 = \text{dry land}$$

$$\text{Pixel value} \leq 78 = \text{water}$$

This classification effectively extracts water pixels within the settlement area, but the main disadvantage of this classification is that it cannot distinguish between dry surface and water under cloud shadow.

Areas under cloud shadow receive only scattered sunlight, but the low illumination results in a suppressed reflectance from all land-cover categories. Therefore, it is very difficult to discriminate between land and water for the areas under cloud shadow (Sheng *et al.*, 1998). Water has a significantly low reflectance in the NIR region of the electromagnetic spectrum compared with dry land surface. Owing to low reflectance of non-flooded areas under cloud shadow, the above-mentioned threshold spuriously classifies it as 'water' (Figure 3). As a consequence, the classified image appears as an underestimation of non-flooded area and an overestimation of water. Actually, the difference in the reflectance between flooded and non-flooded regions becomes so low that it is not possible to separate them by means of a threshold value. Land surface reflects

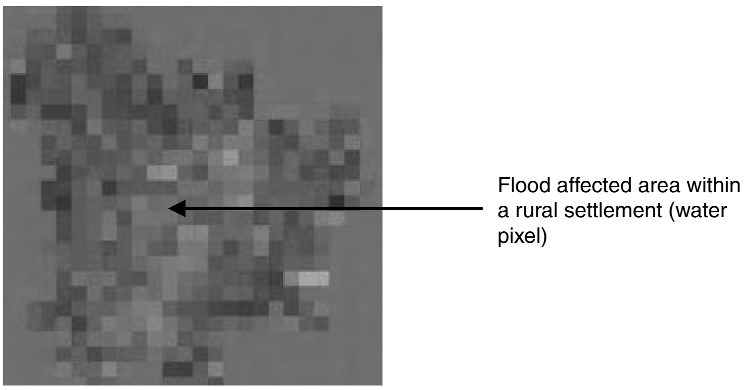

Figure 2. Landsat ETM$^+$ FCC (zoomed eight times from optimum resolution) showing flooded area within a settlement

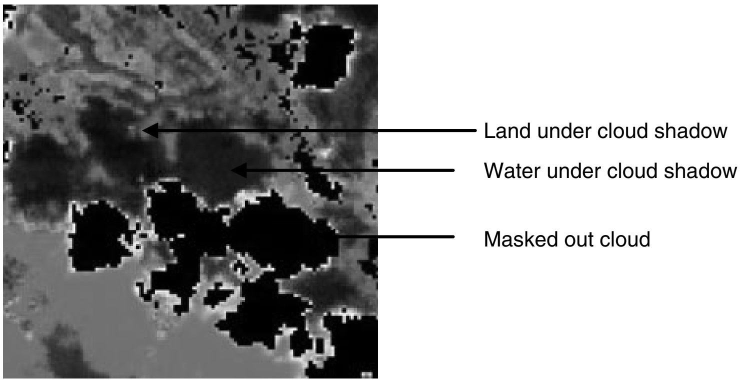

Figure 3. FCC showing flooded and non-flooded areas under cloud shadow

higher energy in the red band (Landsat band 3) compared with water. On the other hand, reflectance from water is significantly lower in the NIR band compared with land. Therefore, the ratio of NIR and red (band 4/band 3) increases the difference in the reflectance of flooded and non-flooded pixels. This difference in the ratio image would be sufficient to distinguish water from land under cloud shadow. After comparison with the FCC it has been found that, although the ratio image is effective in differentiating water and land under cloud shadow, it is not as sensitive as band $(4 + 7)$ to water surface. The problem arises from situations where the 'water' pixels mixed with non-flooded pixels have a higher value than the open water pixels in the ratio image. Sometimes, values of such pixels overlap with the non-flooded pixels surrounded by water under cloud shadow. Hence, the ratio image is not able to identify a cluster of 'water' pixels surrounded by land.

In this paper, an attempt has been made to synthesize the advantage of band $(4 + 7)$ and the ratio image for extracting the non-flooded area from the flood scene. After thorough observation of the DN of the ratio image and Landsat FCC, it has been found empirically that pixel values ranging from 0·52 to 1 in the ratio image represent non-flooded areas under cloud shadow. Pixels that fall within the specified range in the ratio image have been selected and merged with non-flooded pixels derived from band $(4 + 7)$. It should be pointed out that these two regions derived from different manipulation of Landsat bands have not been found mutually exclusive in their spatial coverage. Overlapped areas are dissolved to create one classified layer depicting the non-flooded area. In the synthetic image, the advantages of band $(4 + 7)$ and band (4/3) have been incorporated to extract the non-flooded area in the flood scene accurately.

There is no doubt that the above classification scheme enhances the accuracy. However, it still suffers from limitations. The roofs of submerged houses appear as dry surface, and in areas of compact settlement this error becomes significant. The tree canopy also creates confusion in delineating the boundary between water and land. Flooded areas under canopy appear as non-flooded. Although forest constitutes an insignificant land-cover category in Gangetic West Bengal, rural settlements are often surrounded by trees. This factor constitutes a major source of error in the classification scheme.

Site visits to some selected settlements were made during the fieldwork in 2003 and local people were interviewed to locate specific sites for validation. However, after incorporating that information it was found that it is not scientifically viable to do an accuracy assessment based on people's recollection of a flood event that occurred 2 years ago. In the absence of any aerial photograph during the flood we embarked upon using a digital elevation model (DEM) to assess the accuracy of our classification indirectly. To assess the accuracy of the classified image, pixels classified as 'non-flooded' have been superimposed on a DEM derived from the advanced space-bome thermal emission and reflection radiometer (ASTER) images of the study area. The ASTER DEM has a spatial resolution of 30 m with a relative accuracy of more than 10 m. These digital data are suitable to meet the 1 : 50 000 to 1 : 250 000 map accuracy standard (USGS, 2003). The elevation distribution of the non-flooded pixels is shown in Figure 4. The graph shows a sudden upward

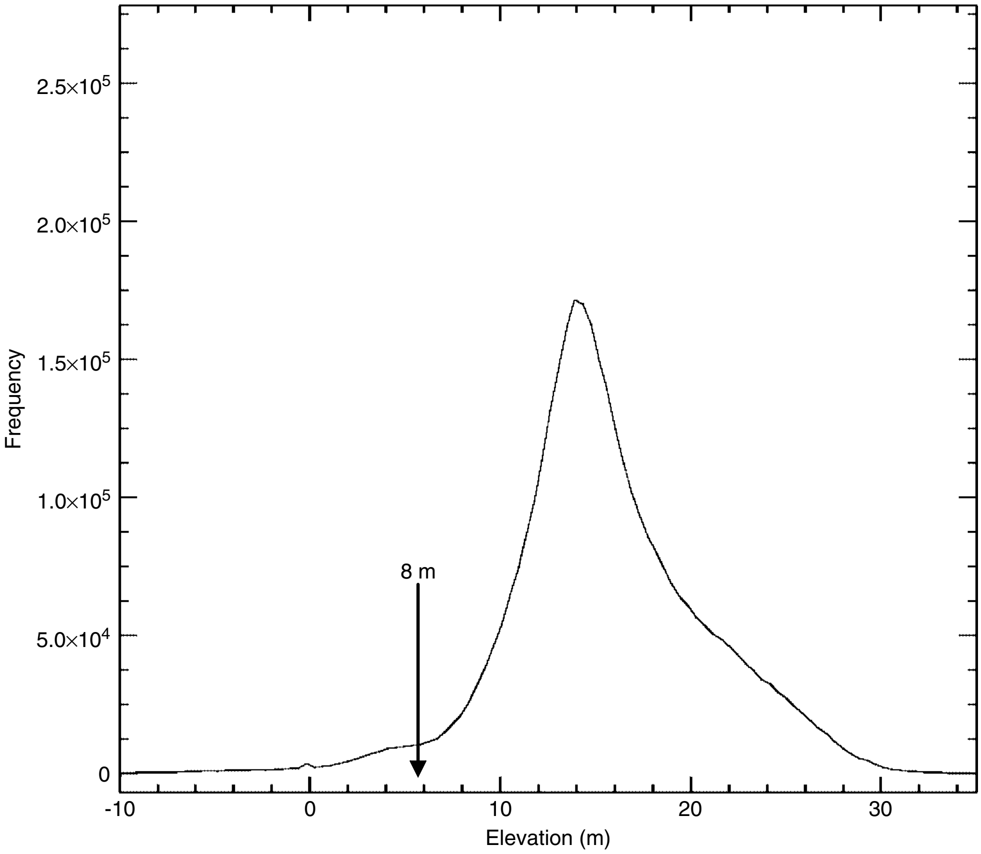

Figure 4. Elevation distribution of the non-flooded area extracted from the ASTER DEM

trend approximately at an elevation of 8 to 9 m. The rest of the non-flooded area has elevation ranging from 10 to 35 m. It is assumed that non-flooded pixels having an elevation of below 8 m are most probably tree canopy or other confusing land cover in the flooded area that has been incorrectly classified as 'non-flooded'. To rectify this error, and to reduce overestimation of dry area, all 'non-flooded' pixels having an elevation of less than 8 m have been eliminated. Since we have used a relative ASTER DEM, the absolute heights of the pixels are not very accurate from sea level, but the relative difference of elevation between two pixels meets our requirement. An 8 m elevation has been used to illustrate the point of inflection in the elevation distribution curve of the non-flooded pixels. The absolute height is not very relevant in this discussion. The elevation distribution of the non-flooded pixels is the purpose of the graph.

After the 'non-flooded' pixels had been imported into ArcGIS, it was realized that any 'non-flooded' surface occupying less than nine Landsat TM pixels is potentially a tree canopy, or part of some road surface that was not submerged by flood water. After considering the general topography of Gangetic West Bengal it was realized that there is the rare possibility of the existence of such a small abruptly higher area. Nine pixels were chosen as a threshold because this presents a block of 90 m × 90 m. Thus, any apparently dry land surface that is less than a 3 × 3 block is not likely to be a suitable place for flood shelter. Anything smaller than three pixels may also contain some inundated area in the sub-pixel level and, therefore, cannot be relied upon as a potential shelter. Considering the population density of rural West Bengal, even if some of these isolated pixels represent actual non-flooded surfaces, then they are too small to provide any effective shelter to the adjacent population. Therefore, patches with less than nine pixels have been eliminated from the non-flooded layer. Table I summarizes the extent of area eliminated at different levels of correction of the non-flooded area.

Anything other than the non-flooded area depicts the actual water area. This classified 'water' layer includes permanent water bodies of the area. For calculating actual flood-affected area, permanent water bodies must be eliminated from the 'water' layer (Yang *et al.*, 1999). Two Landsat ETM$^+$ scenes of 16 April 2003, depicting normal hydrological conditions, have been used to extract the permanent water bodies. These pixels were subtracted from the 'water' layer to obtain the actual flooded area. The classified image is shown in Figure 5.

Delineating the high flood-depth zone

Flood depth is considered as the most important indicator of flood hazard (Wadge *et al.*, 1993; Townsend and Walsh, 1998; Islam and Sado, 2002). A higher depth of flood is associated with a high discharge, which is a determining factor in flood-induced destruction of life and property. During the 2000 September flood the depth of water was as high as 2·5 to 3 m in some areas. The roofs of the majority of the houses in rural West Bengal are not concrete; they are commonly built with hay or thatch and corrugated sheets of iron. This fact excludes the local people from climbing over their roofs to escape inundation. Flood depth determination from remotely sensed imagery is very difficult, but an indirect method exists to classify a flooded area into different flood-depth zones. The amount of radiant energy reflected by water in the visible light portion of the electromagnetic spectrum is essentially determined by the colour of the water and its turbidity. Except for the blue band, all optical bands have very high correlation with the turbidity and sediment concentration of the water. Deeper water has more turbidity than shallower waters because of its high velocity (Islam and Sado, 2000b).

Interband correlation is a major impediment in analysing multispectral data. Principal component (PC) transformation is performed to overcome this problem. This image-processing technique makes the bands less correlated and reduces the dimensionality of the original dataset (Lillesand and Kiefer, 2000). To enhance contrast and facilitate classification, a PC transformation has been applied over bands 2, 3, 4, 5 and 7 of the Landsat ETM$^+$ data acquired during the flood. The first three components, explaining about 99·65% of the total variation, were selected for further analysis. The other components were excluded from further analysis as their noise-to-signal ratio is expected to be very high. Kunte and Wagle (2003) attempted to classify depth of water in the Gulf of Kutch and reported that the PC2 band is particularly sensitive to the concentration of suspended sediments and, therefore, can be effectively used for broad classification of water depth. To enhance the amount of information, different combinations of the three PC bands into RGB were tried to create an FCC, and it was found that PC2, PC1 and PC3 (RGB) generate the best FCC. Figure 6 shows the turbidity/sediment concentration in the flooded zone.

Figure 6 clearly represents at least two turbidity zones in the shades of yellow and violet. The general trend of tonal variation reveals that highly turbid water exists at the core of the flooded zone and the sediment concentration of the water gradually decreases towards the boundary of water and land. After consulting Islam and Sados' (2000b), we have assumed that highly turbid water is associated with high velocity. During a flood, high water flows at high velocity through the highly inundated zone. Highly turbid/deep water appears in yellow and shallow water with less turbidity is presumed to appear in violet tint. The major rivers of the region, namely the Bhagirathi and Jalangi, fall in the yellow zone, which further confirms our visual interpretation of flood depth.

Table I. Correction of non-flooded area under different levels of processing

Processing level	Total area (km^2)	Area reduced (km^2)
Digital image processing of Landsat flood scene	3512·98	–
After eliminating area under 8 m	3324·98	7·41
After eliminating polygons less than nine Landsat pixels	3287·08	37·90

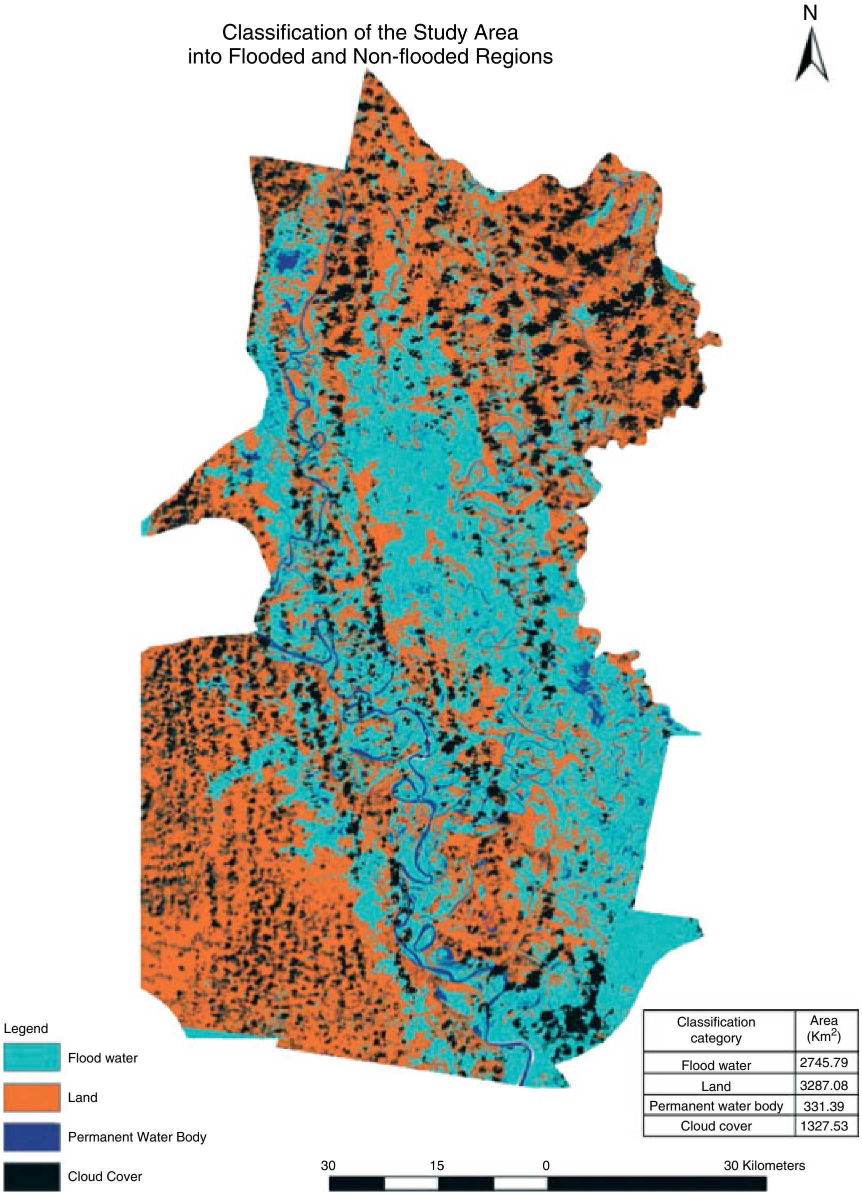

Figure 5. Classified image showing the flooded area on 30 September 2000

It is interesting to note that along the River Bhagirathi the majority of the flooded area falls in the high depth zone and there is little evidence of a shallow depth zone along the margin of the flooded area. This phenomenon is attributed to the presence of extensive embankments along the river. Land use is very intensive in the immediate flood plain of the River Bhagirathi. To protect the land, the West Bengal Government has built hundreds of kilometres of embankment. These embankments, along with other minor flood-control measures, put an abrupt stop to advancing flood water during the flood of September–October 2000. The low-lying active flood plain, located between the embankment and the river, records very high water discharge during floods and, consequently, the area suffers from high water depth. Owing to the existence of the embankments, there

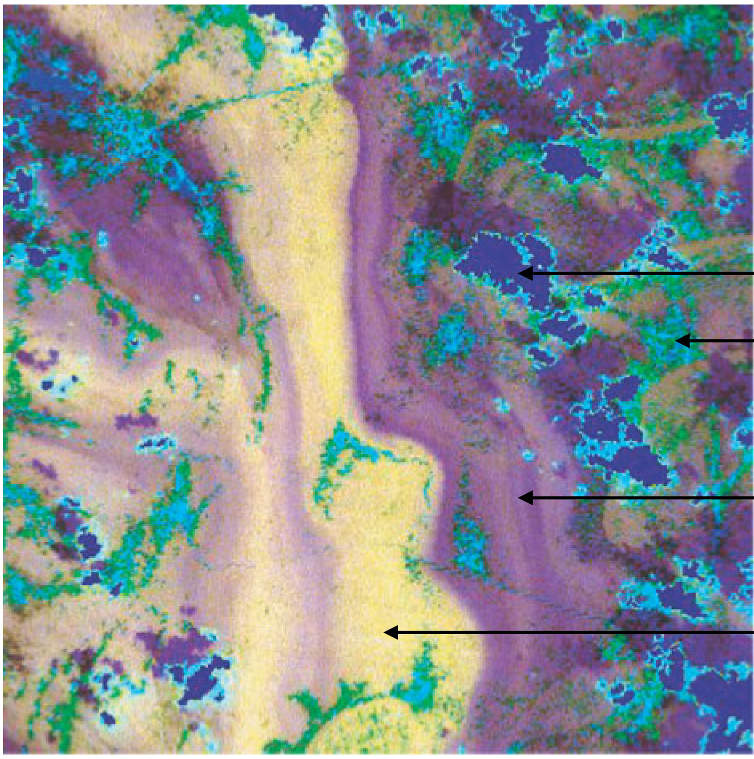

Figure 6. Different flood depth/turbidity zones identified over an FCC (PC2, PC1 and PC3 as RGB)

was hardly any zone of transition from deep to shallow flood water along the bank. The River Bhagirathi represents the most dominant distributary of the River Ganga that flows through the Indian state of West Bengal. The discharge and carrying capacity of this river is much higher than the other smaller rivers in this region. Owing to a higher velocity and discharge, the sediment concentration and consequent turbidity in the River Bhagirathi is much higher than the other smaller rivers, like the Jalangi. This factor also contributes to deep inundation along the Bhagirathi river bank.

After comparing the PC2 band with the FCC, a threshold value has been selected empirically to extract highly turbid/deep water from the PC2 band. Although river flooding is not only a function of terrain configuration, the topography of the flood plain does play an important role in the course of an advancing flood. Therefore, the ASTER DEM has been used to verify the accuracy of the high flood-depth zone. Elevation distribution of the high flood-depth zone has been plotted in Figure 7. This plot ratifies that classification of high flood depth is, on average, accurate and realistic. The area having an elevation of more than 16 m was not likely to experience high flood depth. The fact that the wave drops sharply after an elevation of more than 16 m attests to the high accuracy of the classified high flood-depth zone.

Delineating human settlements

As mentioned in the Introduction, the main focus of this study revolves around identifying the flood-vulnerable settlements. Human settlement is the main input layer of the present study, and all other layers of information are used to estimate its vulnerability to floods. Land-use maps of 1 : 250 000 have been used as the basic source of information for delineating human settlements in this area. These maps were published in 1991 and, therefore, are considered relatively outdated. Satellite imageries have been used to update these maps. Application of remotely sensed data for identifying human settlements is not new. Radar imageries

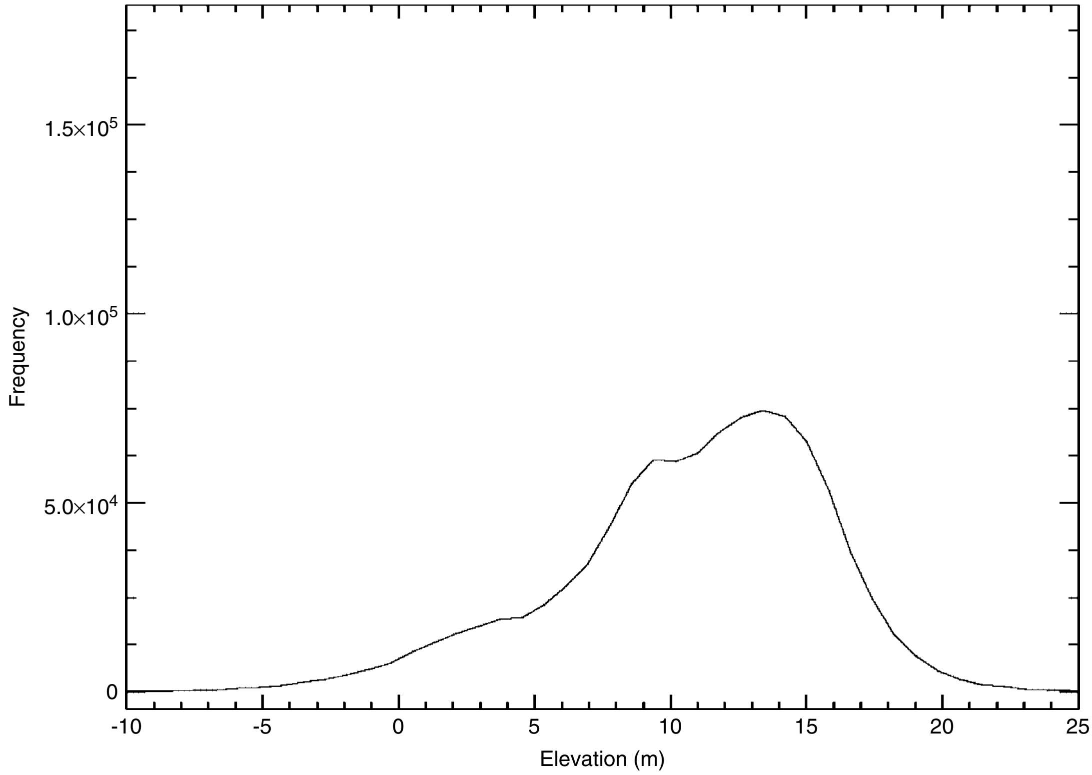

Figure 7. Elevation distribution of the area affected by high flood depth

are more widely used than optical data because settlements are particularly accentuated in radar imageries owing to their geometric shapes and dielectric constants. These parameters of SAR data are different from optical sensors (Henderson and Xia, 1997). Studies in Germany (Henderson, 1995), China (Lo, 1986), and the Ganges Plain (Imhoff *et al.*, 1987) reported that settlements having a population of more than 1000 are generally recognizable from SAR imageries.

The difficulty associated with identifying human settlement from radar data is that most SAR sensors obtain data at a single wavelength with fixed polarization. The best band combination for any classification should cover major portions of the electromagnetic spectrum (visible, infrared and microwave) that are normally used for remote sensing (Haack *et al.*, 2000). Haack and Slonecker (1994) reported that neither Landsat TM nor SAR data can independently locate villages in Sudan. For the present study, Landsat ETM$^+$ data of 13 April 2003 and ERS-1 SAR data of 9 October 1995 have been used to interpret the land cover of the study area visually and to update the settlement layer previously digitized from the land-use map.

The Landsat ETM$^+$ data obtained are geometrically corrected (level-1G) from the USGS. It has been further registered to the 1 : 250 000 land-use maps. The radar data have been obtained as a precision image (PRI). ERS SAR PRI products are projected to ground range and resampled in 12·5 m pixel size. The PRIs have been coregistered with the land-use map with a root-mean-square error (RMSE) of 1·06 pixels. After georeferencing, the SAR imageries have been mosaicked. A low-pass filter of a 5×5 pixel window has been applied to it to reduce speckle and improve visual interpretability in identifying settlements. After stacking TM bands 4, 3 and 2 in RGB, a coloured image was generated by fusing it with the processed radar image. HSV sharpening tools have been used to perform this operation. This function transforms an RGB image to HSV colour space, replaces the value band with the high-resolution radar image and automatically resamples the hue and saturation bands to the high-resolution pixel size (12·5 m) of the SAR PRI using the nearest neighbour method. Finally, it transforms the image back to the RGB colour space. The output was a coloured image of 12·5 m spatial resolution. It is necessary to point out that a coregistration RMSE of more than one pixel is acceptable for SAR scenes because the 12·5 m $\times$ 12·5 m pixels of radar imageries are fused with 30 m $\times$ 30 m ETM$^+$ pixels.

Figure 8 represents part of the study area. The small rural settlements are easily distinguishable by their bright appearance over a reddish background of vast cropland. The vectorized settlement layer, already extracted from the land-use maps, has been superimposed on this image to update the boundaries of settlements manually. Automated classification methods have not been considered, as a majority of studies reported low accuracy for smaller settlements (Dowman and Morris, 1982; Lo, 1984; Liu *et al.*, 1986). It has also been found that the reflectance of fallow land and bare soil are quite similar to the rural settlements in the TM bands. Such substantial error in classifying human settlements would jeopardize the whole gamut of results. This factor induced us to rule out the option of automated classification.

Processing different data layers in a GIS environment

In order to reach the objective of obtaining a flood vulnerability of settlements, it is necessary to make the other information layers compatible with the settlement layer to allow for the analysis of their vulnerability to river inundation. Apart from the hydrological information extracted from the Landsat ETM$^+$ flood scenes, some socio-economic data, like population density, have also been incorporated in the current framework to facilitate vulnerability analysis of the settlements. Considering the scale of this investigation, development blocks have been selected as the appropriate administrative unit for reporting population density. Development blocks are the smallest administrative unit in West Bengal, as well as India, for collecting majority of the data. Boundaries of the development blocks over the study area are shown in Figure 1.

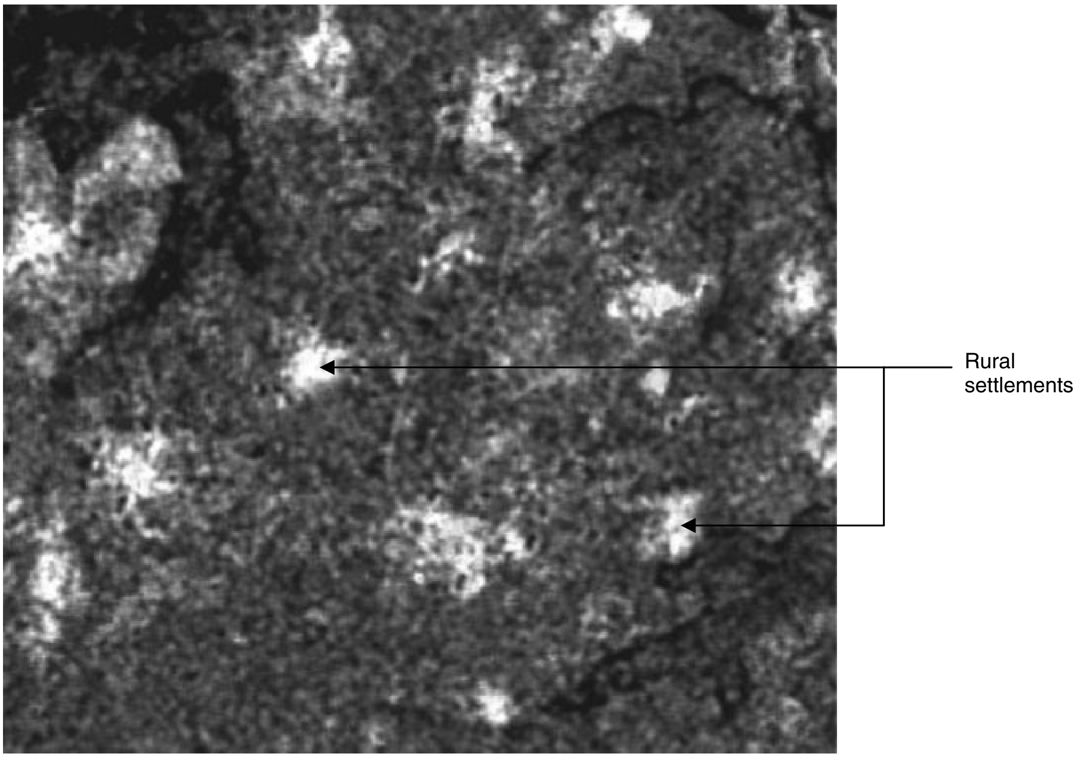

Figure 8. Landsat ETM$^+$ bands 4, 3 and 2 merged with ERS SAR image to identify the rural settlements in Gangetic West Bengal visually

Administrative boundaries of the development blocks have been digitized from a 1 : 500 000 map. Demographic data, collected from the 2001 Indian Census, have been integrated in the attribute table of the development blocks. It has been observed over the study area that human settlements, especially rural hamlets, develop across administrative boundaries of the development blocks. The frequency of flood occurrence in each development block for the decade 1991 to 2000 has been considered to identify those areas that have been suffering from chronic flooding (Sanyal and Lu, 2005). This information was obtained from the Annual Flood Reports of the Irrigation and Waterways Department, Government of West Bengal, India.

To facilitate spatial analysis, the shape of the file containing settlements has been 'clipped' by the administrative boundaries of the development blocks. In the output, individual settlements have been subdivided where an administrative boundary cuts across them. This overlay operation proved very useful in subsequent analysis, as in the output layer each settlement contains certain information about the development block it is located in. This information includes name, population density and frequency of flood occurrence over the last decade. It should be noted that the above-mentioned attributes are assumed to be constant for all individual settlements falling within one development block. It is evident that the accuracy of this information decreases with increasing size of the development blocks. Owing to the monotonous flat terrain of the study area, there is not much scope for large variation of topography within a development block.

RESULTS AND DISCUSSION

The main purpose of this investigation is to analyse the interaction between different flood hazard indicators that contribute to vulnerability of human settlements in the study area. The previous section illustrated a systematic approach to extracting the spatial extent of these hazard factors. A GIS environment is used to evaluate the interaction of these factors in a spatial dimension and locate the settlements that should be given priority in implementing remedial measures. It should be pointed out that this methodology is likely to yield better perception of flood hazard if a very high magnitude flood is taken into consideration. Using a moderate magnitude event would increase the risk of leaving some vulnerable settlements unidentified that have chances of be affected should a very high magnitude event hit the area. We found that an overwhelming majority of the settlements contain some pixels classified as non-flooded. The reason behind this phenomenon is that the local people traditionally build their houses over relatively higher ground, leaving the flood-prone lowlands for paddy cultivation. This settlement pattern evolved as an outcome of local inhabitants' adaptation to living in a flood-prone area. In Gangetic West Bengal, owing to the vast expanse of cropland, trees are only found in and around rural settlements and along roads. Misclassification of some tree canopy as non-flooded pixels also contributes to overestimation of non-flooded areas. It has been decided that any individual settlement having very little intersection with the non-flooded pixels should be practically considered as highly vulnerable to flood. For computational ease we have calculated the hazard indicator as

$$\mathrm{HI}_1 = (A/B) \times 100 \tag{1}$$

where A is the area of intersection between non-flooded pixels and individual settlements and B is the total area of an individual settlement.

The intersection function in ArcGIS is used to calculate the common area between the settlement and non-flooded layer. Part of the attribute table of the output shape file is shown in Table II. This table shows that, very often, each settlement polygon intersects with more than one polygon representing non-flooded areas. Therefore, the area of intersection is distributed in more than one row, which makes it difficult to calculate HI_1. It is noted in Table II that settlement ID 23 is distributed in 13 rows. In other words, settlement 23 contains 13 patches of non-flooded land area. Table II is not the complete table. It only shows a typical portion of it that exemplifies the difficulty in calculating HI_1. The attribute table of the output shape file has been summarized by dissolving the polygons on the basis of their settlement ID number. Part of the processed table is illustrated in Table III. After calculating HI_1 it has been found that out of 921 settlement polygons in the study area 390

Table II. Part of the attribute table illustrating how the intersection of a non-flooded layer with individual settlements is distributed in different polygons. Note that 13 polygons represent the area of intersection between settlement 23 and the non-flooded layer

Settlement polygon ID	Intersection area (m^2)	Block name	District	Settlement ID
56796	21 191·9	Balagar	Hugly	13
56791	35 132·4	Balagar	Hugly	13
56720	1 100 000 000·0	Balagar	Hugly	20
56720	1 100 000 000·0	Pandua	Hugly	22
56720	1 100 000 000·0	Pandua	Hugly	22
56793	27 830·0	Chakdaha	Nadia	23
56789	25 025·9	Chakdaha	Nadia	23
56791	35 132·4	Chakdaha	Nadia	23
56782	9343·0	Chakdaha	Nadia	23
56780	29 386·0	Chakdaha	Nadia	23
56788	84 385·4	Chakdaha	Nadia	23
56778	10 258·0	Chakdaha	Nadia	23
56781	46 821·7	Chakdaha	Nadia	23
56769	9920·9	Chakdaha	Nadia	23
56758	81 549·4	Chakdaha	Nadia	23
56738	37 867·3	Chakdaha	Nadia	23
56707	10 245·9	Chakdaha	Nadia	23
56762	310 000·0	Chakdaha	Nadia	23
56720	1 100 000 000·0	Pandua	Hugly	24
56525	49 332·4	Pandua	Hugly	26
56525	49 332·4	Pandua	Hugly	26
56720	1 100 000 000·0	Pandua	Hugly	26
56720	1 100 000 000·0	Pandua	Hugly	28
56720	1 100 000 000·0	Pandua	Hugly	28
56720	1 100 000 000·0	Pandua	Hugly	28

Table III. Area of intersection between settlement layers and non-flooded area summarized on the basis of individual settlements

Development block	District	Settlement ID	Intersection area (m^2)
Balagar	Hugly	13	45 790·00
Balagar	Hugly	20	8136·84
Chakdaha	Nadia	21	0·00
Pandua	Hugly	22	580 052·79
Chakdaha	Nadia	23	685 687·78
Pandua	Hugly	24	1 200 000·00
Chakdaha	Nadia	25	0·00
Pandua	Hugly	26	160 071·56
Balagar	Hugly	27	0·00
Pandua	Hugly	28	126 631·52

settlements have less than 50% of their area classified as non-flooded. In other words, more than 50% of the area of these settlements was affected by inundation. For more than 75% inundation, the number is still high at 206 (22·36%). There are 124 settlements having more than 90% of their area submerged under flood water (Figure 9). These settlements are located in a high flood-hazard zone and, therefore, are extremely vulnerable

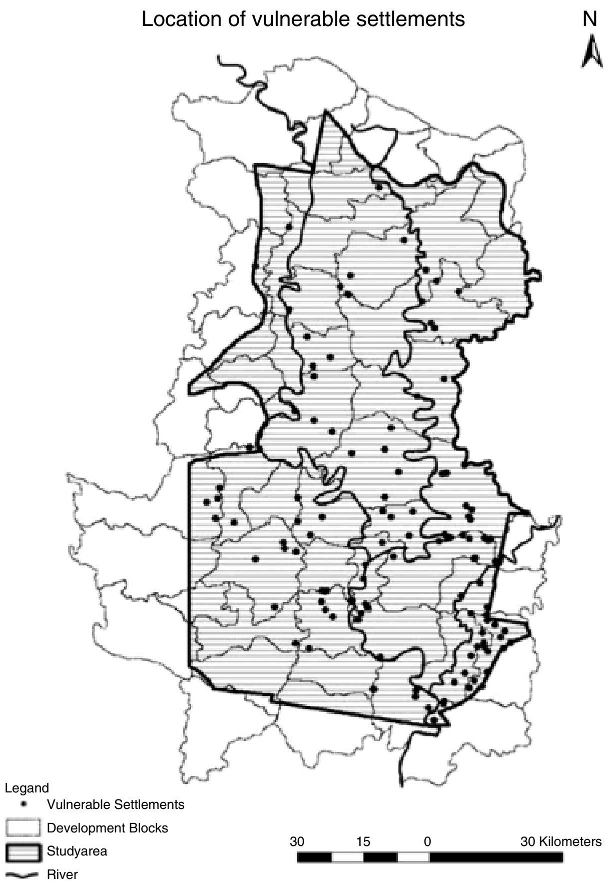

Figure 9. Location of the settlement that does not have access to higher ground as shelter during the flood on 30 September 2000

to inundation. In addition, a buffer operation reveals that 13 of these 124 settlements have no non-flooded areas within a buffer zone of 500 m. For the sake of simplicity, a straight-line distance has been used to calculate the buffer. Since Gangetic West Bengal is remarkably flat there is no conspicuous physical barrier such as, mountain to travel for short distance. It is been also realized that 500 m is the maximum distance that people can travel even at a moderate flood depth. Thus, these settlements have no immediate access to a potential flood shelter. Out of these 13 settlements, five are located in the Ranaghat-II block, four are in the Ranaghat-I block, two are in the Balagar block, one is in the Chakdah block and one is in the Chapra block (see Figure 1 for block names).

The location of a particular settlement with respect to the high flood-depth zone forms the basis for computing the second flood hazard indicator: any settlement having the majority of its area under deep flood water has been considered as vulnerable. The hazard indicator has been calculated as

$$\mathrm{HI}_2 = (C/D) \times 100 \qquad (2)$$

where C is the area of intersection between high flood depth and an individual settlements and D is the total area of an individual settlement.

The vector layer of the high depth zone, extracted from Landsat imageries, has been overlaid with the settlement layer and the intersection between the two layers was calculated using ArcGIS. Attribute tables are compiled in a similar way to HI_1. It is evident from previous discussion that a high HI_2 value for any settlement indicates its vulnerability to flood-induced disaster. More than 50% of the area of 19 settlements was affected by high flood depth. Ten of these 19 settlements are situated in Nadia District and nine are in the Bardhaman District of West Bengal.

Vulnerability analysis for each settlement can be done more meaningfully by synthesizing HI_1 and HI_2 with population density. It was mentioned before that each settlement polygon already contains information regarding average population density as the number of people per square kilometre. An SQL query was built to identify those settlements that have a high population density and which are vulnerable with respect to both hazard factors.

A total of 18 settlements meet the criteria where $HI_1 < 25$, $HI_2 > 50$ and Pop_density >750. These settlements have been classified as extremely vulnerable. The distribution of these settlements in different development blocks is shown in Table IV. Table V shows the location coordinates of the centroids of the polygons. The coordinates will help local planners and administrators easily locate highly vulnerable settlements over the administrative boundaries of revenue villages. During any high-magnitude floods, the local administration should place high priority on providing relief to the population of these settlements.

Table IV. Development block-wise distribution of extremely flood-vulnerable settlements

Development block	District	Flood occurrences (1991–2000)	Vulnerable settlements
Chapra	Nadia	3	2
Kalna-I	Bardhaman	2	3
Krishnanagar	Nadia	3	1
Nabadwip	Nadia	3	2
Purbasthali-I	Bardhaman	4	3
Purbasthali-II	Bardhaman	3	2
Ranaghat-II	Nadia	2	1
Shantipur	Nadia	2	4

Table V. Precise location of centroids of the settlements that are highly vulnerable to flooding

Settlement ID	Block Name	Latitude (°N)	Longitude (°E)
1	Kalna-I	23·296	88·361
2	Kalna-I	23·317	88·291
3	Purbasthali-I	23·336	88·348
4	Purbasthali-I	23·356	88·293
5	Kalna-I	23·354	88·290
6	Purbasthali-I	23·356	88·284
7	Purbasthali-II	23·500	88·340
8	Purbasthali-II	23·509	88·283
9	Santipur	23·218	88·412
10	Santipur	23·296	88·364
11	Ranaghat-II	23·292	88·650
12	Santipur	23·310	88·369
13	Nabadwip	23·329	88·380
14	Santipur	23·332	88·349
15	Nabadwip	23·335	88·349
16	Krishnanagar-I	23·456	88·545
17	Chapra	23·501	88·615
18	Chapra	23·530	88·606

Table V shows that these settlements are not only vulnerable to extreme flood events like in 2000, but most of them are also located in very frequently flooded development blocks because of a rapidly increasing population density and hence rapid expansion of settlements to highly flood-prone zones.

To verify rural settlement increases, an old 1 : 63 360 map of Nakashipara development block, produced between 1917 and 1921, was digitized. This map depicts the pre-independence settlement pattern in the region. A query was built in the current settlement layer to identify the polygons in Nakashipara development block that have more than 50% of their area under flood water. A total of 33 polygons from the current settlement layer met this criterion. The results show that 17 of these 33 settlements have no area of intersection with the settlements present during 1921. Hence, it can be concluded that the majority of vulnerable settlements are of new origin and that people chose to build settlements in a flood-prone zone most probably out of socio-economic compulsion.

CONCLUSIONS

This study demonstrates a cost-effective and efficient way to create a moderate-resolution spatial database for identifying human settlements that are highly vulnerable to monsoon flooding. The results of this study are, as yet, not conclusive. It is an effort to build the architecture of a flood hazard database that can be analysed from a spatial dimension. Any other combination of the attributes based on in-depth local knowledge might prove more rational for implementing mitigation measures. The accuracy of classifying flooded areas from non-flooded areas is limited by partial cloud cover over the area and the predominance of tree canopy in the rural settlements. Unavailability of updated large-scale maps also restricts our ability to undertake a detailed mapping of settlements in the study area. The lack of high-resolution digital terrain data for developing countries also leads to difficulty in assessing the accuracy of the classification results (Sanyal and Lu, 2004). In spite of these constraints, the study has resulted in a reasonably accurate spatial information database that is suitable for generating 1 : 250 000 hazard maps. Since a very high magnitude and low-frequency flood has been studied, there is very little possibility of leaving any potentially flood-vulnerable settlement unidentified. We propose that efforts should be made to generate cost intensive, high-resolution terrain data, at least for areas of high flood-depth zone. Such an effort will immensely enhance our capability to estimate flood hazard and assess the vulnerability of people.

ACKNOWLEDGEMENTS

This research has been funded by the National University of Singapore (NUS) research grant R-109-000-049-112. NUS's support of this research is gratefully acknowledged by the authors.

REFERENCES

Bagchi K. 1945. *The Ganges Delta*. University of Calcutta: Kolkata, India.

Basu PK. 2001. Flood 2000: a comprehensive study with major floods in the state *Sechpatra* **3**(1): 6–12.

Berg A, Gregiore JM. 1983. Use of remote sensing techniques for rice production forecasting in West Africa (Mali and Guinea: Niger–Bani project). In *ESA Satellite Remote Sensing for Developing Countries*, Ispra, Italy; 161–168.

Bhavsar PD. 1984. Review of remote sensing applications in hydrology and water resource management in India. *Advances in Space Research* **4**(11): 193–200.

Cannon T. 2000. Vulnerability analysis and disasters. In *Floods* Vol I, Parker DJ (ed.). Routledge: New York; 45–55.

Deutsch M, Ruggles F. 1974. Optical data processing and projected application of the ERTS1 imagery covering the 1973 Mississippi river valley floods. *Water Resource Bulletin* **10**(5): 1023–1039.

Deutsch M, Ruggles F, Guss P, Yost E. 1973. Mapping the 1973 Mississippi floods from the Earth Resource Technology satellites. In *Proceedings of International Symposium on Remote Sensing and Water Resource Management*. *American Water Resources Association*, No. 17. AWRA: Burlington, Ontario; 39–55.

Dowman IJ, Morris AH. 1982. The use of synthetic aperture radar for mapping. *Photogrammetric Record* **10**(60): 687–696.

England CF, Hunt GE. 1985. A bispectral method for the automatic determination of parameters for use in imaging satellites cloud retrievals. *International Journal of Remote Sensing* **6**(8): 1545–1553.

FEMA. 2003. Federal Emergency Management Agency. http://www.fema.gov/nfip [20 September 2000] (in Bengali).

Ganashakti. 2000. November 3 (In Bengali).

Goswami AK. 1983. Studies on the nature of floods in the depressed belt of southern West Bengal. *Geographical Review of India* **41**(2): 189–201.

Haack BN, Slonecker ET. 1994. Merged spaceborne radar and thematic mapper digital data for locating villages in Sudan. *Photogrammetric Engineering and Remote Sensing* **60**(10): 1253–1257.

Haack BN, Herold ND, Bechdol MA. 2000. *Photogrammetric Engineering and Remote Sensing* **66**(6): 709–716.

Henderson FM. 1995. An analysis of settlement characterization in central Europe using SIR-B radar imagery. *Remote Sensing of Environment* **54**: 61–70.

Henderson FM, Xia ZG. 1997. SAR applications in human settlement detection, population estimation and urban land use pattern analysis: a status report. *IEEE Transactions on Geoscience and Remote Sensing* **35**(1): 79–85.

Hewitt K. 1997. *Regions of risk: A Geographical Introduction to Disasters*. Addison-Wesley Longman: Harlow, UK.

Imhoff ML, Vermillion C, Story MH, Choudhury AM, Gafoor A, Polcyn P. 1987. Monsoon flood boundary delineation and damage assessment using space borne imaging radar and Landsat data. *Photogrammetric Engineering and Remote Sensing* **53**(4): 405–413.

Islam MM, Sado K. 2000a. Flood hazard assessment in Bangladesh using NOAA AVHRR data with geographical information system. *Hydrological Processes* **14**(3): 605–620.

Islam MM, Sado K. 2000b. 1997. Development of flood hazard maps of Bangladesh using NOAA-AVHRR images with GIS. *Hydrological Sciences Journal* **45**(3): 337–355.

Islam MM, Sado K. 2002. Development of priority map flood countermeasures by remote sensing data with geographic information system. *Journal of Hydrologic Engineering* **7**(5): 346–355.

Kunte PD, Wagle BG. 2003. Sediment transportation and depth variation study of the Gulf of Kutch using remote sensing. *International Journal of Remote Sensing* **24**(11): 2253–2263.

Lillesand TM, Kiefer RW. 2000. *Remote Sensing and Image Interpretation*, 4th edn. John Wiley: Chichester.

Liu J, Teng X, Xiao J. 1986. Application of Shuttle imaging radar data for land use investigation. *Remote Sensing of Environment* **19**: 291–301.

Liu Y, Islam MN, Gao J. 2003. Quantification of shallow water quality parameters by means of remote sensing. *Progress in Physical Geography* **27**(1): 24–43.

Lo CP. 1986. Settlement, population and land use analysis of the North China Plain using Shuttle imaging radar-A data. *Professional Geographer* **38**(2): 141–149.

McGinnis DF, Rango A. 1975. Earth Resource Satellite System for flood monitoring. *Geophysical Research Letters* **2**(4): 132–135.

Mitchell JK (ed.). 1999. *Crucibles of Hazard: Mega-Cities and Disaster in Transition*. United Nations Publication: New York.

Morrison RB, Cooley ME. 1973. Assessment of flood damage in Arizona by means of ERTS-1 imagery. In *Proceedings of Symposium on Significant Result Obtained from the Earth Resource Satellite 1*, New Carrollton, Maryland, **Vol. 1**; 755–760.

Rango A, Anderson AT. 1974. Flood hazard studies in the Mississippi river basin using remote sensing. *Water Resources Bulletin* **10**(5): 1060–1081.

Rango A, Solomonson VV. 1974. Regional flood mapping from space. *Water Resources Research* **10**(3): 473–484.

Ruangsiri P, Sripumin R, Polongam S, Kanjanasuntorn P, Wongparn S. 1984. *State of flooding in the Mun-Chi river basin area, N. E. Thailand by digital Landsat data analysis*. Report: Remote Sensing Division, National Resource Council of Thailand, Bangkok, Thailand.

Rudra K. 2001. *The Flood in West Bengal: September 2000*. Jayasree Press: Kolkata, India.

Sanyal J, Lu XX. 2004. Application of remote sensing in flood management with special reference to monsoon Asia: a review. *Natural Hazards* **33**: 283–301.

Sanyal J, Lu XX. 2005b. GIS based flood hazard mapping in Gangetic West Bengal. *Singapore Journal of Tropical Geography* (in press).

Saunders RW. 1986. An automated scheme for the removal of cloud contamination from AVHRR radiance over western Europe. *International Journal of Remote Sensing* **7**(7): 867–888.

Sheng Y, Su Y, Xiao Q. 1998. Challenging the cloud-contamination problem in flood monitoring with NOAA/AVHRR imagery. *Photogrammetric Engineering and Remote Sensing* **64**(3): 191–198.

Spate OHK, Learmonth ATA, Learmonth AM. 1965. *India and Pakistan: A General and Regional Geography*. Methuen: Bungay, Suffolk.

Smith LC. 1997. Satellite remote sensing of river inundation area, stage and discharge: a review. *Hydrological Processes* **11**: 1427–1439.

The Times of India. 2000. Rain reality mired in recording controversy. 18 November.

Townsend PA, Walsh SJ. 1998. Modelling flood plain inundation using integrated GIS with radar and optical remote sensing. *Geomorphology* **21**(98): 295–312.

USGS. 2003. http://edcdaac.usgs.gov/aster/ast14dem.html [6 December 2003].

Varley P (ed.). 1994. *Disaster, Development and the Environment*. Wiley: Chichester, UK.

Wadge G, Wisloki AP, Pearson J, Whittow JB. 1993. Mapping natural hazard with spatial modelling system. In *Geographic Information Handling Research and Application*, Mather PM (ed.). Wiley: New York; 312–324.

Wang Y, Colby JD, Mulcahy KA. 2002. An efficient method for mapping flood extent in a coastal flood plain using Landsat TM and DEM data. *International Journal of Remote Sensing* **23**(18): 3681–3696.

Yang C, Zhou C, Wan Q. 1999. Deciding the flood extent with Radarsat SAR data and image fusion. In *Proceedings of 20th Asian Conference of Remote Sensing*, Hong Kong, 22–25 November.

