J Vis
DOI 10.1007/s12650-017-0453-3

REGULAR PAPER

Sudipto Saha · Shashank Shekhar · Sounak Sadhukhan · Prasun Das

An analytics dashboard visualization for flood decision support system

Received: 30 March 2017 / Revised: 9 September 2017 / Accepted: 27 September 2017

Abstract Nowadays, visual analytics dashboard for decision support system for disaster management is very essential, as the frequency of several disasters is growing rapidly. It is very useful to take decision in time-sensitive environment from diverse rapid data set. Thus, it is critical to reduce the overhead in data integration and visualization for efficient decision making. Geographical map visualization can be an effective solution in such environments by providing an integrated approach to extract, integrate, and view diverse information. In this paper, an analytics dashboard has been developed for identifying and visualizing risk zones and vulnerable areas before disaster has occurred. For this, rescuer will have more time to be prepared and a better plan will be produced to rescue the people and reallocate them into safe areas.

Keywords Visualization · Disaster management · Flood · Risk · Vulnerability · Decision support

1 Introduction

A disaster is a serious disruption of the functioning of a community or a society involving widespread human, material, economic, or environmental loss and impacts, which exceeds the ability of the affected community or society to cope using its own resources. Disaster can be natural or even manmade. Floods account for about one-third of all the geophysical hazards globally affecting a large number of people compared to other natural calamities (Smith and Ward 1998). Some of the remarkable floods in history are flooding in China in the years 1931 and 1938 in Yellow River, 1974 flood of Bangladesh and flooding in Myanmar from cyclone Nargis in 2008, and so on. The International Flood Network indicates that from 1995 to 2004, natural disasters caused 471,000 fatalities worldwide and economic losses totalling approximately $49 billion USD, out of which approximately 20% of the fatalities and 33% of the economic damages were contributed to floods alone. Flooding claims more than 20,000 lives per year over the globe and adversely affects 140 million people on average each year (Smith 1997) (WDR 2003 and WDR 2004). According to the United Nations World Water Development Report 2, India and its subcontinent are mostly affected by flood and about 83% of all the flood-related disasters occur in Asia. According to the International Disaster Database (EM-DAT), since the year 1900, a total of 143 floods have occurred in India with more than 300 million affected people. More recent flood-related disasters in India were 2013 North-India flood which resulted in the death of more than 5700 people and 2015 Chennai flood took more than 500 lives and 1.8 million displaced. Adhikari et al. (2010) have made a compilation of flood-related disasters which have occurred between 1998 and 2010 according to year of occurrence, duration, severity, causes, and location.

S. Saha · S. Shekhar · S. Sadhukhan (✉) · P. Das
Statistical Quality Control and Operations Research Unit, Indian Statistical Institute, No. 203, Barrackpore Trunk Road, Kolkata, West Bengal 700108, India
E-mail: sounaks.cse@gmail.com

Published online: 04 November 2017

Eastern region of India is mostly affected by flood, in particular, West Bengal, Orissa, Bihar, and Assam. The state of Orissa in the eastern coast can be cited as an example, where a large number of severe and devastating floods, endangering both life and property have been recorded (Parthi et al. 2012). Similar parts of Eastern India, including the subject of our case study Kolkata Municipal Development Area (KMDA), have experienced several severe floods causing huge damages to life and property.

Though it is impossible to prevent disasters completely, but necessary preparedness, planning and effective response would mitigate the loss of lives, property damages, and sufferings. To develop better planning and provide a better service at emergency situations, a platform is in need which incorporates all the necessary information from different sources and represents pictorially the current situation to the end user for better understanding. The disaster management organizations handle massive data sets every day from different sources. These data cannot be interpreted through naked eyes, but to take better and efficient decision and rapid assessment of the disaster situation, interpretation of the bulk size data set is required. The idea of visual data exploration is to present the data in some visual formats, allowing the user to gain insight into the data, draw conclusions in proficient way during preparation, planning, and response phase of the disaster management.

Researchers have made numerous studies on visualization system for disaster situation from several aspects. Some of the indispensable studies are Kemec et al. (2009), Liu (2009), Gajdoš and Radecký (2009), Ma et al. (2012), Lu et al. (2015), Xie et al. (2016), Cerutti et al. (2016), Li et al. (2016), and Li et al. (2017a, b) which create a major impact in this area. The studies mentioned above are mainly used to improve situational awareness and surveillance purposes in disaster situations without considering the socio-economic parameters which are the most important parts to identify vulnerable and risk zones. In this article, a decision support system for flood is developed to predict the risk of the areas using vulnerability and hazard. It will help the disaster organization to identify the most vulnerable region even before a flood can occur and they will get reasonable time to prepare and produce a plan to mitigate the effect of flood. Vulnerability and hazards are derived from flood prediction model of Manna (2016). The vulnerability is defined using several socio-economic parameters, and hazard is calculated using inundation depth, river-water-level data, and predicted rainfall.

The objective of our study is threefold: categorizing the municipalities into household, industrial, commercial, and mixed types; showing predicted rainfall for each municipality on a particular day; and segregating the municipalities in various risk zones depending upon the risk calculated using hazard and vulnerability and display temp in geo-spatial map. The proposed methods showed positive results in the identification of areas affected by the flood at regional scale.

The rest of this paper is organized as follows. Section 2 focuses on different studies carried out on visualization for disaster management. Section 3 illustrates system design and databases that include voronoi polygon formulation, database description, risk scale, and detailed design of the proposed system. Section 4 discusses results and Section 5 concludes with limitations and future scope of the study.

2 Related study

Plenty of researches have been carried out on data visualization for disaster management in several aspects, some of which are discussed here.

2.1 Disaster management

(Kemec et al. 2009) offered a 3D visualization for disaster management with the help of a rule-based model to explore the relationships between the types and levels of disaster. These, in turn, helped in achieving interoperability by integrating the semantic and geometric information into a single model using the levels of detail. In another study by (Ma et al. 2012), the approaches of disaster data management and disaster information expression are described with the help of 3D GIS. The authors proposed a novel method to manage disaster data in integrated environment taking account of the multi-score, multi-scale, and other features of disaster data. The ways about how to express the various kinds of disaster data were also explained in this study.

Based on the concept of disaster risk triangle (hazard, vulnerability, and exposure), a comprehensive analysis using Urban Flood Simulation Model and Urban Flood Damage Assessment Model is proposed for urban flood risk analysis by Li et al. (2016), and Xie et al. (2016) have presented a novel web browser-based interactive visualization tool for rapid tsunami evacuation planning. The interface loaded with simulated route-schedule flow data for visual analysis allows users to explore the spatio-temporal evolvement of an evacuation process with a dynamic map. The interface also supports multiple filtering and summary interaction including filtering the roads by their utilization efficiency and summarizing the congested roads during evacuation. The multiple linked views as well as some diagrams examine the evacuation efficiency and verify evacuation schedules. The proposed tool also supports dynamic evacuation parameter settings for making the evacuation strategies and optimizing decision making. The disaster managers and domain experts conducted case study on the possible tsunami evacuation events. This paper received encouraging remarks on usability and performance. Li et al. (2017a) have proposed a visual analytics method to analyze flood risk in cultural heritage from various aspects, including expected flood peak flow, flood propagation, flood impact, and vulnerability. The proposed method can also provide multiple information from numerous scales, containing the basin, site, multi-cave, and single-cave scale levels. The amalgamation of the visualization techniques enables the proposed method to support decision makers to make decisions with respect to extenuation measures.

Lu et al. (2015) have explored the underlying trends in positive and negative sentiments in context of disasters and geographically related sentiment. A visual analytics framework for sentiment visualization of geo-located Twitter data has been proposed in this study. The proposed framework comprises of two components with the first one being sentiment modelling and the second one being geographic visualization. In particular, the authors have devised an entropy-based metric for modelling of the sentiment hidden in social media data. With the investigation of the extracted sentiment, a visualization framework has been further explored for finding out the uncertainty of public opinion. The authors have validated their model with Ebola Twitter data set, which explores the methodology that how the visual analytics techniques and sentiment modelling can reveal interesting patterns in disaster scenarios. Cerutti et al. (2016) have used geo-referenced tweets to define the geo-spatial footmark of a flood event. After cleaning the data, density-based clustering, distance-bounded spatio-temporal event clustering, and data-driven territory tessellation techniques were applied. Visual analysis was used to define the combination of the best parameters. A comparison between the results and ground-truth data was performed. The proposed methods showed positive results in identification of the areas affected by flood at regional scale. Li et al. (2017b) have developed data-driven disaster management using advanced data collection and analysis techniques to achieve more effectiveness and responsiveness. In this study, they have developed methods that help users to understand the situation reports and improve situation awareness; to present the intelligent information delivery techniques for effectively supporting users' information needs; and to discuss how the techniques developed in data mining and machine learning are leveraged to improve the effectiveness of the disaster management systems.

2.2 Visualization Systems

To cope with the hydro-related issues such as flood, quality of the air and water, soil erosion, and ecology (Gajdoš and Radecký 2009) designed a platform-independent prototype of an open modular system of environmental risk model, based on internet technologies with the aim to predict hydrographs and flood lakes in real time. The devised system has an online communicational interface to interact with the system for resolution of crisis, interoperability, and usability issues. Another solution for the data-visualization problem has been proposed by (Gupta et al. 2010) who developed a geo-spatial-programmed application with demonstrative power for management of crisis and disaster. (Malik et al. 2011) have developed a visual analytics system to analyze historic response operations and assess the potential risks in the maritime environment associated with the hypothetical allocation of Coast Guard resources. The system includes linked views and interactive displays that enable the analysis of trends, patterns, and anomalies among the U.S. Coast Guard search and rescue operations and their associated sorties. The developed system allows users to determine the potential change in risks associated with closing certain stations in terms of response time, potential lives, and property loss and provides optimal direction as to the nearest available station.

3 Visual data analytics—system design and related database

This paper limits its study to the pre-disaster phase with respect to flood occurrences. It should be noted that risk assessment of a specific area is an important tool in flood disaster management. A critical study on historic flood events would reveal that a majority of the floods in the Eastern parts of India are due to contribution from downstream uncontrolled catchment beyond the reservoir. Precipitation in an area has perhaps the strongest and direct impact on the flood proneness of that area. It is to be assumed that a certain number of rain-gauge stations have been physically installed at different locations (the same as municipality locations) covering the entire landscape of the area under study. Subsequently, the entire region of the study can be demarcated into as many number of voronoi polygons (Supriya et al. 2015) as the number of rain-gauge stations. Each such polygon would encompass only a single location of the rain-gauge station placed. The following sub-sections describe voronoi polygon formulation, database description, risk scale assessment, and detailed design of the proposed system.

3.1 Voronoi polygon formulation

In this study, the area (area under KMDA) is sub-divided by voronoi polygons (Aurenhammer and Klein 2000) as voronoi polygons are useful in dividing regions into sub-regions. The geographical points considered in this study depict rain-gauge stations which are useful for collecting rainfall measurement; however, it cannot be possible that only a particular point is engulfed with rainfall. There always be a surrounding region to each and every point, where "rain" can happen. In addition, rain-gauge stations are able to measure rainfall of surrounding areas. Therefore, voronoi polygons are used to demarcate the surrounding region of a rain-gauge station.

The voronoi polygons are generated using "sweep-line" algorithm (Fortune 1987). This method suggests drawing a sweep line below the agglomeration of points of location of the sites. A parabolic arc is drawn around each site, as shown in Fig. 1a. The edges of the polygons are extended through the points of intersection between the arcs encompassing the sites (Fig. 1b). The position of the beach line (dotted) is indicative of the minimum of the parabolas defined by sites above the sweep line and the sweep line itself, thus separating the known and unknown parts of the voronoi diagram. The red points of intersection between the parabolic arcs represent the break points in the beach line, which move and trace out the voronoi edges (Fig. 1c). Break points are defined by two sites and the sweep line. The sweep line and the beach lines change continuously whenever a new location or site is augmented to the region of study, and whenever the sweep line reaches a new site, a new parabola starts (Fig. 1d). Furthermore, two new break points appear on the beach line and a new voronoi edge is created, suggesting another voronoi polygon (De Berg et al. 2008; Dobrin 2005).

3.2 Database description

This study requires few databases, and the description of which is given below.

Rainfall data set: This data set contains the predicted rainfall amount on a daily basis for the months of June, July, August, and September across the 2016 using the historical rainfall data set.

River-water-level data set: The data set consists of the river-water level corresponding to the river basin surrounding each of the districts under KMDA and the municipalities. The data set contains water-level data for the monsoon season in India [with special reference to The Flood in September (2000a, b): Ref. PIL No. WP No. 113(W) of 2001—by The Committee of Experts constituted by the Hon'ble High Court, Calcutta].

Geographical attributes data set: This data set contains the data related to the geographical attributes with respect to the river basin surrounding the individual districts and municipalities. The data set consists of several variables which are used to calculate the probability of flood occurrence for the areas under consideration (with special reference to The Flood in September (2000): Ref. PIL No. WP No. 113(W) of 2001—by The Committee of Experts constituted by the Hon'ble High Court, Calcutta).

Socio-economic data set: The data set involves the data about the parameters for hazard and vulnerability estimation with respect to each of the locations (in our case locations are municipalities). The type of area has been identified by the dominant socio-activity in each of the municipalities. The real-life data pertaining to the parameters influencing the vulnerability due to flood occurrence in the region of case study have been taken from source representative of the districts under the KMDA areas. The corresponding data have been taken for Hooghly district, Howrah district, Nadia district, North 24 Parganas district, and South 24 Parganas district from the report, which have been published by the HRDCC, Department of Planning and Development, Government of West Bengal.

Rainfall Classification data set: The data set contains the qualitative classification of rainfall. The rainfall forecast magnitudes fall into various segments on the basis of predicted values of the rainfall data for a municipality for a given date (www.imdkolkata.gov.in).

$$\text{Predicted rainfall index} = \left\{ \begin{array}{ll} 0\,\text{mm} & \text{No Rain} \\ \leq 2.4\,\text{mm} & \text{Very Light Rain} \\ > 2.4\,\text{mm} \; \& \; \leq 7.5\,\text{mm} & \text{Light Rain} \\ > 7.5\,\text{mm} \; \& \; \leq 35.5\,\text{mm} & \text{Moderate Rain} \\ > 35.5\,\text{mm} & \text{Rather Heavy Rain} \end{array} \right\}$$

Geographical information data set: This data set is a shape file which contains the geographical information about all the states of India and further all the districts of each state. These geographical data help to construct the outline of West Bengal along with each of its districts (www.gadm.org).

3.3 Flood Risk—Assessment and Scale

Risk analysis provides an estimation of the degree of interaction between society and the environment at the point, where hazards due to the disaster (flood, in this case) unfold. Hence, its assessment can be based upon the following expression:

$$\text{Risk} = \text{Hazard} \times \text{Vulnerability}.$$

The measure of risk for each unit location in each of the voronoi polygons would be scaled uniformly into formation of a risk index, common to any location in the jurisdiction of this study. Thereafter, any location can be subjected to risk when flood occurs. The risk can be estimated based on the range in the risk scale under which the risk index of the location would fall. The methodology applied for formulating the risk scale is explained in the following sections.

3.3.1 Hazard assessment

The forecasted amount of rainfall in the sample location is recorded as a parameter. Runoff of the excess water onto the floodplain can be attributed to the excess overflow from river basin (river-water discharge)

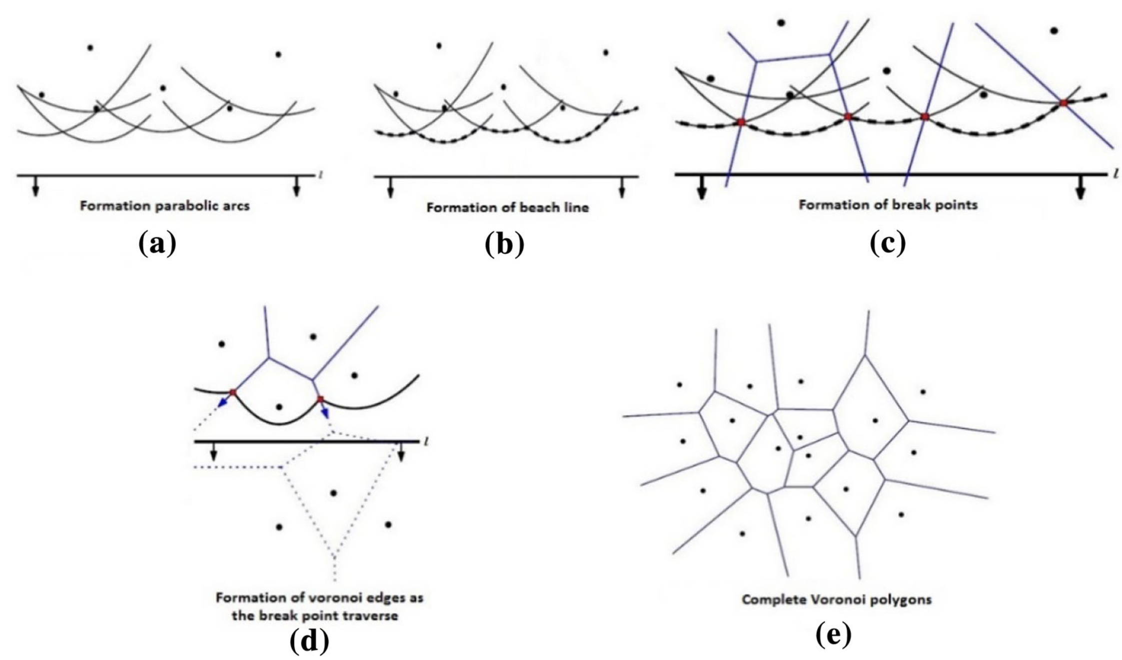


Fig. 1 Different phases of voronoi polygon formation

and to the excess water flow caused due to dam overtopping. This estimated value of runoff along with the forecasted rainfall amount in the unit location act as parameters in determining the inundation depth due to flood in the unit location of study. This inundation depth acts as a primary factor in determining the probability of flood in the location (Manna 2016).

$$\text{Probability of flood occurrence} = P_{flood} = \begin{Bmatrix} \frac{I_{depth}}{I_{depth}+a.h_{tf}} & I_{depth} > 0 \\ 0 & I_{depth} \leq 0 \end{Bmatrix}$$

Here, I_{depth} denotes the inundation depth which, in turn, is dependent upon forecasted rainfall amount and the excess runoff measure, h_{tf} denotes the threshold flood gate of the location, and a is a constant which is to be attributed specific to the location under study (we have considered $a = 1$ in this study).

Expected hazard level of an area is evaluated in accordance with the hazard index which is obtained by fixing certain intervals of flood occurrence probability which, in turn, is directly influenced by the inundation depth. The hazard assessment associated with the chances of flood occurrence is a function of P_{flood}. As P_{flood} reaches a certain range of values, the corresponding hazard index formulates to a certain constant factor measure (Manna 2016):

$$\text{Hazard} = \begin{cases} 0 & I_{depth} \leq 0 \\ 1 & 0 < I_{depth} < 0.25a.h_{\text{tf}} \\ 2 & 0.25a.h_{tf} \leq I_{depth} < 0.5a.h_{\text{tf}} \\ 3 & 0.5a.h_{tf} \leq I_{depth} < 0.75a.h_{\text{tf}} \\ 4 & I_{depth} \geq 0.75a.h_{\text{tf}} \end{cases} .$$

The individual thresholds indicated above suggest the fact that for values of P_{flood} between 0 and 0.2 (20%), the hazard index is 1. For values of P_{flood} between 0.2 (20) and 0.33 (33.33%), the hazard index is 2. For values of P_{flood} between 0.33 (33.33) and 0.4286 (42.86%), the hazard index is 3. For values of P_{flood} greater than 0.4286 (42.86%), the hazard index is 4. This hazard index, when multiplied by the vulnerability measure, would give an assessment of the risk associated with the flood (Cancado et al. 2008).

3.3.2 Vulnerability Assessment

The measure of vulnerability due to the occurrence of flood can be estimated through the specific parameters that form a part of the socio-economic set of elements. Vulnerability assessment can have three specializations, namely, vulnerability to life (V_{Life}), vulnerability to property (V_{Property}), and vulnerability to the area (V_{Area}) as a whole. The assessment of the first specialized category, i.e., V_{Life} in a unit location due to flood, depends on major attributes like forecasted rainfall amount and number of flood-prone sites in the vicinity of the current location. V_{Property} denotes the chances of property and assets of the people and the unit location as a whole to be affected by the flood. V_{Area} is inversely proportional to the awareness expected to be created around the most hazardous locations during flood.

Vulnerability to the Life: The illustrative expression for assessing the vulnerability of flood to life specified as (Manna 2016)

$$V_{\text{Life}} = \left[x_1 \cdot \overline{R_{\text{i}}} + x_2.F_{\text{p}} + x_3 \cdot \frac{\overline{U_{\text{p}}}}{P_{\text{A}}} + x_4 \cdot \frac{\overline{O_{\text{A}} + N_{\text{C}} + N_{\text{H}} + L_{\text{I}} - L_{\text{A}}}}{P_{\text{A}}} + x_5 \cdot \overline{(E_{\text{C}} + C_{\text{C}} + H_{\text{C}} + S_{\text{C}} + H_{\text{p}})}\right] \cdot (E_{\text{A}}), \quad \text{(i)}$$

where

R_{i}: spatially interpolated rainfall amount in respect of the unit location
F_{p}: no. of flood-prone sites in the proximity
U_{p}: unemployment (no. per 1000)
P_{A}: population/area measure
O_{A}: no. of people above 70 years of age/1000 people
N_{C}: non-car ownership
N_{H}: fraction of population with no shelter
L_{I}: no. of people who are victims of long-term illness/1000 people
L_{A}: Life Insurance Availability (no. per 1000)

E_C: no. of electrical connections
C_C: no. of community centres
H_C: no. of health centres
S_C: no. of schools
H_p: no. of hospitals
E_A: dominant economic activity

Vulnerability to the Property: The illustrative expression for assessing the vulnerability of flood to property may be specified as (Manna 2016)

$$V_{\text{Property}} = \left[y_1 \cdot \overline{R_i} + y_2 \cdot \frac{\overline{O_s + G_s}}{P_A} - y_3 \cdot \frac{\overline{P_i}}{P_A} + y_4 \cdot \overline{(T_c + W_c)} \right] \cdot (E_A), \tag{ii}$$

where
$\overline{R_i}$: spatially interpolated rainfall amount in respect of the unit location
O_s: no. of owned structures
G_s: no. of govt. structures
P_i: Property Insurance Availability (no. per 1000)
P_A: Population/area measure
T_c: no. of telephone connections
W_c: no. of water supply connections
E_A: dominant economic activity

Vulnerability to the area: the illustrative expression for assessing the vulnerability of flood to the location may be specified as (Manna 2016)

$$V_{\text{Area}} = \frac{1}{\{(N_M + N_{TV})P_A\}^{z_1}}, \tag{iii}$$

where
N_M: no. of mobile phones/1000 people
N_{TV}: no. of televisions/radios per 1000 people
P_A: population/area measure.

In Eqs. (i), (ii), and (iii), any parametric variable of the form $\overline{Q}$ indicates the ratio:

$$\frac{\text{Actual value of the parameter}}{\text{Maximum value of the parameter for all unit locations}}.$$

In Eq. (i), x_1–x_5, in Eq. (ii), y_1–y_4, and in Eq. (iii), z_1 are the weights to the different parameters.

Each of these vulnerabilities is calculated on the basis of the data provided in the database. The total vulnerability measure of the unit location can thus be assessed by the summation of the three components of vulnerability explained in the previous sections:

$$\text{Vulnerability} = x \cdot V_{\text{Life}} + y \cdot V_{\text{Property}} + z \cdot V_{\text{Awareness}},$$

where the constants x, y, and z represent area weights with respect to each of the three components of vulnerability. In this study, all area weights have been considered equal to one.

3.3.3 Risk Scale

For the formulation of a risk scale, every risk value corresponding to every municipality encompassing the region of study needs to be brought down to a uniform risk index, the value of which ranges from 0 to 1. Furthermore, the risk index corresponding to the polygon, denoted by R_{pg}, can be evaluated as $R_{pg} = \frac{R_p - R_{min}}{R_{max} - R_{min}}$, where R_{pg} is the scaled risk value for a particular polygon (municipality), R_p is the actual assessed risk for a particular polygon, and R_{max} and R_{min} are the maximum and the minimum of all assessed risk values, respectively. The risk index measures of every location or polygon may follow any statistical distribution. Hence, it is expected that there exist certain quartile values such as the upper quartile, Q_3, lower quartile, Q_1, and the median, Q_2, of the particular distribution (Fig. 2). With respect to the allocation of degree of risk under which each unit location is expected to lie, it is worth noting that the distribution of the risk index would definitely follow a positively skewed distribution. This is in cognizance with the fact that

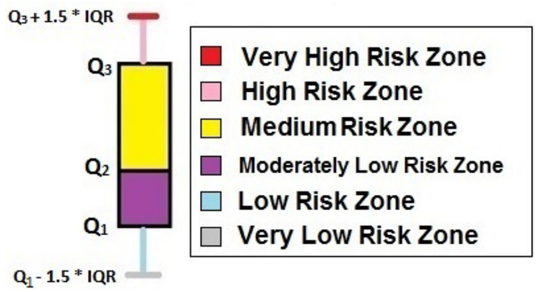


Fig. 2 Box plot on the quartiles of the risk index values

the number of unit locations that would fall in the lower risk zone would be more than the number of locations that would entail very high risk.

This study proposes six zones of risk category under which the unit locations can be allotted, namely, very low-risk zone, low-risk zone, moderately low-risk zone, medium-risk zone, high-risk zone, and very high-risk zone. The specific risk zones are shown in Fig. 2 in the form of a box plot and it is used as risk scale of the proposed system. Each and every location is being coloured according to the category the location falls into, corresponding to the values of their risk index. The higher whisker is marked at a value of $[Q_3 + 1.5 \times IQR]$, while the lower whisker is marked at a value of $[Q_1 - 1.5 \times IQR]$, where IQR indicates the inter-quartile range of values of risk index. If any risk index corresponding to any particular unit location equals or goes beyond any of the values corresponding to the two whiskers, they are termed outliers. These locations would lie in the zone of the highest risk or the lowest risk accordingly. The locations under the highest risk zone would invariably entail the highest priority from the point of view of relief distribution planning or setting up of warehouse locations in anticipation of flood event.

3.4 System design: a schematic diagram

A schematic diagram, shown in Fig. 3, represents the overall design of the system. There are five data sets: geographical attributes, rainfall data set, socio-economic data set, spatial data set, and water-level data set. The grey boxes represent the databases holding these data sets. Associated with the databases, there are corresponding arrows leading to some white boxes. While an arrow basically propagates the effects of applying some processes or operations on the data sets on its left end, the white box at the right end of the arrow holds the results of the operations. For example, a grey box containing the “Rainfall Dataset” has the correspondence with the white box holding “Forecasted Rainfall Amount” with an arrow in-between in the direction from the former to the latter. As some forecasting operations have been applied on the rainfall data set, the arrow is basically carrying the effect of the operation from the grey box to white box resulting in a forecasted rainfall amount (2016 in this case). The amount of forecasted rainfall together with the geographical attributes produces the prediction of probability of flood occurrence. In the diagram, there are other sets of arrows existing between two or more white boxes. In this case, arrows coming from the white boxes transfer their cumulative effects to the resultant white box. The “Socio-economic Dataset” and the forecasted amount of rainfall jointly estimate the degree of vulnerability. In addition, the historical data set of water level corresponding to the river basin surrounding each of the districts under KMDA and the municipalities has been forecasted for the year 2016 from river-water-level data set. Similarly, “Risk Index” is a result that has been found by combining the results “Predicted Probability of flood occurrence”, “Estimation of Vulnerability”, and “Forecasted Water Level”. Therefore, three arrows from the three mentioned results are merged and ended to “Risk Index”. Finally, the “Final Dashboard” is made using both “Risk Index” and “Voronoi Polygon”. In addition, six categories have been formed according to the different ranges of the risk index, which has further been used to demarcate the regions according to their severity level. This would help to prioritize the regions in the post-disaster phase towards supply chain management.

Finally, the “Spatial Dataset” is used to produce the voronoi diagram on the regions considered. It consists of latitudes and longitudes of all the regions and also the geographical information about all the states of India along with all the districts of each state. These geographical data are used to construct the outline of West Bengal along with each of its districts. As a voronoi diagram is actually unbounded, therefore, a boundary for the whole voronoi diagram has been established by merging the computed voronoi

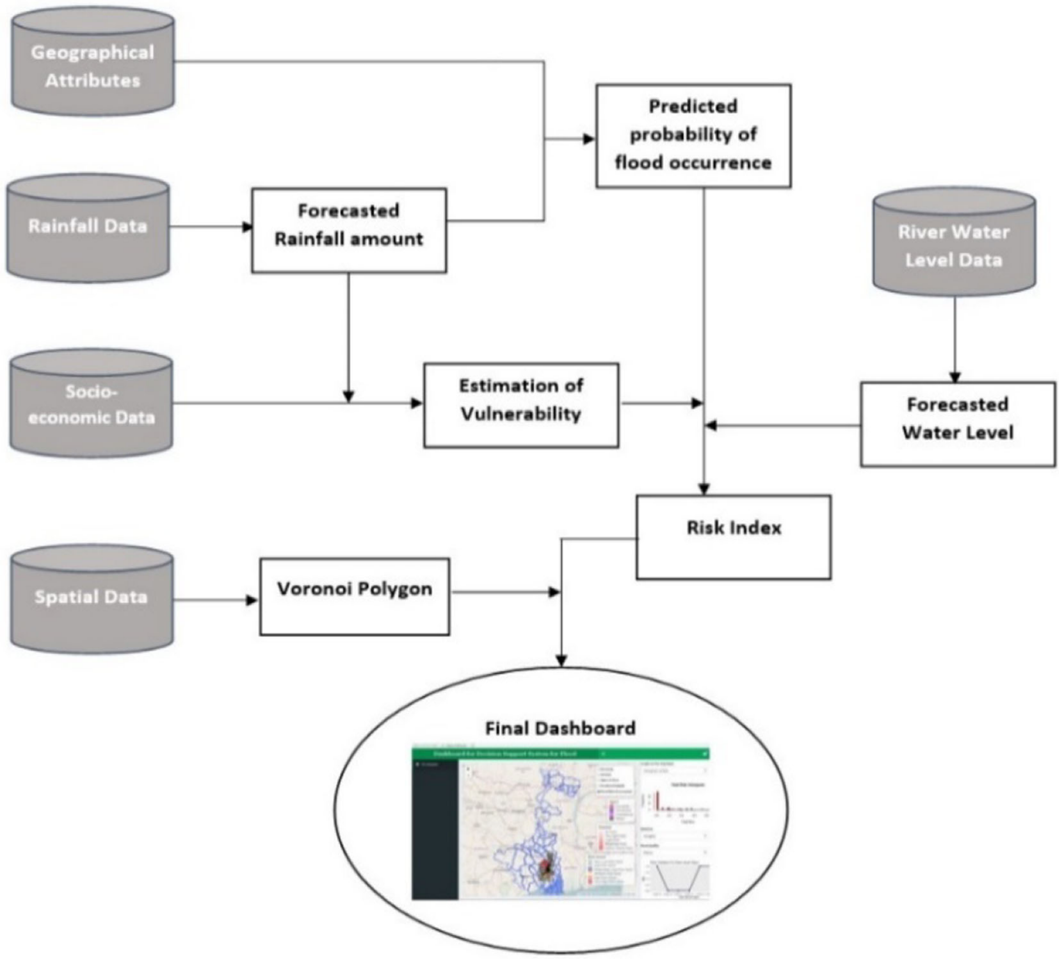


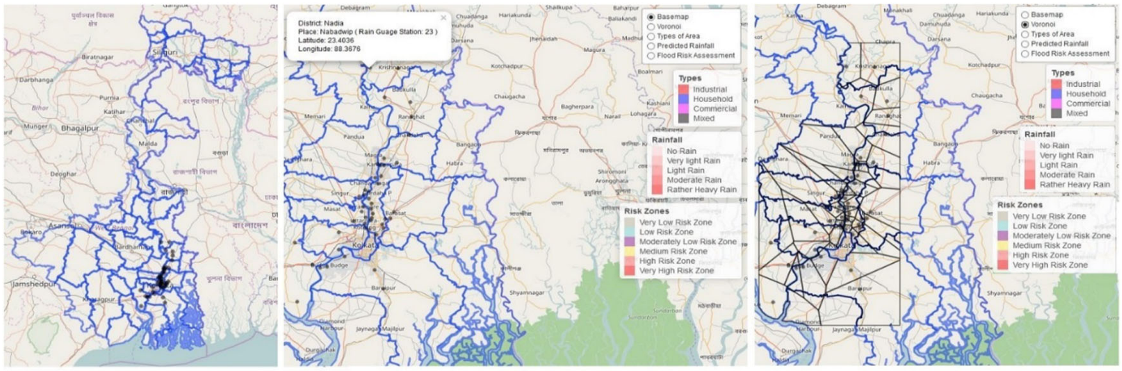
Fig. 3 Schematic diagram of the proposed system

with the geographical outline. Finally, the overall visualization and the respective dashboard are made taking all the above results.

4 Case study

The primary task is to form a digital map of West Bengal state of India demarcating all the districts. This is called a base map. Data in the form of shape file are needed to plot the outline of the state and its districts and are taken from geographical information. After completion of this step, voronoi polygons are overlapped with the base map. Each polygon of voronoi diagram embedded map is further sub-divided into four types of unit areas, namely, industrial, commercial, household, and mixed. Predicted rainfall in all 47 voronoi polygons on a particular day is represented using different colour codes to get an estimate of the areas having the maximum amount of rainfall and thus are more susceptible to flood. Risk levels of each polygon as a cumulative sum of all risk factors affecting the various unit locations in that given polygon is calculated. These risk levels are brought down to a uniform risk index, lying in the range 0–1. Colour codes are used for different risk indices and these polygons are highlighted using these colours to showcase the regions with varying levels of flood risk assessment. Risk variation plots for the next 7 days (from the date entered) have been plotted for all the municipalities under this study. Risk and vulnerability histograms for all municipalities (rain-gauge stations) are plotted to have an overview of the imminent flood danger.

The map of West Bengal has been constructed using the Leaflet package in R (Fig. 4a). The package makes it easy to integrate and control Leaflet maps in R. Features of this package include creating base

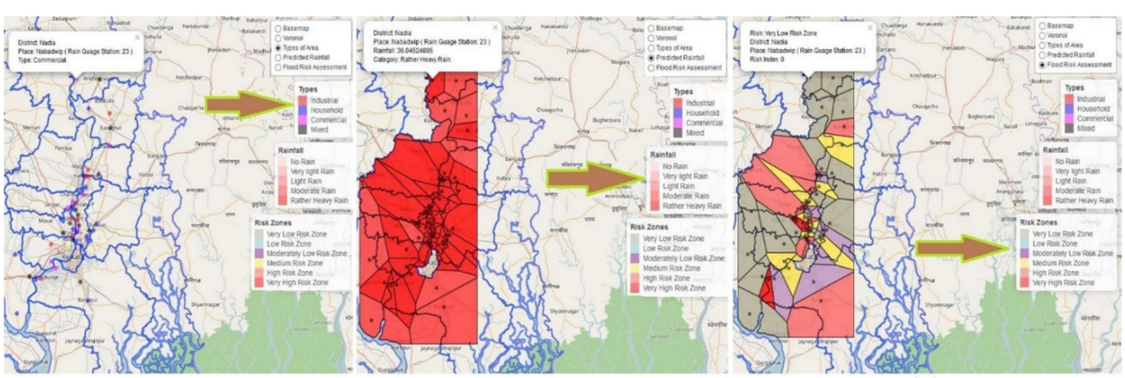


(a) **(b)** **(c)**

Fig. 4 Snapshot of base map of West Bengal (**a**), **b** and different municipalities under KMDA area pointed out with black dots and **c** formation of voronoi diagram on the map

geographical maps with interactive panning/zooming. Maps can be composed using arbitrary combinations of map tiles, markers, polygons, lines, popups, geojson, colours, and legend functions. Moreover, maps can be created right from the R console. Leaflet maps can be easily embedded in Shiny apps. It is easily render spatial objects from the *sp* package, or data frames with latitude/longitude columns on these maps. Knowledge of the exact positions of places on the map through latitude and longitude helps in creating desired popups, labels, markers, etc. To plot the boundary of West Bengal and its districts, data are required in the form of shape file which is taken from GADM database of Global Administrative Areas. GADM is a spatial database of the location of the world's administrative areas (or administrative boundaries) for use in GIS and similar software. Administrative areas in this database are countries and lower level subdivisions such as provinces, departments, districts. After getting the base map, all the 47 municipalities are plot on the map using latitude and longitude and represent them as black dots (Fig. 4b). A voronoi polygon is overlapped with the base map constructed earlier using Leaflet package (Fig. 4c).

Each polygon of voronoi diagram is embedded on the map and further sub-divided these polygons into four primary types of unit areas, namely, industrial, commercial, household, and mixed (Fig. 5a). Polygons can be randomly selected for assessing its flood proneness and corresponding risk involved. Predicted rainfall in all 47 (as the total number of municipalities is 47) voronoi polygons on a particular day is represented (Fig. 5b) using different colour codes to get an estimate of the areas getting the maximum amount of rainfall and thus are more susceptible to flood. Risk levels of each polygon as a cumulative sum of all risk factors affecting the various unit locations in that given polygon is calculated. These risk levels are map to a uniform risk index, lying in the range 0–1. Colour codes are used for different risk indices and these polygons are highlighted using these colours to showcase the regions with varying levels of flood risk assessment (Fig. 5c).



(a) **(b)** **(c)**

Fig. 5 **a** Pointed out different types of areas with different colour codes, **b** predicted rainfall on 3rd September 2016 on different polygons and flood risk of different polygons on that particular day

Fig. 6 Risk variation plot for a particular municipality for consecutive 7 days starting from 3rd September 2016

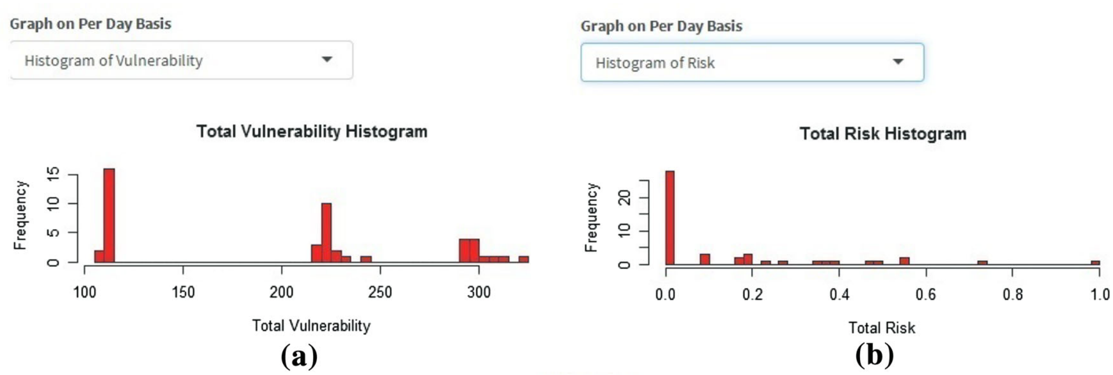


Fig. 7 Histogram of **a** total vulnerability versus frequency and **b** total risk versus frequency on 3rd September 2016

Risk variation plots for the next 7 days (from the date entered) have been plotted for all the municipalities under our study (Fig. 6). Risk and vulnerability histograms for all municipalities are plotted (Fig. 7) to have an overview of the imminent flood danger. The immediate inference that can be drawn from such a prediction of flood risk in the unit locations involves the planning process ahead of the impending monsoon season. In terms of allocating resources for relief distribution of setting up of necessary shelters as part of the disaster management scheme, the effective measures in the pre-disaster phase need to be implemented in the most vulnerable and critical of the locations. In other words, the municipalities or the districts on the larger front which are predicted to lie in the very high-risk zone and the high-risk zone would eventually be in the hit list for immediate action in terms of preparedness for disaster management. The priority for relief distribution and pre-disaster planning should also be done keeping in cognizance the risk zones under which each of the municipalities is predicted to fall over a given period in the impending monsoon season.

5 Qualitative evaluation of the proposed system

The decision support system proposed in this paper is developed keeping in mind in and around Kolkata region, although it is not a flood-prone area, but definitely has severe water logging problem. The system is based on the needs of experts comprising of different government officials from various relevant departments of Government of West Bengal, India. The experts participate in the evaluation of our proposed analytics dashboard as well.

The first step was to present our visual analytics dashboard to the experts for their assessment. Thereafter, they have evaluated the developed system from several aspects such as graphical representation, interactiveness, user-friendly, functionality, and support towards decision making. According to them, all visualization tools in the dashboard are suitable and logical to flood-related scenarios. Moreover, the experts considered this proposed system to be beneficial for risk analysis and decision making at the time of flood. A few recommendations were, for example, (i) the visual design should be improved, so that it looks more interactive and more effective for user understandability; (ii) the system must have some real-time monitoring ability, so that it can help the responder to provide a clear view of the situation in emergency; (iii) some functions should be added for post-disaster situations such as shelter allocation and relief distribution, etc., which might help to configure rescue, reallocation, and relief distribution plan and act as a comprehensive support system for flood disaster. Moreover, the officials have suggested to carefully look into the technical matters during real-life deployment of the system.

In summary, as envisaged by the experts, the proposed dashboard visualization is a systematic and integrated system that can aid the government by predicting flood in advance and finding out vulnerable areas before the actual occurrence. Through the proper use of the proposed system, the effect of the flood can be mitigated to a great extent through proper planning and preparation.

6 Conclusion and future scope

The study has been presented with the view that the proposed models for rainfall forecasting, flood prediction, and flood risk assessment are due for validation and further analysis. The rainfall pattern of any area is a primary component in the field of flood prediction. The inundation depth is a function of various parameters including rainfall pattern. Inundation depth, in turn, decides the risk index of a given area. The subsequent river-water levels and river-water flow discharge vary with significant amount of uncertainty. The phenomenon of dam or reservoir overtopping is also probabilistic in its measure. All of these parameters are help to determine flood proneness of a unit location. The study is based on the rainfall history of five districts in West Bengal comprising of 47 municipalities. Each municipality has one rain-gauge station from where the rainfall data are collected. This article proposes to form subdivisions of the entire region under study using voronoi diagrams, where each subdivision is the representative of a site of rain-gauge station. It is assumed that one rain-gauge station is sufficient enough to capture all rainfall patterns in that voronoi region. It may not be true every time, for example, if the geography of the region cut by the voronoi diagram is wide-ranging. Thus, it forms a big barrier in generalizing this study for regions that have wide landscapes, making it necessary to study the efficacy of the voronoi tessellation methods for various landscapes. We must note that the risk indices of a majority of locations will be near to zero, because the number of locations in the low risk zone will be more than the number of locations in the high-risk zone. In cognizance with the above fact, it can be observed that risk index plot for consecutive 7 days shows a risk value of zero for most of the municipalities under study. Any other index to segregate flood-prone areas can also be studied.

On a separate note, the proposed system reduces the data integration overhead through geo-visualization. In emergency situation, huge amount of spatial and temporal data is generated from diverse sources. Interpretation of these bulk data in an efficient manner is not possible through developing charts or tables or static maps as they have limited exploratory capabilities, and therefore, the decisions can be delayed. Our developed system emphasizes knowledge construction over knowledge storage through communication of geo-spatial information, combines it with other databases, and creates a dynamic environment which helps in decision-making faster. Geo-visualization allows for more interactive maps, including the ability to explore different layers of the map, to zoom in or out, and to change the visual appearance of the map, usually on a computer display.

As a part of future research, the proposed system needs some improvement. Real-time-based variations of risk levels in a given municipality can be incorporated to give the dashboard a more interactive and appealing look. Dynamic dashboard visualization can also be done using D3.js, Javascript, CSS and Bootstrap. Since Javascript forms the basis of any visualization technique, it would, therefore, be better to integrate Shiny R and Javascript for a better display. This proposed system of dashboard visualization of voronoi segregated maps can also be applied to study crop patterns in various regions. The methodology to calculate risk indices using the concepts of hazard and vulnerability can be extended to study the suitability of a crop in a given region. Of course, the metrics and mathematical/statistical methods to calculate these indices would not be the same, but any study in this field will be benefitted from the proposed method discussed in this paper. The risk indices also might help in prioritizing post-flood rescue and relief operation to the high-risk zones in terms of shelter location, establishing central depot, transportation logistics, and evacuation plan. Map visualization of these critical locations through various colour codes might help in demarcating them and preferential treatment is given to these areas.

Acknowledgements We would like to thank Prof. K. Koyamada and Prof. K. C. Kim, Editor-in-Chief of Journal of Visualization for being supportive towards the publication of this paper. Special thanks are due to the two anonymous reviewers whose comments greatly improved the presentation. Thanks are due to the Indian Statistical Institute Kolkata, India for funding this study. We are also thankful to the government personnel of Irrigation & Waterways Department of West Bengal and India Meteorological Department for their cooperation and help.

References

Adhikari P, Hong Y, Douglas KR, Kirschbaum DB, Gourley J, Adler R, Brakenridge GR (2010) A digitized global flood inventory (1998–2008): compilation and preliminary results. Natural Hazards 55(2):405–422

Aurenhammer F, Klein R (2000) Voronoi diagrams. Handb comput geom 5:201–290

Cançado V, Brasil L, Nascimento N, Guerra A (2008) Flood risk assessment in an urban area: measuring hazard and vulnerability. In: 11th International conference on urban drainage, Edinburgh, Scotland, UK

Cerutti V, Fuchs G, Andrienko G, Andrienko N, Ostermann F (2016) Identification of disaster-affected areas using exploratory visual analysis of georeferenced Tweets: application to a flood event. Association of geographic information laboratories in Europe pp 1–5

De Berg M, Van Kreveld M, Overmars M, Schwarzkopf OC (2008) Computational geometry, 3rd edn. Springer, Berlin Heidelberg

Dobrin A (2005) A review of properties and variations of Voronoi diagrams. Whitman College

EM-DAT: http://www.emdat.be/database. Accessed 26 Jan 2017

Fortune S (1987) A sweepline algorithm for Voronoi diagrams. Algorithmica 2(1–4):153–174

Gajdoš P, Radecký M (2009) Web visualization support for disaster management. In Internet Technology and Secured Transactions, 2009. ICITST 2009. International conference for IEEE, pp. 1–6

Gupta S, Knoblock C (2010) Building geospatial mashups to visualize information for crisis management. Proc, ISCRAM

Geographical information data set: www.gadm.org. Accessed 14 July 2016

Kemec S, Zlatanova S, Duzgun S (2009) Selecting 3D urban visualisation models for disaster management: a rule-based approach. In: Proceedings of TIEMS 2009 Annual Conference, pp. 9–11

Li C, Cheng X, Li N, Du X, Yu Q, Kan G (2016) A framework for flood risk analysis and benefit assessment of flood control measures in urban areas. Int J Env Res Pub Health 13(8):787

Li H, Zhang J, Sun J, Wang J (2017a) A visual analytics approach for flood risk analysis and decision-making in cultural heritage. J Vis Lang Comp 41:89–99

Li T, Xie N, Zeng C, Zhou W, Zheng L, Jiang Y, Chen SC (2017b) Data-driven techniques in disaster information management. ACM Comput. Surv (CSUR) 50(1):1

Liu Y (2009) Data-driven geospatial visual analytics for real-time urban flooding decision support. In: AGU fall meeting abstracts

Lu Y, Hu X, Wang F, Kumar S, Liu H, Maciejewski R (2015) Visualizing social media sentiment in disaster scenarios. In: Proceedings of the 24th International Conference on World Wide Web, ACM, pp. 1211–1215

Ma Y, Duan F, Liu H, Tian J, Duan G (2012) Research on the management and expression of disaster data. In: Proceedings of geoinformatics, 20th International Conference on IEEE, pp. 1–5

Malik A, Maciejewski R, Maule B, Ebert DS (2011) A visual analytics process for maritime resource allocation and risk assessment. IEEE Conference on visual analytics science and technology, pp. 221–230

Manna P (2016) Model for flood prediction/forecasting and flood risk assessment. M.Tech Thesis. Indian Statistical Institute, Kolkata, India

Parhi PK, Mishra SK, Singh R, Tripathi VK (2012) Floods in Mahanadi river basin Orissa (India): a critical review.India Water Week-Water Energy and Food Security

Rainfall Classification data set: www.imdkolkata.gov.in Accessed 02 Dec 2017

Review of Floods in West Bengal (with special reference to The Flood in September, 2000): Ref. PIL No. WP No. 113(W) of 2001—by The Committee of Experts constituted by the Hon'ble High Court, Calcutta

Smith K (1997) Environmental hazards: assessing risk and reducing disaster. Env Int 3(23):422

Smith K, Ward R (1998) Floods: physical processes and human impacts. John Wiley, New York

Supriya P, Krishnaveni M, Subbulakshmi M (2015) Regression analysis of annual maximum daily rainfall and stream flow for flood forecasting in Vellar River Basin. Aquatic Procedia 4th Edn, pp. 957–963

The Flood in September, 2000: Ref. PIL No. WP No. 113(W) of 2001—by The Committee of Experts constituted by the Hon'ble High Court, Calcutta

The Flood in September, 2000: Ref. PIL No. WP No. 113(W) of 2001—by The Committee of Experts constituted by the Hon'ble High Court, Calcutta

United Nations World Water Assessment Programme (UN-WWAP), 2006

WDR: World disaster report (2003) focus on ethics in aid. International federation of red cross and red crescent societies, Geneva, p. 240

WDR: World disaster report (2004) focus on community resilience. International federation of red cross and red crescent societies, Geneva, p. 240

Xie C, Ma G, Li Q, Xun J, Dong J (2016) Visual exploration of tsunami evacuation planning. J Vis 19(3):475–487