E3S Web of Conferences 7, 05001 (2016) DOI: 10.1051/e3sconf/20160705001
FLOODrisk 2016 - 3rd European Conference on Flood Risk Management

Application of an Earth-Observation-based building exposure mapping tool for flood damage assessment

Silvia De Angeli[1,2,a], Fabio Dell'Acqua[3] and Eva Trasforini[2]

[1] Institute for Advanced Study of Pavia, Palazzo del Broletto, Piazza della Vittoria, 15, I-27100 Pavia, Italy
[2] CIMA Research Foundation, Via Armando Magliotto, 2, I-17100 Savona, Italy
[3] Dept. ECBE, University of Pavia, Via Ferrata, 1, I-27100 Pavia, Italy

Abstract. Detection and characterization of territorial elements exposed to flood is a key component for flood risk analysis. Land-use description works well for small scales of representation but it becomes too coarse while increasing the scale. "Single-element" characterization is usually achieved through surveys, which become prohibitive as the amount of elements to be characterized increases. Mapping schemes represent a compromise between level of description and efforts for data collection. The basic idea is to determine the statistical distribution of building characteristics inside a homogeneous class starting from a sample area and to apply this distribution to the whole area, realizing a statistical extrapolation. An innovative approach was developed, merging the mapping scheme methodologies developed by the Global Earthquake Model [1] and Blanco–Vogt and Schanze [2], in which homogeneous classes are not development areas but building clusters. The approach was applied to the buildings in the Bisagno River floodplain, Genoa (Italy). Buildings were classified according to a building taxonomy. Once the percentage of basement presence was assigned to each class by surveying a limited subset of the exposed assets, a series of possible basement distributions was simulated to calculate the corresponding damage distributions for a real flood event. The total average damage obtained is very close to the refund claims, with a percentage error lower than 2%.

1 Introduction

A natural hazard disaster risk is the intersection between a natural hazard and a certain human dimension, characterized by a certain exposure, a vulnerability and a resilience able to contain damage. Therefore, the detection and characterization of exposed elements that compose the territorial system affected by a hazard is a key component to develop a risk scenario.

The detection of the different elements on the territory is mainly related to the so-called "exposure" component of the risk equation i.e. the quantification of "people, property, systems, or other elements present in hazard zones that are thereby subject to potential losses" [3]. The characterization of the detected elements is instead related to the "vulnerability" sphere. The vulnerability is a measure of "the degree of loss to a given element at risk (or set of elements) resulting from a given hazard at a given severity level" [4]. The way in which a given external stress turns into a loss and the degree of the loss itself depend strongly on the characteristics of the considered exposed element. One of the main issues in exposure and physical vulnerability modelling is related to built environment detection and characterization. Buildings are one of the most important types of elements at risk because they house the population and their behaviour under a hazardous event determines whether the people in the building might be injured or killed [5].

The most general level of knowledge of the territory can be obtained using a land use map. This type of information can be used as a starting point to evaluate how different territorial patterns respond to a certain hazard and to know a primary areal scale response of our study area. Nowadays remotely sensed data are properly processed to obtain a well detailed land use coverage of the whole globe. This type of approach works quite well for small scales of representation (global, inter-national, national) but becomes insufficient increasing the scale.

At local scale, exposed elements should be identified and characterized element by element. The "single element" characterization is an expensive work, usually achieved collecting different sources of data, among which Remote Sensing (RS) can only partially be applied. The main source of information is related to local knowledge and field survey acquisition. A new increasing source of information is also the one related to virtual surveys, done with expensive and advanced technologies such as remotely controlled cameras or drones. If this type of effort can be sustained at municipal or neighbourhood scale, at regional scale the amount of

[a] Corresponding author: silvia.deangeli@cimafoundation.org

elements to be characterized becomes prohibitive, both in terms of money and time.

The need of a compromise between the level of description of the territory and technical efforts to collect data suggests the introduction of an intermediate level of characterization of territorial elements suitable at regional scale. This compromise could be represented by the development and application of "mapping schemes". The basic idea behind them is to determine the statistical distribution of building characteristics inside a homogeneous area starting from a sample and apply this distribution to the whole area, realizing basically a process of statistical extrapolation.

One of the aims of this study is to apply Remote Sensing tools as main source of data for mapping schemes development, integrated with other information when necessary. Remote sensing is introduced as a powerful tool for exposure and vulnerability estimation, which can be synergistically integrated with data collected in-situ or through virtual surveys to obtain a description of the elements as complete as possible, in a way functional to properly apply vulnerability functions for damage estimation. The systematic application of the methodology, starting from high-resolution optical data as input, will be tested through a study case based on a past event. This choice will give us the opportunity to te.st the accuracy of the final results respect with both assets characterization and final damage evaluation.

2 State of the art on exposed elements characterization

The identification of the different land use (or land cover) classes inside the study area is the first and minimum level of knowledge in order to properly assess the vulnerability of a territory from different hazards. Genovese [6] underlines, for example, that "the flood damage depends on the land use type: in urban areas floods produce as a consequence much more damage than floods in a rural area." Similar considerations can be applied also to all the other natural hazards.

Moving from areal to building scale characterization, two main factors concur to determine the potential losses and degree of damage of buildings that are exposed to a certain type of hazardous event: the type of negative effects that the event might have on the building which is exposed to it and the characteristics of the building that define the degree of damage due to the hazard exposure [5].

Considering first of all the second of the two factors, we can argue that the propensity to suffer damage due to an external stress varies from building to building according to a series of structural and non-structural characteristics. It is necessary to introduce a "single-element" detailed description of exposure in order to properly model damage changes inside a certain land use homogeneous area.

The first factor indeed addresses to the nature of the external stress that acts on the buildings. Different types of hazards are able to generate different impacts, interacting in a different way with the structure of the building. As a consequence, some building attributes become more or less relevant in function of the nature of the threat. We can assume that there is a strong hazard dependency of attributes useful for vulnerability assessment of buildings. Apart from some specific non-structural attributes, such as occupancy, which are important for all the hazards, there are a series of characteristics that becomes essential in case of earthquake, for example, and completely useless for a flood or a fire. An interesting summary of important building characteristics for different hazards is reported by Van Westen *et al.* [5].

2.1 Building taxonomies

Buildings characteristics are the main ingredient to develop a building taxonomy. Taxonomy is generically the practice of classifying contents. A building taxonomy is a classification of buildings according to a series of specific characteristics. Building taxonomies have been developed for various purposes. In the field of natural hazard risk assessment, building taxonomies have been built mainly on the field of earthquake engineering. The criterion for classification, when building inventory is used as input for a risk assessment chain, is suggested by Muthukumar [7]: buildings need to be classified into specific sets that represent adequately the average characteristics and behaviours of all the buildings grouped in those sets, i.e. each defined class of buildings should exhibit substantially different damage behaviours and loss characteristics.

In the field of the seismic risk, a lot of different taxonomies have been developed and applied in order to properly classify buildings for seismic damage assessment. One of the pillars of seismic-oriented building taxonomies is the World Housing Encyclopedia (WHE) [8], an encyclopaedia of housing construction in seismically active areas of the world, with the purpose to develop a comprehensive global categorization of characteristic housing construction types across the world. As it is reported on the website of the project, the WHE Report Database contains 130 reports on housing construction types in 43 seismically active countries. Each housing report is a detailed description of a housing type in a particular country, prepared from a number of standard close-ended questions and some narrative ones.

Another classification has been developed by the U.S. Geological Survey inside the Prompt Assessment of Global Earthquakes for Response (PAGER) program. PAGER database [9] is a global database of building inventories using taxonomy of global building types for use in near-real-time post-earthquake loss estimation and pre-earthquake risk analysis. It draws on and harmonizes numerous sources: UN statistics, UN Habitat's demographic and health survey (DHS) database, national housing censuses and the World Housing Encyclopedia. A series of structure type categories are identified for global building inventory development, such as Wood (W), Steel (S), Reinforced Concrete (RC), Reinforced Masonry (RM), Mobile Homes (MH), Mud Walls (M), Adobe Block Walls (A) etc.

To conclude this brief overview on seismic building taxonomies, we introduce the GEM Taxonomy, which builds on the knowledge base from other taxonomies, including the World Housing Encyclopedia, PAGER, and HAZUS [10]. The purpose of the GEM Building Taxonomy is to describe and classify buildings in a uniform manner in order to assess their seismic risk. The taxonomy has been developed applying the following criteria: relevance to seismic performance of different construction types; comprehensiveness yet simplicity; collapsibility; adherence to principles that are familiar to the range of users; extensibility to non-buildings and other hazards. The taxonomy is organized as a series of expandable tables, which contain information pertaining 13 different building attributes [11].

Moving from seismic risks, there are in literature very few examples of taxonomies developed for other hazards. In the specific case of floods, one of the few attempts to develop a building taxonomy for settlements has been done by Blanco Vogt and Schanze [2]. They propose a building taxonomy approach as a step of a methodology to assess physical flood susceptibility of buildings on a large scale. In particular, they identify seven parameters - Height, Size, Elongatedness (length/width ratio), Roof form, Roof slope (Roof pitch), Index inversely compactness, Adjacency - of the taxonomy and discretize the value they can assume into classes called categories. These parameters are directly extracted or derived from remotely sensed data, specifically VHR optical images and digital surface models. The taxonomy is applied by Blanco Vogt and Schanze to cluster buildings in a proper way. Buildings characterized by the same taxonomy string, obtained as a textual sum of the codes assumed by each parameter, are clustered together. Representative buildings are then selected from each building type as samples for the subsequent assessment of physical susceptibility of buildings to flood.

2.2 Remote Sensing as a tool for exposure and vulnerability

Remote sensing data and methods contribute to natural hazard risk assessment providing indicators for the spatial distribution of natural hazards, as well as identifying physical and demographic aspects of vulnerability. Considering the different phases of the Disaster Risk Management cycle, remote sensing can be a valuable tool not only for pre and post disaster assessments but also for recovery and reconstruction processes and their evolution over time.

Focusing on physical aspects of vulnerability, nowadays high-resolution remote-sensing data and algorithms are available for detection, extraction, and analysis of building features, which can be applied systematically for regional building exposure assessment avoiding time-consuming field surveys. Although generally less accurate than in-situ surveys and less rich in details at individual buildings scale, remote sensing can provide cheaper information in terms of acquisition costs and allows for the capture of large geographical extents that in-situ inspections cannot compete with. [2,12,13].

Using high-resolution satellite data, the complex and heterogeneous urban landscape can be classified automatically, obtaining land cover information. The spectral and structural characteristics of the data and the classification result can be used to extract building characteristics, such as built-up density, building heights, predominant usage of building, building age etc. [12].

Focusing on demographic aspects of vulnerability, population distribution and related characteristics are main attributes to be estimated. The structural building characteristics– building heights, building density and land use – can be used to perform an indirect assessment of population distribution with accuracies of around 90% [12]. Using the assessment of the location of commercial and residential areas, the dynamic spatial population patterns as a function of the time of day can be computed. [12]. The use of physical proxies retrieved from medium to high resolution (HR1 to MR1) optical and very high resolution (VHR1-2) LIDAR data is also proposed by different authors – e.g. Ebert *et al.* [14], Taubenböck [15], and Zeng et al. [16] - for the approximation of socioeconomic vulnerability indicators. [13].

2.3 Mapping schemes

The building scale characterization is an expensive work, achieved collecting different sources of data, not all directly available in an automated way using RS data. All the structural parameters required by the different taxonomies developed in literature, should be attributed to each single building in the study area, in order to model the damage at the scale of the single asset.

Mapping schemes are tools for the statistical inference of structural parameters using input data from different sources (regional defaults, expert opinion, survey data) from a certain area. A mapping scheme is a useful tool to harness the best-available data for inventory generation at a regional scale [17, 18]. Data are collected with the detail of the single element only in a sample area representative of a certain population, reducing drastically the time and the resources required. Mapping schemes can be seen as statistical summaries of building attributes to homogenous zones, or areas with sufficiently similar structure type distribution [1]. Homogenous zones are defined as "areas with sufficiently similar structure type distribution to be characterized by a mapping scheme. These could be geographic regions of relatively uniform use such as a central business district, a manufacturing district, or a residential neighbourhood. Typically, these will correspond with use and occupancy classifications at the block scale but also consider known construction patterns. For example, an area of single family residential construction may be subdivided by era of development" [1].

Among the different approaches described in literature, two of the most significant are those proposed by HAZUS and GEM.

HAZUS is the Federal Emergency Management Agency's (FEMA's) methodology for estimating potential losses from disasters. HAZUS methodology uses mapping schemes "to define properties of attribute information for aggregated data, such as setting the distribution of building types for a specified geographical division" [19]. Building inventory data in HAZUS is stored in two types of tables:

1. occupancy exposure tables: aggregate data on square footage, building count, building

exposure value and content exposure value stored by occupancy at the census tract level

2. occupancy mapping scheme tables: distributions indicating typical construction types by occupancy.

Occupancy mapping tables indicate, by occupancy, the percentage distribution of square footage among various structural or model building types. Examples of T building types supported by HAZUS are: Wood light frame (W1), Steel moment frame, low-, mid- and high-rise (S1L, S1M, S1H), Concrete moment frame, low-, mid- and high-rise (C1L, C1M, C1H), Mobile homes (MH).

Occupancy mapping relationships exist at two levels:

3. a general mapping scheme, which indicates the single, assumed regional distribution of square footage by occupancy across the five basic construction classes or basic building types (Wood, Concrete, Steel, Masonry, and Manufactured Housing)
4. specific occupancy mapping schemes or building type distributions, which indicate, for a given occupancy and material type, the distribution across the detailed model building types, including variations reflecting the various design levels and building quality classes (e.g., "High-Code" = High-seismic design, Code quality). These detailed mapping scheme distributions drive which vulnerability functions will be used to estimate damage and loss for each occupancy class [20].

The Global earthquake Model (GEM) Foundation is a global collaborative developer of projects with the aim to provide organisations and people with tools and resources for transparent assessment of earthquake risk anywhere in the world. Inside the GEM approach, a mapping scheme is simply a statistical summary of the percentage of sampled buildings in each category defined by the GEM Basic Building Taxonomy [11]. Mapping schemes are developed as part of the Spatial Inventory Data Developer (SIDD) tool that is in turn part of the GEM IDCT (Inventory Data Capture Tools) software for developing building exposure data [21].

The methodology ingests 3 key data sets for developing an exposure data set:

5. the footprint database, which will ultimately provide the number and square meter area of buildings;
6. the zones delineating land use and development patterns;
7. the building samples consistent with the GEM taxonomy.

For each type zone defined in the homogenous zone dataset (i.e.: Res-pre 1900, Res-post 1900, Com, Ind, and special), SIDD creates a preliminary mapping scheme. This mapping scheme can be successively adjusted and reviewed by the user. SIDD allows users to define and adjust the following eight different attributes for each zone: Material, Lateral Load Resisting System, Roof, Floor, Height, Date of Construction, Structural Irregularity, Occupancy.

Finally, when the mapping schemes appear satisfactory, they are applied to the exposure. The resulting file allocates the footprints aggregated for each individual homogenous zone into each GEM taxonomy class in the mapping scheme [1].

3 Methodological framework

The final aim of this individual study is to develop and apply to a specific study case a methodology to characterize exposure and vulnerability at regional scale applying mapping schemes.

The innovative approach proposed in this study builds on the "traditional approach" for mapping schemes inherited by HAZUS [20] and GEM [1, 22], merging it with the building taxonomy for settlements proposed as a step of a flood susceptibility assessment by Blanco–Vogt and Schanze [2]. The "traditional" mapping scheme methodology –identified as "approach A"- as proposed by GEM [1, 22], is illustrated in Figure 1.

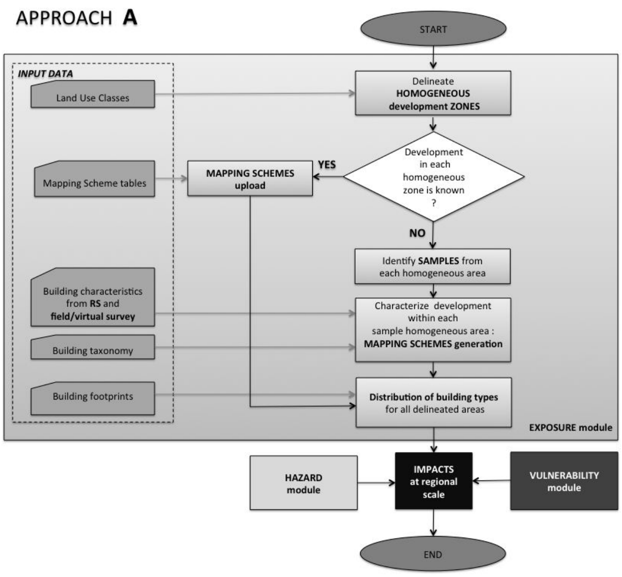

Figure 1 First approach proposed for mapping scheme development, based on GEM approach [1,22]

The innovation of the innovative approach illustrated in Figure 2, with respect to the procedure developed by GEM, is represented by the fact that homogeneous classes are not homogeneous development areas but homogeneous building clusters. In other words, the sample set is not an areal element inside which investigate point elements, but a point element itself. Building clustering is done using only RS data, as well as in the case of "traditional mapping schemes". The idea behind this type of approach is that the only occupancy information obtained from a land use classification is sometimes not enough accurate to be used to properly cluster buildings, particularly in areas where commercial, residential and small industrial occupancies are merged together and sometimes located in different storeys of the same buildings. On the opposite, the identification of some remotely sensed structural parameters together with an information of occupancy can properly group together buildings having - with a higher probability- common structural features, that can useful for example for flood damage assessment.

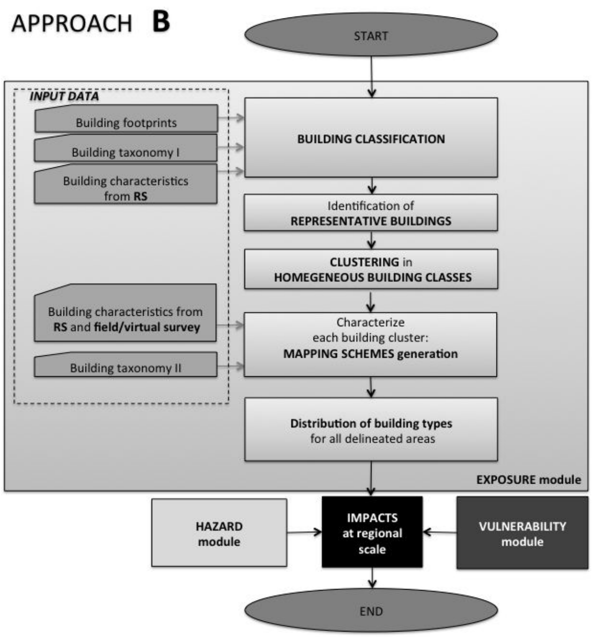

Figure 2 Innovative approach proposed for mapping scheme development, mainly based on Blanco–Vogt and Schanze [2].

Four main steps compose the methodology:

1) Building classification. Mapping schemes infer population parameters from a sample. The population is here represented by a certain homogeneous building cluster, In order to properly group buildings, the first step is to classify them according to a given taxonomy. Starting from a certain building footprints dataset obtained from RS, a series of attributes are attached to each building footprint. These attributes are mainly obtained directly from RS or calculated using RS as a proxy, in order to have an automatized quick procedure of classification. Parameters are discretised into classes called categories. Each category has a numeric code associated to it. The final taxonomy string attached to each building is obtained as a textual sum of the codes assumed by each parameter.

2) Identification of representative buildings and clustering in homogeneous building classes. Theoretically, buildings having the same taxonomic code can be grouped together, forming homogeneous classes. Let's consider now the simple case in which we have n different parameters having an equal k number of classes. The number of possible building codes is n^k. Having, as a realistic example, 7 parameters each with 4 classes, the amount of possible permutations is equal to 2.401 i.e. the homogeneous classes to investigate can be up to 2.401. To reduce the efforts, representative buildings are selected. This particularly means to narrow down the population of buildings with the same taxonomic code to a number of representative buildings, suitable for assessing the buildings attributes useful for vulnerability curves application. A fuzzy clustering method is applied for merging the codes. Analysing the histogram of the codes, those having higher frequency are selected as representative buildings, while the other codes are considered as non-representative. The selection is done fixing a certain threshold. Non representative building codes are clustered to representative ones using a membership function.

3) Characterize each building cluster. This step corresponds with the development of the mapping scheme. Inside each homogeneous building class obtained from clustering, a certain number of buildings is investigated in detail using data sources such as field surveys, virtual surveys, institutional and census data and others. The percentage distribution of building characterising according with a certain building taxonomy inside each homogeneous sample class is collected in matrices.

4) Distribute building types for all building population. The statistical distribution of parameters is applied to the building footprints contained inside each individual homogeneous class. As a default approach we can assume a random distribution of parameters according to the mapping scheme statistical distribution.

The methodology described above is the one that will be applied to a study case discussed in Chapter 5

3.1 Clustering using membership functions

Let's have B as a generic taxonomic code of length n. Let's have $j = 1, \dots n$, $j \in \mathbb{N}$ as the considered parameter inside the taxonomic code i.e. the considered position inside the code, and $i_j = 1, \dots m_j$, $i_j \in \mathbb{N}$ as the possible values that each parameter j can assume, named categories. i_{Rj} is the representative value for the parameter j inside the representative taxonomic building code B_R and i_{NRj} are the non-representative values for the parameter j inside the generic taxonomic building code B. The membership function $\mu(i_{NRj})$, that measures how much i_{NRj} belongs to the class of i_{Rj},, has the following behaviour:

$$\mu(i_{NRj}) = 1 - \frac{|i_{NRj} - i_{Rj}|}{\max(|m_j - i_{Rj}|, |i_{Rj} - 1|)} \qquad \mu(i_{NRj}) \epsilon (0,1) \qquad (1)$$

When $i_{NRj} = i_{Rj}$ we have that $\mu(i_{NRj}) = 1$, while when $i_{NRj} = \max(|m_j - i_{Rj}|, |i_{Rj} - 1|)$- i.e. i_{NRj} is the value as far as possible respect to the representative one – we have $\mu(i_{NRj}) = 0$.. Intermediate values follow a linear behaviour. The level of matching of the complete building code B (i.e. those considering all the order parameters j) with the representative one B_R, considering all the positions of the taxonomic code, is obtained as:

$$(B|B_R) = \sum_{j=1}^{n} \mu(i_{NRj}) * w_j \qquad (2)$$

Where w_j is the weight (i.e. the relative importance) of the parameter j.

This function is calculated for each non-representative building code respect with all the non-representative ones. Each non-representative taxonomic code is associated to the representative one for which a higher level of matching has been computed.

The membership function $\mu(i_{NRj})$, as defined in equation (1), works well in case of quantitative continuous parameters, such as the building height or the area. Quantitative continuous parameters are ordered in increasing or decreasing order, so that it is reasonable to assume that, once a parameter is discretized in categories, also categories maintain the same order. In other terms, for these parameters it is reasonable to assume that category *n* is more similar to category *n+1* (or *n-1*) rather than *n+2* (or *n-2*)), where $n \in \mathbb{N}$ is a generic value of the category for a given parameter.
In case of qualitative parameters, such as the building use, this type of approach has no meaning. For this reason, an alternative membership function has been defined:

$$\iota(i_{NRj}) = \begin{cases} 1 \; if \; i_{NRj} = i_{Rj} \\ 0 \; if \; i_{NRj} \neq i_{Rj} \end{cases} \qquad \mu(i_{NRj}) \, \epsilon \, (0,1 \qquad (3)$$

When $i_{NRj} = i_{Rj}$ - i.e. the parameter for the considered building has the same value of the parameter for the representative one - we have that $\mu(i_{NRj}) = 1$, otherwise we have $\mu(i_{NRj}) = 0$. It is possible to mix both quantitative and qualitative parameters inside a single taxonomic code. The level of matching of the building code B with the representative one, considering all the positions of the taxonomic code, is obtained applying equation (2), where $\mu(i_{NRj})$ is calculated applying either equation (1) or equation (3) according with the type of the considered parameter *j*, i.e. if it is qualitative or quantitative.

4 Study case

In order to test the feasibility of this new hybrid approach, a study case has been carried out. The study case is represented by a real flood event happened in Genoa (Italy), in November 2014. An important building characteristic in case of flood is the presence of basements. Unfortunately, it is not possible to assess the basement presence from remote sensing. Also “low-cost” virtual surveys, using Google Street View, are not performing well, due to the presence of cars parked in front of the buildings (or other visual obstacles) or simply to the unavailability of a virtual view of all the facades of the building.

The basic idea is to compute the taxonomic codes for all the buildings inside the study area, cluster them according to the membership function computation and successively to attribute the percentage of buildings with basement inside each homogeneous class, starting from a properly defined sample area. Each of these steps will be discussed in detail in the following sub-chapters. From the operational point of view this means that the mapping of the basement presence, that is an important element in flood damage assessment, can be virtually obtained only surveying a subset of the exposed assets, if a robust statistical mapping tool is applied. Once the basement presence percentage is assigned to each class, a series of possible basement distributions are simulated and used to calculate the corresponding damage distributions. A high number of simulations ensures an average damage as close as possible to the real damage for each building.

An overview of the applied methodology is proposed inFigure 3.. Inside the so-called “EXPOSURE module”, the mapping scheme methodology is applied in order to properly assign the basement presence to each building.

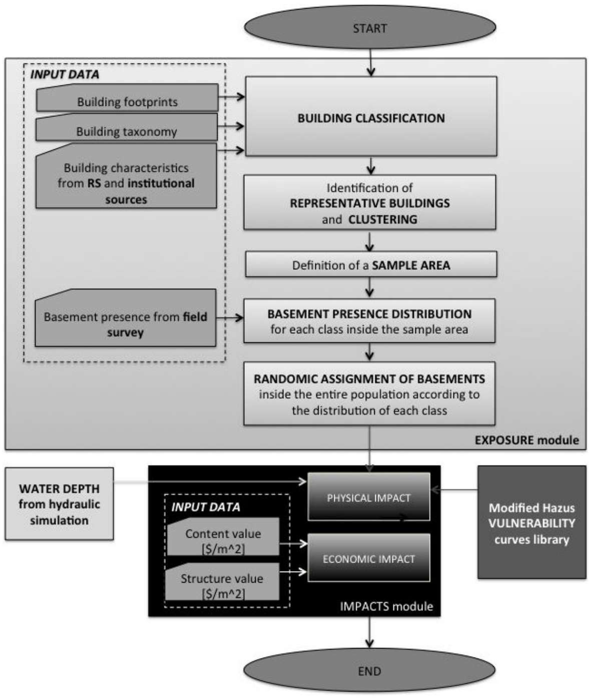

Figure 3 Applied methodology for Genoa flood (2014) case study, using the methodological approach proposed for mapping scheme development based on Blanco–Vogt and Schanze [2]

The exposure layer, properly characterized is ingested, together with the flood scenario (the “hazard module”) and a proper vulnerability curves library (the “vulnerability module”) into the “IMPACTS module”. A first physical percentage damage for both structures and contents is calculated directly from the curves. In a second step, considering a certain economic value per square meter of structures and content, an economic damage is indicatively assessed.

The choice has fallen on this specific event for a series of operational reasons. First of all, the real basement presence has been assessed through a field survey for all the buildings inside the affected area. This information can be used to evaluate if a good basement distribution can be obtained applying the statistical mapping scheme. Secondly, a distributed damage claim mapping is available on the study area. In this way it is possible to roughly evaluate if the order of magnitude of the damage assessed at the end of the mapping scheme chain is coherent with the real damage.

4.1 Genoa 2014 Flood Scenario

A severe thunderstorm affected central Liguria on 9th October 2014. The event was composed of two distinct phases of intense precipitation: the first, between 06 and 10 UTC, with a total accumulated rainfall of between 50 mm and 130 mm; the second, between 17 and 24 UTC, which reached a total accumulated rainfall of between 150 mm and 260 mm, with an hourly peak of approximately 100-130 mm between 20 and 21 UTC. Over the entire duration of the event the cumulative peak precipitation was close to 400 mm. Following this rainfall, the flooding of Bisagno creek (around 21.10 UTC with the flood peak at around 21:45 UTC), Fereggiano brook (22UTC) and Sturla creek (shortly after 22UTC) were recorded. Other smaller brooks (Veilino, Noce) burst their banks in neighbouring areas (Staglieno, San Fruttuoso), in the same time period. [23].

The flood scenario, in terms of maximum water depth, has been computed through a hydraulic model. The applied model [23], with the necessary input information, calculates the water depth at any point in the domain of interest, returning a matrix (discretized according to the input DEM) with the maximum water depth calculated in the simulation. The adopted numerical modelling is a two-dimensional one. It solves a set of equations derived from the Navier-Stokes equations. The spatial domain is modelled by the "storage cell" approach that approximates the spatial domain through a series of discrete cells. The time domain is modelled adopting a "predictor-corrector" explicit scheme in which a first attempt solution (predictor) is subsequently corrected with a second iteration (corrector), that uses the first attempt solution as input. The available data for Genoa are: Bisagno river hydrograph, measured near the subterranean part of the river; DEM of the area from LIDAR with a horizontal resolution of 1 x 1[m], a detailed relief of the riverbed and banks provided by the municipality of Genoa.

The case study hydraulic simulation gives the maximum water-depths illustrated in Figure 4.

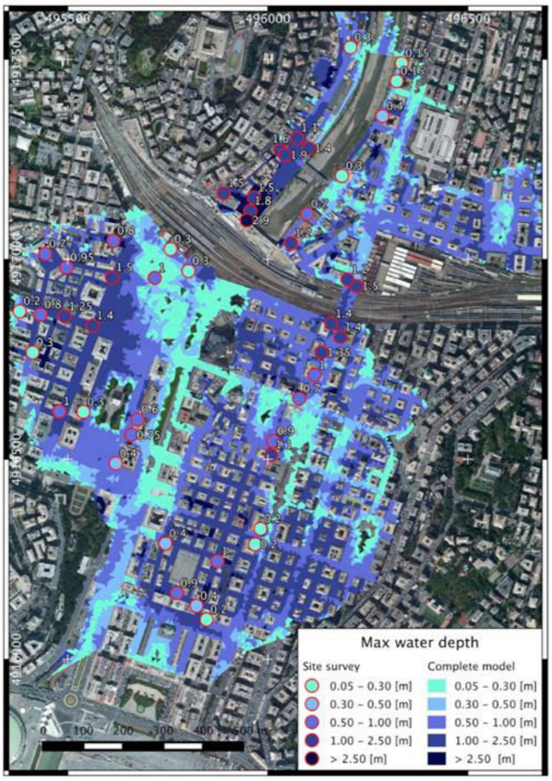

Figure 4 Maximum water depth from complete model compared with site survey for Genoa Flood 2014 [23].

The comparison with in-situ measures shows a good matching of the model, despite an area which was modelled as flooded by the model and which in reality was not touched by water (the easternmost area within the flooded zones, along Via Casaregis), maybe due to errors in the DTM or errors in the initial conditions.

The water depth scenario obtained from the model is used as input hazard in order to calculate, through damage curves, the percentage of damage experienced by each building.

4.2 Classification and clustering of buildings

The methodological procedure reported in the "EXPOSURE module" of Figure 2 is here applied to the buildings inside the study area. The building footprints inside the study area are properly characterized through the following 3 parameters: Filling parameter, Use, Height.

Two of them are geometrical parameters – the filling parameter and the height – while the use is a descriptive one. The filling parameter is calculated through an algorithm directly from the available building footprints. It is obtained as a ratio between the area of the minimum convex polygon containing the footprint and the area of the same footprint.

The height has been detected from RS data trough a simple algorithm, such as those described in chapter 2.2. Specifically, an algorithm implemented inside the SENSUM project has been applied. Once the shadows have been extracted from a very high-resolution optical image through classification, the calculation of the height associated with each shadow is done using the shadow length (computed with the shadow length function) and the sun position (determined from the raster acquisition date) [24]. The algorithm automatically associates footprints with respective shadows.

The use has been determined referring to cadastral data available for the study area. In addition to standard "residential" and "commercial" occupancy classes, a mixed category has been introduced in order to properly classify buildings having both classes together. This is the typical case of a lot of buildings in the city centre of Genoa, where commercial activities are located at the first floor, while the remaining floors are residential. A class named "other" has been added to incorporate all the remaining categories, which are marginal in the study area, such as industrial buildings, museums and monuments, stations, schools and educational buildings, government buildings.

The geometric continuous parameters have been discretized into intervals, while for the descriptive one this passage has been unnecessary. As final result, for each parameter different categories have been defined and a specific code has been associated to each of them. The final taxonomic code for each building is obtained as textual ordered sum of the values assumed by each of the three parameters. The ranges of categories for the parameters of this building taxonomy are reported in Table 1.

For instance, the code "124" describes from left to right: a regular and compact structure (I position: filling parameter); with mixed occupancy (II position: use); with square form in the space (3rd digit: elongatedness); higher than 25 meters (III position: height). Applying this classification to the buildings inside the study area, 30 different taxonomic codes have been identified.

Position	Parameter	Code	Description
I	Filling Parameter	1	<1.1: absence of internal/external cavities, regular structure
		2	>= 1.1: presence of internal/external cavities, irregular structure
II	Use	1	Commercial
		2	Commercial & Residential (mixed)
		3	Residential
		4	Other
III	Height	1	< 15 m
		2	15-20 m
		3	20-25 m
		4	> 25 m

Table 1 Range of categories for the parameters of the building taxonomy.

Representative buildings are here selected as the most frequent inside the study area. Using histograms, the taxonomic codes with a frequency higher than a threshold of 20 buildings are separated and named as representative buildings. The other buildings with a lower frequency are called non-representative buildings. As shown in Figure 6, six taxonomic codes have been initially identified as representative. Applying the procedure described in detail in Chapter 4, a clustering algorithm is performed on the exposed buildings. A membership function is calculated for each non-representative building code respect with all the non-representative ones. Each non-representative taxonomic code is associated to the representative one for which a higher level of matching has been computed.

After the first application of the clustering algorithm it can happen that a certain taxonomic code is so "different" from all the representative ones, that it is attributed to a certain class with a very low value of the membership function. These taxonomic codes can be called "atypical", according to the parameters defined for the code. In order to avoid to cluster them in a class that is not representative enough, they are added to the list of the representative codes and the clustering algorithm is performed again. From the operational point of view, a threshold on the value of the membership function is set, so that the clustering algorithm is performed while all the taxonomic codes are attributed to the representative ones with a membership function higher than the threshold. Using this criterion, a final number of seven clusters was obtained, adding then atypical code '242' to the initial list of six representative codes reported in the histogram of Figure 5.

As it is possible to observe from the histogram, the atypical code which has been added to the list of the clusters is not statistically very representative inside the study area, since only 2 buildings are classified with this code. Nevertheless, once the code has been added to the list and the algorithm has been performed again, a significant number of buildings have been re-assigned to this new class with a higher value of the membership function, which is the most important criterion in the process of creation of the clusters.

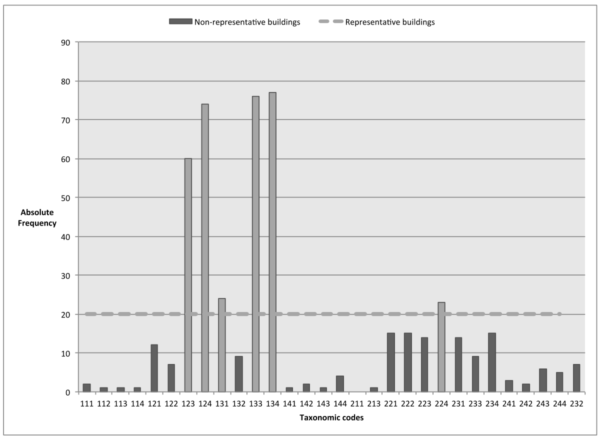

Figure 5 Selection of representative building taxonomies for the Genoa 2014 study case, as most frequent ones i.e. those higher than a threshold of 20 buildings.

The distribution of the final seven taxonomic classes inside the study area is illustrated in Figure 6.

.

Figure 6 Distribution of the final seven taxonomic classes, obtained applying the clustering algorithm.

4.3 Basement presence distribution

Once the buildings have been clustered, it is necessary to properly distribute the basement presence inside the study area. In order to do that it is necessary to define a sample area and determine the basement presence distribution for each class inside the sample area. Only successively it is possible to randomly assign the basements inside the entire population, according to the distribution of each class obtained from the sample.

The selection of the sample area is a crucial step in the whole procedure. The sample area should be as representative as possible of the whole population, in order to have an extrapolation that will be as close as possible to the real distribution. In order to have a representative sample area it is necessary to have in it a reasonable number of buildings belonging to each of the seven classes. The statistical representativeness is not the only important aspect to consider. If the basement presence has to be assessed through field survey, the sample area has to be not fragmented, and preferably concentrated in a certain part of the study area e.g. along one of the main roads. In this way it is possible to optimize the time required for the survey. For the study case, the sample area illustrated in yellow in Figure 7 has been identified. A total of 101 over 482 buildings, corresponding to the 21% percent of the total buildings have been selected.

Figure 7 Buildings selected as sample area (identified in yellow).

The composition of the sample area, both in terms of number of buildings and percentage of buildings with and without basement inside each of the seven classes, is reported in Table 2.

	Class						
	242	**224**	**134**	**133**	**131**	**124**	**123**
Number of buildings surveyed	9	7	9	21	15	20	20
%without basement	67	57	33	19	67	20	40
%with basement	33	43	67	81	33	80	60

Table 2 Composition of the sample area, in terms of number of buildings and percentage of buildings with and without basement inside each of the seven classes.

The percentages derived from the sample are applied to the entire population of buildings. As an example, considering the buildings belonging to class '124', the basement will be randomly assigned to the 80% of them. A series of possible basement distributions, according to the percentages of Table 2, are simulated and used to calculate the corresponding damage distributions. A high number of simulations ensures an average damage as close as possible to the real damage for each building. For this study case, 3000 simulations ensured an average damage with a relative error lower than 2%, respect to the one obtained considering the real distribution of basements.

4.4 Damage assessment

Once the different simulations for basement distribution became available, it was possible to compute the damage of each building associated to each simulation. Building damage is evaluated through the application of flood damage functions, which relate the water depth to the percentage of damage experienced by a specific type of building. The library of curves applied is the flood damage curves library developed inside the RASOR project [25].

RASOR library has initially inherited the HAZUZ curves for flood, which are damage functions in the form of depth-damage curves, relating depth of flooding (in feet), as measured from the top of the first finished floor, to damage expressed as a per cent of replacement cost. In the RASOR platform the functions have been transformed in order to express water depth in I.S. units. Depth-damage functions in the HAZUS library are provided separately for structure (load-bearing system, architectural, mechanical and electrical components, and building finishes) and for content. Different curves are available for different occupancy classes (residential, commercial, industrial, educational and so on). By trying the adoption of such a library to RASOR European test cases, it has been immediately clear that the classification used in HAZUS is not suitable to address the characteristics of European cities; at least it is true for the historical centres where narrow spaces impose the coexistence of different uses in the same building. For this reason, a modified HAZUS library has been developed, building ad hoc vulnerability functions by combining the curves for specific uses that can be found in the original library. The initial HAZUS curves, corresponding to different uses for the building and number of stories, have been incremented by a series of mixed-use curves, creating the final RASOR library named "HAZUS modified".

Considering the original HAZUS library, only "Single Family Dwelling" occupancy type curves are available both with and without basement. As already mentioned, basement presence can significantly influence physical damage in case of flood. For this reason a complete library of curves for buildings with basement have been built starting from the HAZUS modified library. As final result, a complete library was available, having different curves (for structure and content) according with the occupancy classes (single and mixed) and the basement presence.

Once the percentage damage is attributed to each building, the correspondent economic damage *ed*[€] is calculated through the following expression:

$$ed = pd * A + rc * (n + b) \quad (4)$$

where:

- pd [%] is the percentage of damage
- A $[m^2]$ is the area of the building footprint

- rc $[\frac{euros}{m^2}]$ is the replacement cost per square meter
- $n = \max(nf, 3)$, with nf the number of floors
- $b = \begin{cases} 1 \; if\ the\ building\ has\ the\ basement \\ 0 \; if\ the\ building\ has\ not\ basement \end{cases}$

This function can be applied for both economic damage to structure and content. In the first case the percentage of damage is the one calculated with the curve for the structure and the replacement cost per square meter is only referred to the structure. In the second case, both percentage damage and replacement cost are referred only to the content. The final total damage for each building is obtained summing the two contributions. The replacement cost is one of the main sources of uncertainty in the estimation of the final damage. Theoretically this parameter strongly depends on the building use, particularly for the content, which varies consistently passing, for example, from a school to a museum. Practically, it is very difficult to have such a detailed estimation of the replacement cost except for insurance companies databases, which are usually not open for public reference. For this study, fixed replacement costs per unit area are assigned, for structure damage and content damage, respectively equal to 500 €/m2 and 400 €/m2.

Damage claim for the whole study area amounts to around 92 millions of euros. This information cannot be used as a real damage with which assessing the goodness of the methodology. In fact, a discrepancy between damage claims and real damage is expected; firstly, claimed damage are estimated by the affected people and secondly many minor damages are usually not claimed. Nevertheless, this information can be used as a reference order of magnitude to immediately identify if there are huge mistakes in the methodology.

Applying the library of curves inherited by RASOR, a total damage was calculated for each of the 3000 simulations, obtaining an average total damage over the 3000 simulations equal to 104.653.946 euros, which has the same order of magnitude of the damage claim. The average total damage is obtained with a standard deviation of around 986.000 €, which ensures a 95% confidence interval of (104.689.230 €, 104.618.663 €). A spatial damage distribution is reported in Figure 8.

Figure 8 Total economic damage (structure plus content) in euros, calculated per each building applying the mapping schemes methodology.

4.5 Results analysis

In order to evaluate the goodness of the obtained results, it is necessary to compare them with the damage calculated considering the real basement distribution, which will be called "real damage" hereafter. Basement presence, for the 482 buildings inside the study area, has been assessed trough a 3 days' field survey by one person. This type of information gives immediately the idea of the effort required to properly assess the building presence. Using the real basement distribution, a total damage of 106.285.465 € has been calculated. This result is very close to the one obtained applying mapping schemes. The committed percentage error of the methodology is equal to 1,6%. In order to deeply test the robustness of the methodology, the result is also compared with other four basement distribution scenarios:

1. none of the buildings inside the study has a basement
2. all the buildings inside the study area have the basement
3. 50% of the buildings inside the study area has the basement
4. 80% of the buildings inside the study area has the basement.

Each of these options can potentially be chosen by a person who does not know anything about the real distribution of basements in the area. Particularly, the first option gives immediately the idea of the influence of the basement in the economic damage assessment. Ignoring completely the presence of cellars, the committed percentage error on the total damage is on the order of 20%. On the opposite, considering all the buildings with basement, the percentage error is lower but still significant, reaching around 10%. The difference between these two errors is explained considering the real percentage of buildings with basement inside the area, which is around 60%. That is the reason for which the error committed ignoring basements is higher than those committed attributing basements to all buildings. The same reasoning explains also why, betting that 50% of the buildings has the basement, the error decrease till the 3,3% while remaining higher than those obtained with the mapping schemes approach. In fact, the mapping scheme approach allows to properly select a sample area which is as much as possible representative of the whole population. Inside the sample area of Figure 7 the percentage of buildings with basement was around 61%, which is very close to the real percentage of 60%.

Real damage	**Mapping Schemes**	**All buildings without basement**
106.285.465 €	104.653.946 €	88.258.683 €
All buildings with basement	**50% of buildings with basement**	**80% of buildings with basement**
117.657.551 €	102.869.908 €	111.562.199 €

Table 3 Total economic damage obtained considering different basement distributions inside the study area.

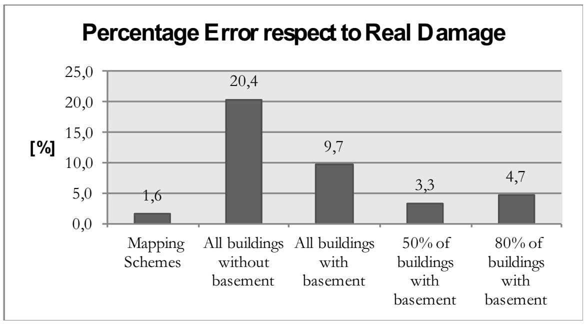

Figure 9 Percentage error committed calculating the damage considering different basement distributions inside the study area, respect to the damage calculated with the real basement distribution.

As a final countercheck, let's suppose that a mapping scheme has not been performed and different classes have not been available to properly select a representative sample area. Another sample area of 101 buildings along Corso Torino, which is one of the main roads inside the study area, has been selected. Inside this area, 80% of the buildings has a basement. Applying this percentage to the whole population, the percentage error becomes of the order of 5%. In conclusion, it is possible to affirm that the approach works quite well, considering the total damage in the whole study area.

As a further analysis, it could be interesting to evaluate the percentage error committed at the scale of the single building. The results of this analysis are illustrated in Figure 9. As expected, the percentage error for each single asset is higher than the percentage error at areal scale. The reason is because the damage is overestimated for some buildings and underestimated for others, so that the total damage is counterbalanced by the two opposite contributions. Anyway, the error at single scale is higher than 30% only for few buildings, while for the majority of the buildings (around 66%) is lower than 20%.

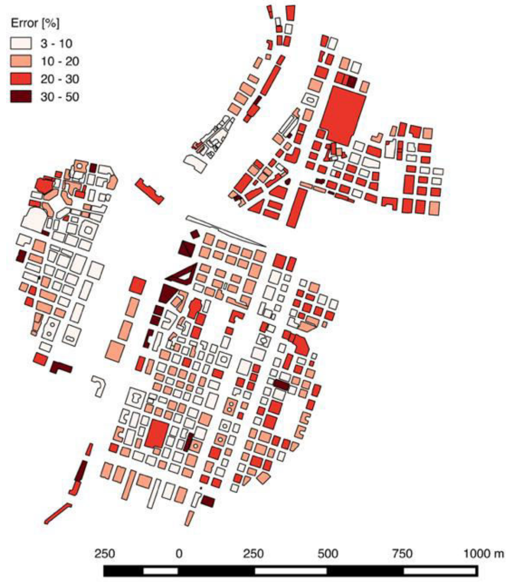

Figure 10 Percentage error committed calculating the damage with the mapping schemes approach, respect to the damage calculated with the real basement distribution at single building detail.

5 Findings

The approach proposed in this study represents an affordable tool to properly characterize buildings exposed to flood, with the final aim to perform a flood damage assessment.

The main issues related to the methodology concern the following aspects. First of all, the generalization of the procedure: the methodology has been developed on the specific case study of the city of Genoa and the identification of the three parameter use to classify buildings has been done considering that for this specific city there is a good relationship between them and the presence of the basement. Of course, this is not necessarily true everywhere. It could be interesting to enlarge the study to other cities in order to understand if a set of common parameters can be identified for a group of cities belonging to a certain common geographic area (northern Italy, all Italy, Europe). The second aspect to be considered is related to the identification of the sample area, which is the most crucial step of the whole procedure. The choice of the sample can significantly influence the accuracy of the results and cannot be easily transformed in an automatic procedure. As a final aspect, also a series of other sources of uncertainty are introduced inside the damage assessment procedure, which are not directly related to the application of the mapping schemes. These uncertainties are related to the hazard scenario used for the analysis, the damage curves applied and the estimation of the replacement cost per square meter of the structure and the content of the buildings.

Nevertheless, the approach proposed in this study allows to characterize exposed elements at "single-element" scale reducing drastically the time required respect with traditional data collection methodologies (i.e. field or virtual surveys). The combined use of remotely sensed data, in-situ data and statistical tools allows to obtain in a shorter time a damage assessment with a high level of accuracy. The methodology, applied to the buildings in the Bisagno creek floodplain (Genoa, Italy), allowed to statistically attribute the basement presence to the buildings, in order to properly quantify the damage caused by the November 2014 flood. The basement presence is one of the most significant characteristics which can influence the damage caused by flood. In this specific study case the error that can be made ignoring the basements can be up to the 20% of the total.

The methodology allows to compute the final damage with a percentage error lower than 2% with a reduction in the time required for field surveys of the order of the 80%.

6 References

1. Hu, Z., Huyck, C., Eguchi, M., Bevington, J. (2014). User guide: Tool for spatial inventory data development. *GEM Technical Report 2014-05 V1.0.0*, GEM Foundation, Pavia, Italy.
2. Blanco-Vogt, A., Schanze, J. (2014). Assessment of the physical flood susceptibility of buildings on a large scale - conceptual and methodological frameworks. *Natural Hazards and Earth System Science*, **Vol. 14**, No. 8, pp.2105-2117.
3. United Nations International Strategy for Disaster Risk Reduction (2009). UNSDR terminology on disaster risk reduction.
4. Coburn, W.A., Spence R.J.S., Pomonis, A. (1994). Vulnerability and Risk Assessment. *UNDP Disaster Management Training Programme*.
5. Van Westen, C.J., Kingma, N., Montoya, L. (2009). Multi-hazard risk assessment Guide Book, Session 4: Elements at risk. United Nations University – ITC School on Disaster Geo-information Management.
6. Genovese, E. (2006). A methodological approach to land use-based flood damage assessment in urban areas: Prague case study. European Commission, Joint Research Centre.
7. Muthukumar, S. (2008). The application of advanced inventory techniques in urban inventory data development to earthquake risk modeling and mitigation in mid-America. Georgia Institute for Technology, GA.
8. EERI (2002). World Housing Encyclopedia. Earthquake Enginnering Research Institute, Oakland, CA, USA.
9. Jaiswal, K., Wald, D., Porter, K. (2010). A Global Building Inventory for Earthquake Loss Estimation and Risk Management. *Earthquake Spectra*, **Vol. 26**, No. 3, pp. 731-748.
10. HAZUS – MH2.1, Multi hazard estimation methodology, Earthquake Model - User Manual. Department of Homeland Security Federal Emergency Management Agency Mitigation Division, Washington, D.C., USA.
11. Brzev S., Scawthorn, C., Charleson, A.W., Allen, L., Greene, M., Jaiswal, K., Silva, V. (2013). GEM Building Taxonomy Version 2, GEM Technical Report 2013-02 V1.0.0. GEM Foundation, Pavia, Italy.
12. Taubenböck, H., Post, J., Roth, A., Zosseder, K., Strunz, G., and Dech, S. (2008). A conceptual vulnerability and risk framework as outline to identify capabilities of remote sensing. *Natural Hazards and Earth System Sciences*, **Vol. 8**, pp. 409-420
13. SENSUM (2013). Deliverable 2.1: Present day and future remote sensing data. Report on globally available present-day and future remote sensing data and products. Framework to integrate Space-based and in-situ sENSing for dynamic vUlnerability and recovery Monitoring (SENSUM) Project.
14. Ebert, A., Kerle, N., Stein, A. (2009). Urban social vulnerability assessment with physical proxies and spatial metrics derived from air- and spaceborne imagery and GIS data. *Natural Hazards*, **Vol. 48**, pp. 275–294.
15. Taubenböck, H. (2011). The vulnerability of a city - diagnosis from a bird's eye view. In: Mörner N.A. (ed) *The Tsunami threat - research and technology.* InTech, Croatia, pp. 107–128.
16. Zeng, J., Zhu, Z.Y., Zhang, J.L., Ouyang, T.P., Qiu, S.F., Zou, Y., Zeng, T. (2011). Social vulnerability assessment of natural hazards on county-scale using high spatial resolution satellite imagery: a case study in the Luogang district of Guangzhou, South China. *Environmental Earth Science*, **Vol. 65**, pp. 173–182.
17. Huyck, C., Hu, Z., Bevington, J., Ghosh, S., Eguchi, R. (2011). Inference of Structural Characteristics for Regional Building Inventories Using Remotely Sensed Data and Ground Observations. *Proceedings of the Ninth International Workshop on Remote Sensing for Disaster Response*, Stanford, CA.
18. Bevington, J., Eguchi, R., Huyck, C., Crowley, H., Dell'Acqua, F., Iannelli, G., Jordan, C., Morley, J., Wieland, M., Parolai, S., Pittore, M., Porter, K., Saito, K., Sarabandi, P., Wright, A., Wyss, M. (2012). Exposure Data Development for the Global Earthquake Model: Inventory Data Capture Tools. *Proceedings of the 15th World Conference on Earthquake Engineering*, Lisbon, Portugal.
19. Kaveckis, G. (2011). Potential Contribution of Hazus-MH to Flood Risk Assessment in the Context of the European Flood Directive. *Master Thesis*, Carinthia University of Applied Sciences, School of Geoinformation, Austria.
20. Seligson, H. (2008). The ShakeOut Scenario Supplemental Study: HAZUS Enhancements and Implementation for the ShakeOut Scenario, USGS.
21. Dell'Acqua, F., Iannelli, G., Pittore, M., Wieland, M. (2012). Inventory data capture tools - remote sensing (Draft V0.3), report published by Global Earthquake Model Foundation.

22. ImageCat (2011). Indonesia Building Exposure Development Report. Internal report to Australia-Indonesia Facility for Disaster Reduction (AIFDR) and the Indonesian Government, ImageCat, Inc., USA.
23. Fiorini, M. (2015). Flood damage assessment assisted by satellite data, *PhD Thesis*, Scuola di Dottorato in Scienze e Tecnologie per l'informazione e la Conoscenza, XXVI Ciclo in Monitoraggio dei sistemi e gestione dei rischi ambientali , University of Genoa, Italy.
24. Harb, M.M., De Vecchi, D., Dell'Acqua, F. (2015). Phisical Vulnerability Proxies from Remotes Sensing: Reviewing, Implementing and Disseminating Selected Techniques. *Geoscience and Remote Sensing Magazine*, IEEE, **Vol.3**, No.1, pp.20-33.
25. Rapid Analysis and Spatialization of Risk (RASOR) project (2014). Working Package 8.1: Definition of vulnerability, functionality and criticality functions.