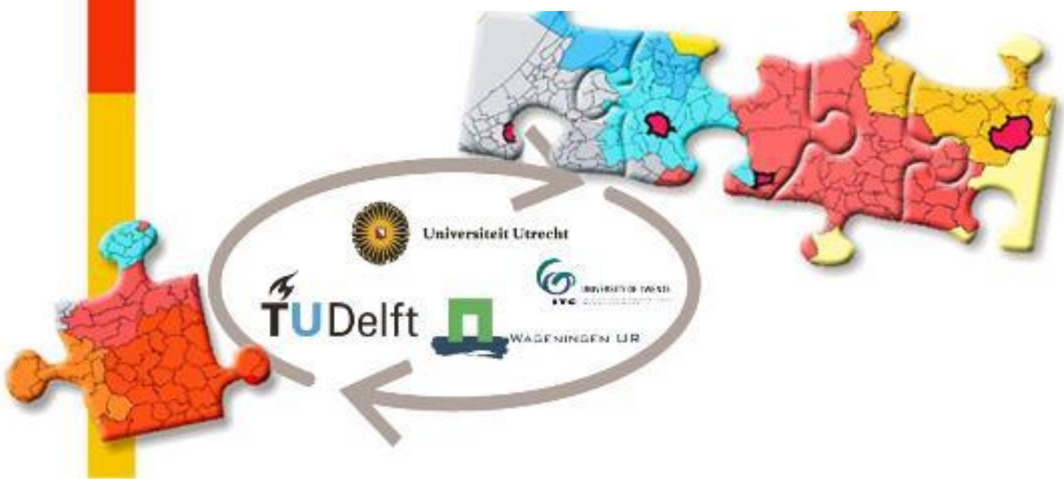
GIMA

Geographical Information Management and Applications

Shelter evacuation in relation to demand characteristics in Dominica

Alina Korman

July, 2015



Shelter evacuation in relation to demand characteristics in Dominica

Written by

Alina Korman

Thesis submitted to Utrecht University, Delft University of Technology, Wageningen University and the International Institute for Geo-Information Science and Earth Observation in a partial fulfillment of the degree of Master of Science in Geographical Information Management and Applications (GIMA).

Date:
Autor's name: Alina Korman
Student number: s6009972
E-mail: alina.korman@gmail.com

First supervisor:
ir. P.W.M. (Ellen-Wien) Augustijn

Second supervisor:
dr.ir. W. (Wietske) Bijker

Professor:
Prof. Dr. M. J. (Menno-Jan) Kraak

Abstract

Reducing disaster vulnerability is one of the main objectives of government actions in countries prone to natural hazards like hurricanes. Evacuation policy planners in such areas need precise and up to date information about population location, roads, rivers etc. Moreover, they aim to employ the most efficient methods to identify vulnerable areas and suggest possible intervention solutions. This research proposes a hybrid approach using object-based image analysis, self-organising maps and network analysis to determine shelter evacuation areas. First, a method of extracting building data from multi spectral and panchromatic images is presented. Data extraction process is performed using an object image based analysis and ends up with assigning to each building attributes that are necessary for determining hurricane vulnerability. Second, obtained features are classified using self-organizing maps to identify the most vulnerable groups. Finally, extracted and classified buildings are used in network shelter analysis. The whole method is highly extendable and modifiable so it can serve not only for evacuation modelling in the study case of Dominica, but also for other areas at risk.

Key words: Buildings extraction; object-oriented classification; data mining; SOM clustering; network analysis; evacuation planning; hurricane risk.

Acknowledgments

In this place I would like to thank all the people who made writing this thesis possible. Obtaining Master's degree has been a long and demanding journey. Until three years ago I would have never dreamt about studying abroad. Now, when everything is finished, I am deeply grateful that God gave me the opportunity to study in The Netherlands, write this thesis and work with such fantastic people.

First I want to give my sincere thanks to Mrs Ellen-Wien Augustijn for supporting me all the time, from the very beginning, when I told her that I would have to study fully on distance, from Poland, through the long and hard time of work. We exchanged many e-mails and met almost always online, but until the end, she was encouraging me to finish writing and defend my thesis on time. Without her devotion to the subject and motivating Skype sessions it would not be possible to have such a happy-ending-master-thesis story.

I would also like to thank Mrs Wietske Bijker for the whole support, especially helping me at the difficult beginning, when the thesis was just started, for the time that she spent explaining eCognition and guiding me through my first steps in working with satellite data.

Special thanks to Mr Corne van Elzakker, the most devoted study director I have ever known, a great man, who was always open to solve problems, answer questions and help with any issues concerning GIMA. Although very busy every day, he was always available, understanding "international's" problems and finding time to sit and talk with students during lunch of coffee breaks. What a gift to have such a Study Director!

Here, I would also like to thank one very special classmate, Kas Kroese, my dear friend, with whom I spent many hours working in almost all group projects. Kas, thank you for all the work that we did together, you were the best project-buddy that I could ever get! Thank you for your patience and the best help I could get - I could always count on you. ☺ Big thanks also to all students from my year, very close for me internationals: Maarja, Stanislav and Emmanuel and all the Dutch students who helped me during these studies – for always trying to answer my questions about dealing with living in The Netherlands and helping me a lot with administration. Special thanks for Stanislav for printing this thesis for me in The Netherlands!

Great thanks also to couchsurfers who hosted me in Enschede, Wageningen and Delft. I especially want to thank Jan Willem Alfenaar, who always had a place for me in Enschede, encouraged me to work hard and always believed that I would succeed.

Big thanks to all administration staff, from Utrecht, Enschede, Wageningen and Delft, for making my studying abroad not so scary.

At the end I would like to thank my mum and my best friends from The Hague:

Mamo, dziękuję Ci za to, że zawsze we mnie wierzyłaś i dopingowałaś moje naukowe działania. Rozumiałaś moje zmęczenie, nerwy i starałaś się pomagać jak tylko mogłaś. Bez Twojego wsparcia napisanie tej pracy byłoby niemożliwe!

Carin and John, you were like parents for me for all the time that I was living in The Netherlands. Your support and love was strengthening me every day when I was so far away from my country. With you I could not only celebrate my successes but also cope with all difficulties. I could not dream about meeting such great people in my life!

Table of Contents

List of figures

List of tables

1. Introduction

Islands in the Caribbean Sea are exposed to a range of weather-related natural hazards like hurricanes and floods. According to The National Weather Service Organization (NOAA, 2014) the typical hurricane season there runs from June 1st to November 30th. Nevertheless, tropical systems may occur outside this period (like e.g. 2013 "Christmas" storm). For 2014, 12 tropical storms, 6 hurricanes and 2 major hurricanes were predicted for this area (MS, 2014). As even one tropical system can cause significant devastation, it is very important that the inhabitants of the islands are prepared for future emergency situations.

In general, a disaster occurs when a natural hazard hits vulnerable people who live in unsafe conditions and lack social protection. A vulnerability can be defined as "*a combination of factors that determine the degree to which someone's life, livelihood, property and other assets are put at risk by a discrete and identifiable event (or series or 'cascade' of such events) in nature and in society*" (Wisner, Blaikie, Cannon, & Davis, 2003). To prevent disastrous events, not only a specific hazard should be recognized, but also living conditions of people at risk should be well-known. One of the problems related to emergency evacuation planning is the exact location of the population, including characteristics of houses and resistance of these structures. Unfortunately, in many places in the world, this kind of up-to-date data is not available or difficult to obtain. What is present are satellite images, which can serve as the most recent information source from which a lot of useful data can be extracted.

Automatic feature extraction from digital images is a broad scientific topic constantly investigated by many research teams. The approach depends primarily on the type of data that needs to be extracted. Commonly used techniques of automatic building extraction from satellite images range from pixel-based to an object-oriented solutions.

There is a limited number of compiled (digital) datasets available for the Caribbean Islands. However, many of them are outdated or incomplete. In some cases, but mostly only for locational characteristics, some hazard maps exist. Nevertheless, they are not always up-to-date and they do not include any information regarding structural elements of houses. Next to this, there is a lot of satellite data recently obtained for a World Bank project.

Hurricane resistance of houses can be indicated by a wide range of locational and structural characteristics. These characteristics include environmental elements like flood hazard, wind hazard or landslide hazard and structural elements like building construction quality, its shape or the type of the roof. Methods for enhancing buildings with data vary from automatic classification of previously segmented maps to a manual attributes assignment.

Shelter locations are now rather randomly distributed in Dominica (which will serve in this research as a case study area) and other Caribbean Islands. To find more relevant locations of shelters, a network analysis could be conducted using the buildings as an input, making sure that all inhabitants trying to reach a shelter will have access to such a facility within a limited amount of time. However, not all buildings are in the same risk categories. People living in houses insusceptible to hazards are far less likely to move to shelters compared to people living in areas with high environmental hazard or living in poorly constructed houses. Such factors allow for a differentiation between houses. This data should be extractable in an automated way to allow for frequent updates.

1.1 Background

1.1.1 Hurricanes

Initially, hurricanes start as tropical cyclones and can develop further into hurricanes and major hurricanes. Tropical cyclones do not exceed 38 mph to 73 mph. Hurricanes reach wind speeds of 74 mph or higher, while major hurricanes have a maximum of sustained winds of 111 mph or higher which is corresponding to a Category 3, 4 or 5 on the Saffir-Simpson Hurricane Wind Scale. The Saffir-Simpson Hurricane Wind Scale is a rating from 1 to 5, based on a hurricane's sustained wind speed. The table below includes a short description of each category including the impact on buildings and other built-up structures that they are likely to cause.

Category	Sustained winds	Types of damage due to hurricane winds
1	119-153 km/h	Some: Some well- constructed houses may have roofs damages. Tree branches will snap, some poorly rooted may be toppled. Power outages can occur due to extensive power lines damages.
2	154-177 km/h	Extensive: Major roof and siding damages. Roads will be blocked by many uprooted trees. Several days or weeks without electricity power.
3 (major)	178-208 km/h	Devastating: Roofs may be removed. Numerous roads blocked by laying trees. Lack of electricity and water for several days to weeks.
4 (major)	209-251 km/h	Catastrophic: Houses may lose not only roofs, but also exterior walls. Most trees will be uprooted. Residential areas will be isolated by fallen trees and power poles and uninhabitable for weeks to months.
5 (major)	252 km/h or higher	Catastrophic: Total roof and wall failure. Total trees collapse. Isolation of residential areas. Most of the area uninhabitable for weeks or months.

Table 1-1 Saffir-Simpson Hurricane Wind scale, own elaboration based on source: http://www.nhc.noaa.gov/aboutsshws.php, obtained: 10.09.2014

In Dominica, tropical cyclones (storms) and hurricanes are rather rare. To show it, the maps from http://stormcarib.com/ - the website about tropical systems threatening the Caribbean islands were obtained. In 160 years, counting from 1851-2010, there were 32 tropical storms, eleven 1st category hurricanes, six 2nd category hurricanes, four 3rd category hurricanes, five 4th category hurricanes and none 5th category hurricane. Together there were 58 storm events. The plots are show in Figure 1-1.

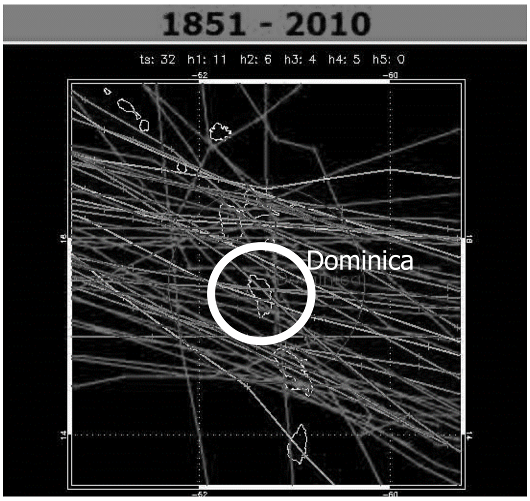


Figure 1-1 Tropical systems in Dominica from 1851-2010, source: from http://stormcarib.com/, obtained: 11.09.2014

Besides the severity of the storm, the direction is also of interest. Since the storms are developed in the low-level of the Northern Hemisphere system of clouds, they are guided by the global winds. For the Atlantic Ocean the most common place of hurricanes origin is the coast of Africa, from where they are moving westward towards the Caribbean Sea and the North American coasts (Society, 2015).

Figure 1-2 The most common hurricanes track in the Atlantic Ocean, source: http://www.hurricanescience.org/science/science/hurricanemovement/, obtained: 20.05.2015

For the period 2010-2014, additional data were taken from the National Hurricane Center and the National Weather Service. In 2011, there were two tropical storms (Emily and Irene), in 2012 one tropical storm (Isaac), in 2013 Chantal, and till now, in 2014, only one storm caused by hurricane Bertha. All tropical systems (for Dominica) from 1851 to 2014 are shown in the Figure 1-3. It is noticeable that in the recent 10 years there is an increase in the number of storms hitting Dominica. The last time this happened was during the 1930-1934 period. Although the number of storm event is the highest in 50 years, the severity of the storms is low (all are tropical storms). During the periods 1965-1969 and 1975-1979, the last hurricanes of category 3-5 took place.

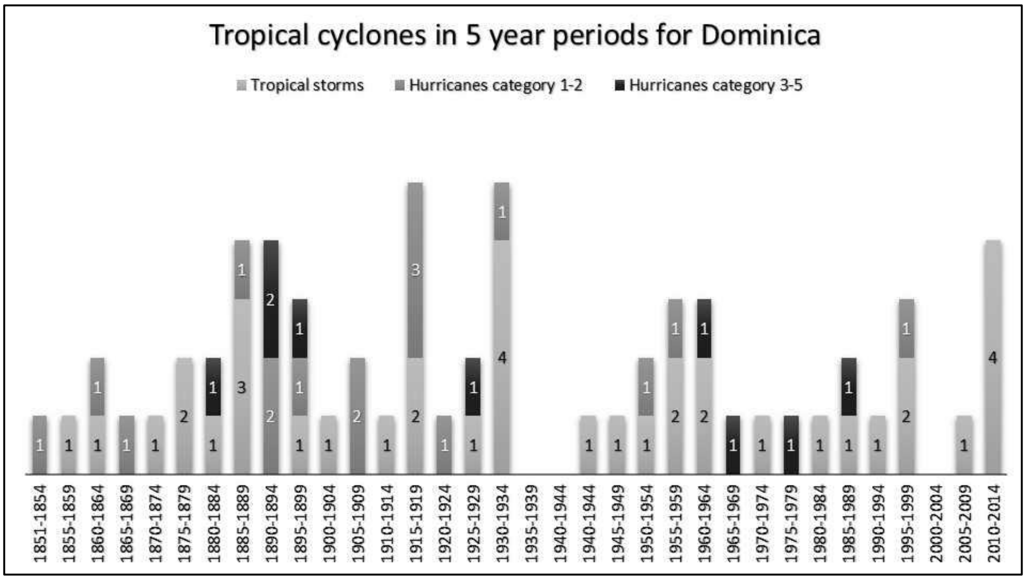


Figure 1-3 Tropical cyclones in 5 year periods for Dominica, own elaboration based on: http://stormcarib.com/ and National Hurricane Center of National Weather Service, data from: 12.09.2014

The fact that the last strong hurricane is already some decades ago does not mean that no danger of more significant cyclone hitting Dominica exists. Data about reported losses express the best the impact of the hurricane hazard on the population.

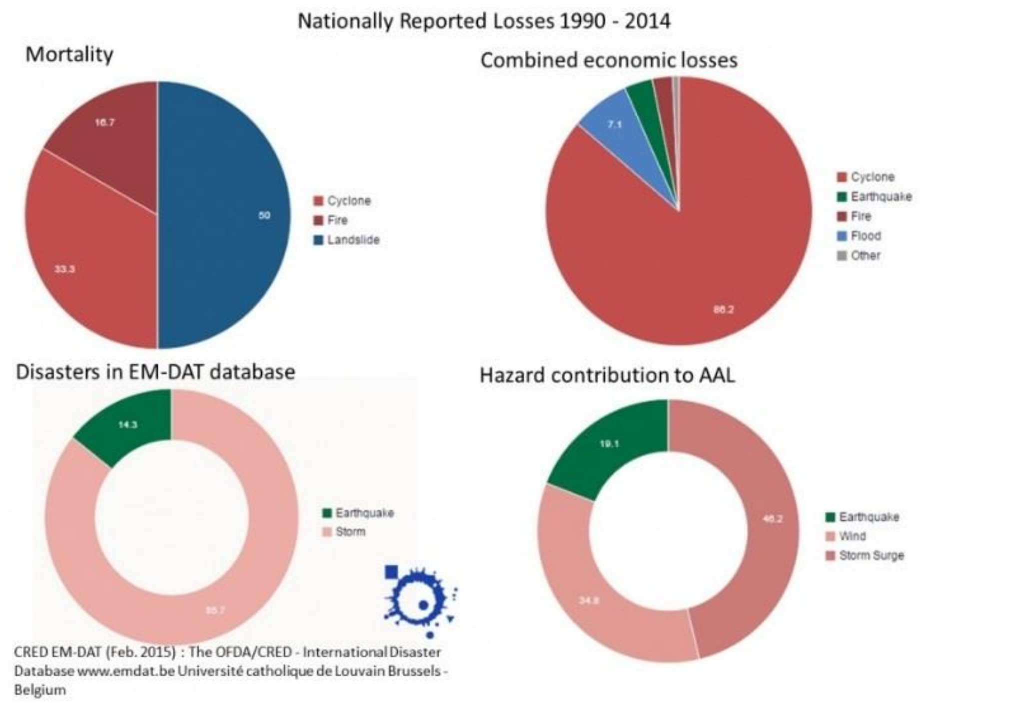


Figure 1-4 Nationally reported losses, 1990-2014 for Dominica, source: (CHARIM, 2015)

The cyclones and the landslides (as a result of the rain during the cyclones) caused the biggest mortality in Dominica, while the contribution to the Average Annual Loss (AAL) of wind and storm surge was more than 80% from all hazards.

1.1.2 Disaster vulnerability reduction projects

In Dominica and other Caribbean islands, several projects (national and regional) connected with reducing disaster vulnerability are currently going on. The Government and the national society agreed that:

"given the current and projected socio-economic factors coupled with the vulnerability of the island, particularly as a result of climate change; it is necessary to implement measures to mitigate against the magnitude of natural disasters on the lives of people" (C. o. Dominica, 2014).

Some of the projects are small and the initiative is coming from outside the area, as for example the Canadian Red Cross project: Disaster Risk Management - Community Resilience, addressing vulnerability at the local level by enhancing community resilience to disaster risk (Cross, 2014). Other projects that come from local sources for example the Government of Dominica is working the creation of several risk maps, which are all available on the national web-site (G. o. t. C. o. Dominica, 2015a). There are flood, wind and landslides maps showing the areas that might be affected by these hazards. The combined map is presented in Figure 1-5, showing that almost half of the island is vulnerable to natural hazards.

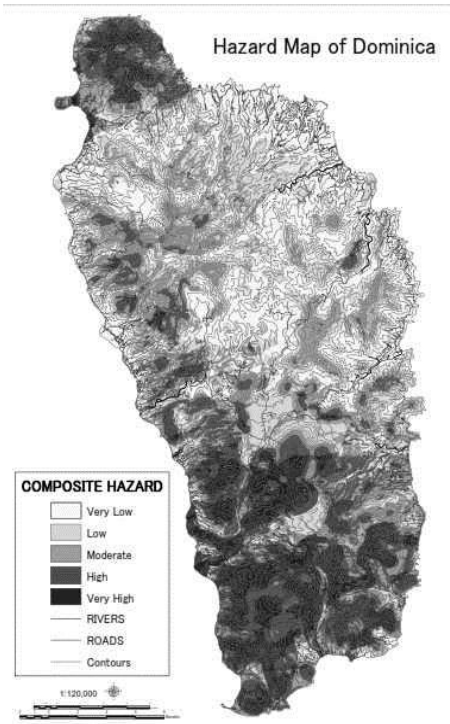


Figure 1-5 Composite hazard map of Dominica, source: http://physicalplanning.gov.dm/land-use-and-development/maps, obtained: 17.10.2014

The most vulnerable are the inhabited areas on the west and south coast, so cities like Portsmouth, Roseau and Baracoa. The Government explains in evacuation procedures and information how to prepare for the disasters via the national web-sites. The Office of Disaster Management (ODM) and the Caribbean Disaster Emergency Management Agency (CDEMA) give useful information regarding disaster management, containing e.g. hurricane Kit Check List, preparedness description for Businesses and Homes and a list of shelters. To reduce houses vulnerability, there are also the requirements for planning and building as describe in the building guidelines (G. o. t. C. o. Dominica, 2015b).

One of the biggest project for Disaster Risk Vulnerability in Dominica was the World Bank Project: Disaster Risk Vulnerability Reduction Project for Dominica. Its main objective is to reduce vulnerability to natural hazards and climate change impacts in Dominica through investment in resilient infrastructure, as well as improved hazard data collection and monitoring systems. The project consists of four components of which the second: Capacity Building and Data Development, Hazard Risk Management and Evaluation is addressing issues that are connected with this research. It supports the "creation of relevant core data and data collection systems as well as the integration of analytical tools to permit improved decision making and engineering design for risk reduction and climate change adaptation" (Trohanis, 2014).

In 2014, The World Bank started the Caribbean Risk Information Program, a consortium run by the Faculty ITC of the University of Twente. A name of the project is CHARIM, which stands for Caribbean Handbook on Risk Information Management (CHARIM, 2015). Already during the 2014, many activities including: collecting all available geospatial data, making several hazard mapping studies, developing a handbook and data management strategy were conducted. As far for 2015, the project is still running. Almost every month some new datasets are added and made available online on the project website itself and in the GeoNode - a web-based application and platform for developing and sharing geospatial information (GeoNode, 2015).

1.2 Motivation and problem statement

1.2.1 Evacuation planning

Disaster management, a complex process in which evacuation planning plays a major role, can be expressed by a general schema. This schema is based on *Framework for disaster operations and associated facilities and flows* made by Caunhye, Nie, & Pokharel (2012). Firstly, the pre-disaster operations have to be undertaken. In order to perform evacuation planning as a pre-disaster operation,

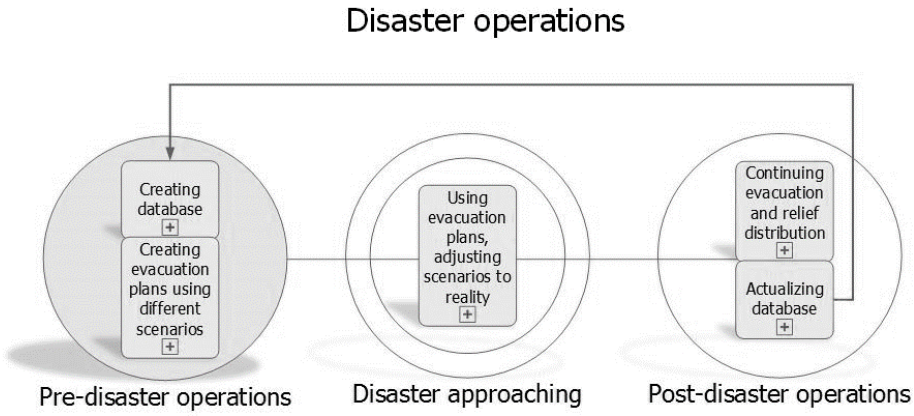


Figure 1-6 Schema of disaster operations, own elaboration based on Framework for disaster operations and associated facilities and flows, made by Caunhye et al. (2012)

two elements are essential: data to perform these analysis (1) and selection of the algorithm (analysis to be performed) given possible limitations in data (2). So in this step several scenarios of evacuation plans are created, using appropriate data. The next step is performed when the disaster is approaching. Then, one of the best fitting scenario is applied, with an adjustment to real conditions. The third step, post-disaster operations include continuing conducting the evacuation plan, relief distribution and actualizing existing database. New, actualized data allows planners to perform new possible scenarios in location-allocation analysis and finally to construct more adjusted evacuation plans. These plans show where the shelters should be placed and how to get to the closest one.

When under hurricane thread, not only data about the damaging power of storm itself is needed, but also about the vulnerability of its target, that is people and their possessions. Also information about obstacles like flooding or landslides that can damage houses or close the way of escape are needed. In many simple evacuation models, shelter locations are planned based on the distribution of houses. Yet, is it too restricted to think that all people will move to evacuation shelters. Some people live in houses of better quality then the construction of the shelter. People will only move to shelters when there is no other option. Moreover, housing distribution data alone is not enough to evaluate what is the probability that houses will be damaged and people will have to move to evacuation shelters. Additional attribute data is needed to identify the type of structures, information about the neighbor, the surface on which it is standing etc. The process of data collection has to be quick and precise, in order to provide the most up to date information.

The problem for Dominica and many other Caribbean islands is that there is no evacuation plan based on actual data. There is no verified existing transport network of the whole island. No complete and up to data digital house layers exist. There is also no data about the exact location of the population, including characteristics like quality of houses and resistance of their structure. The Office of Disaster Management (ODM) and the Caribbean Disaster Emergency Management Agency (CDEMA) are in need of the most recent data about the population location, state of the shelters and all other data that can be helpful in disaster management policy. As it is recently underlined on the CHARIM website, in Dominica: "*ODM currently does not have a GIS mapping capability and lacks equipment, software and trained GIS professional staff. As a result, ODM is unable to use the hazard mapping tools developed.*" (CHARIM, 2015).

1.2.2 Research objectives

The main objective of this research is: *Develop a semi-automated method for identifying the location of houses with a high hurricane vulnerability in order to enhance adequate hurricane shelter planning.*

This splits into three sub-objectives:

- How can building location be extracted from satellite images in an automated way?

- How can hurricane vulnerability characteristics of these buildings be determined in an (semi)automated way?

- How can house vulnerability data be used in shelter location and evacuation planning?

This research will present a method to extract building locations from satellite images in order to fill the gap of missing building data. However, knowing the location of a building does not provide any information on the vulnerability of this structure and the extracted data is only useful if a method exists to enrich the data with descriptive attributes related to hazard vulnerability. Vulnerability of houses can be split into two different types of vulnerability: structural characteristics and locational characteristics.

The first type is related to the structure itself (age of the building, type of construction etc.) the other to the location of the building (on a slope, close to the waterfront, in a densely built-up area). During this research a number of building characteristics related to both the structure of the house and the location will be extracted using GIS techniques. To be able to use the attribute information it is necessary to cluster buildings with similar attributes into more or less hazard sensitive classes. This will be done using the data mining technique called SOM (Self-organizing Maps). The big advantage of this technique is that it can be applied in an un-supervised manner. At the end, the network analysis based on the previous vulnerability classification will be done, to provide information on the most desirable shelter locations.

Presented method will be applied to Dominica example, but the kind of approach could be useful in every (also fast changing or out of reach) inhabited areas, where in every moment satellite images can be captured. For this, greater scientific research is recommended in order to create a more versatile solution.

The workflow of the research is divided into three phases and five steps. In the brackets, after the research question there is an indication in which step it falls. The first and the last secondary research questions are more general and create a closing buckle for the whole research.

- How can buildings be extracted from satellite images using object-based methods? (STEP 1)
- How can the location of the extracted buildings be validated? (STEP 2)
- What kind of data on building characteristics would be indicators of hurricane vulnerability? (STEP 3)
- How can building vulnerability data be extracted from the available data sources? (STEP 3)
- How can buildings with similar vulnerability be clustered into vulnerability classes? (STEP 4)
- How can housing information regarding the vulnerability be used to determine shelter locations and evacuation routes? (STEP 5)

The research is limited to the extraction and use of the housing data. It is not about the network data, the flooding (or other hazard) data and connected with it transport network vulnerability or creating evacuation plans itself. The outcome will provide a useful method for evacuation planners and decision-makers to identify shelter evacuation areas in Dominica. It will not be an ideal solution that can be applied in every location in the world, but some proposals for creating more versatile method will be included in recommendation part of the thesis.

The structure of the whole workflow is presented below, in Figure 1-7. The rows present steps and the columns respectively: literature, software, action and result. Each step demands proper literature review which prepares researcher for further work. The literature review helps in showing the possibilities how to solve certain problems, e.g. in step one, literature review helps answer at the question: "How can building vulnerability data be extracted from the available data sources?" In each step also different software is needed. For buildings layer extraction eCognition software was chosen, but in next steps ArcGIS or R is used. Column "Action" tells what exactly will be done, and the "Result" describes the final outcome.

For a better structuring of this thesis the methodology, results and reflection are described for each separate phase. This will make a clear arrangement, allowing the reader to follow the whole research as it was actually carried out. The whole methodology is holistic and leads in few steps from pure satellite data to identification of demanded locations. In this case it will be shelter locations based on the quality of buildings and their sensitivity to hurricane damage, but the core of the method can be used also in other instances.

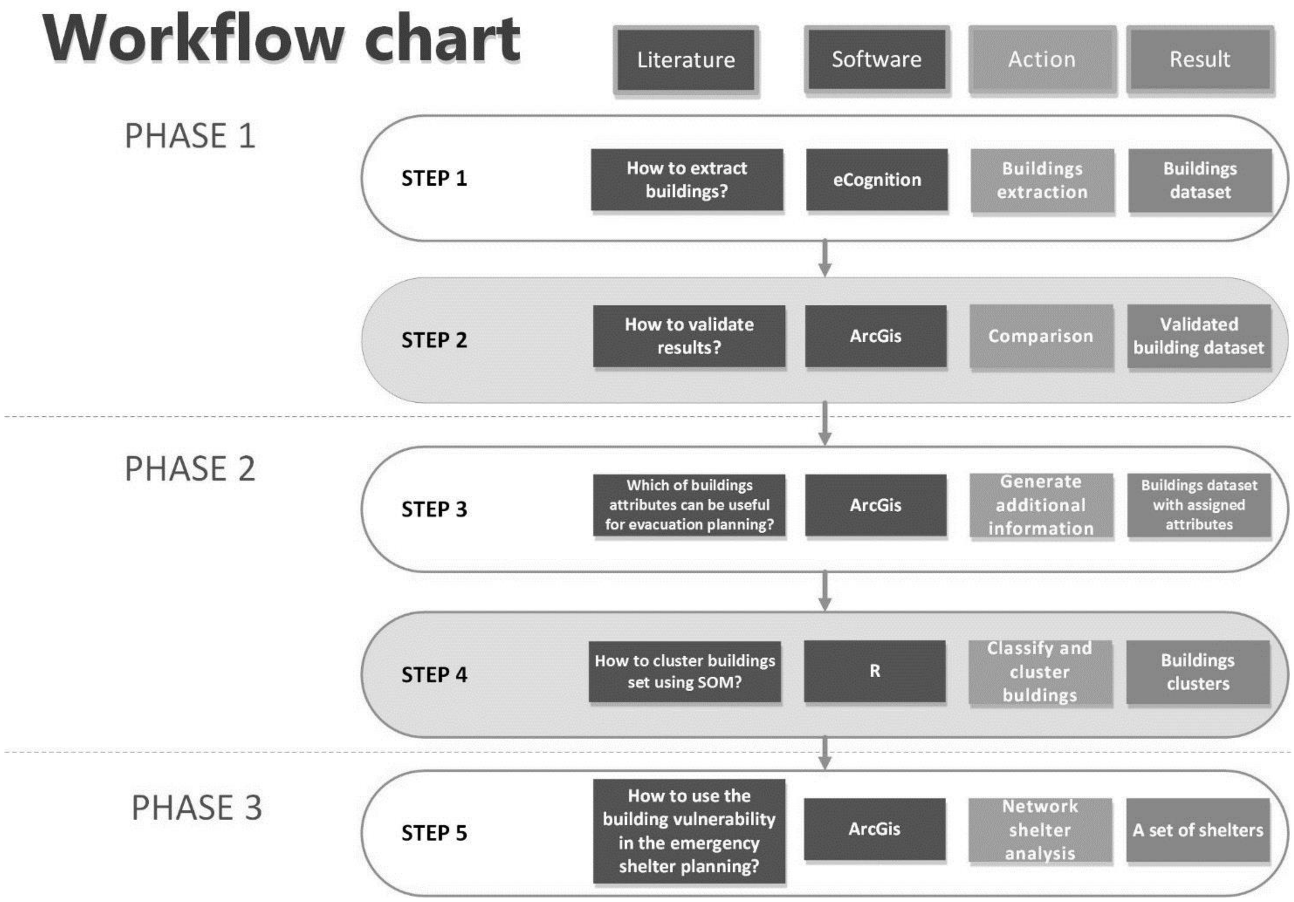


Figure 1-7 Methodology of the whole research, workflow chart; own elaboration

1.3 Thesis structure

This thesis is divided into seven chapters. The introduction was given in chapter 1. Overview of relevant literature will be presented in chapter 2. In chapter 3, case study area and data description will be provided. The building extraction procedure and results (steps 1 and 2) will be discussed in chapter 4. Chapter 5 will discuss the process and results of determining the vulnerability data and creating the vulnerability classes (steps 3 and 4) and chapter 6 will give an overview of the shelter planning. Chapter 7 contains the conclusions and recommendations.

2. Literature review

2.1 Extracting information from satellite images – object-based image analysis

Remote sensing plays an important role in geographic information science. Thanks to airborne, VHR and radar data one can describe and analyze very fast big parts of the earth surface for example to detect land cover. Information extracted from these digital sources can be used in many areas like regional and local planning, policy development, monitoring the effects of climate change, disaster management, vulnerability assessment and many more. Remote sensing techniques improve every year, giving more and more accurate representation of reality. It is not only thanks to a better spectral or spatial resolution of the data, but also thanks to the way researchers are now looking at the images. A general approach changed from pixel-based to object-oriented processing which is in accordance with the principles of human observation. Object oriented processing is now considered to be one of the best image processing methods (Uça Avcı, 2014), especially for the considered especially suitable for very high resolution (VHR) images (Blaschke, 2010). Object-oriented classification allows to see parts of the image as meaningful, comparable objects. It consists of two steps: segmentation and classification.

Figure 2-1 Process of object-based classification, source: http://flash.lakeheadu.ca/~ndewar/Dewar/ecognition.htm, obtained: 14/5/2015

2.1.1 Segmentation

Segmentation divides an image into parts that share certain characteristics, according to previously defined parameters. Segmentation is considered to be well-chosen, only if it can be later used for further processing (Hofmann, Puzicha, & Buhmann, 1998). There are many algorithms and techniques developed for image segmentation, but in general one can say that they can be divided into top-down and bottom-up methods. In the top-down approach, the user provides a model representation of the desired objects and the program treats this model like a samples, looking for the best approximation in the image. In the bottom-up approach the program uses some criteria to evaluate which groups of pixels are likely to fall into similar classes and which should be considered as a background. The whole process can be fully- or semi-automated, depending on the amount of control that a user wishes to have. In general, one can distinguish three main segmentation techniques that are commonly used: pixel-based, edge-based and region-based segmentation.

The pixel-based techniques use the fact, that each pixel has its own properties, e.g. color, so one can distinct groups of pixels that have similar characteristics. In this case, each pixel can be subjected to an unsupervised classification. For this purpose e.g. a k-means algorithm may be used. For monochromatic images, the most efficient technique is thresholding. It creates histograms that are produced based on the number of pixels occurring under certain grey level condition. A threshold value (or values when having multiple-levels) has to be carefully selected and a noise cleaning has to be performed. In thresholding few methods can be used, including Otsu's method (maximum variance) or maximum entropy method (Gliwiński, 2009). After successful classification, adjacent pixels of the same class can be grouped into one object.

Edge-based techniques can be used for both, monochromatic and colored images. Algorithms from this group determine the boundary between areas instead of detecting objects directly. Some of the most common methods are for example: Laplace operator, gradient operator, etc. (Gliwiński, 2009). The advantage of this algorithm is that it is rather fast quick and does not demand a determination of a certain threshold. On the other hand, it has a disadvantage of being much more sensitive to noises and fuzzy edge detection (Krawiec, 2012).

Region-based techniques include i.e. region growing, region merging, region splitting, and split & merge method. In general these techniques are based on the fact that searched objects can be homogenous, but not necessarily limited in space to a single point. In region-growing technique the "growing" starts at some point and ends when it reaches the border set of points with given properties. It is similar to sequential clusters creation. Another approach that is used is the region splitting method, where starting from the whole image, the algorithms divides the input layer into sub-images as long as the resulting sub-images meet the content uniformity test. Individually, splitting or merging algorithms have many disadvantages, for example, the splitting algorithm may not detect certain details of the image and merging algorithms are strongly dependent on the properties of the "home" pixels. The solution is to use hybrid algorithms, so split & merge algorithms. Their idea is to start from the intermediate level of the tree (for example 4x4 pixel blocks) (Krawiec, 2012).

2.1.2 Classification

After a successful segmentation, classification of the data can be performed. In image analysis the aim is to extract a meaningful set of objects that can be easily classified according to their semantics. So for example, from one satellite image, one would like to be able to extract separately classes like: roads, buildings, rivers, vegetation, etc. Many researchers are constantly working on improving such extraction methods. For every kind of object, different algorithms and rules apply. The development of the rule-set will depend on analysts' decisions considering spatial, spectral or textural characteristics of the object. For example, vegetation will have a certain NDVI index, roads will be comparatively elongated, and buildings will have a roofs of different textures and colors (Hamedianfar, Zulhaidi Mohd Shafri, Mansor, & Ahmad, 2014).

The principle of classification is to assign each pixel from the digital image to a class, according to certain conditions. There are many classification techniques which include for example Neural network, Decision tree or Fuzzy classifier, but two of the most common methods are: pixel-based and object-oriented.

Pixel-based classification and object-based classification

Pixel-based techniques, considered to be conventional, are mostly using Maximum Likelihood (ML) and Support Vector Machine (SVM) classifiers. ML classifier is a supervised method, which compares vectors from n-dimensional data to the preliminary model of each pixel's vector. For each cell a statistical probability is compute in order to determine if the cell will be the member of a class. The function can be presented as follows:

$$L_k = \frac{\mathrm{P(k)} * \mathrm{P}(\frac{\mathrm{X}}{\mathrm{k}})}{\sum P(i) * P(\frac{X}{i})}$$

Where $\mathrm{P(k)}$ is a prior probability of class k and $\mathrm{P}(\frac{\mathrm{X}}{\mathrm{k}})$ is a conditional probability to observe **X** from class k, or probability density function (Sensing, 1996).

SVM classifier is also a supervised method, but based on the statistical training idea, which assigns during learning each pixel to one of two categories. After some learning "rounds" points are to classes between which there is a clear gap. For non-linear classification SVM uses a kernel and maps inputs to high-dimensional feature spaces (Fradkin & Muchnik, 2006).

In general, because pixel-based approaches are taking into consideration only spectral information (missing spatial and some textural characteristics) they can result in noise, so called "salt-pepper" effect (Hamedianfar et al., 2014). That is why as a single solution they are considered to be not sufficient, but they can very well serve as an addition to other methods.

There are also some measures like e.g. NDVI or texture measures that can be used in both, pixel-based and object-based classification.

The object-based approach assumes that an image is already segmented and features of each segment like size, shape, texture etc. can be compared among themselves. The topological context together with spatial, spectral and textural characteristics leads to creating a rule-set that may be applied for classification of certain objects.

Besides many obvious characteristics like size, brightness or density that can be used, one of the additional characteristic that is commonly used to texture recognition is Texture after Haralick, which is based on the gray level co-occurrence matrix (GLCM) or gray level difference vector (GLDV). GLCM says about the frequency of different combinations of pixel gray levels and GLDV says about the references to pixels differences in all directions. One of the varieties: GLDV can measure the homogeneity of neighboring objects (Mhangara, Odindi, Kleyn, & Remas, 2014a).

$$\text{GLDV Angular 2nd moment} = \sum_{k=0}^{N-1} V_k^2$$

Where: N is the number of rows or columns and V_k is the image object level and k is

Very important customized feature is NDVI, which stands for Normalized Vegetation Index and is a common measure expressed by an equation:

$$NDVI = \frac{\mathrm{NIR} - \mathrm{Red}}{\mathrm{NIR} + \mathrm{Red}}$$

Where NIR is a spectral reflectance measurements acquired in the near-infrared and Red in the visible regions.

In general, object-based classification is closer to the human perception of the world, since we recognize things rather in a holistic way than as individual cells. It is proven that object-based classification recognizes better the diversity of urban areas (Hamedianfar et al., 2014). Of course the final result of the classification depends not only on the method itself, but also on factors like the quality of data acquisition, resolution of satellite images, errors during data fusion, different geographic properties of the land, etc. It should be noted that the overall accuracy of object-based classification is strongly dependent on the previous segmentation quality. Segmentation though should be adjusted accordingly to each study area, considering characteristics of occurring land cover classes.

Methods for building classification

Even when considering only building extraction, there are many approaches that are nowadays in use. Some recent results show that for example using Very-high-resolution satellite imagery, object-based classification can lead to feature extraction (also buildings roofs) with more than 90% accuracy. What is very important for further users is the explanation of the rule-set that was used.

Some authors propose certain indexes to enhance the level of heterogeneity of surface materials Hamedianfar et al. (2014) uses four indexes: NDRR – Normalized difference red and red edge, NDNB- Normalized difference NIR and blue, NDRY – Normalized difference red edge and yellow and NDGR – Normalized difference green and red edge, to take into account spectral reflectance sensitivity of certain types of roofs. When developing a rule set, also many additional texture and spatial attributes are taken into consideration. For WorldView-2 (WV-2) and combination of WV-2 with lidar data, using object-oriented classification Hamedianfar (2014) achieved 85% and 92.84% of accuracy, respectively (with the overall accuracy of 75.69% for pixel-based approach). Still, some misclassification occurred, especially considering wrong classification of roads as roofs or wrong roof type classification.

Combination of spectral, spatial, textural, contextual and semantic information was also used in the work of Du (2008). There, the most attention has been devoted to distinct classes of buildings with different roof colors. Two levels of classification were performed, using that kind of indexes like e.g. NDVI (to separate vegetation and water bodies), textural dissimilarity (to separate houses from their neighbor), spectral values of roofs (colors like e.g. red or blue), and others. The overall accuracy was 98.61% (user's accuracy) and 63.39% (producer's accuracy). However this work does not provide the exact threshold values that were used, it shows what kind of characteristics can be taken into consideration when using object-oriented approach.

There is also a possibility to select a set of optimal object class features with their threshold for subsequent classification. This approach was presented in work of Mhangara et al. (2014a), where program called SEaTH (for feature recognition and analysis) was used. After choosing samples from the area with features like brightness, spectral bands, NDVI, geometric characteristics, etc. a SEaTH algorithm choses features and thresholds that are the most useful to compare. Apart from NDVI, NDWI Texture after Heralick and Density was primarily used to discrimination of buildings from other objects. The overall accuracy was 93.2%, the misclassification occurred especially between bare areas and buildings, water and tarred surfaces. This research proves that object-oriented, rule-based classification is very much applicable in urban mapping.

2.1.3 Obtaining a good quality input data – satellite images

Apart from using the best techniques for object extraction, having the best possible input data is of high importance. In this research only processing of satellite data is taken into consideration. The quality of satellite images (multispectral and panchromatic) varies a lot, but there are fusion techniques that allow to merge these images and preserve a high resolution. One of it is a pansharpening process.

Pansharpening is an image fusion method based on extracting spatial details from PAN images and putting them into the resampled MS image. The panchromatic data is of high-resolution, while a MS image is colored but with lower resolution. There are two main approaches: Component Substitution techniques (CS), using spectral combination of bands, without filtering the PAN image and Multi Resolution Analysis (MRA) that is filters the PAN image (Vivone et al., 2014).

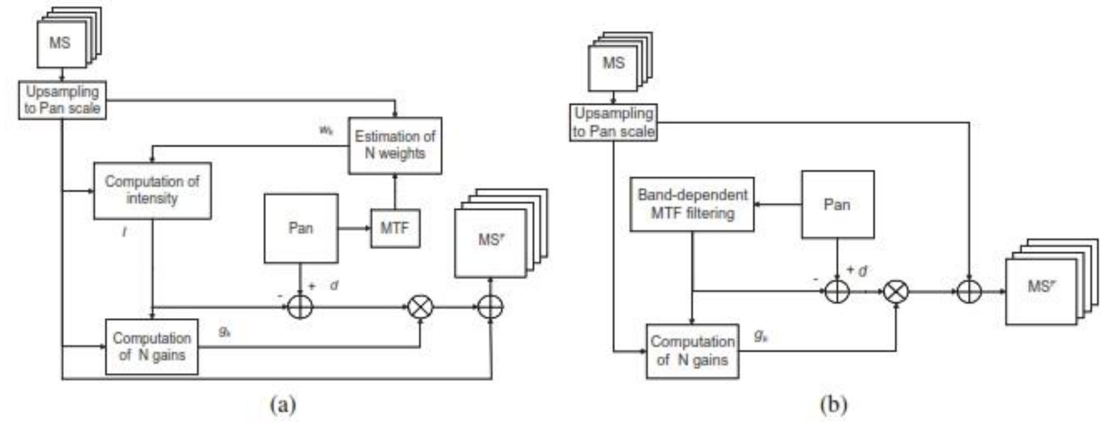


Figure 2-2 Two main pansharpening approaches, a): based on spectral combination of bands, without filtering the PAN image; (b): based on filtering the PAN image (MultiResolution Analysis); source: Vivone, G, et al.; A critical comparison of pansharpening algorithms.

For the purpose of this research the Gram-Smith algorithm, which is one of the CS technique was chosen. This algorithm is the most recommended, since it uses for estimation the spectral response function of a given sensor. In ArcGIS, it is one of the algorithms available through the geoprocessing tool. It is based on orthogonalization of vectors are in the image case, spectral bands. It aims at matching the spectral response of the PAN image and preserves the spectral information of the MS original in the final product. The method itself is patented, and the details of it can be found in: *Laben, Craig A., and Bernard V. Brower. Process for Enhancing the Spatial Resolution of Multispectral Imagery using Pan-Sharpening. U.S. Patent 6,011,875, filed April 29, 1998, and issued January 4, 2000. Eastman Kodak Company, Rochester, N.Y.*

2.2 Hazard Mapping

2.2.1 Risks associated with hurricanes - vulnerability factors

This research deals only with the hurricane hazard and it's occurrence in the coastal zones. There are many researchers that already examined numbers of factors that can influence coastal zones vulnerability. Factors that are mostly mentioned are not only connected with climate, environmental aspects like topography or elevation, etc. but also with housing condition, building materials, and even with educational level of population (Paquette & Lowry, 2012). Some scientists, when comparing the disaster risk coastal counties in USA, proposed even a hurricane index. As an example can serve Hurricane Disaster Risk Index (HDRI) of Davidson and Lambert (2001). Pompe and Haluska (2011) developed a Hurricane Vulnerability Index (HVI), which tells about the susceptibility of coastal areas to a hurricane damage. Quantifying a hurricane vulnerability assessment is a very complex task. It can be based on the examination of historical data, be very diverse and depending on the approach of the certain researcher.

In general, based on the other researches, one can list two main groups of hurricane vulnerability factors: socioeconomic and physical. The socioeconomic factors include the buildings themselves, their types, numbers and condition. The physical factors are connected with three main risks: wind (blowing off roofs, and uprooting trees) flood (damaging houses and other structures) and landslides (demolishing houses).

Buildings subjection to hurricanes – structural factors

When describing building subjection to a hurricane, the most important factor is its vulnerability to the wind forces. It includes most of all the building size, geometry and its construction quality, roof condition and its shape, and shielding by surrounding structures (Khanduri & Morrow, 2003).

SIZE AND SHAPE

What is also of importance is the shape of the building. It is proven, that compact, symmetrical shapes, on the square plan are performing better than these with a rectangle or L-shaped contour. The rectangle is better than the L-shaped plan (Gibbs, 2001). The area of the building is connected with the number of openings and doors that are present in the building structure. Bigger buildings tend to have more windows and doors. Every opening increases the difference in pressurization inside and outside of the buildings, thus decreases the overall wind-safety of the building envelope (Smith, 2010). Additionally, the more windows and doors, the more fastening connections are needed to ensure the strength of the building walls under the lateral forces. Considering the shape of the building, the more irregular, the more prone to wind forces. Every additions to the walls like dormers or chimneys, can cause a local turbulences. Turbulences are causing wind speed-up, which increases the wind loads (Smith, 2010).

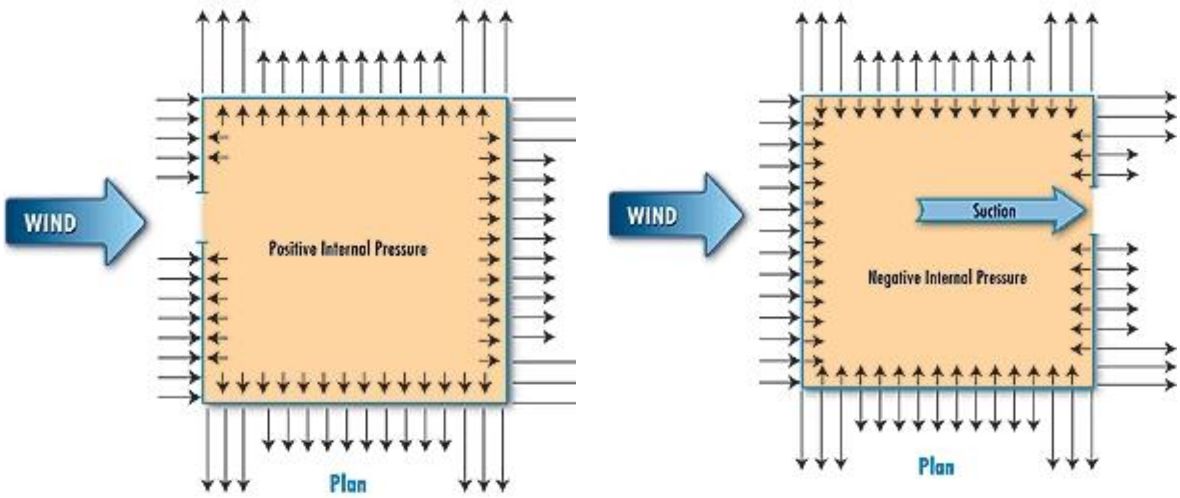


Figure 2-3 Pressure changes during the wind strike on the building, Positive internal pressure and Negative internal pressure; source: (Smith, 2010)

ROOF

The roof provides the main lateral support for walls of the buildings. When the roof is damaged, the ability to withstand the pressure on the walls is significantly reduced. The material of which the roof is constructed is very important, but proper fastening and framing is even more consequential for the final performance of the entire roof structure. It also has been found, that the steeper roof system, the more it is subjected to the wind forces. There are suggestions to exchange gabled roof structures, which are more failure-prone to hip roofs that in general perform better (Crosbie et al., 1997). These hipped shape roofs should be steeply pitched (25 to 30 degrees), having only little or no overhanging elements, sloping in four directions (Gibbs, 2001).

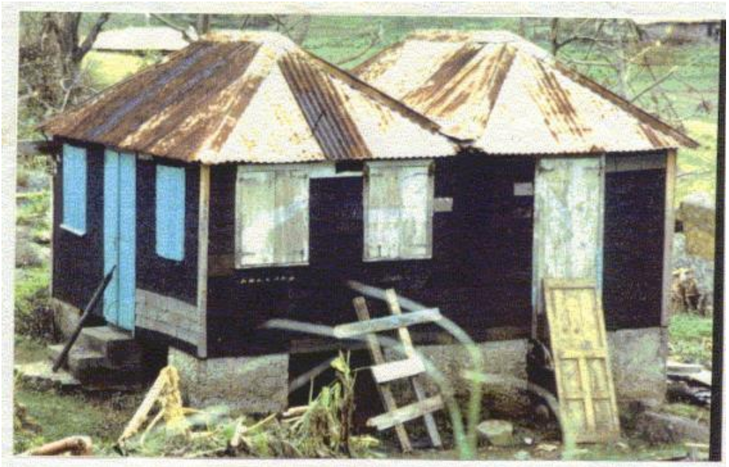


Figure 2-4 Thanks to a favorable geometry and roof shape, this building survived the hurricane Hugo, source: (Gibbs, 2001)

CONSTRUCTION QUALITY

During the hurricane, when wind is hitting the building diaphragm, the lateral forces must be very well transferred from roof to the walls and from the walls to the foundation. Paying attention to the connection details can significantly reduce the damage of the house and is underlined as the most important issue when constructing hurricane-resistant buildings (Crosbie et al., 1997), (Smith, 2010), (Gibbs, 2001). The pictures show how the lateral forces can affect the house during such a weather phenomenon like hurricane.

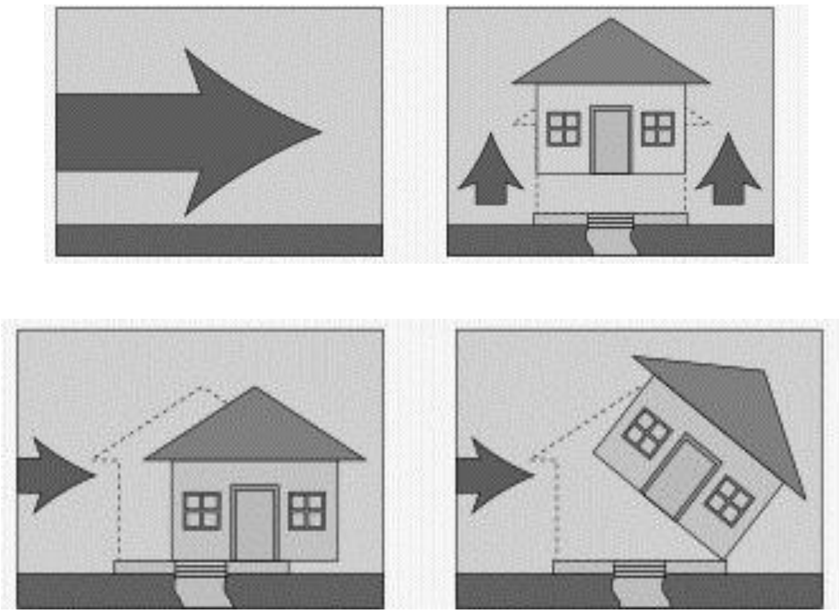


Figure 2-5 Lateral forces that hit the building during a hurricane, source: (Crosbie, Perry, & Smith, 1997)

The information about the condition of the house and construction quality is not visible at sight, and cannot be easily assessed from the satellite images. What can be very helpful in assessing the state of the house is the information in what kind of area it is situated (The Country Poverty Assessment, 2010). The number of existing buildings, changes of housing units and additional information like population density will give some leads to grade the area's poverty. Of course there can be exceptions from this assumption, and the most valuable results would be when performing the fieldwork.

DEVELOPMENT DENSITY

What can be also of importance is the urban pattern of the houses, or more precisely how close they are from each other or if they have a compact structure, or are they more standing alone. The exposure of the building influence the wind loading. The more flat, open terrains, close to oceans regions the more the greater the wind load, so the bigger wind hazard vulnerability (Smith, 2010).

Area subjection to hurricanes – environmental and locational factors

From the section above, one can understand why the wind hazard, as a component of every hurricane occurrence is affecting the overall hurricane vulnerability of the house itself. The wind is of course a nature phenomenon, with properties that can be discussed as a separate factors itself. But the hurricane damage as a whole is caused by at least two additional major components: flooding and landslides (Prevatt, Dupigny-Giroux, & Masters, 2010). Wind, flood and landslides can influence each other, like it is shown on the diagram below (only the environmental and locational factors were taken into consideration).

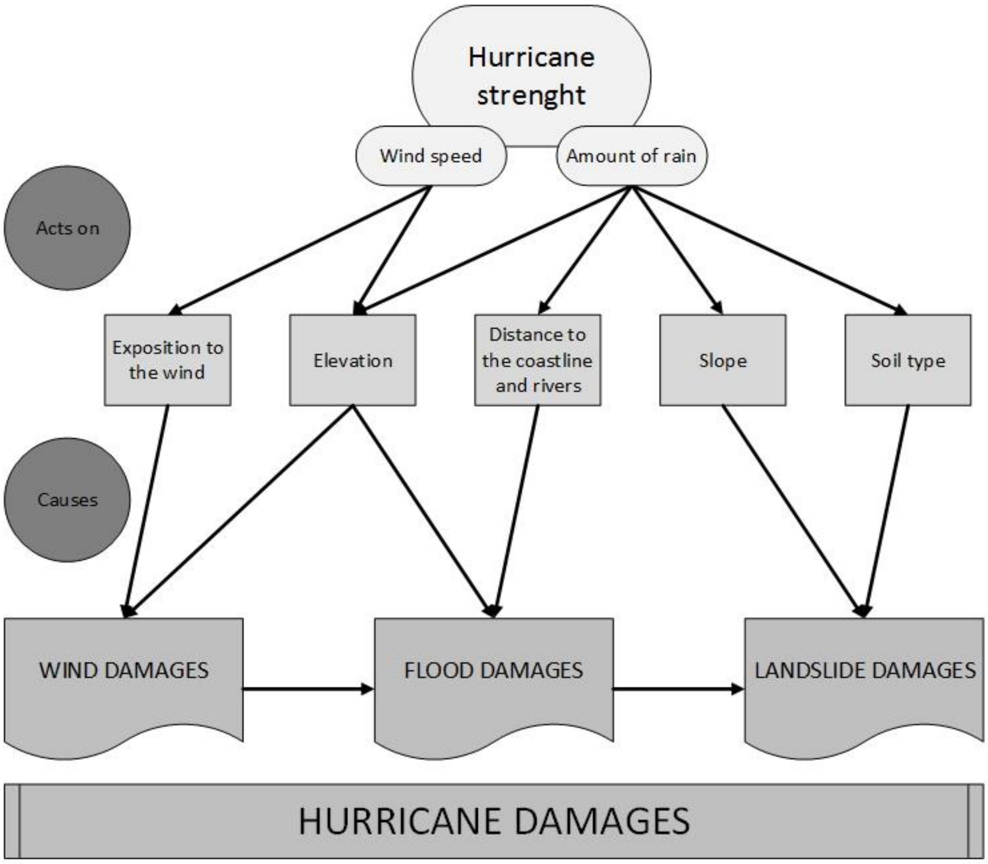


Figure 2-6 Possible environmental factors effects on hurricane damages, source: own elaboration

During a storm, strong winds and large amounts of rain reach the land surface. In general, when the wind hits areas that are very much exposed to the wind and/or of high-altitude, it causes wind damages. Low-altitude areas, close to the coastline and rivers, during large amount of rain are very much prone to flood damages. When strong rain occurs on slopes with poor soil conditions, it is very likely that there will be landslide damages. Damaged by the wind buildings are much more vulnerable to the penetration of the rain, thus flood damages. Buildings destroyed by the wind and floods are much more likely to slide from the slope.

It is also possible that some locations could be sensitive to flooding but not to wind or landslides, or can be sensitive to landslides (uphill) but not to flooding etc. For example, the building which is placed on a big slope can be very well constructed and standing on a very concrete soil. That is why all these factors can be also observed separately.

Many researchers list also some additional physical characteristics like: sea level rise or mean wave height as contributing to a flood damage from hurricanes (Pompe & Haluska, 2011). For wind damages factors of importance are among others also wind direction and the storm duration (Khanduri & Morrow, 2003).

2.2.2. Methods for hazard vulnerability mapping

The first hazard models were developed by insurance companies to assess the loss projection (Pita, Pinelli, Gurley, & Hamid, 2013), however, nowadays they are widely used by spatial planners (Bajabaa, Masoud, & Al-Amri, 2014), (Khanduri & Morrow, 2003). In areas vulnerable to any risks such as floods, landslide, earthquakes, hurricanes, etc. disaster prevention is one of the most important policy. People want to better know the world in which they live and all natural processes that occur. This, together with assessing human vulnerability can give them a clue how to prevent disasters. Identifying, measuring and presenting such information is the main objective of hazard mapping. The final goal of this process is to disseminate information about vulnerable areas to residents (Udono & Kishor Sah, 2002).

Initially risk mapping was based on the knowledge of historic data and statistical predictions. Archival data is very useful for assessing the frequency of the event, measuring the scale of people's preparedness to possible disaster and finally for assessing the scale of the former damages. Unfortunately, it is limited as it only concerns part of the area or a certain time frame. Many times it is also incomplete, or does not include any information about previous buildings characteristics (Pita et al., 2013). That is why that kind of modelling approach is not enough to fully represent current vulnerability of the area. Moreover, constant development of urban areas and continuous environmental changes demand more accurate and more actual mapping techniques.

To map hazard vulnerability of an area, a technique that will be able to perform on a given, multi-factor data its classification and clustering is needed. In the past, many classical multivariate statistical techniques were used for this purpose. Nowadays, there are some alternative solutions (Ferentinou, Karymbalis, Charou, & Sakellariou, 2011).

One of the approach how to quickly map potentially hazardous areas, having a large scale data derived from a remote sensing observation is to use the unsupervised learning algorithms and clustering techniques.

Neural network techniques like SOMs allow users to classify multidimensional data and identify possible patterns (Augustijn, 2013). SOMs have already been employed to evaluate for example: concentration and risk assessment of atmospheric emissions (Cervone, Franzese, Ezber, & Boybeyi, 2008), discrimination of seismic events (Kuyuk, Yildirim, Dogan, & Horasan, 2011), exploring spatiotemporal diffusion patterns (Augustijn, 2013) and many others. Moreover, SOMs can be used in many engineering fields, image analysis or pattern recognition. It is applicable inter alia in analyzing economic stability (Resta, 2012), text clustering (Yuan-Chao, Ming, & Xiao-Long, 2012) or cloud detection and land cover mapping (Angeli, Quesney, & Gross, 2012). In this work SOM will be applied for the purpose of hurricane hazard mapping.

2.3 Data mining using Self Organizing Maps and other clustering techniques

2.3.1 Clustering techniques

In general, when analyzing a high-dimensional large data set and searching for a possible patterns, the analyst looks for the most optimal method of data mining. Clustering is very often chosen solution. There are many clustering techniques, based on competitive learning or statistics. They can be very broadly divided into partitional, hierarchical and density based clustering. Mountain and subtractive clustering, fuzzy clustering, k-means clustering or SOM are some of the partitional methods.

Mountain (Yager & Filev, 1994) and subtractive (Stephen, 1994) clustering are simple and little sensitive to noise algorithms that allow to detect the number of clusters and locations of their centers. However, they produce only some and require a long training time, especially for bigger data sets. Nevertheless, similarly to SOM, they can be used as an initialization procedure for another clustering (K. L. Du, 2010).

K-means clustering is the most common clustering algorithm. It is changing the pattern of the dataset according to nearest-neighbor rule and at each step recalculates the mean of the cluster samples. It demands pre-selection of the cluster prototypes and setting the initial cluster numbers. It has however a drawback of working separately on each cluster, which causes the overall control problem (K. L. Du, 2010).

Fuzzy clustering (Bezdek, 1973) is based on the condition that each vector is assigned not to one, but to multiple clusters. There are many kinds of fuzzy clustering algorithms, using different networks (e.g. Kohonen network), but all of them are based on the fuzzy membership concept.

In hierarchical clustering, one can tell about agglomerative or divisive approach. Agglomerative technique is a bottom-up-grounded and initiate from one cluster for each data point. A merging series are performed until all points are assigned to a new clusters. A divisive approach starts from the whole data set and splits it accordingly. Results of both techniques can be presented on a dendrogram (a clustering tree). Hierarchical is a static clustering, of big complexity, making it not practical for big data sets (K. L. Du, 2010).

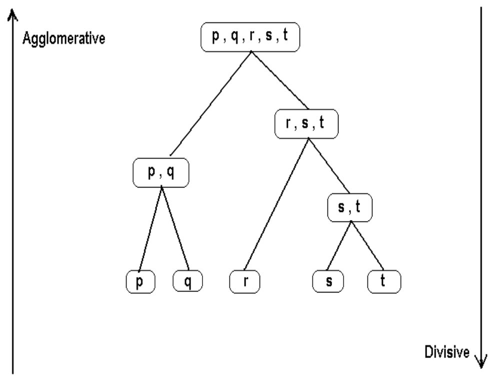


Figure 2-7 Hierarchical clustering, difference between agglomerative and divisive approach, source: http://www.solver.com/xlminer/help/hierarchical-clustering-intro, obtained: 22/05/2015

Density based clustering is grouping data set into clusters according to density conditions. It is based on the assumption of high density of a data within a cluster and low density in other regions. It is very well suited for data with outliers and random clusters shape. It has the same order complexity as hierarchical clustering (K. L. Du, 2010).

2.3.2 SOM

A Self Organizing Map (T. Kohonen, 1990) and (T. Kohonen, 2001) is one of the method that helps to present multidimensional information in a two – dimensional space. The visualization aspect is one of the biggest assets of SOMs, but it is built for the greater purpose. The algorithm itself is of neural network kind, and it is closely related to other unsupervised learning methods like: vector quantization, principal component analysis, principal curves or auto associative neural networks (Vesanto, 1999).

In general, SOMs are based on training neurons that are competing with each other and organizing themselves into a network. To explain more the SOM's algorithm one particular kind, a Kohonen network will be used.

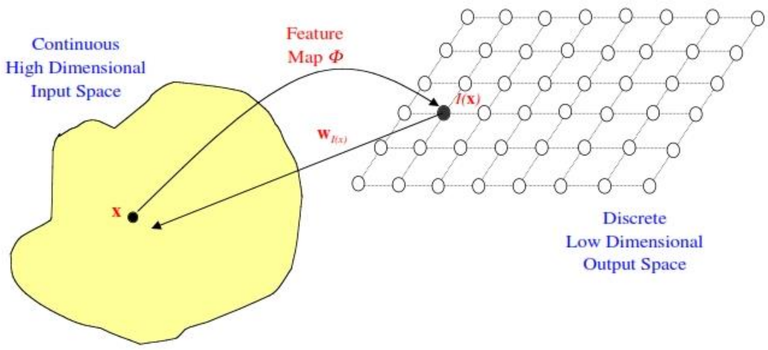


Figure 2-8 Organization of the mapping in SOM, source: (Bullinaria, 2004)

At the beginning of the process there are two main components present: the high dimensional input layer and the low dimensional output lattice. The input layer can be continuous or discrete, while the output space is arranged in a neuron's grid.

The mapping of each point from the input space proceeds in four major steps:

- Initialization: when all connections weights get some small random values.
- Competition: when the discriminant function of Euclidean distance between the input and output vector is defined and neurons compete to have the closest value of the weight vector to the input vector. The one that has the most similar value is consider as a winner, so called BMU (Best Matching Unit). The winner moves to the output data point.
- Cooperation: when all adjacent neurons are also "moving", because their spatial location is determined by a winning vector.
- Adaptation: when the weights of the winning and neighborhood vectors are updated.

This process continues until the nonlinear transformation no longer changes. At the end the output grid is a map that represents the best the input data (Bullinaria, 2004).

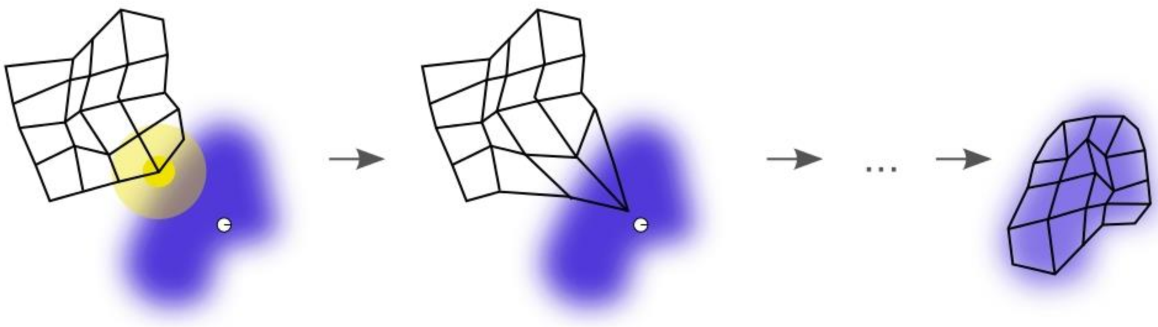


Figure 2-9 SOM process visualization, source: http://en.wikipedia.org/wiki/Self-organizing_map#/media/File:Somtraining.svg, author: Mcld, Created: May 19, 2010, obtained: 13/05/2015

On the picture above the violet area is the distribution of the training data, while the white point is the current training sample. On this picture the whole SOM process is clearly visible, especially the moment when the neighbors of the winning neuron are also moving (in a lesser extend though) to a training input.

In general, the SOM identifies similar point's structures and organize them in an unsupervised manner. It is considered as a one of the clustering techniques that organize data sets and preserve its topology. It enables the user to make a visual distinction of data characteristics and helps to approximate the probable number of clusters. That is why the SOM can be used as a first-stage clustering to produce data prototypes which can be later on subjected to secondary clustering algorithm. That kind of approach is proven to greatly reduce computation time (Palamara, Piglione, & Piccinini, 2011) (Vesanto & Alhoniemi, 2000).

2.4 Shelters location planning using network analysis

2.4.1 Planning shelters locations as a part of disaster management

Disaster management can be defined as a collection of all systematic activities that help to anticipate, prevent and control all emergency situations that are possible to occur in a hazardous area, together with providing a framework for a post-disaster recovery (UNISDR, 2007). Evacuation planning plays a major role in disaster management. Evacuation plans are created in pre-disaster operations, to be used when the hazard is approaching, and to be actualized by a new data in post-disaster actions (Caunhye et al., 2012). Shelters planning is a part of pre-disaster evacuation planning. People that are living in vulnerable areas should be prepared to evacuate to a potential shelter. The decisions that they will make, which shelter location to choose, can be crucial for overall evacuation plan performance (Kongsomsaksakul, 2005).

Figure 2-10 Planning shelters location as a part of disaster management possible schema, own elaboration

2.4.2 Evacuation modelling

Models that deal with the movement of people being at risk of hurricanes, floods, etc. towards safe zones are aiming at creating the most efficient routing plan. Such a modelling needs to take into consideration primarily factors like: number of evacuees (i.e. demand) and their location, network structure and its capacity and number of existing shelters and its location, but also the difference between the exerted demand and the actual network and shelter capacity or possible distortions caused by accidents and misfortune human behavior (Church & Cova, 2000). Modelling results provides not only information of a possible routing solutions, but also show which areas may be difficult to evacuate. Recognition of such an evacuation-risk areas could be used to develop more efficient planning maps (Cova & Church, 1997).

There are many analytical models that are used in emergency management and evacuation modelling. These are for example: Critical Cluster Model - CCM (Cova & Church, 1997), Capacity Constrained Route Algorithm – CCRP (Qingsong, 2006), Intelligent Load Reduction and Incremental Data Structure (Sangho, Betsy, & Shashi, 2007) or Answer Set Programming (Zepeda & Sol, 2007). There are also some methodologies that aim to develop a decision support system (DSS) as an extension of GIS: CEMPS (de Silva, Eglese, & Pidd, 2003) or ALEP (Zepeda, Osorio, & Sol, 2005). In a common use there are also GIS solutions provided by ArcGIS software, with a Network Analyst extension as the principal.

2.4.3 Shelters location planning

Evacuation planning can be modelled by different functions, for example based on shelters location. This approach, taking into consideration the impact of shelter locations on evacuation time was carried out by inter alia Sherali, H. D. et al. (1991), ElDessouki (1998), Kongsomsaksakul et al. (2005) and Sharawi (2007). In general, it can be described as a combined distribution (in this case: of evacuees) and assignment (here: to shelters) problem, called shortly CDA or location-allocation modelling.

ElDessouki (1998) divides his approach into a pre-disaster and post-disaster case. For pre-disaster evacuation plans he assigns each shelter (of a previously set capacity) to a certain amount of affected population in the shortest possible time. A post-disaster case is a special case of multi-period network design.

Sherali (1991) focuses on a flood and hurricane shelter planning and uses a location-allocation model that selects some candidate shelters from a possible set and propose a plan that will minimize the total evacuation time. Among this study an experiment of superimposing a flow of evacuees to the network without using previously identified destinations is performed.

Kongsomsaksakul (2005) studied an optimal shelters location in case of flooding hazard occurrence. The evacuees have a choice to which shelter to go and which route to choose, while the authority deals with the traffic that is present in the network. The whole planning aims to minimize the total evacuation time. The problem is presented using a Stackelberg game and formulated as a bi-level programming. It differs from works of Sherali and ElDessouki in the way the leader can control the evacuee's behavior. The choice of the number and location of the shelters can be designated by the authority.

Sharawi (2007) addresses the problem of shelter planning for a hurricane disaster as a joint location-allocation-inventory problem. In addition to the previously conducted researches, he includes emergency supply inventory determination problem. The model that is used in this case is called a mixed integrated non-linear programming (MINLP). A significant characteristic of this study is its stochastic approach. It takes into consideration possible change in hurricane behavior and enables to accommodate it within the plan, without causing a negative impact on the evacuation cost and demand at shelter locations.

In general, from a technical point of view, for every location-allocation model, there are few common parameters that have to be set. These are:

- Origin locations (for example houses)
- Destination locations (shelters)
- Network (commonly a road network)
- Impedance parameter (the cost, for example driving time; the higher the value, the closest shelter will be chosen)
- Demand (number of people to be evacuated)
- Number of shelters and their capacity

Of course many more parameters and assumptions can be set up, specified for different cases and model demands.

3. Case study and data description

3.1 Study area

At the beginning of this research the study area was set for the whole Dominica Island. Nevertheless, later on during the work, it was discovered that because of time-limited method assessment, the smaller area had to be chosen. The choice fell on the south-west part of the island, containing the city of Rousseau. This subset was chosen also because only for this region the additional vector buildings data was available. When comparing the results of proposed, automated land cover analysis output with this additional vector layer, the assessment of proposed method could be performed.

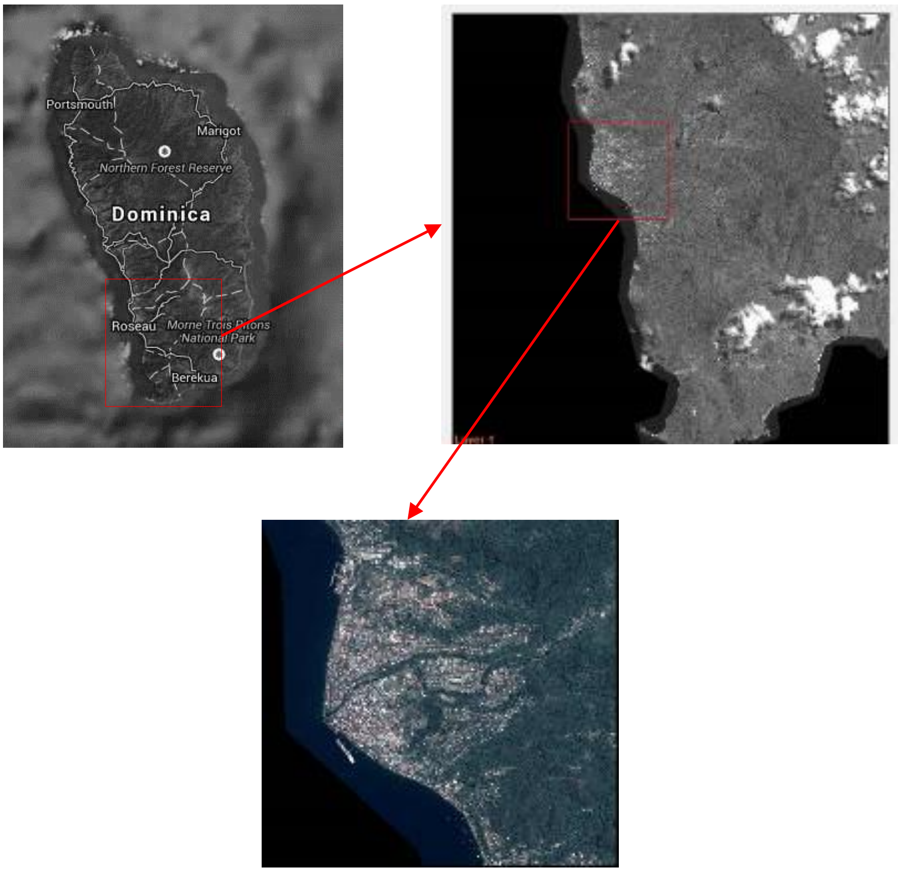


Figure 3-1 The choice of the study area subset for the city of Roseau, three scales

3.1.1 Geography and population of Dominica

Dominica, officially named: Commonwealth of Dominica, is the third biggest island of the Lesser Antilles, placed between Guadeloupe and Martinique (Rad et al., 2013), in the east region of Caribbean Sea with exact location: 15.53N 61.30W. It was discovered by Christopher Columbus in 1493, but remained isolated for many years. Being first French and then a British colony, it became an independent country in 1978. Its largest city and the capital is Roseau, placed on the south-west coast. The whole island has a total area of 750 km2; it is 47 km long with a maximum width of 26 km. The number of inhabitants is around 72 000 (Jury, 2014). In the 90' the population was around 85 000 (Ulrich, 1998), so a significant process of emigration took place in last 20 years.

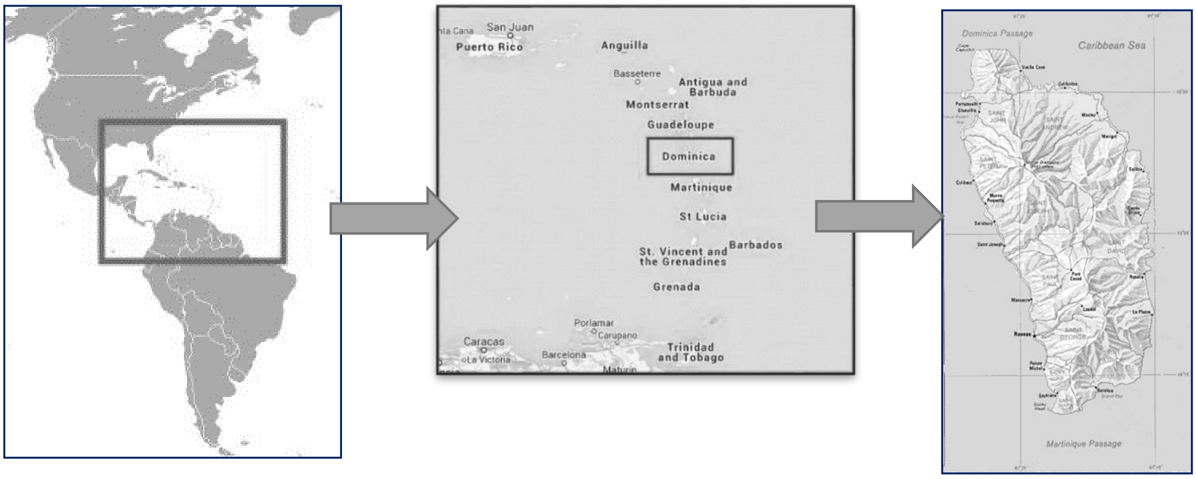


Figure 3-2 Dominica location, own elaboration based on data from Google Maps, 28.10.2014

Dominica has an oval shape, augmented in north-south direction and it is of volcanic origin with active hot springs. Cities and villages are located on the coast and the middle of the island is covered by jungle, with sharp peaks ranging over 1400m, rivers, waterfalls, slopes and gorges. On the north, coastal cliffs are more than 200 m high (Ulrich, 1998). Figure 3-3 shows 14 watershed locations (Rad et al., 2013). The climate is tropical, with summer temperatures around 32°C and winter temperatures from 29 to 30°C. Rainfall is most significant in the mountains (even up to 6 350 mm). On the coast the average

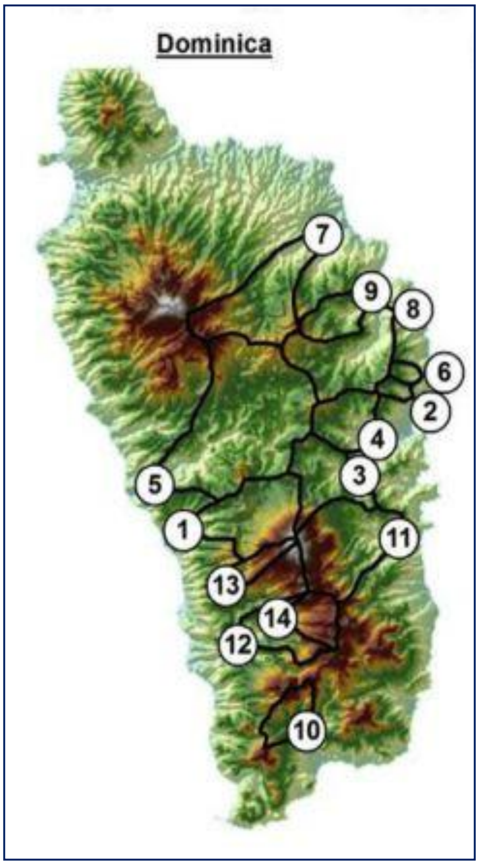


Figure 3-3 Watershed location in Dominica, (Rad, Rivé, Vittecoq, Cerdan, & Allègre, 2013)

annual rainfall is between 1 500 to 3 700 mm. Most of the island is covered by crops, and around 32% of the population lives in rural areas (data for 2013, FAO, Aquastat, http://www.fao.org/nr/aquastat/, Generated: 18 Sep 2014 at 17:08 CEST).

3.1.2 State of the construction sector

In Dominica, like in most of the small islands of The Caribbean Sea, a constant building development is observed and construction sector plays a critical role in the stability of the economy. According to The Physical Planning Division of Dominica (which assists The Ministry of Housing), each year, around 500 development applications are submitted (in fact, less than 80% of them are accepted, and only in 20% of the cases the constructions is starting). The majority of these applications referred to residential use type of building. In 2011 it was around 77% for single-family units, double dwellings, multiple family dwellings and apartments. The biggest building development is observed in southwestern part of the country (Division, 2011).

There are construction practices kept on the island, including regulations like e.g. *Building Code*, which contains all rules and guidelines that have to be fulfilled when constructing a new building. It emphasizes the importance of developing standards that would help to prevent a damage caused by hurricanes and earthquakes invasions. It describes demands for construction of the building frame and the installation of firm holding down mechanisms for the roof, which is important especially for light timber buildings which may be overturned by high winds. The document describes also roof and floor construction with some alternative holding down acceptable methods, but the main emphasis is put on the fire-resistance (Authority, 1996).

In general, Physical Planning Division monitors the whole buildings development, but there is still a need of appreciating the importance of this institution in the minds of people. Many of them, especially with the low income are not building according to approved plans. There are also many cases when developers are subdividing lands contrary to previously established by the law divisions (Limited, 2010). All these unlawful activities has to be monitored and the more awareness should be raised among residents.

In Dominica, like in other Caribbean islands, one can distinguish three main types of houses:

- The informal constructions located in steep hill-sides
- The early 20th century wood constructions
- The contemporary concrete block wall structure

These three kinds of houses mostly have very similar roofs. Typically "*they have a light gauge (24–26 gal), corrugated galvanized iron roof sheeting fastened to 25 mm by 100 mm wood purlins (laid flat) and spaced approximately 0.9–1.2 m apart. The purlins are nailed to wood rafters (nominally 50 mm by 100 mm) spaced 0.9 m apart and spanning from exterior walls to a roof ridge board. Wood collar ties are sometimes used. The rafters are typically toe nailed to a wood wall plate anchored to the concrete ring beam. The rafters may be constructed also as site-built wood trusses with a horizontal bottom chord to support the ceiling. The corrugated metal panels are the primary water barrier and serve as the structural diaphragm in the roof plane. Roof overhangs on residential housing vary by island from the relatively short 0.3–0.45-m lengths seen in some islands where hurricane occurrences are more frequent to as long as 0.9- or 1.2-m*" (Prevatt et al., 2010).

The pictures below, taken in Roseau can give a general overview of the construction state in Dominica:

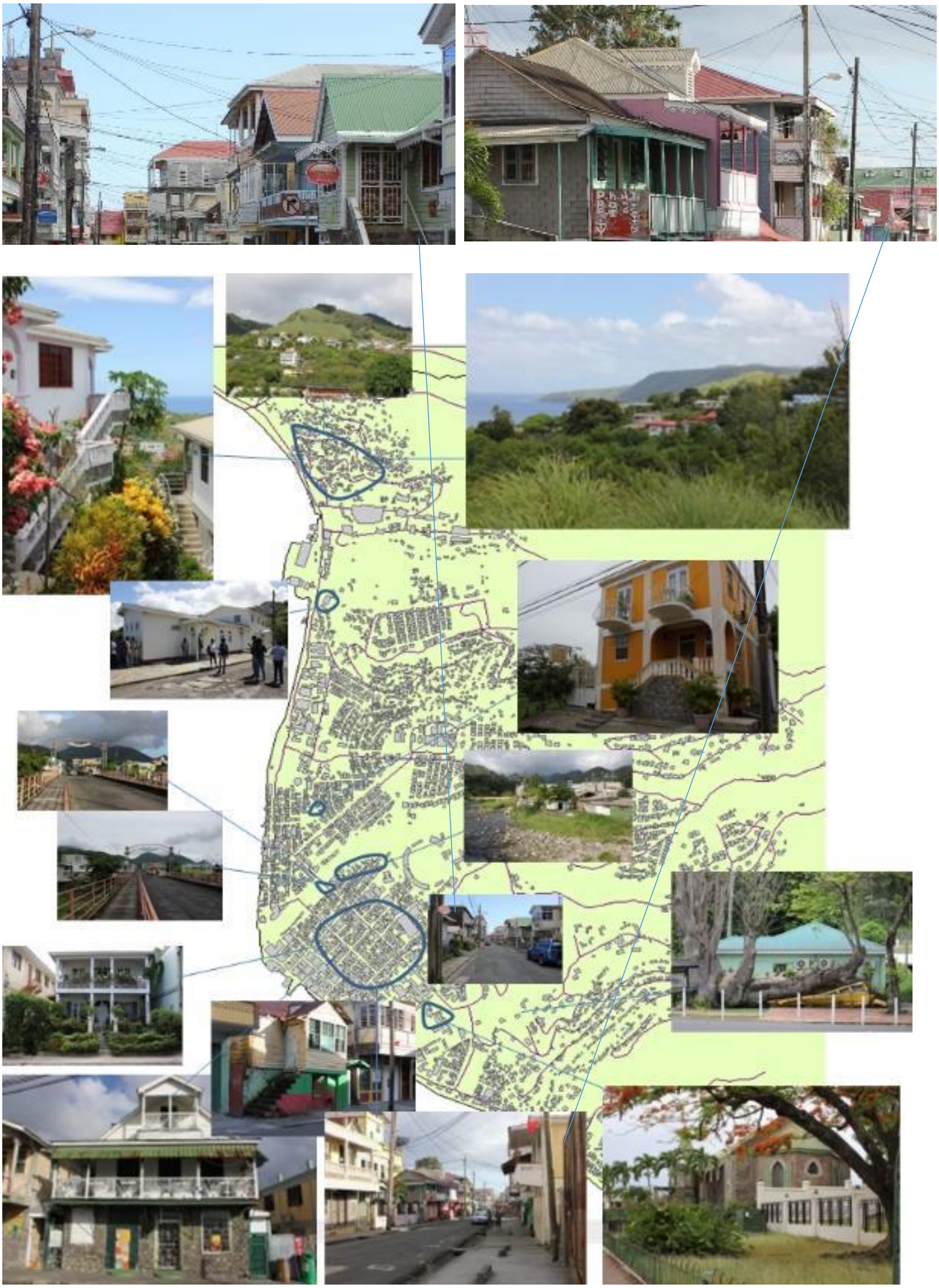


Figure 3-4 Dominica, Roseau, photos taken in 2014 and owned by Mrs Wietske Bijker and Mrs Ellen-Wien Augustijn, researchers in ITC University

3.2 Software

Besides Microsoft Office software programs like Word, Visio or Excel, analytical, graphical and statistical programs are used:

ArcGIS – it is a commonly known software made by ESRI, that let to visualize, analyze and interpret maps and geographic information by providing editing, classifying, trends and patterns finding, etc. It is a base in geographic information system handling in this research. ArcGIS was used in steps 2, 3 and 5.

ILWIS – Integrated Land and Water Information System is a software dedicated for GIS and Remote Sensing. It integrates pictures, vectors and thematic data in one, big package for desktop. It provides such functions like i.e. digitalizing, editing, analyzing, exporting and presenting data. It was developed by ITC Enschede, but from 2007 it became a free software. ILWIS was used in step 3 in this research.

R – Programming language and environment for statistical computing and graphics. It is free software, on GNU General Public License. R provides a variety of statistical: linear and nonlinear modelling, classical statistical tests, classification, clustering etc. and graphical techniques. It is very much extensible and allows users to add additional functionality by defining new functions. It offers:

- *an effective data handling and storage facility,*
- *a suite of operators for calculations on arrays, in particular matrices,*
- *a large, coherent, integrated collection of intermediate tools for data analysis,*
- *graphical facilities for data analysis and display either on-screen or on hardcopy*
- *well-developed, simple and effective programming language which includes conditionals, loops, user-defined recursive functions and input and output facilities (Foundation, 2014).*

R was used in step 4 in this research, for implementation of SOM, hierarchical clustering and to export the results in csv. files.

eCognition – more precisely: eCognition Developer is an object-based commercial software environment made by Trimble. It uses remote sensing data and helps with developing rule sets for the automatic image analysis. It offers a broad collection of segmentation algorithms such as multiresolution segmentation, quad tree or chessboard and classification algorithms like sample-based nearest neighbor, fuzzy logic membership function or specialized context-driven analysis. Extracted features can be exported in raster or vector format allowing integration with GIS software. Rule sets and applications developed for one task can be reused over large areas (Trimble, 2014). In this research eCognition was implemented to perform step 1 – extract buildings from satellite images.

3.2.1 eCognition implementation

Segmentation

In this thesis, for segmentation processing an eCognition Developer software was used. There are few segmentation techniques available in eCognition: chessboard, quadtree-based, contrast split, multiresolution, spectral difference, multi-threshold and contrast filter segmentation, of which some are described below.

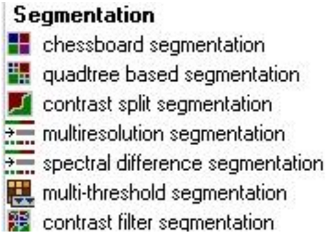


Figure 3-5 List of segmentation types available in eCognition software, version 9.0

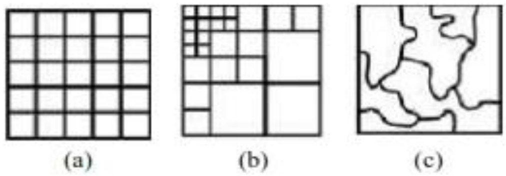


Figure 3-6 Three common segmentation types: (a) Chessboard, (b) Quadtree, (c) Multiresolution, source: Uça Avcı, Z., D., Parameter tests for image segmentation of an agricultural region

A chessboard segmentation produces a simple squared objects of the same, given size. It is the simplest top-down segmentation type. Quadtree divides images into squares of different sizes, depending on the homogeneity of the image. That means that the more heterogeneous areas, the smaller image objects will be distinct. *It is mostly used for images with a well-separated background or foreground and images with a significant proportion of background that is easy to separate* (Trimble, 2014).

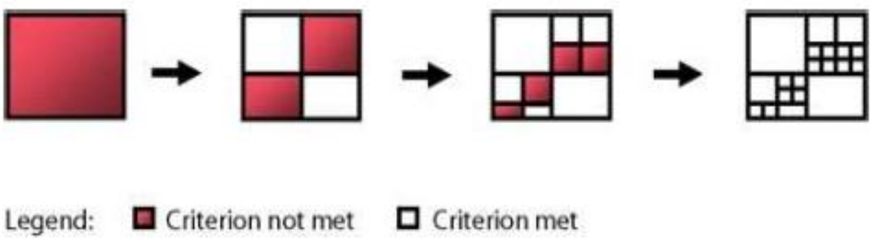


Figure 3-7 Quadtree segmentation, source: eCognition User Guide (Trimble, 2014)

One of the mostly used bottom-up approach is multiresolution segmentation. It is *identifying single image objects of one pixel in size and merges them with their neighbors, based on relative homogeneity criterion based on combination of spectral and shape criteria.* In the multiresolution segmentation the important parameter that has to be set is scale, which determines the size of the objects. The bigger the scale value, the more objects can be merged and the larger objects can be obtained (Uddin, 2010). The multiresolution segmentation is *mostly used for extracting features that are characterized not purely by color but also by shape homogeneity and extracting land cover or man-made features from remote sensing imagery* (Trimble, 2014).

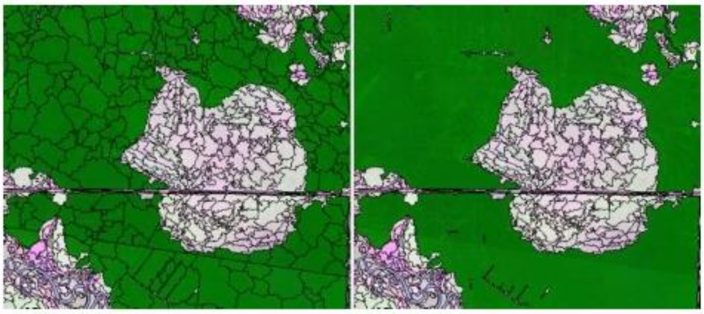


Figure 3-8 Region merging performance – here for vegetation class, source: eCognition User Guide (Trimble, 2014)

There are also algorithms that are close to the segmentation idea, but they cannot be used to identifying unidentified objects. These are: merge region and grow region algorithms. The merge region is reducing the number of objects, leaving the classification unchanged. The grow region algorithm extends all objects that are specified in the domain by neighboring image objects of defined class.

Classification

Classification can be performed using a various features, of which a list is presented below. This list is extendable and new algorithmic and other features can be very simply added. The simplest classification algorithm is called Assign Class and it uses a simple, single threshold condition to decide whether the object belongs to a class or not. More complex is a Classification Algorithm, which evaluates the class description and decides if an image object can be a member of a class. The value of membership is calculated for the three best-fitting classes and stored in an object classification window. Therefore one can see which classes could possibly fit predefined settings. The third is a Hierarchical Classification Algorithm, which can activate or deactivate classes based on their hierarchical setting. *If the membership value of an image object is lower than the pre-defined minimum membership value, the image object remains unclassified. If two or more class descriptions share the highest membership value, the assignment of an object to one of these classes is random (Trimble, 2014).*

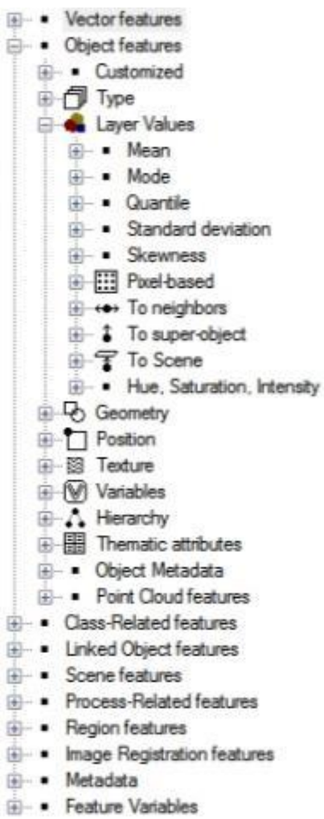


Figure 3-9 Available features list in eCogntion version 9.0, source: eCognition User Guide (Trimble, 2014)

3.3 Data description

The table below presents data that was used in this research. Each data type has its format and attributes description. In brackets it is indicated in which step the data was used.

Type	Format	Available: yes/ not	Description of attributes and comments	Step in which it was used
Satellite images	TIFF	yes	Pleiades images (2014) with around 15% cloud cover, Panchromatic (8) and Multi-spectral (2). Almost full coverage, in North-West the small part of the island is missing.	1
DEM	TIFF	yes	Two files: "DEM10.dem" and "dem_30gt" cover the whole island.	3, 4
Buildings location	Shape file	yes	Two layers: one, "Buildings_Barriet" is bigger, contains Rousseau and surrounding (area of around 80^2 km), but does not contain any attributes information; second one "Buildings_final2013" covers only Rousseau, but has attributes like: Object ID, Shape_lenght, Shape_area and name.	2, 3, 4, 5
Roads location	Vector data	yes	New data, shared on the CHARIM website on 6th of March 2015. Contains also road types.	5
Drains and rivers location	Vector data	yes	New data, shared on the CHARIM website on 6th of March 2015.	3, 4
Shelters location	Vector Data	Yes	New data, obtained from ITC researchers. Contains information like: exact location, shelter code and its type, the name of manager and the community.	5
Landslide susceptibility	Raster	yes	New data, shared on the CHARIM website on 8th of March 2015. Contains three susceptibility classes: low, moderate, high.	3, 4
Flood hazard	Raster	yes	New data, from CHARIM project, not yet available online, but obtained from ITC researchers. Contains 5 classes: no flood, low, moderate, high and very high flood hazard.	3, 4
Wind hazard	Raster	yes	New data, from CHARIM project, not yet available online, but obtained from ITC researchers. Contains 5 classes: very low, low, moderate, high, very high hazard.	3, 4

Table 3-1 Research data

3.3.1 Pleiades satellite images

The satellite images that were used in this research were obtained from ITC University (Enschede, The Netherlands) sources. ITC is currently (February - December 2014) conducting a project for Dominica with the topic: Development of Evacuation Routes and Determination of Shelter Accessibility using High Resolution Satellite Imagery, Dominica, which main objective is to: "*train Government of Commonwealth of Dominica employees in the use of satellite imagery to accurately map assets and perform analysis to determine access and evacuation routes*" (http://www.itc.nl/projects/Bas/Bas.aspx?Id=1772). For performing this project, executed on behalf of The World Bank, Panchromatic, Multispectral and Pansharpened images of Dominica were used. World Bank acquired these images from Spot Image, which is an official distributor of Pleiades products.

Pleiades is a constellation of two satellites with sun-synchronous orbits with 180^0 offset. It enables to get the very-high-resolution (VHR) images of the land with the daily accuracy over any chosen point of the globe. Pleiades 1A was launched in December 2011 and Pleiades 1B one year later. The mission is planned for 5 years and to meet both civil and commercial needs. (Wikipedia, 2014) It is very well suited for many remote sensing applications. Both satellites are characterized with the following properties:

- Resolution: 50 cm (Panchromatic), 2m (Multispectral), 50 cm (Pansharpened) and
- Footprint: 20 km at nadir, up to 100m x 100m mosaics single pass
- 5 acquisition modes: Target, Strip Mapping, Tri-stereo, Corridor and Persistent Surveillance
- Maximum acquisition capacity of 1 million sq.km per day (for one satellite)
- Work plan updated every 8 hours (Space, 2013).

4. Phase 1: Extracting buildings dataset from satellite image

4.1. Methodology

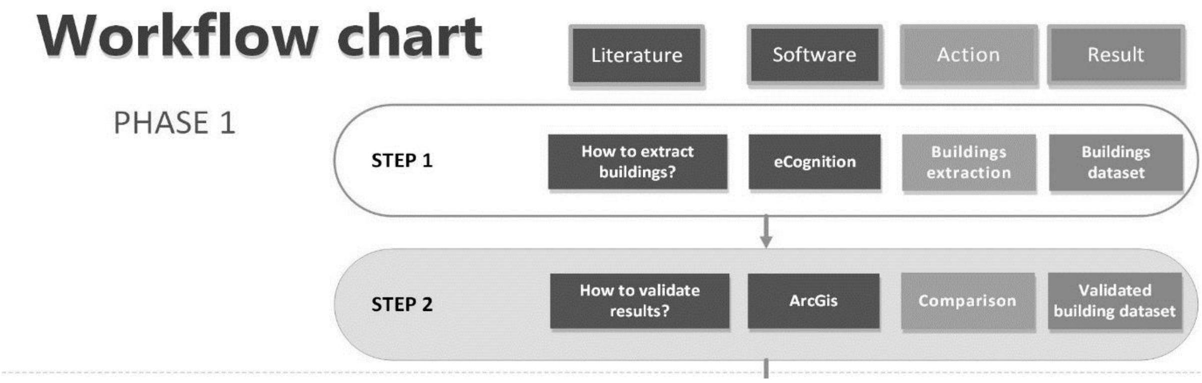


Figure 4-1 Workflow, phase 1

4.1.1 Step 1 - Rule-based buildings extraction

The purpose of this step is to extract buildings dataset from satellite images. It demands learning eCognition software, finding which functions are fitting the best problem solution and knowing what kind of approaches were already applied in similar studies. For this, papers of i.e.: Mhangara (Mhangara et al., 2014a) – about roads extraction, Du (Y. Du, 2008) – about buildings extraction, Hamedianfar et.al (Hamedianfar et al., 2014), Karakış (Karakış, Marangoz, & Büyüksalih, 2014), Baatz & Schäpe (Baatz, 2000) and Damla (Uça Avcı, 2014) about the segmentation procedures were reviewed. Additionally, video tutorials about eCognition software were used as useful source of knowledge about the software possibilities.

At the beginning data preparation has to be done. Multispectral and panchromatic images of the area of the interest are chosen to be used in pansharpening process. Obtained image has a high resolution for visible and NIR bands. The next part is the automated object-based analysis. It uses segmentation and classification procedures which enable to recognize and classify buildings objects. Segmented image is subjected to rule-set classification in order to distinct land cover elements like vegetation, roads, buildings, water etc. For wrongly classified or unclassified parts of the image an additional segmentation and/or classification has to be applied. At the end buildings objects are merged into one layer, which is exported to the shape file with some object feature information.

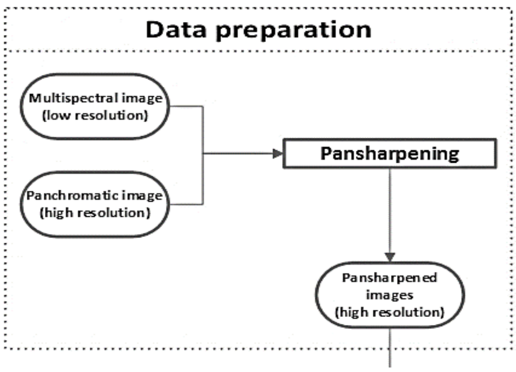


Figure 4-2 Rule-based buildings extraction, data preparation

4.1.2 Step 2 - Accuracy evaluation

In this step the validation procedure on the extracted shape file is performed. It is done by comparing object-based analysis results with an existing buildings layer obtained in a fieldwork (here performed in 2013 by Dominica's researchers). The accuracy assessment is using object-based approach, which is based on works by Zhan et.al (Zhan, Molenaar, Tempfli, & Shi, 2005) and Du (Y. Du, 2008). It is done in two steps, using first size and later on shape condition. Area condition uses CLIPPING tool for polygons from previously exported shapefile and reference data. The clipped areas are recorded as correctly detected buildings. For both layers additional area field is calculated. Then the percentage of the area of these correctly recognized buildings over the reference data is calculated using TABULATE INTERSECTION tool. Shape condition uses MINIMUM BOUNDING GEOMETRY tool, which is sometimes also called bounding box. It is simplifying the polygons of extracted buildings features by creating new polygons based on their spatial extend. These new polygons represent a specified minimum bounding geometry enclosing each input feature. In this research as most optimum bounding geometry type the CONVEX_HULL is chosen. It creates a rectangle of the smallest width enclosing an input feature. For this features and for the original buildings polygons the length and the perimeter is calculated. At the end these two parameters are compared for each object.

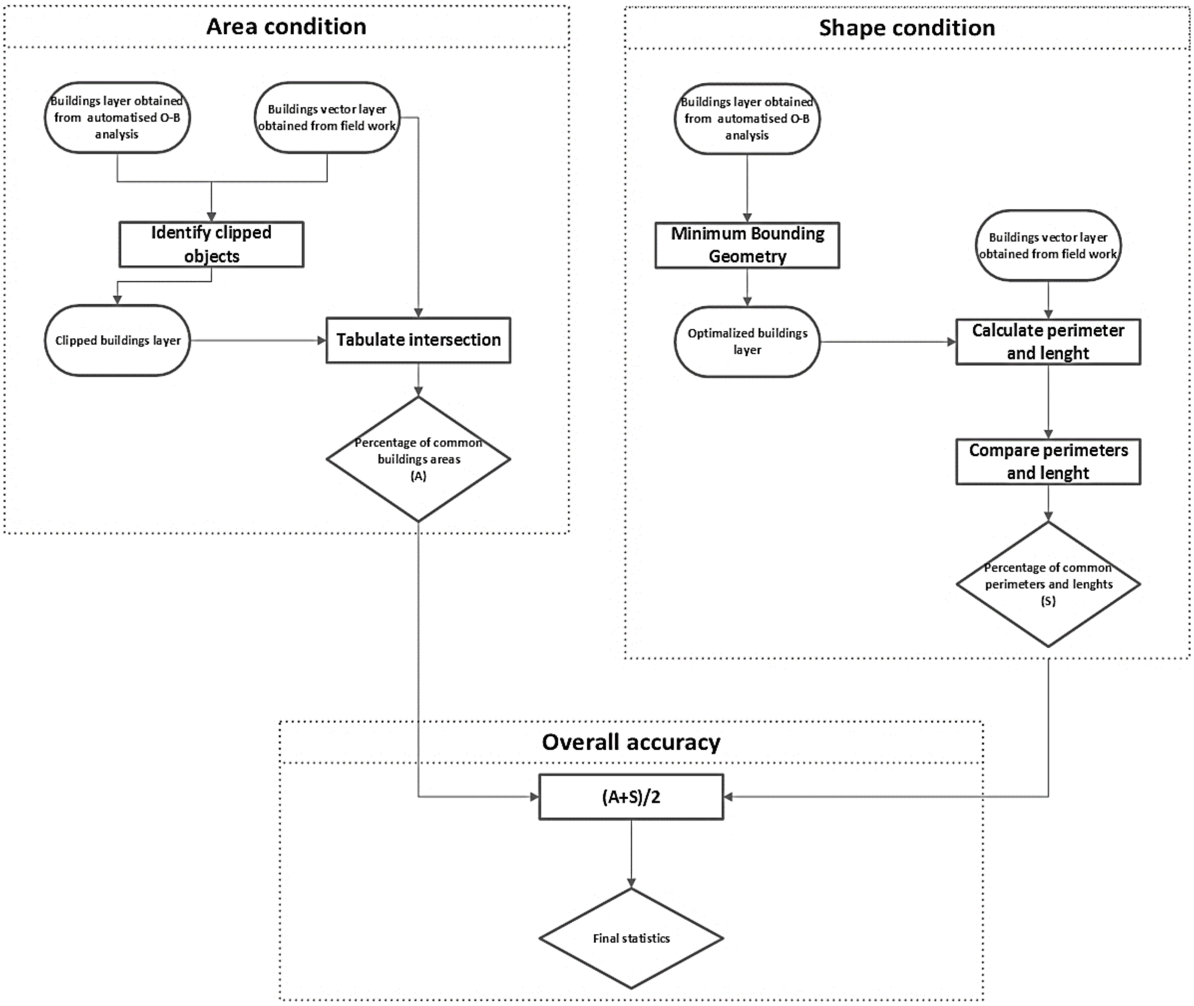


Figure 4-3 Accuracy assessment, area and shape condition, overall accuracy

4.2. Results

4.2.1 Rule-based buildings extraction results – step 1

Data preparation

The available data were high resolution panchromatic image and lower resolution multispectral image of the whole island. Since all images were acquired at the same time, the pansharpening could be used to convert the multispectral image to the same high resolution as the panchromatic image. Additionally, the information about the NIR layer (so the fourth layer), which is very useful for e.g. the vegetation detection, was also included in this pansharpening process. Because of the computation time limitation only the subset of the area was chosen. The cutting was done using a selection of 7324x7182 pixels in eCognition software. The panchromatic image was pansharpened in ArcGis software, using Gram-Smith method, with weights for Red-band: 0.9, for Green-Band: 0.75, for Blue-Band: 0.5 and for NIR-Band: 0.5, using wavelength. The resolution was set to 0.5x0.5 pixel. High resolution, 4-layer image (with NIR) was obtained. This image was saved and uploaded to eCognition software, where the analysis of the land cover could be performed.

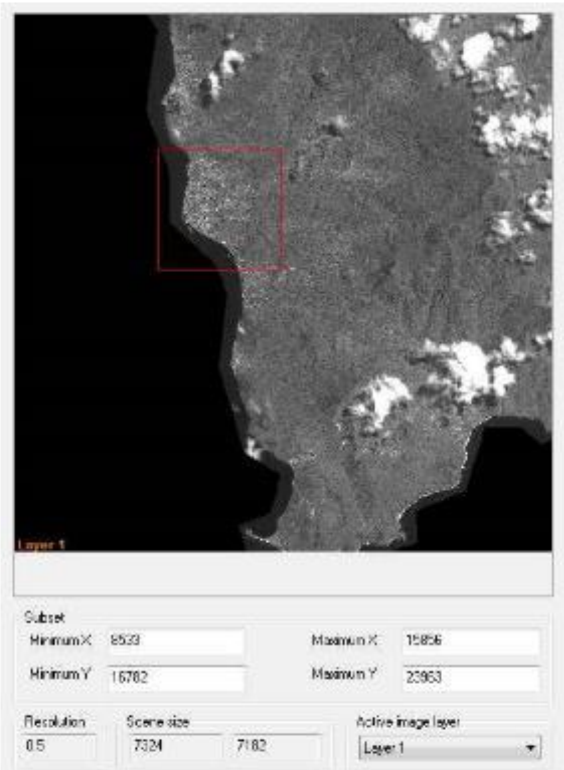


Figure 4-5 Pan-chromatic image of North-West part of Dominica

Figure 4-4 Pansharpened image with 4 layers

Automated object-based analysis

Segmentation is the first step that has to be done before the classification. The settings of each segmentation can vary, and it has to be assign to each project separately. Based on the literature and after conducting several trials, the multi-resolution segmentation and spectral difference segmentation with following settings were chosen as the most optimal:

Segmentation number	1	1a	2	3
Segmentation type	Multispectral segmentation	Spectral difference	Multispectral segmentation of unclassified	Multispectral segmentation of vegetation
Scale	15	12	50	30
Shape	0.9	0.5	0.9	0.9
Compactness	0.9	0.5	0.9	0.9
Weight: Red	1	1	1	1
Weight: Green	1	1	1	1
Weight:Blue	1	1	1	1
Weight: NIR	2	2	2	2

Table 4-1 Segmentation settings

Considering classification, there are many ways of how to classify the land cover from satellite images. In this research a classification as a rule set in eCognition software was developed. It was mainly based on the findings of others e.g.: roads classification by Mhangara et.al. (2014a) and buildings and vegetation extraction by Du (2008) but also included a long trial and error procedure, with adding new rules. Conditions dealing with thresholds on spectral bands, object features like geometry, texture, brightness were applied. In total 9 classes were identified: Vegetation, Water, Shadows, Roads_or_rivers and 5 Buildings classes: Buildings (with dark roofs), Buildings_very_bright (with roofs with Brightness value greater than 1700), Red_buildings (with red roofs), Blue_buildings (with blue roofs) and Gray buildings or ground (for these part of an image that can be recognized as buildings with gray roofs or ground – gray soil). Some explanation of the applied conditions is presented below, while the whole list of the rule set can be found in the attachment nr 1. It was decided to use the top-down approach that is to recognize buildings via gradually detecting other objects and by this assuming that the remaining yet unclassified objects are buildings.

Segmentation 1
In a first step a very basic segmentation process for the purposes of delineating roads were chosen. As first, the multispectral segmentation, with a scale of 100 and with shape and compactness factor 0.9 and the spectral difference segmentation with a scale of 12 were performed. These settings were chosen to be the best to delineate roads (Mhangara, Odindi, Kleyn, & Remas, 2014b).

Classification of roads and rivers, vegetation and water
The Density feature together with Mean NIR values were used to classify roads. Secondly, few classification procedures, using feature object information like Brightness, NDVI or GLDV Ang. 2nd moment (all directions) were performed. The vegetation class was assigned to objects with a high NVDI. Moreover, mean Blue and mean Red values were additionally used for vegetation classification. All objects with NIR values smaller than 80 got a class Water and later on from this class also Shadows were extracted. This all resulted in classification of roads, vegetation and water (see picture below).

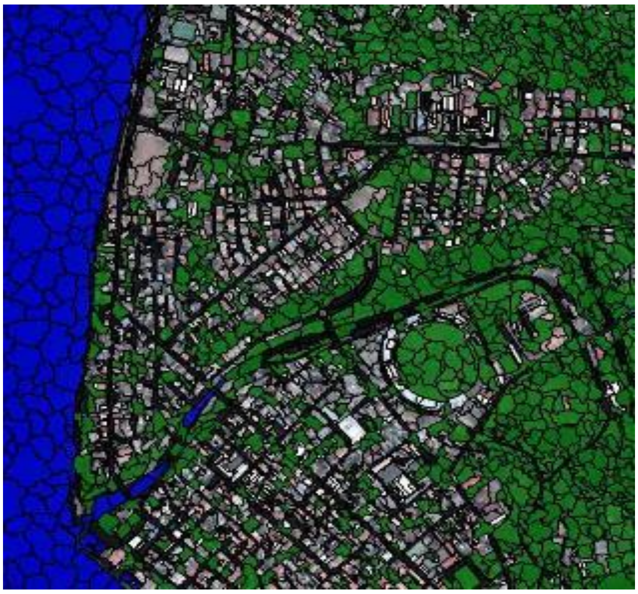


Figure 4-6 First classification results - vegetation, water and roads

Segmentation 2
It has been noticed, that in order to extract each building more precisely, more narrow segmentation process has to be done to obtain smaller segments. With current results, too many single buildings were merged as one object and there was quite a lot of vegetation and roads remaining wrongly classified. Hence, at the beginning of this section, second multispectral segmentation was done, with settings: scale: 50, Shape and Compactness: 0.9. It was done only for the objects that were still unclassified. It resulted in a better object recognition, especially for smaller buildings (see picture below).

Figure 4-7 More narrow segmentation results, smaller segments

Classification of buildings, part 1

After this, to recognize first buildings classes, feature information like Brightness and Mean Values of Blue, Green and Red were used. Additionally also geometric features like Volume, Asymmetry and Rectangular Fit were considered. Red, Gray, Blue and Very Bright buildings were distinguished. Exact list of rules and settings that were used can be found in Appendix 1.

Figure 4-8 First classification of buildings, different roofs

Reclassification

This step was needed to adjust wrongly assigned objects. Grow region and merge region function was used for objects that fall into same class. Most of the buildings were correctly classified; nevertheless smaller buildings surrounded by the vegetation were not recognized.

Segmentation 3

This segmentation was done to help with the latter issue. The Vegetation area was divided into smaller parts using multispectral segmentation with a scale of 30, shape and compactness factors: 0.9. One could distinguish now additional, small buildings.

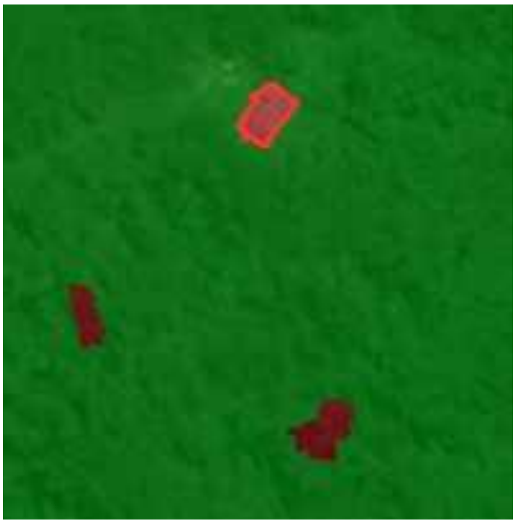


Figure 4-9 Detection of smaller buildings – brown colour

Classification of buildings, part 2

The brightest and bright objects were assigned to Very Bright and Gray buildings, respectively. After performing additional growing region function objects with area bigger than 3000 Pxl and very long objects were assigned back as Vegetation. This all helped to extract small buildings from densely vegetated areas.

Figure 4-10 Final classification of the image with 5 classes of buildings, water, vegetation, shadows and roads or rivers

4.2.2 Accuracy evaluation results – step 2

Area condition
To check the overall statistics which says how much the extracted and known objects are intersecting, a Tabulate Intersection function of ArcGis was calculated. The input for this function was: an original known object layer and a clip of the original and the extracted objects layers. It calculated in percentage how much of the produced in the object-based analysis buildings areas are intersecting with the original buildings areas.

As first the clipping function was performed, which shows how many of the extracted buildings are common with the original Buildings2013 layer.

- *Buildings_final2013* (original buildings layer, obtained from the fieldwork) clip with *Roseau_all* (buildings layer extracted in this research) → *Buildings_final2013_Clip* (common parts of the buildings)

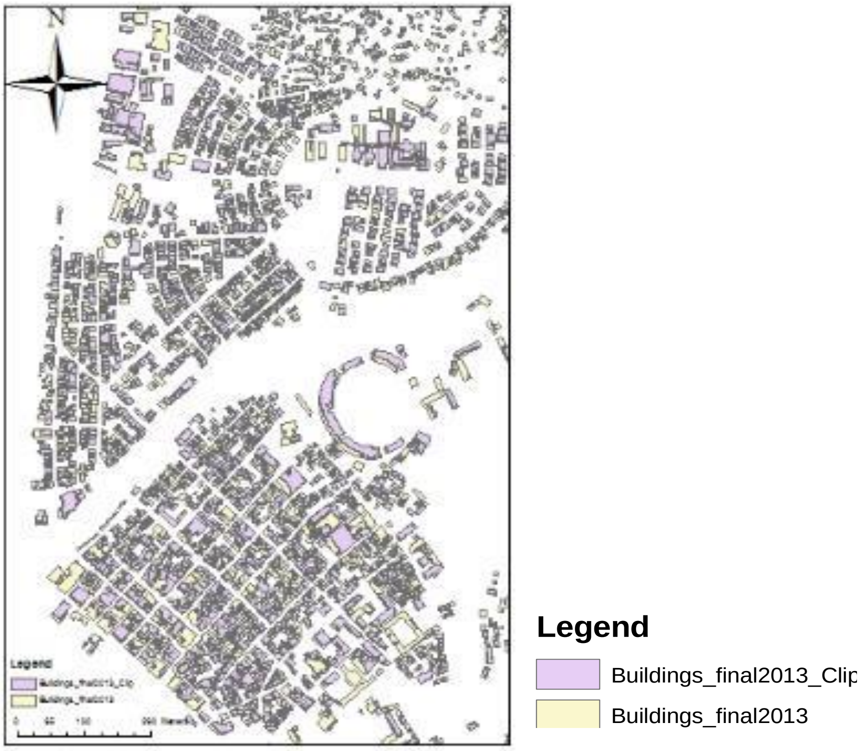


Figure 4-11 The comparison of extracted features and given, original data file obtained during a fieldwork

For both layers additional area field was created and calculated. Later on the Tabulate Intersection was performed, to show the percentage how much of the area of the clipped buildings is common with the original buildings area (layer obtained via fieldwork: Buildings_final2013). Because the Input Zone Features and the Input Class Features are the same dimension (both polygons) the output field records the percentage of the zone feature that is intersected by the class. Here it is a percentage of Buildings_final2013 that is intersected by Buildings_final2013_Clip.

- *Buildings_final2013_Clip* (common parts of the buildings) with *Buildings_final2013* (original buildings layer, obtained from the fieldwork) → *Buildings_final2013_Tabulate2*

The result shows that from 5691 objects of Buildings_final2013 only 4605 are intersecting with the extracted objects. The area which is common for these two layers can be expressed by a box plot (Figure 21). The mean value shows i.e. that over 55% of the study area of extracted buildings is common with the original obtained via fieldwork results.

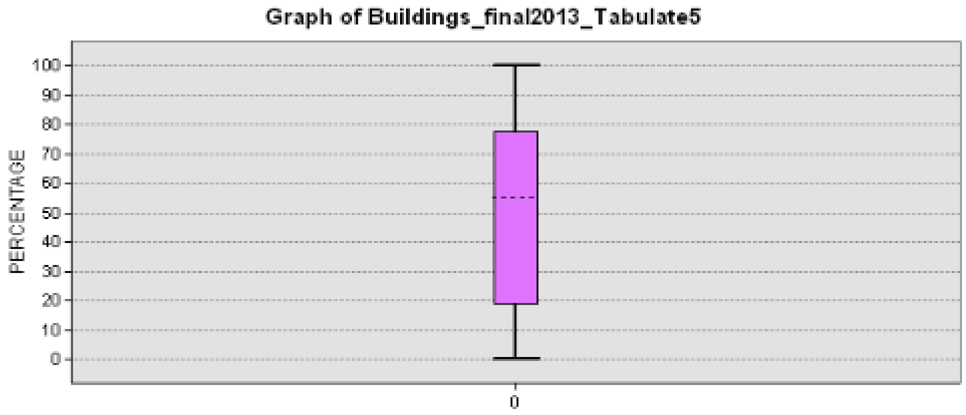


Figure 4-12 The cohesion of extracted features and given fieldwork results, Y axis shows the percentage of overlap between extracted and reference data, measured per building

Shape condition

As the input for this condition the extracted buildings layer was used (*Roseau_all*). The MINIMUM BOUNDING GEOMETRY tool was applied with the optimum bounding geometry type: CONVEX_HULL.

Figure 4-13 Applying Minimum Bounding Geometry tool on Roseau_all layer (layer with extracted buildings)

This resulted in creating rectangles of the smallest width enclosing an input features. New layer called *Roseau_all_Minimum_Bounding_Geometry* was created. Secondly, to the attribute tables of layers: *Roseau_all_Minimum_Bounding_Geometry* and *Buildings_final2013* the Perimeter field was added and its value was calculated using CALCULATE GEOMETRY function. Some of the shape length fields were already available in both tables, but in the Buildings_final2013 layer exactly 470 objects were missing this attribute value. After joining these two tables, the statistic with the percentage difference of the common objects perimeters and lenghts (only for present values so excludng the 470 records) was calculated.

For Perimeter value the equation for each record was performed as below:

(Abs ([Perimeter] - [Perimete_1]) / ([Perimeter] + [Perimete_1])) * 100

Where :

Perimeter – was a perimeter for objects in Buildings_final2013 layer and
Perimete _1 - was a perimeter in Roseau_all_Minimum_Bounding_Geometry layer.

The result was a mean of 31% difference, so 69% of similarity. Similarly for the Length value the equation was:

(Abs ([Shape_Leng] - [MBG_Length]) / ([Shape_Leng] + [MBG_Length])) * 100%

Where :

Shape_Lenght – was a length for objects in Buildings_final2013 layer and
MBG_Length was a length in Roseau_all_Minimum_Bounding_Geometry layer.

This gave the mean of 43% of difference, so 57% of similarity. To sum up, the final accuracy for the shape condition was quite satisfying and reached (69+57)/2 = 63%.

Overall accuracy

The overall accuracy was counted as a sum of the area and the shape conditions. The result was:

(55%+63%)/2 = 59%

4.3 Reflection

An object–based classification aims to overpass disadvantages of pixel-based approach, using spatial and textural information in a way that it can perform the closest to the human seeing image recognition. There are constant improvements of what can be still incorporated in image segmentation (Baatz, 2000), object rule-based detection (Hamedianfar et al., 2014) or accuracy assessment (Zhan et al., 2005). The number of techniques may vary, as well the number and kind of obstacles of each method.

In this research several issues considering both the segmentation technique and the classification that were used can be raised. The satellite data itself was a high quality, containing only a very small percentage of cloud cover. This cloud cover was visible above all mostly only over the forested, not very much inhabited areas, and so was not an impediment in the analysis at all.

For first segmentations parameters, both for multispectral and spectral difference segmentation the settings were chosen to the best delineate roads, following the findings of Mhangara (2014). Different choice of extracting for example first vegetation would lead probably to different final result. Upon a visual inspection, the result of second segmentation seemed to match very well with what could be detected visually.

The most arbitrary part of the analysis was creating a rule-based classification. The fact that rules are created by the interpreter, arbitrarily and on trial an error, hence to easily transferable to other images and areas, is a common critique on eCognition style object based image analysis with rule sets. Several objects features, with different parameters were taken into consideration, as well as the sequence of made operations. The general approach was to classify buildings based on their roofs parameters, as in the work of Du (2008), but at the end, for the purpose of this research only one, merged buildings layer was needed. In general, during performing rules, some reclassification had to be done for wrongly classified objects. Adding an additional parameters helped for example when roads, which were close to the buildings, were wrongly classified as buildings.

The drawback of distinguishing buildings basing on the kind of their roof is that the shadows occurring on the image also have to be taken into account. At the end the buildings areas should be merged in that way that they will show the best the actual buildings shapes, but in fact sometimes the parts of the same roof were considered as two or more independent buildings, whereas two or more different roofs (but having very similar textures and colors) were considered as a one building.

In general one can say, that considering the shape and area condition the result of only more than half of the very well classified buildings is not very satisfying. Comparing to literature based results from other researches (up to 90% and more), this number could be much increased. Low accuracies could be the result of both missing entire buildings in classification, and detecting only parts of buildings, which lead to less overlap although the correct building was detected. That is why the further development of the rule-set, for more decent classification would be very much appreciated and could improve a lot not only this phase of the research, but also following steps and the final result.

5. Phase 2: Using Self Organizing Maps for clustering buildings with various characteristics

5.1 Methodology

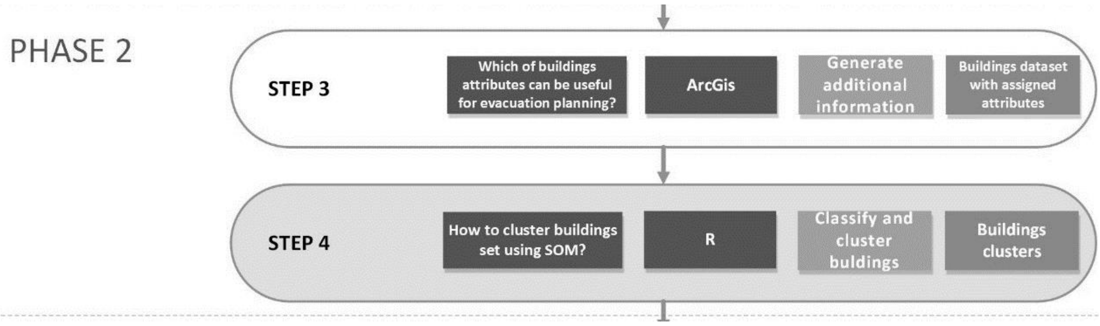


Figure 5-1 Workflow, phase 2

5.1.1 Step 3 – Choosing hurricane vulnerability factors

At this point it should be explained why this research takes into account all buildings that could be extracted in the previous step. The assumption is made that when the hurricane approaches, people that have strong and safe houses will remain at their place and not go to shelters. People can be also afraid that the houses that they will left may be subjected to crime or burglary. They do not want to leave their possessions or animals. Moreover, also the shelter itself is considered as not safe and convenient place. In Dominica, according to the report from last fieldwork of ITC researchers in Dominica, there was even a situation when after the last evacuation warning no people come at all to one of the shelter. It is also believed that this can be forced by the fact that especially major cities, serving as main ports and being located in the low-laying coast areas, attract many poor people seeking an employment and building their poor-constructed houses in a flood-prone regions. Since flooding is one of the major hazard occurring together with a hurricane, these people would have to be firstly allocated to the shelters, which number is only 135 for the whole country *(data for 2015-04-13, source: ITC research database for CHARIM project).* That is why, creating a hypothesis that possibly the majority of people will stay at home, in this research all buildings are treated as a possible people locations, so all of them are taken into consideration in the analysis.

This research focuses on both locational and structural aspects to identify the houses most vulnerability to hurricane damage. It should be decided what makes a structure of the house instable and if some similarly situated houses can be more vulnerable than others. For this purpose, a deep literature review was done, together with obtaining as much as possible additional datasets for Dominica. The previously extracted building layer is enhanced with additional attributes. The final outcome is the buildings dataset with assigned attributes that indicate houses sensitivity to hurricane damage.

It was not possible to include all factors listed in the literature review chapter (both: structural and environmental and locational) in this research mainly due to the lack of the data (for example about the construction quality of buildings, erodability, sea level rise, etc.). Nevertheless, some of the features were present or could be produced from available data, and these were used as an input for further analysis. The list of included factors and the extraction methods are described below.

Structural factors used in this research

The structural factors that are taken into consideration in this study include:

- Building's area – the size factor
- Building's complexity – the shape factor
- Distance to the nearest building – together with density tells about the buildings exposure on the wind
- Buildings density

Buildings area was calculated for each polygon, using ArcGIS AREA tool. The values were assigned to the building points.

Building's complexity was here expressed as ratio: perimeter/area. This is the most common measure for the shape complexity, where higher values express more diverse building shape and low values are close to the simple Euclidean geometry (Institute et al., 2004), (de Smith, Goodchild, & Longley, 2015). The values of the ratio were assigned to each building point.

$$Complexity = \frac{\text{perimeter}}{area}$$

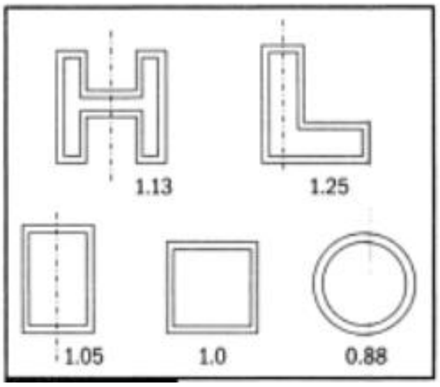


Figure 5-2 Different perimeter to area ratio, source: (Institute, d'Energia, & Programme, 2004)

Distance to the nearest building was calculated using PROXIMITY, near function tool, which determines the Euclidean distance from each feature in the input features to the nearest feature in the neighborhood, within the search radius. The values (the distance to the nearest building, in meters) were assigned to each building point.

Finally, the **Buildings density** was calculated using a POINT DENSITY tool in ArcGIS, with parameter settings: population field: none, a neighborhood of a circular radius 20 m around each cell and square meters area units. The results are values of points per square meter in the neighborhood. From the raster that was produced, the value for each building point was assigned (EXTRACTION TOOL).

Environmental and locational factors used in this research

In this work from three main groups of factors: for wind, for flooding and for landslides are distinct. These locational and environmental factors that are taken into consideration in this phase include:

- Slope and Elevation
- Distance to the coastline and rivers
- Wind hazard data

Slope and Elevation

In the Caribbean area, besides the obvious climatic risk that occurs, topographic features are of high importance. The volcanic spine of Dominica causes a substantial wind acceleration. Moreover, hilly land of Dominica with poor soils in the valleys and floodplains is prone to dangerous power of hurricanes (CHARIM, 2015).

For each building point, the values of already existing data (from CHARIM researchers) of slope and elevation were calculated. This was done using the spatial analyses tool EXTRACTION. The extraction tool, extracts values by points, assigning slope and elevation values for each building point. The bilinear interpolation at the point's location was done, using adjacent cells with valid values. NoData values were ignored in the interpolation unless all adjacent cells were NoData.

Distance to the coastline and rivers

The most hurricane vulnerable areas for islands like Dominica seem to be located along the coastlines. Strong winds, blowing from the ocean are hitting the coast first. Later on, the water is moved into the land, using the existing fluvial discharge systems, that is rivers and channels. This can cause flooding of the area. Both, the distance from the coastline and from the rivers can be calculated in ArcGIS. The river layer for Dominica was present, so it was only the matter of determining the distances that will be later on considered as significant when assessing the vulnerability.

The classification of the closest areas that will be affected by the flood was based on the research of Fernández and Lutz (2010) where intervals used were: <100 m, between 100 and 500 m, between 500 and 1000 m, and >1000 m and Paquette and Lowry (2012) where distance intervals were: <100 m, between 100 and 200 m, between 200 and 1000 m and >1000 m. For calculation the distance values Euclidean Distance tool of ArcGIS was used. The classification was manual, setting the breaks in 100, 200, 500, 1000 and >1000m. Later on, the raster values were assigned to each building, using the same EXTRACTION tool as for slope and elevation values.

Wind hazard data

It should be explained at first, why the wind hazard map was used at this stage. Although wind hazard data is already a result of some analyzes and not a raw data itself, it was chosen to be used, since the wind susceptibility is one of the crucial factor in a hurricane vulnerability assessment and no other wind-correlated data were available. This data is obtained from CHARIM project, thanks to ITC researchers. For 5 classes: very low, low, moderate, high, very high wind hazard five values of 0, 1, 2, 3, 4 were assigned into a raster. Like with previous layers, a single value was assigned for each point that represents a building.

There is also some additional data available about the hazard susceptibility of Dominica, which was not included in the SOM analysis. Is was decided to use this data later on to check if there will be any correlation with the results of the SOM. This additional data include:

- Flood hazard map
- Landslide susceptibility map

Flood hazard data is obtained from CHARIM project, thanks to ITC researchers. It is assumed that probably it was calculated using several risk factors like: distance to the coastline, distance to the rivers, elevation, etc. For 5 classes: no flood, low, moderate, high and very high flood hazard five values of 1, 2, 3, 4, 5 were assigned in a raster. The landslide susceptibility data was obtained from CHARIM project, shared on the CHARIM website on 8th of March 2015. For three values: low, moderate and high, accordingly values of 1, 2 and 3 are present in a raster.

5.1.2 Step 4 – Creating SOM's, clustering, interpretation and validation

The fourth step is dealing with the issue how to define buildings pattern and how to cluster buildings using Self Organizing Maps. For this, amongst others literature about SOMs by Kohonen (1990 and 2001), Vesanto (1999), Vesanto & Alhoniemi, (2000) and Augustijn and Zurita-Milla (2013) was reviewed. The algorithm of SOM was run according to the structure as presented in a Literature review chapter 2, section 2.2. The whole procedure was done in R (a software environment for statistical computing and graphics).

In this step, using Self Organizing Maps helps with mining, clustering and visualizing buildings data which is high dimensional (have many assigned attributes) by projecting it into a 2D space. After this, it is easier to see the possible pattern. The result is satisfying when the clustering will be dependent not only on location, but also on generated attributes and other previously obtained information. At the end the result can be compared with already existing hazard maps.

The procedure for this step was done in few points:

- Training of the SOM lattice based on an input dataset
- Mapping back of the data (either the training dataset or a subset of this dataset) onto the trained SOM
- Secondary clustering
- Interpretation of results, assigning vulnerability classes and comparison with existing hazard maps

Training of the SOM lattice based on an input dataset

In R environment, after loading two libraries: 'kohonen' and 'lattice' one can read an input file, which is in this case a csv. file with building's attributes data set. Next, several parameters had to be established. First, it is an input set – telling which columns and rows have to be chosen from the file. Neurons will be plotted on the two dimensional grid, which size also has to be chosen. In this research the reasoning for choosing certain value was as follows: When analyzing the results it will be stated that each variable had a big, medium or small value (the size of the fan), so in general, one can say that there will be 3^4 possibilities of results for 4 factors. That is why a lattice 9*9 was chosen to be the grid for neurons, because it will represent 3^4 permutations. The number of iterations, which tells how many times the complete data set was presented to a network was chosen according to a "rule-of-thumb" which is "*to use the number of cluster items or 500 times the number of nodes, whichever is greater (IOS, 2015)"* (Bullinaria, 2004). The map size is 9*9, so the number of iterations was chosen to be 9*9*500 = 40500.

SOM generates presents a combinations of given factors to the lattice (a hexagonal grid) where resulted values can be visualized for example as a fan diagram (when the number of input variables is low). On the diagram one circle represents one neuron. The size of each fan inside the circle represents the magnitude of each variable in the weight vector. The corner's locations represent the most unique results.

Mapping back of the data onto the trained SOM

The mapping back function allows to project the initial dataset onto the SOM lattice. The plot that can be produced is a "codebook" for vectors that are assigned to a neurons in the lattice. One can see how many of them were assigned to each neuron. The results can be printed in a csv. file, which will consists of columns like: the vector number (without a name for the column), unit.classif (which is the number of neuron to which the vector was assigned) and some additional columns: distances, whatmap, weights and scale.distances. The most important here is the unit.classif column.

Secondary clustering

Next, the secondary clustering procedure can be done. As it is known from the literature, SOM allows only for a qualitative and visual distinction of data characteristics, but in the same time this information is useful for assumption of the quantitative description of data properties and possible number of clusters (Vesanto & Alhoniemi, 2000). In this case, the number of clusters should correspond with classes that will be in the future assigned to neurons that represent identified buildings vulnerability types. These parameters have to be set independently for different kinds of loaded data. When this number is already approximated, the secondary clustering on the produced prototypes can be performed. It is proven that this kind of two-stage approach can significantly reduce a computation time (Palamara et al., 2011). There are many kinds of possible clustering, but here the agglomerative hierarchical clustering method was chosen. This clustering assigns first neurons to their clusters, then calculates the distance between clusters and later on, iteratively, joins the most similar ones. Results were presented on a colored plot with adding cluster boundaries.

When training the SOMs, performing secondary clustering on the trained neurons and mapping the complete data set, one can be sure that all existing patterns are represented. Thanks to a printing option in R one can find which buildings are assigned to a certain neuron and which neurons are falling into certain classes. Using a simple VLOOKUP function in Excel, user can quickly assign back also the number of cluster to each vector (here: building).

Interpretation and validation

The whole procedure of training, mapping back and clustering was performed accordingly for three datasets: only environmental and locational factors and only structural factors. All results were written to csv. files. After the interpretation of produced clusters, all csv. files were loaded into ArcGis where the visual validation of produced clusters, assigning a vulnerability class and comparison with existing hazard maps was done.

5.2 Results

5.2.1 Assigning buildings attributes results – step 3

The next step, after extraction of buildings is to assign to them attributes that will serve for the further analysis. Attributes were taken firstly from locational and environmental factors and secondly from structural factors. For environmental factors, raster values were assigned to every building. The produced maps are presented in Figure 5-3. For the structural factor of density, also its raster values were assigned to each building, while for factors like: area, near distance and complexity numeric values were calculated at once and written into the csv. file.

Incomplete records were removed. Some rows had to be deleted, because of an incorrect "HouseID". This was caused by the fact, that eCognition for example recognizes objects like ships (that were moored next to the coast) as well as buildings.

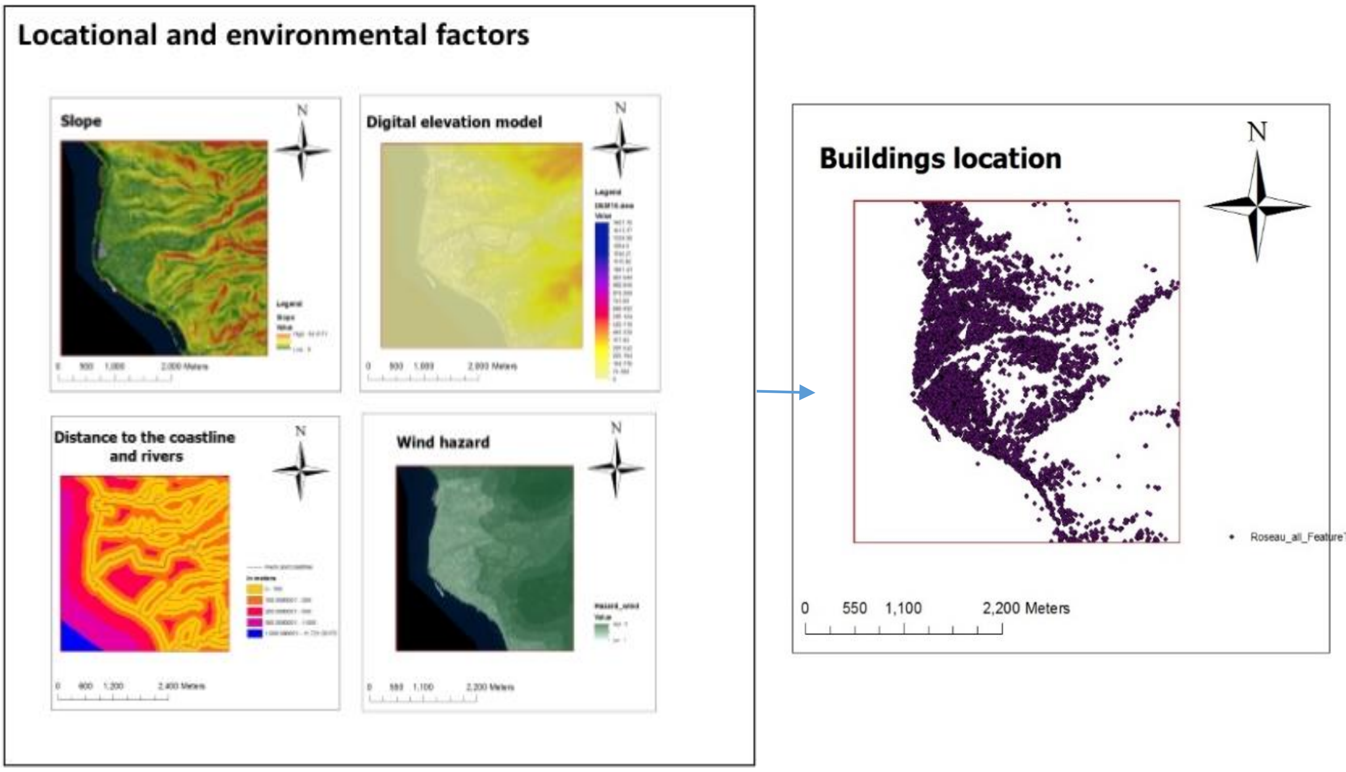


Figure 5-3 Location and environmental factors - assigning to buildings points.

5.2.2 Creating SOM's, clustering, interpretation and validation – step 4

Locational and environmental factors

As first, only environmental factors were taken into account in the SOM, so in this case it was run for first 4 columns in csv. file. The trained SOM lattice is shown in Figure 5-4. Neurons at the top left corner of the training lattice represent buildings with big values of distance to the coast and rivers, small or very small values of slope and very small other values. The more to the right, the more the wind hazard value is visible. The most right and upper corner has the biggest values of DEM. Going down, more impact of the Slope value is visible. By a visual identification one can distinct at least 9 very much visible classes:

* All values small
* Distance to the coastline and rivers values big, other small
* Distance to the coastline and rivers values medium, other small
* Wind hazard values medium, other small
* Wind hazard and DEM values big, other medium
* Wind hazard and slope values medium, other small
* All values medium
* Slope values big, other medium
* Slope values big, other small
* Distance to coast and rivers and slope medium, other small.

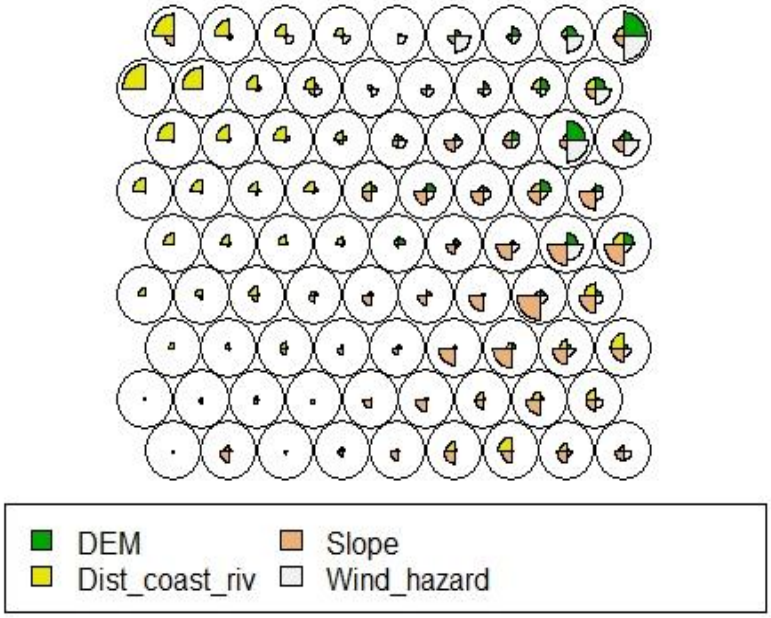


Figure 5-5 Plotting SOM for environmental variables, the size of the fan represents the magnitude of each variable in the weight vector

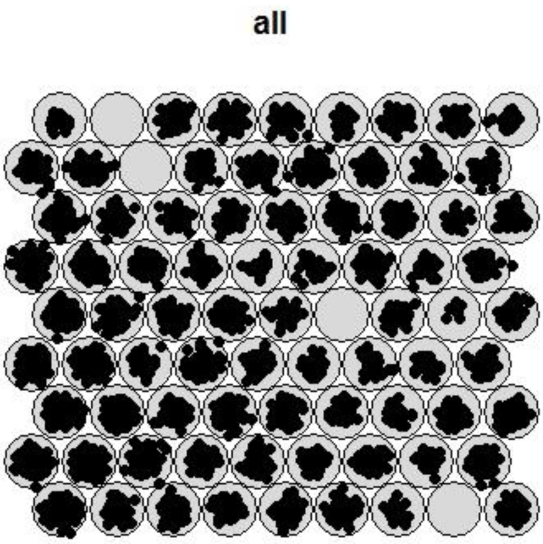


Figure 5-4 Mapping back environmental variables, each dot represents one vector

When mapping back results into space, one can perform a secondary clustering, to distinct automatically different classes. The result of this clustering, which was set here to 9 classes is shown below in Figure 5-6. One can see that the clustering performed by the algorithm is very similar to the previous visual distinction.

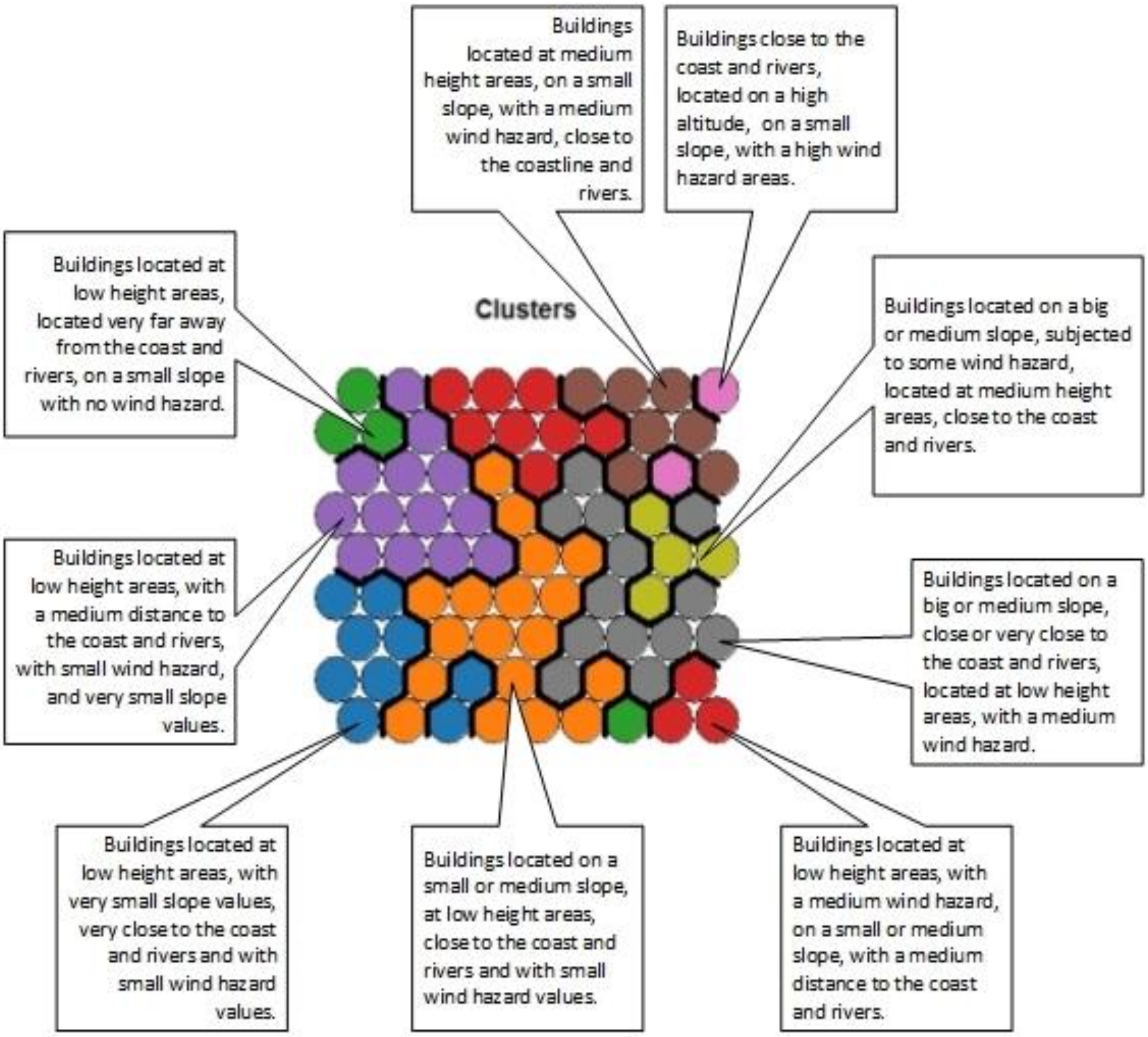


Figure 5-6 Results of hierarchical clustering for environmental and locational factors

When printing all 5988 vectors and assigned to them neuron's values, together with the number of class to which they are falling, the result can be loaded to ArcGIS environment and compared with a present, ready hazard maps.

Structural factors

The SOM was run also for structural factors. These were: area, distance to the nearest building, density value and complexity value, rows from 5:8 in the same csv. file. The input was set like in the previous computation, to 1:5988. The size of the grid and the clustering was also set with the same values. The resulting plot is shown below together with the mapping back plot, showing how many of vectors "fall" into each neuron.

For structural factors, using visual identification one can distinct classes like:

* All values small
* Only Near_dist values big, the rest small
* Compactness values medium, the rest small
* Compactness values big, area values big

* Compactness values and density values medium, the rest small
* Density values big, the rest small
* Compactness and area values medium, the rest small
* Density values medium, the rest small
* Only Near_dist value medium, the rest small

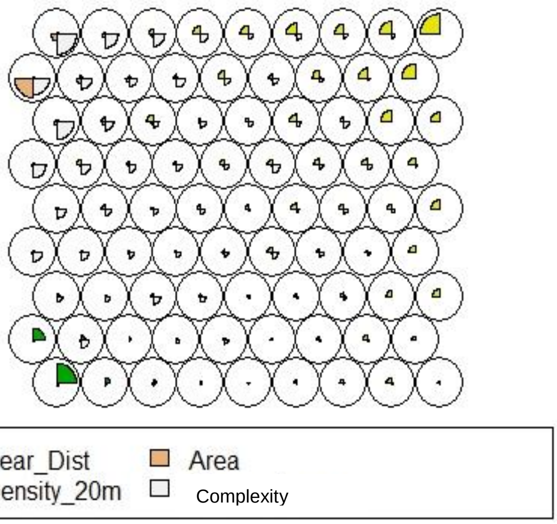


Figure 5-7 Plotting results of SOM for structural factors, the size of the fan represents the magnitude of each variable in the weight vector

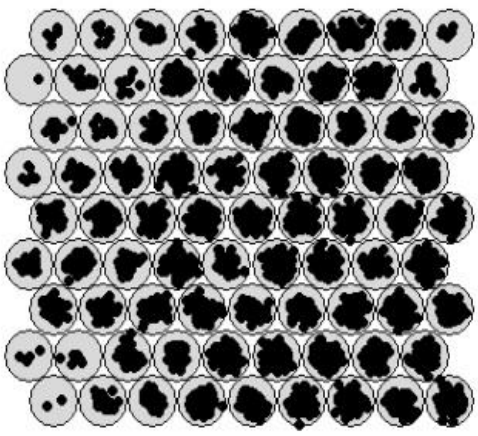


Figure 5-8 Mapping back of structural factors, each dot represents one vector

For structural factors also the hierarchical clustering was used, set similarly at separate 9 classes. The result of this clustering is shown below:

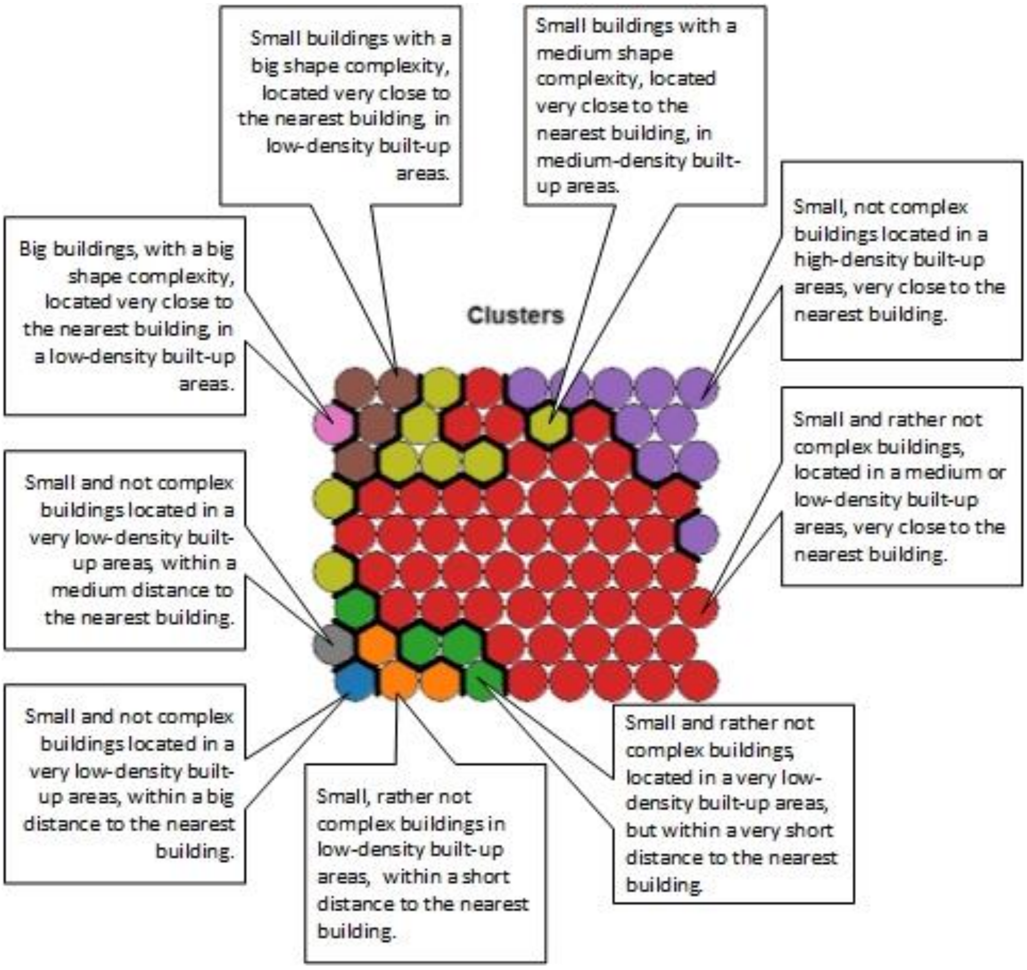


Figure 5-9 Results of hierarchical clustering for structural variables

At the end, the results of mapping and clustering were assigned, like for the environmental and locational factors, to the csv. file.

Interpretation and validation

The interpretation was performed in ArcGIS, where the results were visualized on the previously extracted buildings layer. The further validation procedure was based on a visual comparison of generated clusters with the existing hazard maps. For this, deep analysis of obtained clusters was needed. This was done by describing each cluster according to its vulnerability. As first, for clusters with environmental and locational factors were assessed according to their vulnerability to: wind, flood and landslide hazard, using three assumptions:

- The bigger the slope and the higher DEM, the bigger landslide vulnerability.
- The bigger the wind hazard value, the bigger the wind vulnerability.
- The smaller the distance to the coastline and rivers (and then also smaller DEM) the bigger the flood vulnerability.

The overall vulnerability was calculated using a simple formula of assigning to each factor a level of vulnerability and then calculating the overall result. Below an example is made, just to show the method.

ENVIRONMENTAL FACTOR	VULNERABILITY		
	SMALL (1)	MEDIUM (2)	HIGH (3)
Dist_to_coastline _and_rivers (+ influence of DEM)	x		
Slope (+ influence of DEM)	x		
Wind hazard		x	
	OVERALL VULNERABILITY: 1+1+2=4		

Table 5-1 Example vulnerability calculation; environmental factors

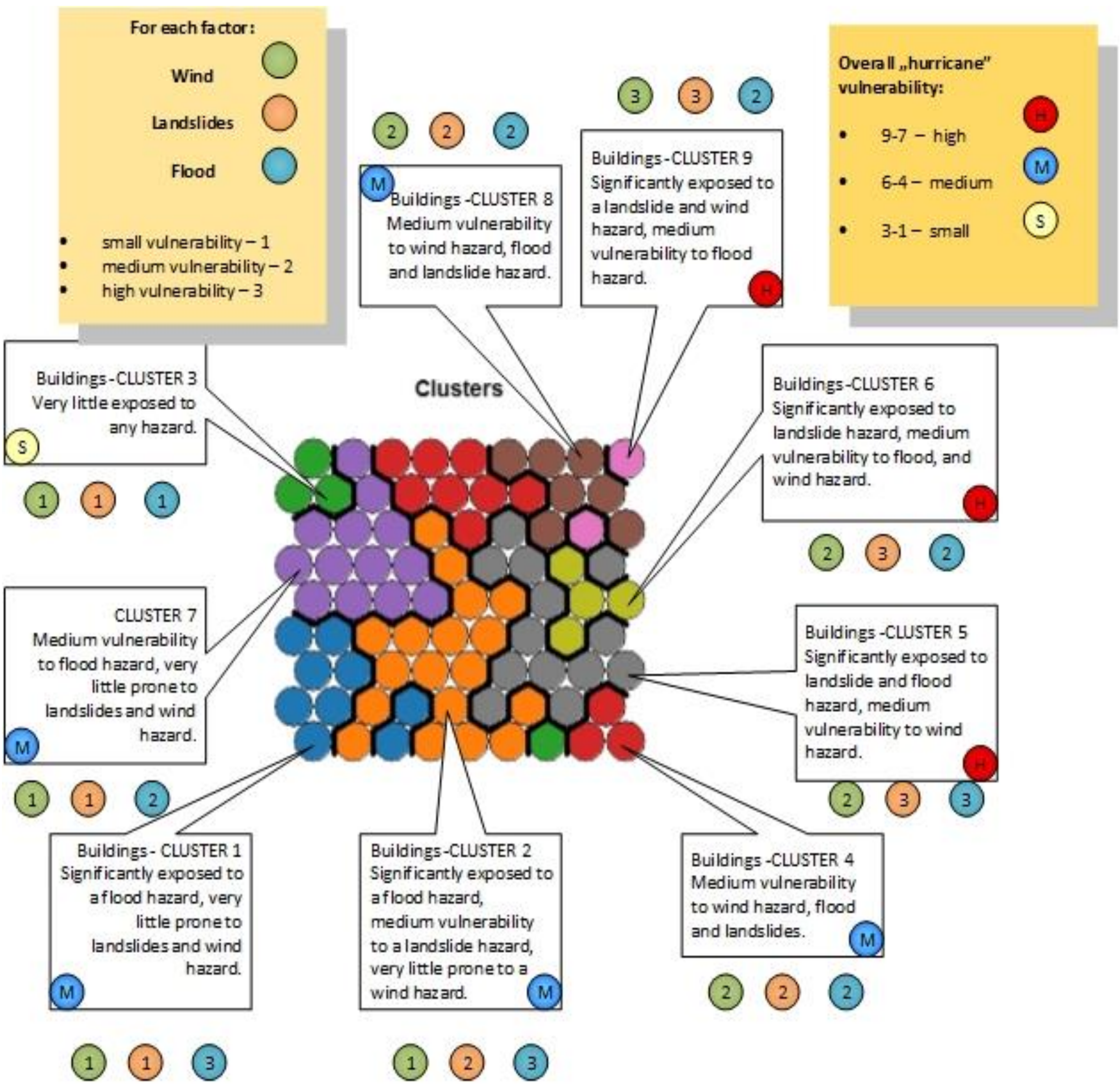


Figure 5-10 Clusters interpretation - environmental and locational factors

For the validation of obtained results clusters are mapped on top of the three (already existing) hazard maps. Separately for each hazard, clusters with three vulnerability measures (for each factor): high, medium and low are presented. By a visual comparison it is possible to see if created clusters are indicating the same or at least similar values for areas as they are presented in ready raster maps.

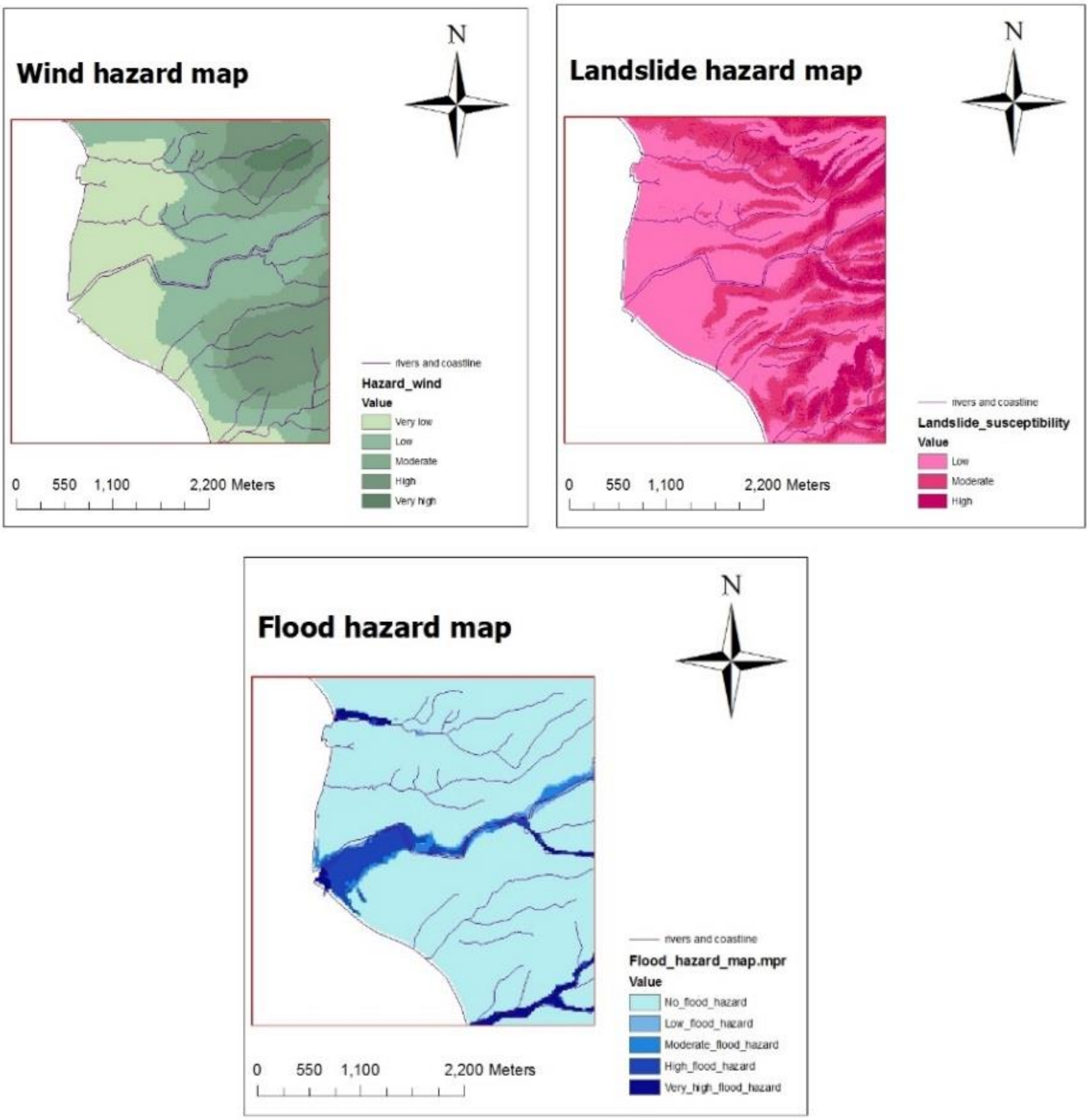


Figure 5-11 Already existing hazard maps, with their legends containing the vulnerability values per hazard (obtained thanks to CHARIM project)

Mapping on top of the existing map: WIND

High vulnerability clusters (3) Medium vulnerability clusters (2) Small vulnerability clusters (1

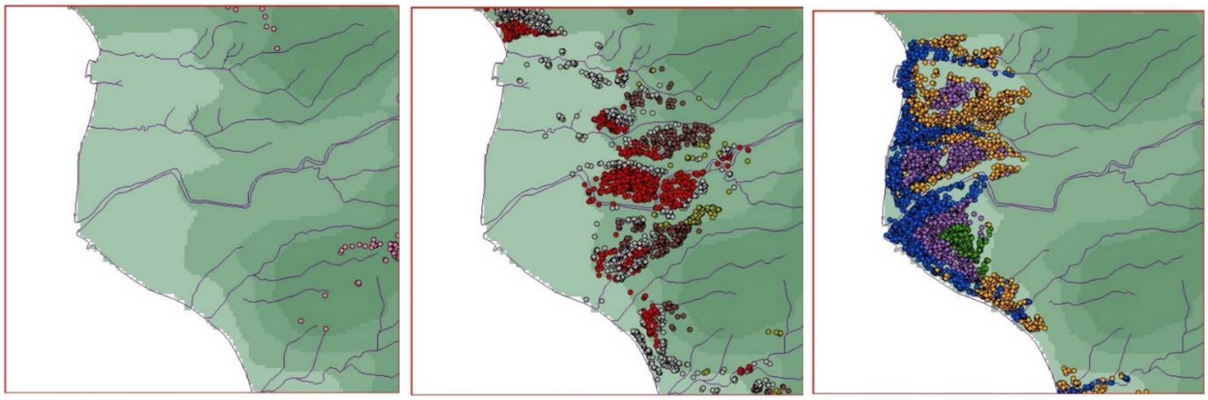


Figure 5-12 Mapping vulnerability lusters for environmental factors on top of the existing map of wind hazard, high vulnerability clusters, medium vulnerability clusters and small vulnerability clusters

From this mapping one can see that high and low vulnerability values are correctly recognized by the clustering. For medium values, most of them are wrongly identified as very little or little prone to the wind hazard.

Mapping on top of the existing map: LANDSLIDE

High vulnerability clusters (3) Medium vulnerability clusters (2) Small vulnerability clusters (1)

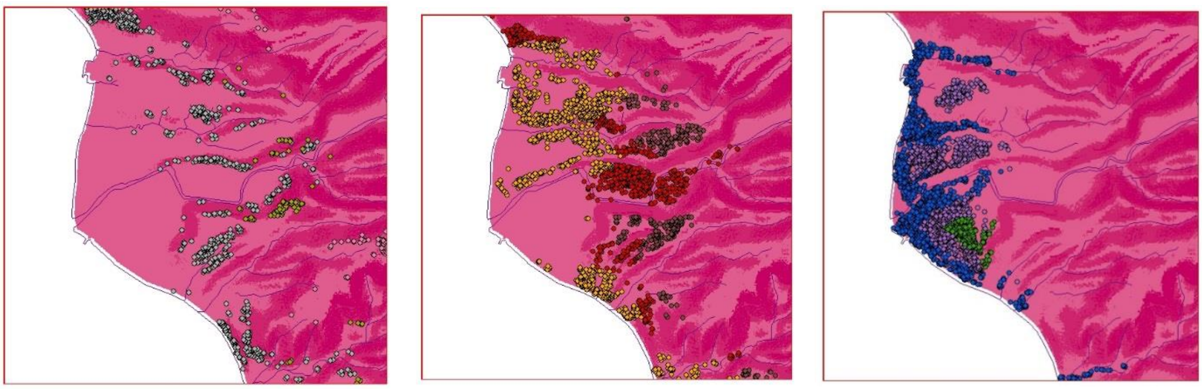


Figure 5-13 Mapping vulnerability clusters for environmental factors on top of the existing map of landslide hazard, high vulnerability clusters, medium vulnerability clusters and small vulnerability clusters

Thanks to this mapping one can see that especially high and low vulnerability values were correctly recognized by the clustering. The only lack of conformity is that medium vulnerability clusters are presented in the raster as very little prone to landslide hazard.

Mapping on top of the existing map: FLOOD

High vulnerability clusters (3)   Medium vulnerability clusters (2)   Small vulnerability clusters (1)

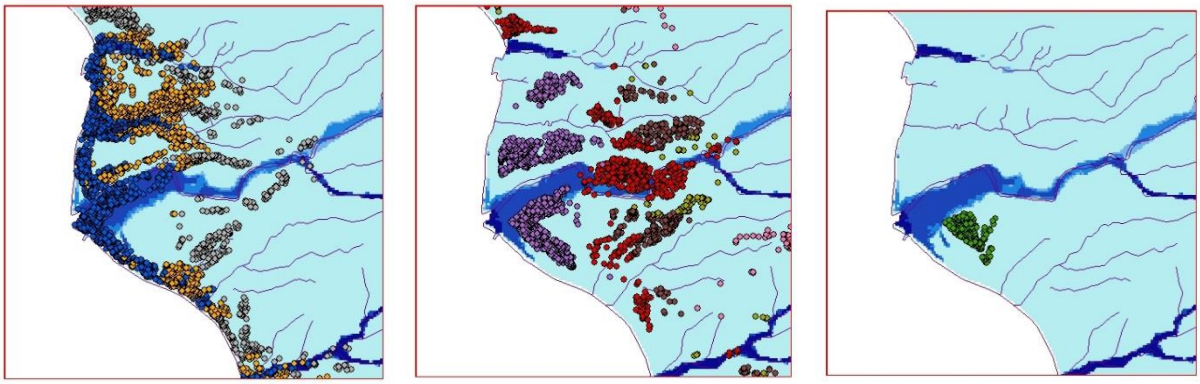


Figure 5-14 Mapping vulnerability clusters for environmental factors on top of the existing map of flood hazard, high vulnerability clusters, medium vulnerability clusters and small vulnerability clusters

When mapping on the top of the flood raster, one can see that not all areas indicated by the clustering (high vulnerability areas) are considered prone to flood hazard in the raster map. There are actually very big areas where there is no flood hazard at all in a raster (bright blue color in a raster and gray color for cluster). Medium vulnerability is also mapped in different places than the raster predicts. Only clusters with a low vulnerability seems to be well presented.

Similar interpretation was done for structural variables. The following assumptions were based on the vulnerability to the wind hazard, which is considered the biggest power during a hurricane together with flooding and to which the structure of the building is the most prone.

- The bigger the complexity of the building, the more it is vulnerable.
- The bigger the area of the building, the more it is vulnerable.
- The smaller the distance to the nearest building the less it is vulnerable.
- The bigger the density of the built-up area, the less it is vulnerable.

Also in this case, an overall vulnerability was calculated by formula of assigning to each factor a level of vulnerability and then calculating the overall result. Below once more an example is made, just to show the method.

STRUCTURAL FACTOR	VULNERABILITY		
	SMALL (0.75)	MEDIUM (1.5)	HIGH (2.25)
Complexity	X		
Size		X	
Density			X
Distance to the closest building		X	
	OVERALL VULNERABILITY: 0.75+1.5+2.25+1.5 = 6		

Table 5-2 Example of overall vulnerability calculation; structural factors

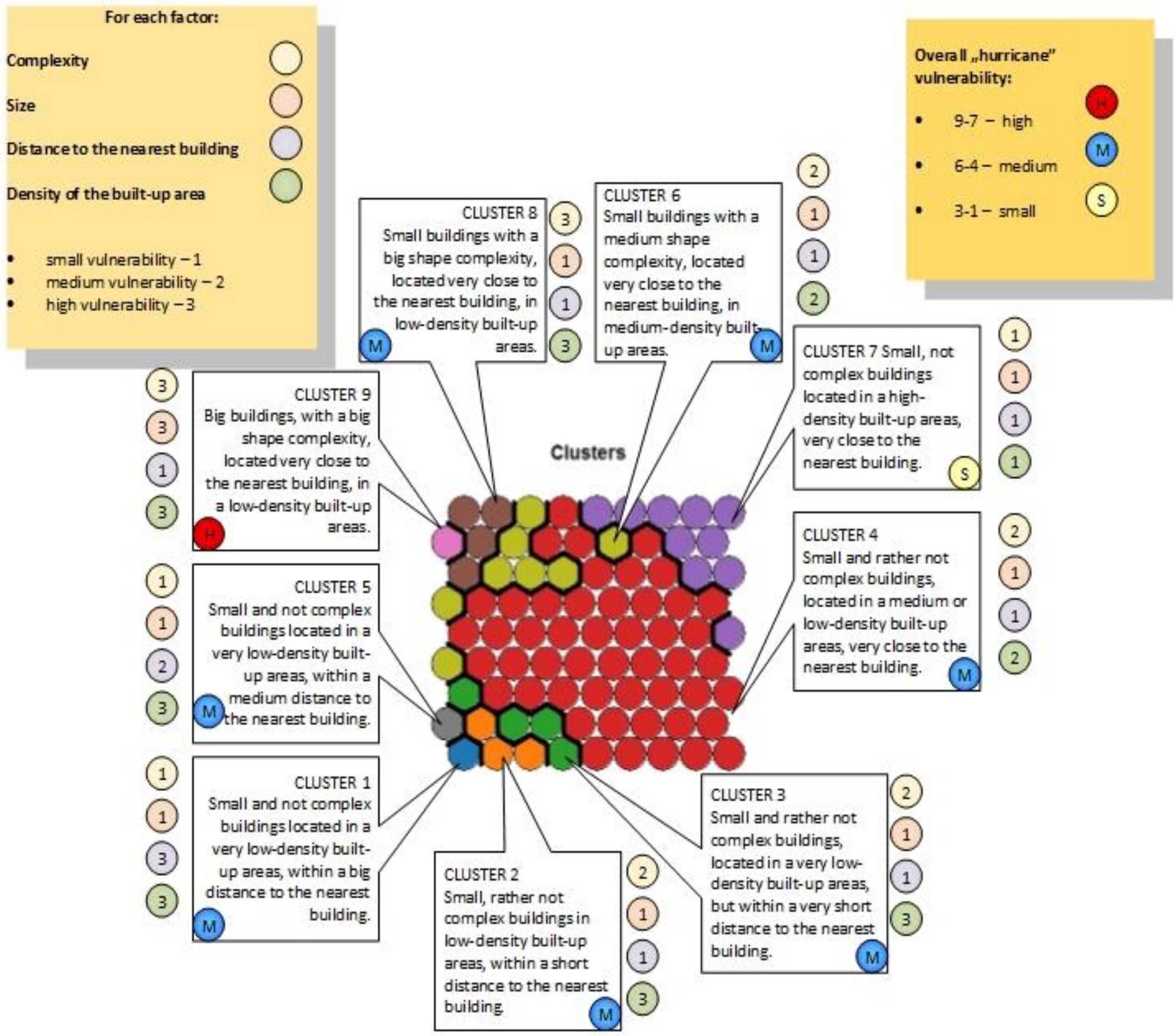


Figure 5-15 Clusters interpretation - structural factors

Here the validation of results was not really possible, since there are no existing maps that contain any information about the real buildings structure. Nevertheless, some observations can be made. When looking for a possible connections, clusters can be also mapped on top of the three (already existing) hazard maps. This was done separately for each hazard, for clusters with three overall vulnerability measures: high, medium and low are presented. By a visual comparison it is possible to see if the created clusters are indicating the same or at least similar values for areas as they are presented in ready raster maps.

Mapping on top of the existing map: WIND

High vulnerability clusters (3)   Medium vulnerability clusters (2)   Small vulnerability clusters (1)

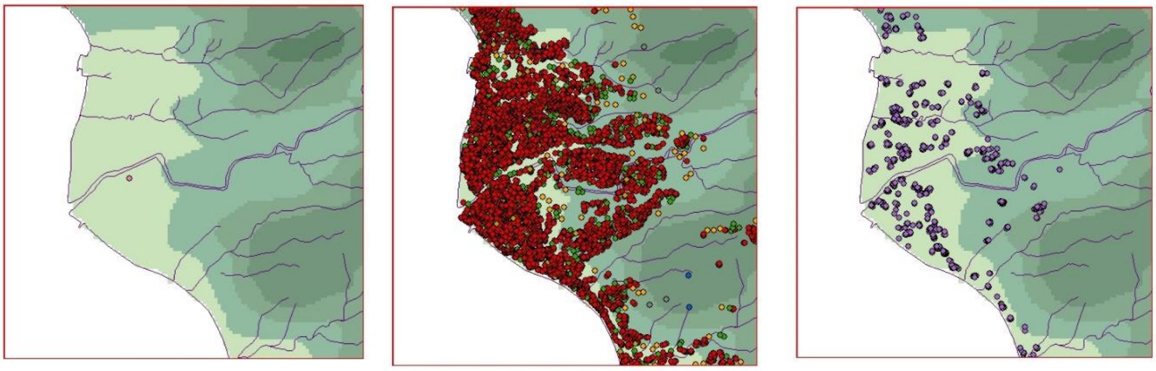


Figure 5-16 Mapping vulnerability clusters for structural factors on top of the existing map of wind hazard; high vulnerability clusters, medium vulnerability clusters and small vulnerability clusters

It seems that there is quite weak correlation between the vulnerability of the house structure and the wind hazard risk indicated in a raster. High vulnerability class (according to the classification here it is only one building) is actually placed in area that is very little prone to a wind hazard. Medium vulnerability classes have very random compatibility with a raster values. The biggest compliance can be seen in case of small vulnerability cluster, which are mostly placed in low risk wind hazard areas.

Mapping on top of the existing map: LANDSLIDE

High vulnerability clusters   Medium vulnerability clusters (2)   Small vulnerability clusters (1)

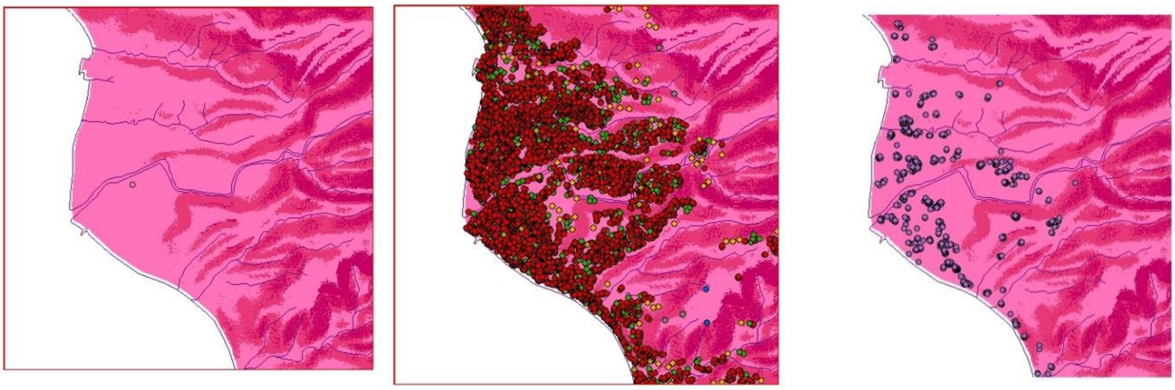


Figure 5-17 Mapping vulnerability clusters for structural factors on top of the existing map of landslide hazard; high vulnerability clusters, medium vulnerability clusters and small vulnerability clusters

From this mapping one can see that there is no correlation between high vulnerability clusters (once more it is only this one building) and high landslide risk. It is observed also that medium vulnerability clusters are mostly falling in the small risk raster values. The best mapped class seems to be the one with small vulnerability clusters, which are mostly located at the low landslide risk areas.

Mapping on top of the existing map: FLOOD

High vulnerability clusters (3) Medium vulnerability clusters (2) Small vulnerability clusters (1)

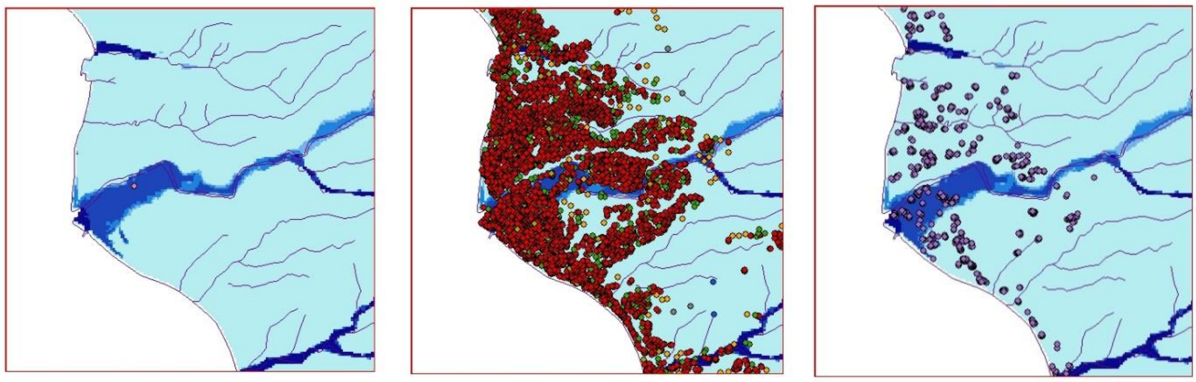


Figure 5-18 Mapping vulnerability clusters for structural factors on top of the existing map of flood hazard; high vulnerability clusters, medium vulnerability clusters and small vulnerability clusters

When mapping on top of the flood risk raster, one can clearly see that there is no correlation between the structural vulnerability of buildings and flood risk level presented by the raster. Many cluster with small vulnerability are located in the very much flood prone areas (e.g. in the delta of the river), while medium vulnerability clusters seem to be randomly correlated with the risk raster values. Only high vulnerability cluster is located in the high risk zone.

Quantitative results are presented in a table. The table 5-4 shows the percentage of the compliance for environmental and structural cluster values and three raster hazard maps. For environmental variables calculation is done separately, first classifying clusters vulnerability as for each risk factor and secondly taking into consideration an overall vulnerability of each cluster (as presented in Figure 5-10 Clusters interpretation - environmental and locational factors).

Although the average compliance for each kind of variable is not so high (from 42.3% to 46.4%) there are very interesting results for certain vulnerability classes. The highest compliance is seen for the low hazard risk areas (for all three risks) and clusters with small vulnerability, irrespective of the kind of variables (from 81.4% to 100%). For example buildings with low vulnerability to the wind are in 100% located in the low wind risk area indicated by a hazard raster. Very interesting is also a high result for structural variables, where low vulnerable clusters are also mostly all located in low risk hazard areas.

Location of clusters with medium vulnerability have a little correlation to the medium risk hazard areas indicated by a raster maps. The best result for the medium vulnerability is seen for landslide hazard, from 9% to 10.6%, which is still very small compliance.

In case of clusters with high vulnerability the highest compliance is seen for wind hazard and environmental variables - 89.8% (with cluster classification as per each factor) and for flood hazard and structural variables -100%. For other cases of high risk, the location of clusters does not correspond with the raster value (very small compliance from 13% to 0%).

The highest average compliance per hazard is seen for wind risk and environmental variables (with cluster classification as for overall vulnerability) – 72.1% and for flood risk and structural variables – 61%.

Assuming that all rasterized hazard risk maps are correct and well classified, one can summarize that proposed SOM and clustering approach find the best buildings with small vulnerability, located within low risk areas.

Compliance		Clusters values (cluster number)											
		ENVIRONMENTAL VARIABLES Cluster numbers - assessment of the vulnerability for each factor separately				ENVIRONMENTAL VARIABLES Cluster numbers - overall vulnerability assessment				STRUCTURAL VARIABLES Cluster numbers – overall vulnerability assessment			
		Low (various per risk)	Medium (various per risk)	High (various per risk)	AVG = AVERAGE	Low (3)	Medium (1,2,4,7,8)	High (5,6,9)	AVG = AVERAGE	Low (7)	Medium (1,2,3,4,5,6,8)	High (9)	AVG = AVERAGE
Hazard maps risk values (exact raster value)	WIND low (1,2)	100%				100%				97.1%			
	WIND medium (3)		8.3%				3.5%				3%		
	WIND high (4,5)			89.8%	AVG= 66%			4.6%	AVG=72.1%			0%	AVG=33.7%
	LANDSLIDE low (1)	90.2%				86%				86.4%			
	LANDSLIDE medium (2)		10.4%				9%				10.6%		
	LANDSLIDE high (3)			0.2%	AVG=33.6%			0.2%	AVG=31.7%			0%	AVG=32.3%
	FLOOD low (0,1)	92.1%				92%				81.4%			
	FLOOD medium (2)		2.5%				2%				1.6%		
	FLOOD high (3,4)			13%	AVG=35.9%			12%	AVG=35.3%			100%	AVG=61%
OVERALL AVERAGE COMPLIANCE PER KIND OF VARIABLES					44.8%				46.4%				42.3%

Table 5-3 Validation of clustering results, matrix of compliance

5.3 Reflection

There are many points in this section of the research that require further discussion. First of all, the step of choosing vulnerability factors should be discussed.

According to the literature the construction quality and proper details fastening is the most crucial factor of the house vulnerability to the hurricane risk. But since this kind of data was not available for this research (it couldn't be easily assessed from satellite images) other house characteristics had to be chosen. These were building's size, its complexity, distance to the nearest building and buildings density. Additionally, it could be very much useful to recognize for example the kind of the roof material. Knowing development practices on an island, this may possible help with an indication of the whole roof construction quality. For buildings density factor, trying out of different parameters could be also of improvement. In general, it would be advisable to include more factors saying about the vulnerability of the building structure to the analysis.

For environmental and locational factors: slope, elevation, distance to the coastline and rivers and wind hazard data was chosen. Excluding distance to the coastline map, other maps were obtained from the government dataset. The classification performed on the distance map was done manually, so setting certain intervals, even when based on the literature, may be subjected to discussion. The most subjective was to include a ready wind hazard map values as a separate factor. It would be advisable not to use it if any other wind hazard connected data would be available (like for example: wind direction and storm strength). In general, also for environmental factors it is advisable to look for more data that could be possible helpful to better assess the vulnerability of the area to the hurricane hazard.

When calculating an overall vulnerability, the impact of each hazard factors (from both groups of environmental and structural one) was consider to be equal, which is probably not correct. Nevertheless, that kind of assumption has to be made because no other proven and reviewed solution for assigning weights for these available factors was found in the literature. In the future, more research has to be conducted in order to develop more accurate weighting for all factors.

When reflecting on next step one can discuss all of the following procedures: Creating SOM's, clustering, interpretation and validation.

When training SOM, several parameters had to be established. One of it was the number of permutations, chosen to be 40500. This was done according to the "rule-of-thumb" based on the literature, however, it would be recommended to research if this is the most optimal setting for that kind of data. For performing secondary clustering an agglomerative, hierarchical clustering was chosen, although it would be also recommended to try-out with other clustering techniques, like e.g. k-means clustering. The number of cluster which was decided to set could also have had an impact on the final result of classification.

Considering validation, the quality of the raster maps which were used may be also subjected to the discussion. Especially flood hazard risk map does not seem to be really reliable (having no flood risk in the areas next to some rivers). It has been found, that there is existing also an additional flood hazard map, not in the CHARIM project website, but in the Physical Planning Division website (Dominica, 2015a). Both maps are very different, with very little overlapping vulnerable areas. It seems that mapping on top of the map from Physical Planning Division would be probably more compliant with the clustering results.

The weak correlation between calculated cluster vulnerability maps and raster maps can be also caused by the fact, that the classification of clusters was made using only three classes: high, medium and low, while it would be probably more beneficial to add more classes.

Since a wind hazard data, computed from risk hazard map is later on compared to the same risk map values, this is can be a reason why there is the biggest compliance with the cluster vulnerability representation and the wind hazard raster map. Small compliance with medium classes can also indicate a possible problem with correct classification, since the medium classes are the most numerous. In case it will be assumed to look only for the least vulnerable clusters, the obtained results are very much satisfying.

For structural data no validation was possible, since no appropriate maps were available. Without having that kind of real maps with buildings structure assessment, it could be perhaps possible to validate this data using maps which describes e.g. the average income of people in certain areas and assume that in the richer areas buildings have a better construction quality.

In this research, the validation of mapping vulnerability for environmental and for structural factors was considered separately. When combining both vulnerability factors group one has to be aware of what kind of weighting to choose. For the overall hurricane vulnerability assessment environmental factors are much more important than the structural factors. A building which is situated in a very prone to wind zone, even when very well constructed, would be still very vulnerable for the hurricane hazard.

6. Phase 3: Network analysis for hurricane shelters locations

6.1 Methodology

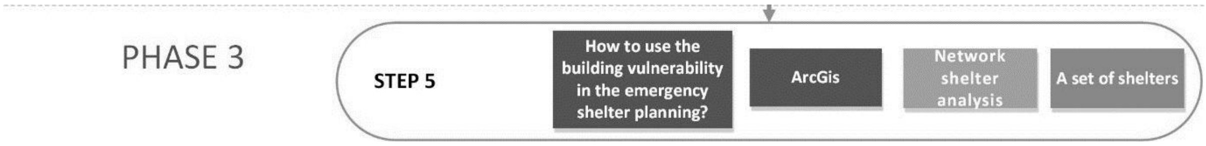


Figure 6-1 Workflow, phase 3

6.1.1 Step 5 – Location-allocation network analysis

In step 5, several different experiments were conducted, varying the number of buildings included in the analysis. Buildings were selected based on their vulnerability – experiment one included only the most sensitive group of buildings, the other experiments consider more vulnerability groups. For each experiment, the number and location of the needed shelters was determined, including the travel distance of the evacuees and the capacity of the shelters. The whole procedure was performed in ArcGIS, using the network analysis, location-allocation tool. This network analysis shows if the current shelters are correctly positioned and have adequate capacity. Additionally, it evaluates where possibly additional shelters would be needed. Since, in general good location of the facility ensures a possible high-quality service at a low cost, this information can be very useful for evacuation policy planners.

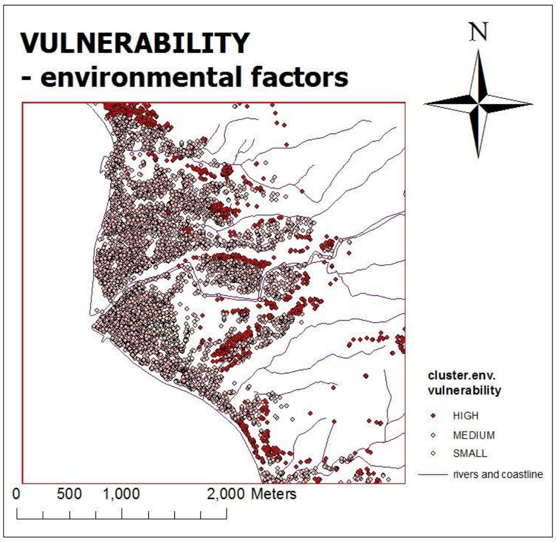


Figure 6-3 Vulnerability – clustering classification for environmental and locational factors (overall)

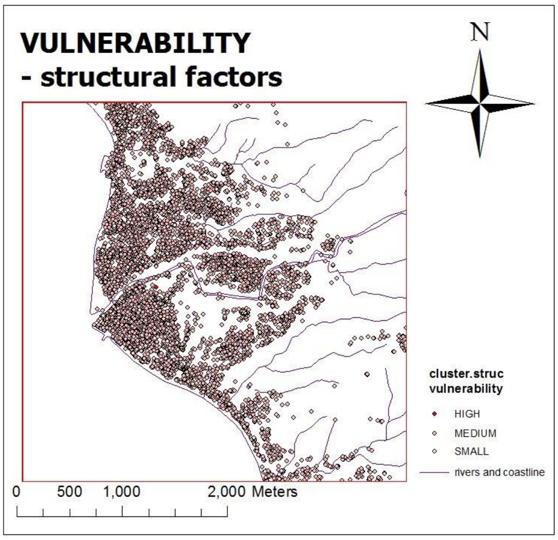


Figure 6-2 Vulnerability – clustering classification for structural factors

The model that was developed in Model builder in ArcGIS is presented below. First, the network is loaded to make a location-allocation layer. Secondly shelters (for this research 9 available shelters) and buildings locations are added. At the end the location-allocation can be solved.

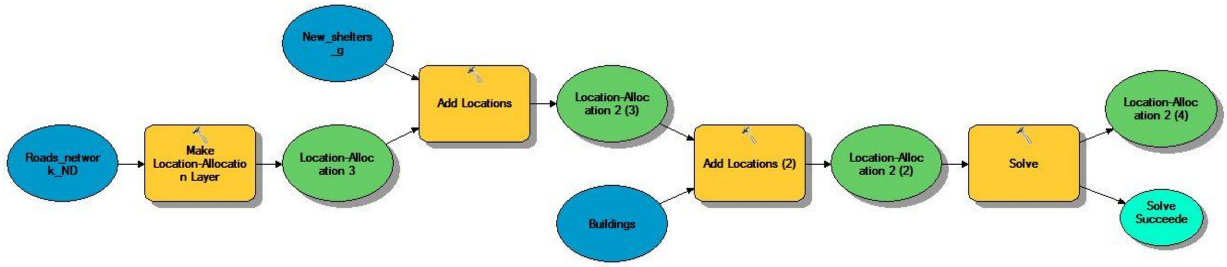


Figure 6-4 Location-allocation model

To perform this analyses the following assumptions were made:

- Just before the evacuation starts, all people are equally distributed over the buildings

- The location of buildings and available shelters is known

- People will walk to the shelter

- The network is given, as well the impedance costs (length)

- The analysis is based on location-allocation model, with type: MAXIMIZE_CAPACITATED_COVERAGE

This type of the analysis can take into account the given impedance cut-off (that is the biggest distance that people will be willing to travel) and ensure that the weighted demand allocated to a facility would not exceed the facility's capacity.

- Origin locations and demand

The buildings used in the experiments are based on the point buildings layer obtained in Phase 1 and enriched by clusters assignment in Phase 2. Each building has an attribute for the number of people present in this house during evacuations. The number of inhabitants of Roseau is set up to 11509 people (data according to the Census population report (Finance, 2011)). It was assumed that population density is equally distributed over the different residential areas (all buildings), and the weight (that is the distribution of people per building) is set equally, which means that it is 1.8 (≈11509 people/6228 buildings). This is a lower number that the one that can be found in the same data from 2011 about the population per dwelling (which is there 2.8). This is caused by the fact, that here it is assumed that the demand is assigned also to non-residential buildings like shops, offices, etc.

- Destination locations, number of facilities (shelters) and their capacity

Shelter point's layer obtained from CHARIM project. In total there are 135 official shelters for the whole Dominica. For the case study area there are 9 available shelters. It was assumed then that the capacity of each shelter will be set to 250. This assumption was made on the field information that was collected during the Spatial Data Management and Identification of the most Vulnerable Schools and Shelters in Dominica (part of the World Bank project, WB 7170161, (C. o. Dominica, 2014)). However, some shelters are known to have a smaller capacity, especially for longer stays. Additionally, new, 9 proposed shelters (bigger buildings) are chosen as a possible short-periods solutions, which will increase an overall capacity possibilities.

- Network

A network data set was build based on the road network of Roseau, consisting of with 5936 polylines, having 5308 junctions.

- Impedance attribute (cost) and cut off

There impedance cost is length, calculated based on the road network. The cut off (distance threshold) was set to 3000m, assuming that people walk at a speed of 5 km/hour (83 meters per minute) and they will be able to walk no more than around 35 minutes.

At the beginning, the model was run for all buildings (experiment 0). In next experiments only some buildings were taken into consideration. The classification was based on the previous phase results, for both clustering: for environmental and locational (overall vulnerability) and for structural factors. In general, in all cases buildings were chosen as a combination for environmental and structural factors with exclusion of buildings from the least vulnerable class. These cluster had the best compliance with the risk map, so it is assumed that this classification is the most correct. Additionally, in few cases, also medium vulnerability class was excluded.

Experiments were conducted according to the list presented below:

<u>Experiment 0</u>

This basic experiment was performed for existing shelters and all people. No classification of buildings is considered. It was conducted only to see how the model works in general. It is unrealistic in respect to the capacity of the shelters in relation to the complete population of Rousseau. However, it is very useful in assessing the extreme situation of all inhabitants evacuating to the currently available shelters.

<u>Experiment 1</u>

The first experiment excluded all low vulnerability classes, both from the perspective of locational and structural factors, so it was conducted based on <u>medium and high vulnerability classes of both kind of factors</u> (using alternative operator 'or'). This experiment shows if the existing shelters are located correctly in relation to these evacuee groups and if the capacity of the current shelters is sufficient.

<u>Experiment 2</u>

The second experiment excluded medium vulnerability classes for environmental and for structural factors. It was conducted only for high vulnerability classes. This experiment shows if the existing shelters are located in places that are reachable for people having <u>the highest potential overall hurricane risk</u>.

<u>Experiment 3</u>

This experiment excluded medium vulnerability classes of environmental factors, so it was run only for these buildings that had high vulnerability class of clustering environmental factors. This experiment shows how many of the buildings <u>at the highest potential hurricane risk for environmental factors</u> are well connected with existing shelters.

<u>Experiment 4</u>

This experiment was excluding low and medium vulnerability classes of structural factors, which means run only for the high vulnerability class of clustering structural factors. According to the classification, this was the case only for one building, so it was decided to not perform this experiment at all, since it will not bring any significant result.

Experiment 5

Fifth experiment was excluding low and medium vulnerability classes of environmental factors. That means that only high vulnerability classes of environmental either high end medium vulnerability classes of structural factors were taken into consideration. It shows how many of the buildings at the highest potential hurricane risk for environmental factors either at the highest and medium potential hurricane risk for structural factors have a good connection with existing shelters.

Experiment 6

This experiment was excluding low and medium vulnerability classes for structural factors, so was run only for high and medium vulnerability of environmental either high vulnerability of structural classes. It shows how many of the buildings at the highest and medium potential hurricane risk for environmental factors either at the highest potential hurricane risk for structural factors have a good connection with existing shelters.

	Included values of vulnerability clusters for environmental factors	Logical operator	Included values of vulnerability clusters for structural factors
Experiment 0	all	or	all
Experiment 1	Medium and High	or	Medium and High
Experiment 2	High	or	High
Experiment 3	Only buildings with high environmental vulnerability (without considering the vulnerability for structural factors)		
Experiment 4	Only buildings with high structural vulnerability (without considering the vulnerability for environmental factors)		
Experiment 5	High	or	High and Medium
Experiment 6	High and Medium	or	High

Table 6-1 Explanation on which groups of values of vulnerability clusters were chosen for each experiment in the network analysis

The first 6 experiments (numerated 1.1, 2.1, 3.1, etc.) will take into consideration only the existing shelters, whereas the next 6 experiments, with the same settings (but numerated 1.2, 2.2, 3.2, etc.) will include randomly chosen additional 9 shelters. For now, in general, shelters are located rather close to each other (7 out of 9 shelters are located in the central part of the city). Two shelters, situated on the south and north end, are very close to the coastline.

New shelters were selected randomly based on their size. In real world, spatial planners, knowing certain area will also know exactly which buildings are suitable to be a shelter, so it would not be the random choice.

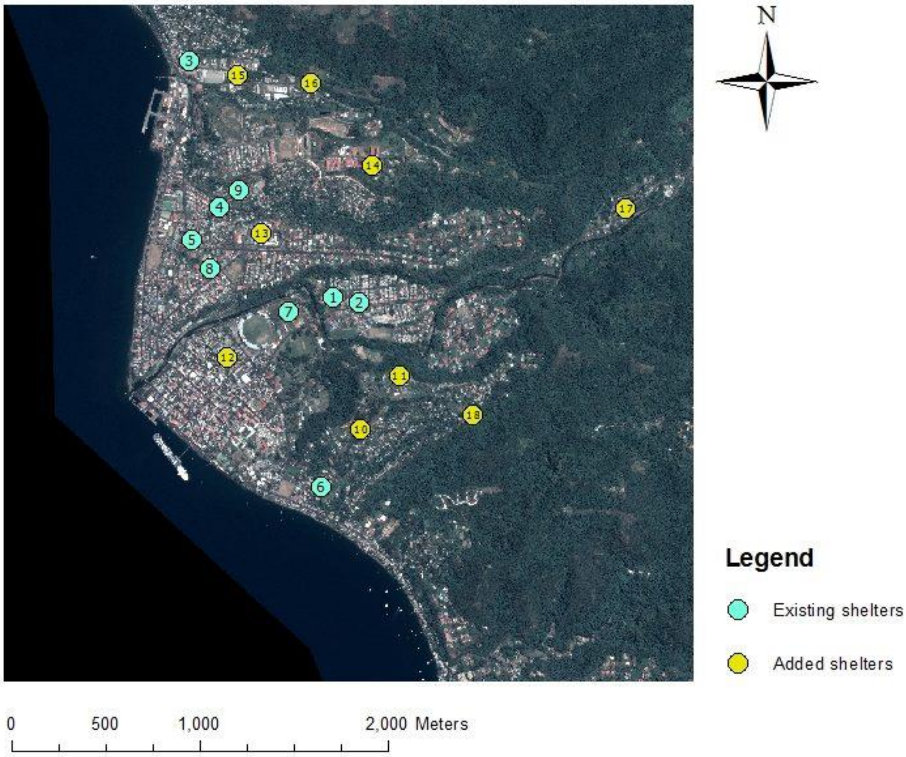


Figure 6-5 Existing and added shelters, Rousseau

6.2 Results

<u>Experiment 0</u>

This experiment resulted in choosing all 9 shelters as opportune facilities. From the 6228 buildings in the Rosseau area, only 1242 buildings were allocated to a shelter, which is around 20% of the total demand. The average travelled distance was 199.72 meters, with the minimum of 0 meters and maximum of 525.39 meters. All shelters were full (equally filled by 248.4 people).

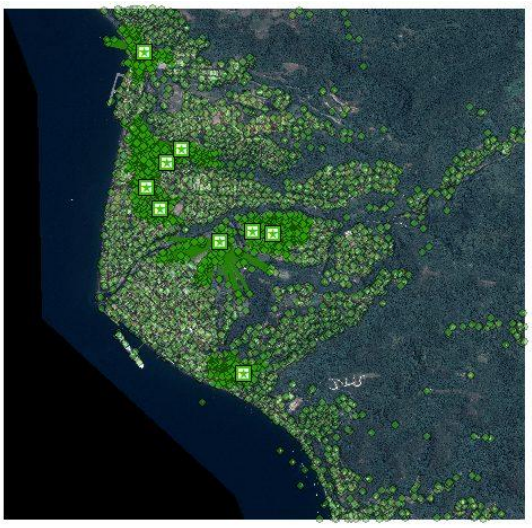


Figure 6-6 Network analysis, experiment 0, for existing shelters

Experiment 1.1

This experiment resulted in choosing all 9 shelters as opportune facilities. From the 5960 buildings that were chosen according to the experiment's setting (high and medium vulnerability for environmental either for structural factors), only 1242 buildings were allocated to a shelter, which is around 21% of the demand. The average travelled distance was 200.16 meters (very close to the result of experiment 0), with the minimum of 0 meters and maximum of 525.39 meters. All shelters were full (equally filled by 248.4 people).

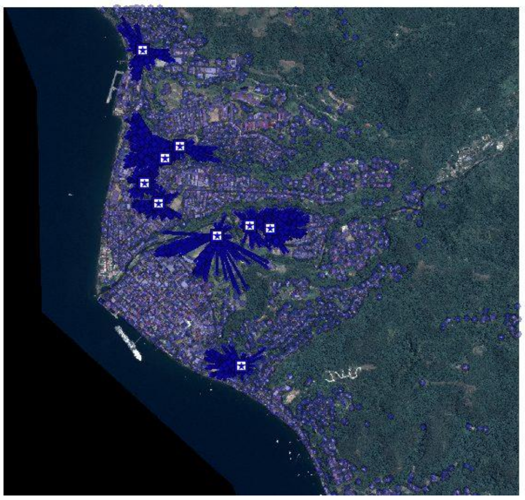


Figure 6-7 Network analysis, experiment 1.1, for existing shelters

Experiment 2.1

This experiment resulted in choosing only 8 shelters as opportune facilities (within a set distance threshold). From the 960 buildings that were chosen according to the experiment's setting (high vulnerability for environmental factors either high vulnerability for structural factors) 836 buildings were allocated, which is around 87% of the demand. This is a very big improvement comparing with previous experiments. It can be caused by the fact that shelter planning was based on these areas. But can be only a guess, since no archival information about shelter planning is available. The average travelled distance was 1281.84 meters, with the minimum of 11.35 meters and maximum of 2851.35 meters. Not all shelters were equally full, the average was 167.2 people, with the maximum of 248.4 people.

Figure 6-8 Network analysis, experiment 1.2, for existing shelters

Experiment 3.1

This experiment resulted in choosing all 9 shelters as opportune facilities. From the 1918 buildings that were chosen according to the experiment's setting (only buildings with high vulnerability for environmental factors), 1242 buildings were allocated, which is around 65% of the demand. The average travelled distance was 1174.69 meters, with the minimum of 11.35 meters and maximum of 2385.17 meters. All shelters were full (equally filled by 248.4 people).

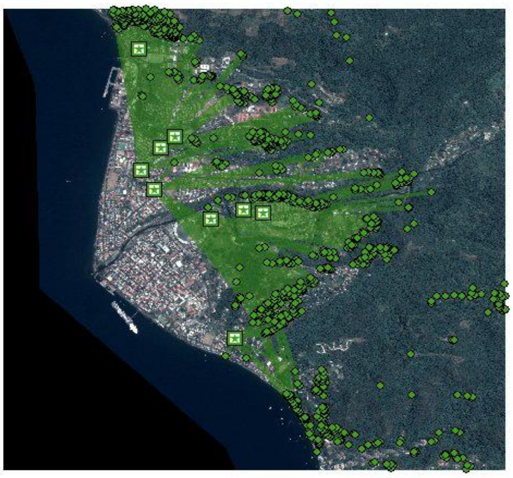


Figure 6-9 Network analysis, experiment 1.3, for existing shelters

Experiment 5.1

This experiment resulted in choosing all 9 shelters as opportune facilities. From the 5392 buildings that were chosen according to the experiment's setting (high vulnerability for environmental factors either high and medium vulnerability for structural factors), 1242 buildings were allocated, which is around 23% of the demand. The average travelled distance was 221.60 meters, with the minimum of 0 meters and maximum of 630.78 meters. All shelters were full (equally filled by 248.4 people).

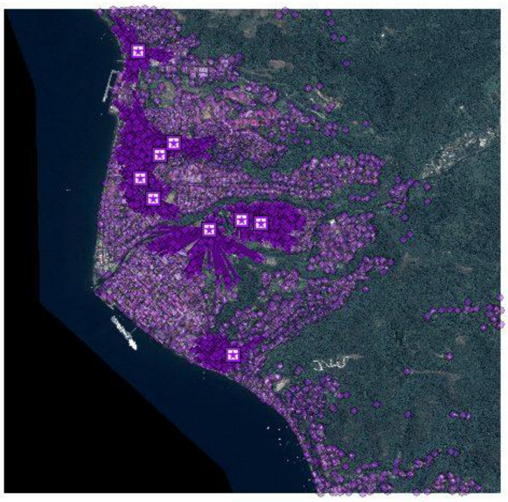


Figure 6-10 Network analysis, experiment 5.1, for existing shelters

Experiment 6.1

This experiment resulted in choosing all 9 shelters as opportune facilities. From 5760 buildings that were chosen according to the experiment's setting (high and medium vulnerability for environmental factors either high vulnerability for structural factors), 1242 buildings were allocated, which is around 23% of the demand. The average travelled distance was 221.60 meters, with the minimum of 0 meters and maximum of 630.78 meters. All shelters were full (equally filled by 248.4 people).

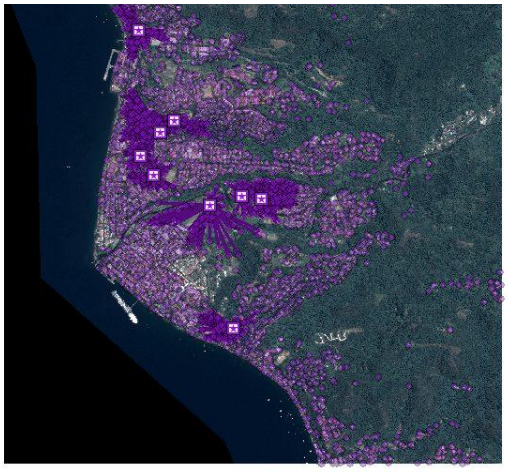


Figure 6-11 Network analysis, experiment 6.1, for existing shelters

Experiment number	0	1.1	2.1	3.1	5.1	6.1
Number of shelters	9	9	8	9	9	9
Buildings chosen	6228	5960	960	1918	5392	5760
Allocated demand	1242	1242	836	1242	1242	1242
% of allocated demand	20	21	87	65	23	22
Mean travel distance per person (meters)	199.72	200.16	1281.84	1174.69	221.60	200.74
Minimum travel distance per person (meters)	0	0	11.35	11.35	0	0

Maximum travel distance per person (meters)	525.39	525.39	2851.35	2385.17	630.78	525.81
Mean demand weight per facility	248.4	248.4	167.2	248.4	248.4	248.4
Minimum demand weight per facility	248.4	248.4	0	248.4	248.4	248.4
Maximum weight per facility	248.4	248.4	248.4	248.4	248.4	248.4
Facility with the min average distance	5	5	8	2	5	5
Facility with the max average distance	7	7	7	8	4	7

Table 6-2 Network analysis for existing shelters - quantitative results

The percent of allocated demand varied from 20% to 87%, but in general most of the people will not be able to reach any of the present shelters. The biggest problem that arises is the allocation of people from the medium vulnerability buildings. A possible solution could be adding new shelters. For now, the maximum distance that a person would have to travel is 2851.35 m, which is close to the threshold of 3000 m. Shelter number 7 is the least accessible, and shelter number 5 has the best accessibility. In most of the experiments all facilities have a demand weight of 248.4, which means that they are full and well located in relation to the demand.

<u>Experiment 0'</u>

This experiment resulted in choosing all 18 shelters as opportune facilities. From all 6228 buildings 2484 buildings were allocated, which is around 40% of the demand (compared to 20% in experiment 0). The average travelled distance was 262.90 meters, with the minimum of 0 meters and maximum of 1142.32 meters. All shelters were full (equally filled by 248.4 people).

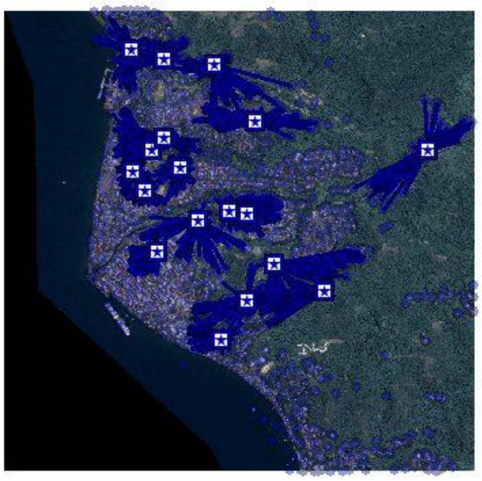


Figure 6-12 Network analysis, experiment 0', for existing and added shelters

Experiment 1.2

This experiment resulted in choosing all 18 shelters as opportune facilities. From the 5960 buildings that were chosen according to the experiment's setting (high and medium vulnerability for environmental either for structural factors), only 2484 buildings were allocated, which is around 42% of the demand (compared to 21 % in experiment 1.1). The average travelled distance was 307.48 meters, with the minimum of 0 meters and maximum of 1523.95 meters. All shelters were full (equally filled by 248.4 people).

Figure 6-13 Network analysis, experiment 1.2, for existing and added shelters

Experiment 2.2

This experiment resulted in choosing only 16 shelters as opportune facilities (within a set distance threshold). From 960 buildings that were chosen according to the experiment's setting (high vulnerability for environmental factors either high vulnerability for structural factors) 874 buildings were allocated, which is around 91% of the demand (compared to the 87% in the experiment 2.1). The average travelled distance was 867.13 meters, with the minimum of 0 meters and maximum of 2997.09 meters. Not all shelters were equally full, the average was 87.4 people, with the maximum of 248.4 people.

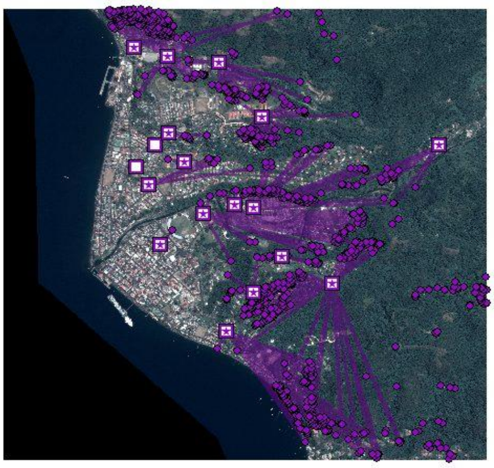


Figure 6-14 Network analysis, experiment 2.2, for existing and added shelters

Experiment 3.2

This experiment resulted in choosing only 17 shelters as opportune facilities (within a set distance threshold). From 1918 buildings that were chosen according to the experiment's setting (only buildings with high vulnerability for environmental factors), 1692 buildings were allocated, which is around 88% of the demand (compared to the 65% in the experiment 3.1). The average travelled distance was 945.18 meters, with the minimum of 0 meters and maximum of 2843.30 meters. Not all shelters were equally full, the average was 169.2 people, with the maximum of 248.4 people.

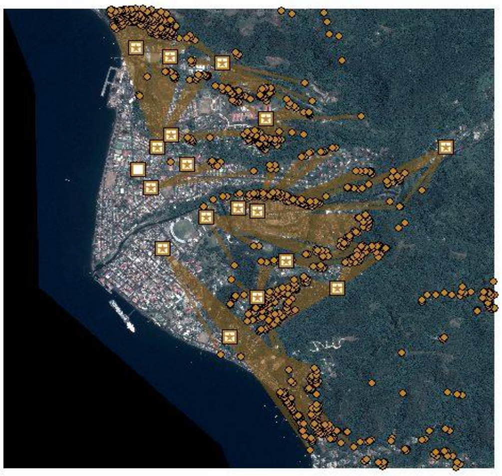


Figure 6-15 Network analysis, experiment 3.2, for existing and added shelters

Experiment 5.2

This experiment resulted in choosing all 18 shelters as opportune facilities. From the 5392 buildings that were chosen according to the experiment's setting (high vulnerability for environmental factors either high and medium vulnerability for structural factors), 2848 buildings were allocated, which is around 53% of the demand (compared to the 23% in the experiment 5.1). The average travelled distance was 338.28 meters, with the minimum of 0 meters and maximum of 1583.10 meters. All shelters were full (equally filled by 248.4 people).

Figure 6-16 Network analysis, experiment 5.2, for existing and added shelters

Experiment 6.2

This experiment resulted in choosing all 18 shelters as opportune facilities. From 5760 buildings that were chosen according to the experiment's setting (high and medium vulnerability for environmental factors either high vulnerability for structural factors), 2484 buildings were allocated, which is around 43% of the demand (compared to the 22% in the experiment 6.1). The average travelled distance was 309.48 meters, with the minimum of 0 meters and maximum of 1523.95 meters. All shelters were full (equally filled by 248.4 people).

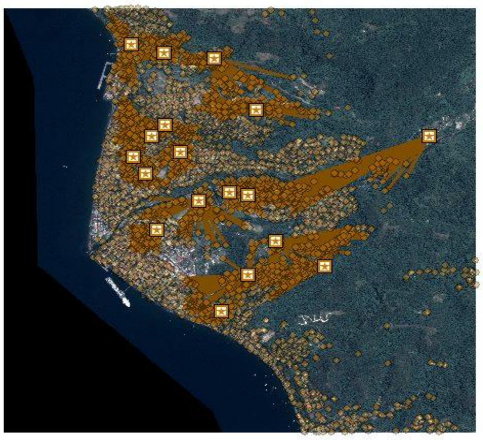


Figure 6-17 Network analysis, experiment 6.2, for existing and added shelters

Experiment number	0'	1.2	2.2	3.2	5.2	6.2
Number of shelters	18	18	16	17	18	18
Buildings chosen	6228	5960	960	1918	5392	5760
Allocated demand	2484	2484	874	1692	2848	2484
% of allocated demand	40	42	91	88	53	43
Mean travel distance per person (meters)	262.90	307.48	867.13	945.18	338.28	309.48
Minimum travel distance per person (meters)	0	0	0	0	0	0
Maximum travel distance per person (meters)	1142.32	1523.95	2997.09	2843.30	1583.10	1523.95
Mean demand weight per facility	248.4	248.4	87.4	169.2	248.4	248.4
Minimum demand weight per facility	248.4	248.4	0	0	248.4	248.4
Maximum demand weight per facility	248.4	248.4	248.4	248.4	248.4	248.4

Facility with the min average distance	5	5	13	13	5	5
Facility with the max average distance	10	17	6	12	17	17

Table 6-3 Network analysis for existing and added shelters - quantitative results

For the experiments with additional shelters, the percentage of allocated demand increased, and varied from 40% to 91%. More evacuees are able to reach the shelters, but the problem of allocation of people from medium vulnerability buildings still remains. Compared to the experiments with existing shelters only, the average travelled distance varied depending upon each case. It significantly decreased for experiments 1.2 and 2.2 which included only the high vulnerability buildings (from 1281.84 to 867.13), but increased for example for the basic experiments including all buildings (from 199.72 to 262.90).

The maximum distance travelled is here 2997.09 m. Shelter number 17 is the least accessible, and still shelter number 5 has the best accessibility. In most of the experiments all facilities have a demand weight of 248.4, which means that they are full and well located in relation to the demand.

When comparing the two groups of experiments one can see that even a 100% increase in the number of shelters does not significantly improve the fulfillment of demand in all cases. It seems that a good allocation of people depends not on the number, but on the location of the shelters. The problem of the right location is seen especially in experiment 2.2, where even doubling the number of shelters resulted in very various allocations of people over the shelters, but did not fulfilled (in this research the lowest) entire demand.

According to the general assumptions of the model, most of the shelters are now located close to the least vulnerable buildings. Within the set threshold of 3000 m most of the people from medium vulnerable areas will not reach any available (and additional) shelter. This result shows that in order to improve an overall efficiency of the evacuation capacity aspect significant actions in shelter development should be undertaken by evacuation planners in this region.

6.3 Reflection

In general, when assuming that a classification made in Phase 2 of this research is correct, a general state of the existing shelters performance evaluated in the phase 3 is very poor. Nevertheless, here a deep reflection should be made not only on the variables, but also on the parameters of the analysis.

The assumption that people are equally distributed over the buildings made an analysis easy to perform but in the same time very unrealistic. Normally, people are unevenly distributed over area, especially during the day, where people are constantly moving and they cannot be strictly assigned to buildings location. This distribution can be improved by assigning people over residential buildings only and calculating their number based on the e.g. number of people per district.

A difficult parameter to establish is the threshold for maximum travelling distance. In this research it was set to 3000 meters, assuming that people are only able to walk and they will not walk longer than around 35 minutes. In general, this should not be taken for granted, since it is possible that some people will use their cars, some will be willing to walk longer, etc. Also, a parameter length should be recalculated taking into account slope values. The real length of the road can be very different, when situated on a sloping hillside. It would be recommended to consider these additions in future analysis.

There are also possible other parameters that were not included in these experiments. These are polygon, line and point barriers. During hurricanes, roads can become impassable due to flooding or landslide. Also some areas may be completely cut off from the land. Analyzing archival data and current conditions of the roads planners can make a prediction where such a barriers may occur.

To conclude, presented network analysis showed that it can be a useful tool for evacuation policy planners. The location of the present shelters can be assessed according to their capacity and accessibility. The possibility of including new location of shelters allows to evaluate an increase in the overall efficiency. It must be emphasized that the type and parameters of modelling have to be carefully chosen and adapted to specific characteristics of a given region.

Using the presented analysis for the area of Roseau in Dominica, for the given input data, the number of shelters for the most vulnerable houses and areas is considered to be currently not sufficient. Additional shelters are needed, but their location should be very well thought out. One should not only ensure that demand will be met, but also whether shelters are located in vulnerable areas. Placing even the best shelter, but in the area very much subjected to the risk makes no sense.

7. Conclusion and recommendations

When hurricanes occur, people have to evacuate to shelters. In many cases, shelters are located based on the total population, however, people living in safe areas in well-constructed houses are not likely to move to shelters. It seems that it would be more realistic to determine the most vulnerable areas and houses and base shelter locations accordingly. In Caribbean islands, which are frequently under hurricane risk, the amount of up to date, digital data is limited. To create an up to date digital building map showing vulnerable houses, automatic extraction based on satellite images is the best option.

This research presents an innovative approach of incorporating object-based image analysis, self-organizing maps and network analysis as a hybrid method to determine shelter evacuation areas. It fulfils the main objective which was formulated as: *Develop a semi-automated method for identifying the location of houses with a high hurricane vulnerability in order to enhance adequate hurricane shelter planning.* The research was divided into three phases, each of which answered a research sub-question.

PHASE 1

The first phase (addressing the sub-objective: *How can building location be extracted from satellite images in an automated way?)* shows how eCognition software can be used for the purpose of object-based oriented analysis. Applying different segmentation and classification procedures resulted in obtaining a buildings layer from a satellite image with an accuracy of 55% for area condition and 63% for shape condition. The results could be improved by incorporating a better adjustment of segmentation parameters and creating a more developed classification rule set. Such improvement is recommended, since characteristics like area or perimeter length are used in the next phase of the research. The accuracy evaluation could also consider evaluators like for example kappa coefficient or error matrix and not only the shape and area condition.

PHASE 2

The second phase *(How can hurricane vulnerability characteristics of these buildings be determined in a (semi) automated way?)* was performed using data mining technique called SOM (Self Organizing Map). Firstly, values of environmental and locational characteristics like: slope, DEM, distance to the coastline and rivers, wind susceptibility, and structural characteristics like: buildings' area, building's complexity, distance to the nearest building and buildings density were assigned to each building. Buildings with assigned characteristic were clustered and classified according to their vulnerability to a hurricane risk using three classes: high, medium and low. Finally, the obtained results were compared with the existing hazard maps. Calculated percent of compliance showed that the most correctly classified group (from 81.4% to 100%) are buildings from the low vulnerability class. Even though for medium and high vulnerability classes the compliance result was not so good, it was assumed that general outcome it is very much satisfying. This was due to the fact that during the classification it was decided to use only three vulnerability classes. The medium and high vulnerable ones can stand in opposition to the low one. Since the low one is calculated very well, then by exclusion one can ensure that the rest on the buildings will fall into medium or high vulnerable class.

There are also some recommendation made for second phase. For the SOM analysis, it would be desirable to incorporate more variables. Choosing the correct variables is difficult and deserves attention. From this research it can be concluded that the best approach is to train the SOM for locational and structural factors separately, and later combine the results. Regarding the locational factors it is advisable to develop one, general formula for assessing the vulnerability of the area. For this research, flood hazard, wind hazard and landslide hazard were regarded as equally important, yet the weights greatly influence the results.

At the moment, no validation layer was available for the structural elements of buildings. It is advisable to find additional data (for example an income layers, house tax layers) that can be used as potential validation data.

Also collecting further data on structural and environmental factors is important. These should be easy to obtain from satellite images (like structural factors presented in this research) or very likely to be present for any area (like environmental factors). At the same time, certain weight for each factor should be indicated.

PHASE 3

In the third phase (*How can house vulnerability data be used in shelter location and evacuation planning?)* the network analysis tool of location-allocation was applied. Buildings, from grouped in previous phase three vulnerability classes were set as a demand, while a shelters served as facilities. The aim of this step was to evaluate if shelters are located correctly in relation to the most vulnerable buildings and have a capacity that matches the combined need. For this research threshold for maximum distance to facility was set to 3000 meters, while the capacity of the shelter was set to 250 people. Few experiments were conducted varying the number of included classes. The first results showed that existing shelters are easily accessible for people from the most vulnerable buildings, but they cannot fulfill the whole demand (various results per experiment, from 20% to 87%). Adding new, randomly chosen shelters showed moreover that even when demand is better satisfied (from 40% to 91%), the location of the shelter, connected with the distance that an evacuee has to cover is much more important.

For this phase, it is recommended to perform more experiments, differing with input parameters, with additional values for possible barriers that can occur during the hurricane and for more random distribution of population. Policy evacuation planners that will use this tool for shelter planning should consider not only the relationship between location of the shelters and their accessibility, but also reflect on their location in relation to vulnerable areas.

SUMMARY

Data available for this research covered nearly the whole island of Dominica, but only the part of the Roseau area was used. It would be beneficial to do a similar research for the whole area of Dominica (or at least for all coastal areas). This would help to evaluate the overall accuracy of the proposed method. By taking up this challenge, however, one needs to be aware of some technical difficulties. The first problem can be a computation time of extraction and classification in eCognition. Secondly, much more time would be also needed for SOMs creation. Additionally, data availability and quality may vary per each area, making it more difficult to compare.

To sum up, this research shows that data extracted from satellite images, enriched by descriptive attributes related to hazard vulnerability and clustered in an unsupervised manner can provide useful information for network analysis, performed to find the best shelter locations. The research was done for the hurricane vulnerability assessment on the example of Dominica, but the core of the method can be applied as a generic tool for other areas and hazards. In order to do this, further research is strongly recommended.

References

Angeli, S., Quesney, A., & Gross, L. (2012). Graph Mining Based SOM: A Tool to Analyze Economic Stability. In D. M. Johnsson (Ed.), Applications of Self-Organizing Maps: InTech. doi: 10.5772/51352

Augustijn, E.-W., Zurita-Milla, R. (2013). Self-organizing maps as an approach to exploring spatiotemporal diffusion patterns. *International Journal of Health Geographics, 12*(60). doi: 10.1186/1476-072X-12-60

Authority, P. P. a. D. (1996). Building code. The Commonwealth of Dominica.

Baatz, M., Schäpe, A. (2000). Multiresolution segmentation – an optimization approach for high quality multi-scale image segmentation. *Angewandte Geographische Informations-Verarbeitung*

Bajabaa, S., Masoud, M., & Al-Amri, N. (2014). Flash flood hazard mapping based on quantitative hydrology, geomorphology and GIS techniques (case study of Wadi Al Lith, Saudi Arabia). *Arabian Journal of Geosciences, 7*(6), 2469-2481. doi: 10.1007/s12517-013-0941-2

Bezdek, J. C. (1973). Cluster Validity with Fuzzy Sets. *Journal of Cybernetics, 3*(3), 58-73. doi: 10.1080/01969727308546047

Blaschke, T. (2010). Object based image analysis for remote sensing. *Isprs Journal of Photogrammetry and Remote Sensing, 65*(1), 2-16. doi: 10.1016/j.isprsjprs.2009.06.004

Bullinaria, J. A. (2004). *Self Organizing Maps: Fundamentals*. Retrieved from http://www.cs.bham.ac.uk/~jxb/NN/l16.pdf

Caunhye, A. M., Nie, X., & Pokharel, S. (2012). Optimization models in emergency logistics: A literature review. *Socio-Economic Planning Sciences, 46*(1), 4-13. doi: http://dx.doi.org/10.1016/j.seps.2011.04.004

Cervone, G., Franzese, P., Ezber, Y., & Boybeyi, Z. (2008). *Risk Assessment of Atmospheric Hazard Releases Using K-Means Clustering*. Paper presented at the 2008 IEEE International Conference on Data Mining Workshops, Pisa.

CHARIM. (2015). Caribbean Handbook on Risk Management. *Dominica.* Retrieved 06/05/2015, from http://www.charim.net/dominica/information

Church, R. L., & Cova, T. J. (2000). Mapping evacuation risk on transportation networks using a spatial optimization model. *Transportation Research Part C: Emerging Technologies, 8*(1–6), 321-336. doi: http://dx.doi.org/10.1016/S0968-090X(00)00019-X

Cova, T. J., & Church, R. L. (1997). Modelling community evacuation vulnerability using GIS. *International Journal of Geographical Information Science, 11*(8), 763-784. doi: 10.1080/136588197242077

Crosbie, M. J., Perry, D., & Smith, T. (1997). *Buildings at Risk; Wind Design Basics for Practicing Architect* Retrieved from http://www.aia.org/practicing/akr/AIAB100834

Cross, C. R. (2014). Project profile: Disaster Risk Management - Community Resilience. from http://www.acdi-cida.gc.ca/cidaweb/cpo.nsf/vWebProjSearchEn/762B62848726F48985257C1E003C027B#h2sectors

Davidson, R., & Lambert, K. (2001). Comparing the Hurricane Disaster Risk of U.S. Coastal Counties. *Natural Hazards Review, 2*(3), 132-142. doi: doi:10.1061/(ASCE)1527-6988(2001)2:3(132)

de Silva, F. N., Eglese, R. W., & Pidd, M. (2003). Evacuation planning and spatial decision making: designing effective spatial decision support systems through integration of technologies. In M. Manuel, A. F. Guisseppi, & N. D. G. Jatinder (Eds.), *Decision making support systems* (pp. 358-373): IGI Publishing.

de Smith, M. J., Goodchild, M. F., & Longley, P. A. (2015). Geospatial Analysis 5th. Retrieved 20/04/2015, from http://www.spatialanalysisonline.com/HTML/?introduction_and_terminology.htm

Division, P. P. (2011). Statistical report on the performance of the construction sector in The Commonwealth of Dominica.

Dominica, C. o. (2014). *Disaster Vulnerability Reduction Project*. Dominica.

Dominica, G. o. t. C. o. (2015a). Physical Planning Division, Maps. Retrieved 11/03/2015, from http://physicalplanning.gov.dm/land-use-and-development/maps

Dominica, G. o. t. C. o. (2015b). Physical Planning Division. Planning and Building Guide. Retrieved 11/03/2015, from http://physicalplanning.gov.dm/requirements/planning-and-building-guide

Du, K. L. (2010). Clustering: A neural network approach. *Neural Networks, 23*(1), 89-107. doi: http://dx.doi.org/10.1016/j.neunet.2009.08.007

Du, Y. (2008). *Verification of tsunami reconstruction projects by object-oriented building extraction from high resolution satellite imagery.* (Master), ITC, Enschede, Netherlands.

ElDessouki, W. M. (1998). Some developments in transportation network analysis and design with application to emergency management problems.

Ferentinou, M., Karymbalis, E., Charou, E., & Sakellariou, M. (2011). Using Self Organising Maps in Applied Geomorphology. Self Organizing Maps - Applications and Novel Algorithm Design. *InTech*. Retrieved from: http://www.intechopen.com/books/self-organizing-maps-applications-and-novel-algorithm-design/using-self-organising-maps-in-applied-geomorphology doi:10.5772/13265

Fernández, D. S., & Lutz, M. A. (2010). Urban flood hazard zoning in Tucumán Province, Argentina, using GIS and multicriteria decision analysis. *Engineering Geology, 111*(1–4), 90-98. doi: http://dx.doi.org/10.1016/j.enggeo.2009.12.006

Finance, M. o. (2011). *2011 Population and Housing Census*. Roseau.

Foundation, R. (2014). What is R? , from http://www.r-project.org/

Fradkin, D., & Muchnik, I. (2006). Support Vector Machines for Classification. *Discrete Mathematics and Theoretical Computer Science, 70,* 13–20.

GeoNode. (2015). CHARIM GeoNode for Dominica. Retrieved 06/05/2015, from http://charim-geonode.net/

Gibbs, T. (2001). Hurricanes and their Effects on Buildings and Structures in the Caribbean. Retrieved from: http://www.oas.org/pgdm/document/bitc/papers/gibbs/gibbs_01.htm

Gliwiński, J. M. (2009). *State-of-the-art in image segmentation.* Retrieved from sirius.cs.put.poznan.pl/~inf74839/proceedings/binwr.pdf

Hamedianfar, A., Zulhaidi Mohd Shafri, H., Mansor, S., & Ahmad, N. (2014). Improving detailed rule-based feature extraction of urban areas from WorldView-2 image and lidar data. *International Journal of Remote Sensing, 35*(5), 1876-1899. doi: http://dx.doi.org/10.1080/01431161.2013.879350

Hofmann, T., Puzicha, J., & Buhmann, M. J. (1998). Unsupervised texture segmentation in a deterministic annealing framework. *IEEE Transactions on Pattern Analysis and Machine Intelligence., Vol. 13*, 478-482.

Institute, E. R., d'Energia, I. C., & Programme, A. U. (2004). *Sustainable Building - Design Manual: sustainable building design practices*: Energy and Resources Institute.

IOS. (2015). Performing a SOM Experiment. Retrieved 08/05/2015, from http://www.improvedoutcomes.com/docs/WebSiteDocs/SOM/Performing_a_SOM_Experiment.htm

Jury, M. R. (2014). Weather - climate interactions in the eastern Antilles and the 2013 Christmas storm. *Earth Interactions*. doi: 10.1175/EI-D-14-0011.1

Karakış, S., Marangoz, A. M., & Büyüksalih, G. (2014). *Analysis of Segmentation Parameters in eCognition Software Using High Resolution Quickbird MS Imagery*. Paper presented at the ISPRS Workshop on Topographic Mapping from Space, Ankara, Turkey.

Khanduri, A. C., & Morrow, G. C. (2003). Vulnerability of buildings to windstorms and insurance loss estimation. *Journal of Wind Engineering and Industrial Aerodynamics, 91*(4), 455-467. doi: http://dx.doi.org/10.1016/S0167-6105(02)00408-7

Kohonen, T. (1990). The self-organizing map. *Proceedings of the IEEE, 78*(9), 1464-1480. doi: 10.1109/5.58325

Kohonen, T. (2001). *Self-Organizing Maps*. Berlin: Springer.

Kongsomsaksakul, S., Yang, C. & Chen, A. . (2005). Shelter location-allocation model for flood evacuation planning. *Journal of the Eastern Asia Society for Transportation Studies, 6*(1), 4237–4252.

Krawiec, K. (2012). *Segmentacja obrazu*. Retrieved from www.cs.put.poznan.pl/kkrawiec/wiki/uploads/PO/po5.pdf

Kuyuk, H. S., Yildirim, E., Dogan, E., & Horasan, G. (2011). An unsupervised learning algorithm: application to the discrimination of seismic events and quarry blasts in the vicinity of Istanbul. *Nat. Hazards Earth Syst. Sci., 11*(1), 93-100. doi: 10.5194/nhess-11-93-2011

Limited, K. C. (2010). Country poverty assesment - Dominica.

Mhangara, P., Odindi, J., Kleyn, L., & Remas, H. (2014a). Mapping urban areas using object oriented classification. Retrieved from

Mhangara, P., Odindi, J., Kleyn, L., & Remas, H. (2014b). Road extraction using object oriented classification. Retrieved from

MS, D. M. S. (2014). 2014 Atlantic hurricane season. Retrieved 20.10.2014, from http://weather.gov.dm/2014_atlantic_hurricane_season.php

NOAA. (2014). National Hurricane Center. Retrieved 20.10.2014, from http://www.nhc.noaa.gov/climo/

Palamara, F., Piglione, F., & Piccinini, N. (2011). Self-Organizing Map and clustering algorithms for the analysis of occupational accident databases. *Safety Science, 49*(8–9), 1215-1230. doi: http://dx.doi.org/10.1016/j.ssci.2011.04.003

Paquette, J., & Lowry, J. (2012). Flood hazard modelling and risk assessment in the Nadi River Basin, Fiji, using GIS and MCDA. *The South Pacific Journal of Natural and Applied Sciences., 30*, 33–43. . doi: http://dx.doi.org/10.1071/SP12003

Pita, G. L., Pinelli, J. P., Gurley, K. R., & Hamid, S. (2013). Hurricane vulnerability modeling: Development and future trends. *Journal of Wind Engineering and Industrial Aerodynamics, 114*(0), 96-105. doi: http://dx.doi.org/10.1016/j.jweia.2012.12.004

Pompe, J., & Haluska, J. (2011). *Estimating the Vulnerability of U.S. Coastal Areas to Hurricane Damage, Recent Hurricane Research - Climate, Dynamics, and Societal Impacts*: InTech.

Prevatt, D. O., Dupigny-Giroux, L. A., & Masters, F. J. (2010). Engineering perspectives on reducing hurricane damage to housing in CARICOM Caribbean Islands. *Natural Hazards Review, 11*(4), 140-150.

Qingsong, L. (2006). *Capacity constrained routing algorithms for evacuation route planning.* University of Minnesota.

Rad, S., Rivé, K., Vittecoq, B., Cerdan, O., & Allègre, C., J. (2013). Chemical weathering and erosion rates in the Lesser Antilles: An overview in Guadeloupe, Martinique and Dominica. *Journal of South American Earth Sciences, 45*(0), 331-344. doi: http://dx.doi.org/10.1016/j.jsames.2013.03.004

Resta, M. (2012). Graph Mining Based SOM: A Tool to Analyze Economic Stability. In D. M. Johnsson (Ed.), Applications of Self-Organizing Maps: InTech. Retrieved from http://www.intechopen.com/books/applications-of-self-organizing-maps/graph-mining-based-som-a-tool-to-analyze-economic-stability. doi: 10.5772/51240

Sangho, K., Betsy, G., & Shashi, S. (2007). *Evacuation route planning: scalable heuristics*. Paper presented at the Proceedings of the 15th annual ACM international symposium on Advances in geographic information systems, Seattle, Washington.

Sensing, J. A. o. R. (1996). Maximum Likelihood Classifier. Retrieved 13/05/2015, from http://wtlab.iis.u-tokyo.ac.jp/~wataru/lecture/rsgis/rsnote/cp11/cp11-7.htm

Sharawi, A. T. (2007). *Optimization Models for Emergency Relief Shelter Planning for Anticipated Hurricane Events*: University of Central Florida.

Sherali, H. D., Carter, T. B., & Hobeika, A. G. (1991). A location-allocation model and algorithm for evacuation planning under hurricane/flood conditions. *Transportation Research Part B: Methodological, 25*(6), 439-452. doi: http://dx.doi.org/10.1016/0191-2615(91)90037-J

Smith, T. (2010). Wind Safety of the Building Envelope. Retrieved from: http://www.wbdg.org/resources/env_wind.php?r=envelope

Society, H. S. a. (2015). Hurricane Movement. Retrieved 12/05/2015, from http://www.hurricanescience.org/science/science/hurricanemovement/

Space, A. D. (2013). *Pleiades. Spot the Detail.* Retrieved from www.geo-airbusds.com

Stephen, L. C. (1994). Fuzzy Model Identification Based on Cluster Estimation. *J. Intell. Fuzzy Syst., 2*(3), 267-278.

Trimble. (2014). eCognition. from http://www.ecognition.com/suite

Trohanis, Z. E. (2014). Integrated Safeguards Data Sheet (Appraisal Stage) - Third Phase Disaster Vuln.Reduction APL for Dominica - P129992. Washington, DC: World Bank.

Uça Avcı, Z. D. (2014). Parameter tests for image segmentation of an agricultural region. *INTERNATIONAL JOURNAL OF ELECTRONICS; MECHANICAL and MECHATRONICS ENGINEERING, Vol.3*(2), 515-524.

Uddin, K. (Producer). (2010). Object Based Image Classification. Retrieved from http://www.slideshare.net/kabiruddin/object-based-classification

Udono, T., & Kishor Sah, A. (2002). *Hazard Mapping and Vulnerability Assessment*. Paper presented at the Regional Workshop on Total Disaster Risk Management Japan.

Ulrich, S. (1998). Dominika (J. R. Roman Kazior, Iwona i Leszek Wawrzyniukowie, Trans.). In M. S.A. (Ed.), *Ameryka, obraz kontynentu. Od Alaski do Ziemi Ognistej.* (pp. 190-195). Warszawa.

UNISDR. (2007). The United Nations Office for Disaster Risk Reduction, Terminology. Retrieved 13/05/2015, from http://www.unisdr.org/we/inform/terminology

Vesanto, J. (1999). SOM-based data visualization methods. *Intelligent Data Analysis, 3*(2), 111-126. doi: http://dx.doi.org/10.1016/S1088-467X(99)00013-X

Vesanto, J., & Alhoniemi, E. (2000). Clustering of the self-organizing map. *Neural Networks, IEEE Transactions on, 11*(3), 586-600. doi: 10.1109/72.846731

Vivone, G., Alparone, L., Chanussot, J., Dalla Mura, M., Garzelli, A., & al.., e. (2014). *A critical comparison of pansharpening algorithms*. Paper presented at the IEEE International Geoscience and Remote Sensing Symposium (IGARSS 2014), Quebec, Canada.

Wikipedia. (2014). Pleiades (satellite). Retrieved 28.11.2014, from http://en.wikipedia.org/wiki/Pleiades_%28satellite%29

Wisner, B., Blaikie, P., Cannon, T., & Davis, I. (2003). *At Risk: natural hazards, people's vulnerability and disasters* (2 ed.).

Yager, R. R., & Filev, D. P. (1994). Approximate clustering via the mountain method. *Systems, Man and Cybernetics, IEEE Transactions on, 24*(8), 1279-1284. doi: 10.1109/21.299710

Yuan-Chao, L., Ming, L., & Xiao-Long, W. (2012). Graph Mining Based SOM: A Tool to Analyze Economic Stability. In D. M. Johnsson (Ed.), Applications of Self-Organizing Maps: InTech. doi: 10.5772/50618

Zepeda, C., Osorio, M., & Sol, D. (2005, 28-02 Feb. 2005). *Modeling Evacuation Planning Using A-Prolog.* Paper presented at the Electronics, Communications and Computers, 2005. CONIELECOMP 2005. Proceedings. 15th International Conference on.

Zepeda, C., & Sol, D. (2007). Evacuation Planning using Answer Set Programming: An initial approach. *Engineering Letters*.

Zhan, Q., Molenaar, M., Tempfli, K., & Shi, W. (2005). Quality assessment for geo-spatial objects derived from remotely sensed data. *International Journal of Remote Sensing, 26*(14), 2953-2974. doi: 10.1080/01431160500057764

Attachments

Attachment 1: Code of rule-set performed in eCognition (used in Phase 1 of the research)

Classes:
Blue_buildings
Buildings_very_bright
Buildings
Gray buildings or ground
Red_buildings
Roads_or_rivers
Shadows
Vegetation
Water

Customized Features:
NDVI: [Mean NIR]/[Mean Red]

Process: Main:
Object-based analysis; domain: Map="From Parent"
Segmentation1; domain: Map="From Parent"
multiresolution segmentation: 100 [shape:0.9 compct.:0.9] creating 'Level1'; domain: Map="From Parent"; params: Overwrite existing level="True", Level Name="Level1", Image Layer weights="Blue=1,Green=1,NIR=2,Red=1", Scale parameter="100", Shape="0.9", Compactness="0.9"
spectral difference segmentation: at Level1: spectral difference 12; domain: Level="Level1", Class filter="[Disabled]", Map="From Parent", Region="From Parent", Max. number of objects="all"; params: Level Usage="Use current", Maximum spectral difference="12", Image Layer weights="Blue=1,Green=1,NIR=2,Red=1"
Roads_rivers; domain: Map="From Parent"
assign class: with Density < 1.8 and Mean NIR < 1000 at Level1: Roads_or_rivers; domain: Level="Level1", Class filter="[Disabled]", Threshold condition="Density < 1.8 ", Second condition="Mean NIR < 1000 ", Map="From Parent", Region="From Parent", Max. number of objects="all"; params: Use class="Roads_or_rivers"
assign class: Roads_or_rivers with Mean NIR <= 300 and Border to Vegetation >= 1 Pxl at Level1: Water; domain: Level="Level1", Class filter="[Roads_or_rivers, User defined]", Threshold condition="Mean NIR <= 300 ", Second condition="Border to Vegetation >= 1 Pxl", Map="From Parent", Region="From Parent", Max. number of objects="all"; params: Use class="Water"
Vegetation_and_water; domain: Map="From Parent"
assign class: with NDVI >= 3 at Level1: Vegetation; domain: Level="Level1", Class filter="[Disabled]", Threshold condition="NDVI >= 3 ", Map="From Parent", Region="From Parent", Max. number of objects="all"; params: Use class="Vegetation"
assign class: unclassified with Mean Blue <= 900 and Mean Red <= 700 at Level1: Vegetation; domain: Level="Level1", Class filter="[Unclsfy, User defined]", Threshold condition="Mean Blue <= 900 ", Second condition="Mean Red <= 700 ", Map="From Parent", Region="From Parent", Max. number of objects="all"; params: Use class="Vegetation"
assign class: all with Mean NIR < 80 at Level1: Water; domain: Level="Level1", Class filter="[All Classes]", Threshold condition="Mean NIR < 80 ", Map="From Parent", Region="From Parent", Max. number of objects="all"; params: Use class="Water"
assign class: Water with Brightness >= 250 at Level1: Shadows; domain: Level="Level1", Class filter="[Water, User defined]", Threshold condition="Brightness >= 250 ", Map="From Parent", Region="From Parent", Max. number of objects="all"; params: Use class="Shadows"
assign class: Shadows with GLDV Ang. 2nd moment (all dir.) > 0.023 at Level1: Water; domain: Level="Level1", Class filter="[Shadows, User defined]", Threshold condition="GLDV Ang. 2nd

moment (all dir.) > 0.023 ", Map="From Parent", Region="From Parent", Max. number of objects="all"; params: Use class="Water"

assign class: unclassified with NDVI >= 2 at Level1: Vegetation; domain: Level="Level1", Class filter="[Unclsfy, User defined]", Threshold condition="NDVI >= 2 ", Map="From Parent", Region="From Parent", Max. number of objects="all"; params: Use class="Vegetation"

assign class: Vegetation with Volume < 1000 Pxl and Brightness >= 1000 at Level1: unclassified; domain: Level="Level1", Class filter="[Vegetation, User defined]", Threshold condition="Volume < 1000 Pxl", Second condition="Brightness >= 1000 ", Map="From Parent", Region="From Parent", Max. number of objects="all"; params: Use class="unclassified"

Merging_water; domain: Map="From Parent"

merge region: Water at Level1: merge region; domain: Level="Level1", Class filter="[Water, User defined]", Map="From Parent", Region="From Parent", Max. number of objects="all"; params: Fusion super objects="False", Use Thematic Layers="False"

Segmentation2; domain: Map="From Parent"

multiresolution segmentation: unclassified at Level1: 50 [shape:0.9 compct.:0.9]; domain: Level="Level1", Class filter="[Unclsfy, User defined]", Map="From Parent", Region="From Parent", Max. number of objects="all"; params: Level Usage="Use current", Compatibility mode="None", Image Layer weights="Blue=1,Green=1,NIR=2,Red=1", Scale parameter="50", Shape="0.9", Compactness="0.9"

Buildings; domain: Map="From Parent"

assign class: unclassified with Brightness >= 1700 at Level1: Buildings_very_bright; domain: Level="Level1", Class filter="[Unclsfy, User defined]", Threshold condition="Brightness >= 1700 ", Map="From Parent", Region="From Parent", Max. number of objects="all"; params: Use class="Buildings_very_bright"

assign class: unclassified with Rectangular Fit >= 0.89 and NDVI < 1.4 at Level1: Buildings; domain: Level="Level1", Class filter="[Unclsfy, User defined]", Threshold condition="Rectangular Fit >= 0.89 ", Second condition="NDVI < 1.4 ", Map="From Parent", Region="From Parent", Max. number of objects="all"; params: Use class="Buildings"

assign class: unclassified with Volume >= 1500 Pxl and NDVI >= 1.5 at Level1: Vegetation; domain: Level="Level1", Class filter="[Unclsfy, User defined]", Threshold condition="Volume >= 1500 Pxl", Second condition="NDVI >= 1.5 ", Map="From Parent", Region="From Parent", Max. number of objects="all"; params: Use class="Vegetation"

assign class: unclassified with NDVI >= 1.7 and Brightness <= 1000 at Level1: Vegetation; domain: Level="Level1", Class filter="[Unclsfy, User defined]", Threshold condition="NDVI >= 1.7 ", Second condition="Brightness <= 1000 ", Map="From Parent", Region="From Parent", Max. number of objects="all"; params: Use class="Vegetation"

assign class: unclassified with Mean Red >= 1200 and Asymmetry <= 0.6 at Level1: Red_buildings; domain: Level="Level1", Class filter="[Unclsfy, User defined]", Threshold condition="Mean Red >= 1200 ", Second condition="Asymmetry <= 0.6 ", Map="From Parent", Region="From Parent", Max. number of objects="all"; params: Use class="Red_buildings"

assign class: Red_buildings with Asymmetry > 0.6 at Level1: unclassified; domain: Level="Level1", Class filter="[Red_buildings, User defined]", Threshold condition="Asymmetry > 0.6 ", Map="From Parent", Region="From Parent", Max. number of objects="all"; params: Use class="unclassified"

assign class: Vegetation, unclassified with Mean Blue >= 1500 at Level1: Blue_buildings; domain: Level="Level1", Class filter="[Vegetation, Unclsfy, User defined]", Threshold condition="Mean Blue >= 1500 ", Map="From Parent", Region="From Parent", Max. number of objects="all"; params: Use class="Blue_buildings"

assign class: Red_buildings, unclassified with Number of pixels >= 1000 at Level1: Vegetation; domain: Level="Level1", Class filter="[Red_buildings, Unclsfy, User defined]", Threshold condition="Number of pixels >= 1000 ", Map="From Parent", Region="From Parent", Max. number of objects="all"; params: Use class="Vegetation"

assign class: unclassified with Mean Blue >= 1000 and Mean Red <= 1000 at Level1: Vegetation; domain: Level="Level1", Class filter="[Unclsfy, User defined]", Threshold condition="Mean Blue >= 1000 ", Second condition="Mean Red <= 1000 ", Map="From Parent", Region="From Parent", Max. number of objects="all"; params: Use class="Vegetation" assign class: unclassified with NDVI >= 1.5 at Level1: Vegetation; domain: Level="Level1", Class filter="[Unclsfy, User defined]", Threshold condition="NDVI >= 1.5 ", Map="From Parent", Region="From Parent", Max. number of objects="all"; params: Use class="Vegetation"

assign class: unclassified with Mean NIR > 1500 and NDVI <= 1.5 at Level1: Buildings; domain: Level="Level1", Class filter="[Unclsfy, User defined]", Threshold condition="Mean NIR > 1500 ", Second condition="NDVI <= 1.5 ", Map="From Parent", Region="From Parent", Max. number of objects="all"; params: Use class="Buildings"

assign class: unclassified with Mean Blue < 1200 and Mean Green < 1200 at Level1: Buildings; domain: Level="Level1", Class filter="[Unclsfy, User defined]", Threshold condition="Mean Blue < 1200 ", Second condition="Mean Green < 1200 ", Map="From Parent", Region="From Parent", Max. number of objects="all"; params: Use class="Buildings"

assign class: unclassified with Max. diff. >= 0.25 at Level1: Buildings; domain: Level="Level1", Class filter="[Unclsfy, User defined]", Threshold condition="Max. diff. >= 0.25 ", Map="From Parent", Region="From Parent", Max. number of objects="all"; params: Use class="Buildings"

assign class: unclassified at Level1: Gray buildings or ground; domain: Level="Level1", Class filter="[Unclsfy, User defined]", Map="From Parent", Region="From Parent", Max. number of objects="all"; params: Use class="Gray buildings or ground"

Reclassification; domain: Map="From Parent"

assign class: Buildings with Density < 1.5 at Level1: Roads_or_rivers; domain: Level="Level1", Class filter="[Buildings, User defined]", Threshold condition="Density < 1.5 ", Map="From Parent", Region="From Parent", Max. number of objects="all"; params: Use class="Roads_or_rivers"

merge region: Buildings_very_bright with Rel. border to Buildings_very_bright >= 0.1 at Level1: merge region; domain: Level="Level1", Class filter="[Buildings_very_bright, User defined]", Threshold condition="Rel. border to Buildings_very_bright >= 0.1 ", Map="From Parent", Region="From Parent", Max. number of objects="all"; params: Fusion super objects="False", Use Thematic Layers="False"

merge region: Blue_buildings, Buildings_very_bright with Rel. border to Buildings_very_bright >= 0.1 at Level1: merge region; domain: Level="Level1", Class filter="[Blue_buildings, Buildings_very_bright, User defined]", Threshold condition="Rel. border to Buildings_very_bright >= 0.1 ", Map="From Parent", Region="From Parent", Max. number of objects="all"; params: Fusion super objects="False", Use Thematic Layers="False"

merge region: Buildings with Number of pixels <= 300 and Rel. border to Buildings >= 0.1 at Level1: merge region; domain: Level="Level1", Class filter="[Buildings, User defined]", Threshold condition="Number of pixels <= 300 ", Second condition="Rel. border to Buildings >= 0.1 ", Map="From Parent", Region="From Parent", Max. number of objects="all"; params: Fusion super objects="False", Use Thematic Layers="False"

grow region: Blue_buildings, Buildings_very_bright, Gray buildings or ground, Red_buildings at Level1: <- Blue_buildings, Buildings_very_bright, Gray buildings or ground, Red_buildings; domain: Level="Level1", Class filter="[Blue_buildings, Buildings_very_bright, Gray buildings or ground, Red_buildings, User defined]", Map="From Parent", Region="From Parent", Max. number of objects="all"; params: Candidates classes="[Blue_buildings, Buildings_very_bright, Gray buildings or ground, Red_buildings, User defined]", Fusion super objects="False", Use Thematic Layers="False"

merge region: Blue_buildings, Buildings, Buildings_very_bright, Gray buildings or ground, Red_buildings with Area <= 300 Pxl at Level1: merge region; domain: Level="Level1", Class filter="[Blue_buildings, Buildings, Buildings_very_bright, Gray buildings or ground, Red_buildings, User defined]", Threshold condition="Area <= 300 Pxl", Map="From Parent", Region="From Parent", Max. number of objects="all"; params: Fusion super objects="False", Use Thematic Layers="False"

grow region: Buildings_very_bright, Buildings with Area <= 300 Pxl at Level1: <- Buildings_very_bright, Buildings; domain: Level="Level1", Class filter="[Buildings_very_bright, Buildings, User defined]", Threshold condition="Area <= 300 Pxl", Map="From Parent", Region="From Parent", Max. number of objects="all"; params: Candidates classes="[Buildings_very_bright, Buildings, User defined]", Fusion super objects="False", Use Thematic Layers="False"

grow region: Gray buildings or ground, Red_buildings at Level1: <- Gray buildings or ground, Red_buildings; domain: Level="Level1", Class filter="[Gray buildings or ground, Red_buildings, User defined]", Map="From Parent", Region="From Parent", Max. number of objects="all"; params: Candidates classes="[Gray buildings or ground, Red_buildings, User defined]", Fusion super objects="False", Use Thematic Layers="False"

assign class: Blue_buildings, Buildings_very_bright, Buildings, Gray buildings or ground, Red_buildings with Area <= 200 Pxl at Level1: Roads_or_rivers; domain: Level="Level1", Class filter="[Blue_buildings, Buildings_very_bright, Buildings, Gray buildings or ground, Red_buildings, User defined]", Threshold condition="Area <= 200 Pxl", Map="From Parent", Region="From Parent", Max. number of objects="all"; params: Use class="Roads_or_rivers"

assign class: Buildings_very_bright with Length\Width >= 2.5 and Rel. border to Buildings >= 0.2 at Level1: Buildings; domain: Level="Level1", Class filter="[Buildings_very_bright, User defined]", Threshold condition="Length/Width >= 2.5 ", Second condition="Rel. border to Buildings >= 0.2 ", Map="From Parent", Region="From Parent", Max. number of objects="all"; params: Use class="Buildings"

assign class: Blue_buildings with Length\Width >= 2 and Rel. border to Buildings >= 0.2 at Level1: Buildings; domain: Level="Level1", Class filter="[Blue_buildings, User defined]", Threshold condition="Length/Width >= 2 ", Second condition="Rel. border to Buildings >= 0.2 ", Map="From Parent", Region="From Parent", Max. number of objects="all"; params: Use class="Buildings"

assign class: Red_buildings with Area >= 3000 Pxl and Rel. border to Vegetation >= 0.3 at Level1: Vegetation; domain: Level="Level1", Class filter="[Red_buildings, User defined]", Threshold condition="Area >= 3000 Pxl", Second condition="Rel. border to Vegetation >= 0.3 ", Map="From Parent", Region="From Parent", Max. number of objects="all"; params: Use class="Vegetation"

Segmentation3; domain: Map="From Parent"

multiresolution segmentation: Vegetation at Level1: 30 [shape:0.9 compct.:0.9]; domain: Level="Level1", Class filter="[Vegetation, User defined]", Map="From Parent", Region="From Parent", Max. number of objects="all"; params: Level Usage="Use current", Compatibility mode="None", Image Layer weights="Blue=1,Green=1,NIR=2,Red=1", Scale parameter="30", Shape="0.9", Compactness="0.9"

Buildings2; domain: Map="From Parent"

assign class: Vegetation, unclassified with Brightness >= 1700 at Level1: Buildings_very_bright; domain: Level="Level1", Class filter="[Vegetation, Unclsfy, User defined]", Threshold condition="Brightness >= 1700 ", Map="From Parent", Region="From Parent", Max. number of objects="all"; params: Use class="Buildings_very_bright"

assign class: Vegetation, unclassified with Brightness >= 1000 at Level1: Gray buildings or ground; domain: Level="Level1", Class filter="[Vegetation, Unclsfy, User defined]", Threshold condition="Brightness >= 1000 ", Map="From Parent", Region="From Parent", Max. number of objects="all"; params: Use class="Gray buildings or ground"

grow region: Gray buildings or ground with Rel. border to Gray buildings or ground >= 0.01 at Level1: <- Gray buildings or ground; domain: Level="Level1", Class filter="[Gray buildings or ground, User defined]", Threshold condition="Rel. border to Gray buildings or ground >= 0.01 ", Map="From Parent", Region="From Parent", Max. number of objects="all"; params: Candidates classes="[Gray buildings or ground, User defined]", Fusion super objects="True", Use Thematic Layers="False"

assign class: Gray buildings or ground with Area >= 3000 Pxl at Level1: Vegetation; domain: Level="Level1", Class filter="[Gray buildings or ground, User defined]", Threshold condition="Area >= 3000 Pxl", Map="From Parent", Region="From Parent", Max. number of objects="all"; params: Use class="Vegetation"

assign class: Gray buildings or ground with Brightness <= 1155 at Level1: Vegetation; domain: Level="Level1", Class filter="[Gray buildings or ground, User defined]", Threshold condition="Brightness <= 1155 ", Map="From Parent", Region="From Parent", Max. number of objects="all"; params: Use class="Vegetation"

assign class: Gray buildings or ground with Length\Width >= 2.5 at Level1: Vegetation; domain: Level="Level1", Class filter="[Gray buildings or ground, User defined]", Threshold condition="Length/Width >= 2.5 ", Map="From Parent", Region="From Parent", Max. number of objects="all"; params: Use class="Vegetation"

assign class: Buildings_very_bright with Length\Width >= 5 at Level1: Roads_or_rivers; domain: Level="Level1", Class filter="[Buildings_very_bright, User defined]", Threshold condition="Length/Width >= 5 ", Map="From Parent", Region="From Parent", Max. number of objects="all"; params: Use class="Roads_or_rivers"

Attachment 2: R code used for calculating SOM and clustering (used in the Phase 2 of the research)

```
# load libraries
require(kohonen)
require(lattice)
lattice.options(default.theme = standard.theme(color = FALSE))
#read your input file
# use change dir to make sure your directory is correct
input <- as.matrix(read.csv("all_without_missing.csv", header=TRUE))
inputset <- input [1:5988, 1:4]
data.matrix(inputset) ### not assigned to a new object
inputset.sc <- scale(inputset)
set.seed(7)
input.som <- som(data=inputset.sc, grid=somgrid(9,9,"hexagonal"),rlen=40500, keep.data=TRUE)
plot(input.som)
# Mapping back the complete dataset
mapping <- map(input.som,inputset.sc)
plot(input.som, type="mapping", classif=mapping, col="black", pch = 19, main="all",
bgcol=gray(0.85))
# hierarchical clustering
input.hc <- cutree(hclust(dist(input.som$codes)), 9)
pretty_palette <- c("#1f77b4", '#ff7f0e', '#2ca02c', '#d62728', '#7f7f7f', '#bcbd22' ,'#9467bd',
'#8c564b', '#e377c2')
plot(input.som, type="mapping", bgcol = pretty_palette[input.hc], main = "Clusters")
add.cluster.boundaries(input.som, input.hc)
```