UNIVERSITA' DEGLI STUDI DI PAVIA
DIPARTIMENTO DI INGEGNERIA INDUSTRIALE E DELL'INFORMAZIONE
DOTTORATO IN INGEGNERIA ELETTRONICA, INFORMATICA ED ELETTRICA

VULNERABILITY ESTIMATION AND SEISMIC DAMAGE ASSESSMENT THROUGH COMBINATION OF REMOTE SENSING AND CROWDSOURCING

Doctoral Dissertation of:
Daniele De Vecchi

Tutor:
Prof. Prof. Fabio Dell'Acqua

The Chair of the Doctoral Program:
Prof. Prof. Paolo Di Barba

2016 XXIX Cycle

There are many things in life that will catch your eye, but only a few will catch your heart...pursue those.

Michael Nolan

Three Rules of Work: Out of clutter find simplicity; From discord find harmony; In the middle of difficulty lies opportunity

---

Albert Einstein

Abstract

THIS document summarizes all the outcome of the three years PhD programme at the University of Pavia. As suggested by the title, the focus is on mutual combination of remote sensing and crowdsourcing for vulnerability estimation and damage assessment. In particular, collected pre-event information should be used to integrate and improve seismic damage extraction from satellite data. 2015 was the year of the Sendai framework for Disaster Risk Reduction (2015-2030). In this framework, the proposed work contributes in the enhancement of disaster prevention by producing relevant risk information and in the enhancement of effectiveness, meaning a quick response in the aftermath of the event. *Chapter 1* gives a brief introduction to all the concepts included in the document (e.g. risk, vulnerability) and the phases of the disaster cycle. Datasets used are also described, including spectral and spatial features. *Chapter 2* is related to information extracted before a disaster. In particular, the focus here is on the contribution provided by remote sensing. First, a list of indicators derived during the SENSUM project and linked to vulnerability is presented. Next, a set of algorithms is proposed in order to process satellite data and extract those indicators. Built-Up Area extraction is one of the main topics and different methods are proposed; the underlying assumption is that, due to different environmental and spectral conditions, it is really difficult to define a one-size-fits-all method capable of considering all the possible variations involved. A tailored and improved version of the algorithm is also available as service in the ESA GPOD system and directly linked with Landsat 8 and Sentinel-2 repositories. Another proposed workflow takes care of the results obtained from each single processed year and produces a map of the evolution in time of the area of interest. Other algorithms are also explained, focusing for example on the extraction of building footprints and their density, therefore requiring very high resolution optical data as input. Crowdsourcing is addressed in *chapter 3* where a generic framework for the collection of geo-tagged reports is proposed. The basic idea is that the great diffusion of smartphones has created a large and dense network of observers; data provided by volunteers can actively integrate what is derived from remote sensing. Two different examples are proposed. The first one, called SEGUICI Vegetation report, was created in the framework of the SEGUICI project to collect data related to different

crops and their stage of growth in order to feed a water consumption and requirements model. The app is released on the Google Play store. The second mobile app, called CLOOPSy (Copernicus Land cOver crOwdsourcing Platform for Sentinel-based mapping) has been developed in the framework of the MyGEOSS contest and is designed to collect data related to land cover for validation purposes. The product is available from the Google Play store and soon it will be available on the Apple App store. *Chapter 4* gives an insight on the methods developed to extract seismic damage from SAR imagery. In particular, two different approaches have been investigated: the first one uses post-event-only VHR SAR data combined with pre-event information. The other one is based on change detection, a widely used technique to extract damage from different sensors; different attempts were made in order to determine the sensitivity to different spatial resolutions and sensors. The combination with pre-event data has also been defined and tested. Conclusions are illustrated in *chapter 5*.

Contents

CHAPTER 1

Introduction

This document is a summary of the research work accomplished in three years as part of the PhD program at the University of Pavia in the Telecommunication and Remote Sensing laboratory. The title, "Vulnerability estimation and seismic damage assessment through combination of remote sensing and crowdsourcing", emphasizes the focus on the complementarity of Earth Observation products and crowd-generated data in the field of natural disasters, looking at earthquakes as the main target but also considering other types of peril. In particular, the aim is to provide a set of tools, mostly automatic, for the estimation of exposure before the event and to fuse these measures with the standard post-event damage assessment in order to improve the response after a seismic event. User-generated reports will fill the gaps and complete the pre-event information. Regarding the former topic, most of the work described was carried on under the framework of several European projects, mainly SENSUM and RASOR, collaborating with the EUCENTRE Foundation; for the latter, the proposed methods were based on previous experience acquired by the Remote Sensing group in Pavia combined with the expertise acquired from Prof. Yamazaki's group at the University of Chiba, Japan.

The manuscript is a blend of all the scientific publications produced within the three years; to clarify sources, every paper will be mentioned at the beginning of each chapter.

This chapter introduces the main concepts and terminology used within the dissertation. Standard definitions are provided in order to avoid misunderstandings and doubts to the reader.

1.1 Disaster Risk Reduction

2015 was the year of the Sendai framework for Disaster Risk Reduction 2015-2030, proposed by United Nations International Strategy for Disaster Reduction (UNISDR) and endorsed by the UN General Assembly [1]. Seven global targets and four priorities have been identified and are hereby reported as from the official UNISDR document.

These are the identified global targets:

1. Substantially **reduce global disaster mortality** by 2030, aiming to lower average per 100,000 global mortality rate in the decade 2020-2030 compared to the period 2005-2015.

2. Substantially **reduce the number of affected people** globally by 2030, aiming to lower average global figure per 100,000 in the decade 2020 -2030 compared to the period 2005-2015.

3. **Reduce direct disaster economic loss** in relation to global gross domestic product (GDP) by 2030.

4. Substantially **reduce disaster damage to critical infrastructure** and disruption of basic services, among them health and educational facilities, including through developing their resilience by 2030.

5. Substantially **increase** the number of countries with national and local **disaster risk reduction strategies** by 2020.

6. Substantially **enhance international cooperation** to developing countries through adequate and sustainable support to complement their national actions for implementation of this Framework by 2030.

7. Substantially **increase the availability of and access to multi-hazard early warning systems** and disaster risk information and assessments to the people by 2030.

And these are the actions planned:

1. **Understanding disaster risk** Disaster risk management should be based on an understanding of disaster risk in all its dimensions of vulnerability, capacity, exposure of persons and assets, hazard characteristics and the environment. Such knowledge can be used for risk assessment, prevention, mitigation, preparedness and response.

2. **Strengthening disaster risk governance to manage disaster risk** Disaster risk governance at the national, regional and global levels is very important for prevention, mitigation, preparedness, response, recovery, and rehabilitation. It fosters collaboration and partnership.

3. **Investing in disaster risk reduction for resilience** Public and private investment in disaster risk prevention and reduction through structural and non-structural measures are essential to enhance the economic, social, health and cultural resilience of persons, communities, countries and their assets, as well as the environment.

4. **Enhancing disaster preparedness for effective response and to "Build Back Better" in recovery, rehabilitation and reconstruction** The growth of disaster risk means there is a need to strengthen disaster preparedness for response, take action in anticipation of events, and ensure capacities are in place for effective response and recovery at all levels. The recovery, rehabilitation and reconstruction phase is a critical opportunity to build back better, including through integrating disaster risk reduction into development measures.

In this framework, our work contributes to:

- the **enhancement of disaster prevention** by producing relevant risk information, which should enhance risk understanding and enable policy makers and administrators to make better informed decision in building their risk reduction strategies;

- the **enhancement of effectiveness** in disaster response by providing earlier information in the aftermath of an event.

1.2 Risk concepts

A standard set of definitions is provided in order to avoid terminology confusion.

Hazard is a potentially damaging physical event, phenomenon or human activity that may cause loss of life or injury, property damage, social and economic disruption or environmental degradation.

Exposure is determined by the presence of people, property, systems or other elements in hazard zones that are thereby subject to potential losses.

Vulnerability The extent to which a community, structure, services or geographic area is likely to be damaged or disrupted by the impact of particular hazard, on account of their nature, construction and proximity to hazardous terrains or a disaster prone area.

Capacity The combination of all strengths, attributes and resources available within a community, society or organization that can be used to achieve the agreed goals.

Risk Measure of expected losses due to a hazard event occurring in a given area over a specific time period. Risk is a function of the probability of particular hazardous event and the losses it would cause.

The concept of risk has been evolving during the last five decades with risk models moving from hazard-only to considering more components, e.g. vulnerability, exposure and capacity. Risk can be formulated as a result of the multiplication between the potential damaging event -the hazard component- and the degree of susceptibility of the exposed elements, as expressed in equation 1.1 [2] and represented in figure 1.1.

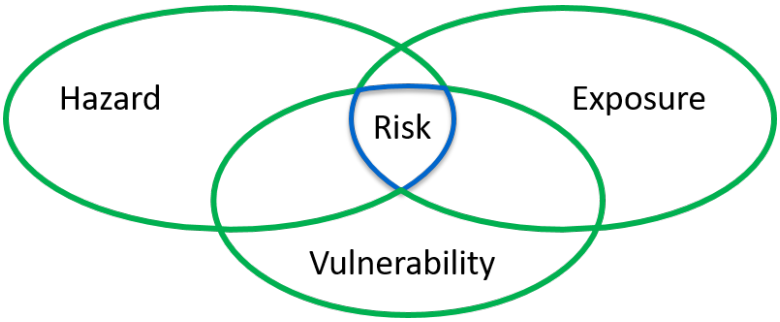


Figure 1.1: *Schematic representation of the risk concept as intersection of hazard, vulnerability and exposure.*

$$Risk = Hazard \cdot Exposure \cdot Vulnerability \tag{1.1}$$

In this dissertation hazard will not be treated directly. Focus will be on exposure and vulnerability, as explained in the following chapters.

1.3 Disaster management

The so-called "disaster management cycle" [3], a reference scheme often used in risk management, represented in figure 1.2, is made up of 2 main sections: before and after an event. This dissertation covers both, from risk management -part of the pre-event phase in mitigation- to damage assessment -related to the emergency stage just after an event.

Figure 1.2: *Schematic representation of the disaster cycle. Picture taken from [4]*

Disaster A serious disruption of the function of a community or a society involving widespread human, material, economic or environmental losses and impacts, which exceeds the ability of the affected community or society to cope using its own resources.

1.3.1 Pre-event phase

The main scope of all the activities and decisions taken before a disaster is to reduce the vulnerability factor of risk, contributing to lower fatalities and economic losses. For example, fatalities caused by earthquakes are strictly related to buildings construction features and population density.

A proper estimation of all the terms involved in risk assessment is necessary in order to reduce the effect of disasters.

Mitigation

Activities done with the aim to eliminate or reduce the probability of a disaster occurrence. For example, land-use management and planning can help to mitigate disasters by discouraging settlements in hazard-prone areas.

Preparedness

Objective of this phase is to improve the readiness to response, with focus on the managerial capacity of governments, organizations and communities.

1.3.2 Post-event phase

After a disaster, the main focus is on response and recovery. This thesis will focus on the former, promoting semi-automatic techniques for the extraction of damage combining data obtained before the event.

Response

This phase starts just after the disaster with search&rescue efforts and damage assessment. Coordination and timing are fundamental aspects.

Recovery

It starts just after the closing of the response stage. The aim is the resumption of services and structures while, at the same time, reducing risk by acting on the vulnerability factor.

1.4 Spaceborne data sources

Different data sources can be considered to feed the proposed algorithms. On one side, remotely sensed data is a fundamental input, thanks to the periodic revisit time and extension of the monitored area offered by spaceborne systems. Medium and high resolution data have been used for this dissertation, trying as much as possible to take advantage of the open data policy that is becoming more and more common recently. However, satellite imagery cannot provide all the "measures" required. A new source of information, known as crowdsourcing, comes into play; its role is complementary, trying to fill the gaps and the technical limitations of satellite.

The main input sources for the algorithms proposed in this dissertation are remote sensing and crowdsourcing; the former is well-known and widely used in exposure estimation and damage assessment using respectively optical and radar (Synthetic Aperture) data from spaceborne platforms. On the other hand, crowdsourcing is considered

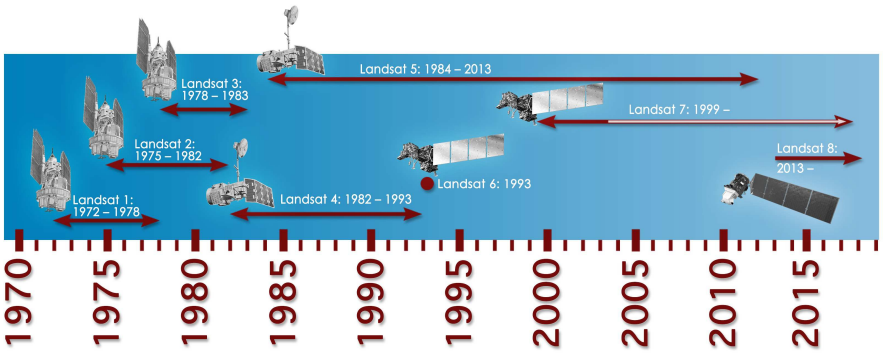


Figure 1.3: *Landsat program timeline. Picture taken from [9]*

as a possible additional source of data, given the increasing attention it is receiving as a complement to Earth Observation.

1.4.1 Optical data

With every sunrise, the Earth's surface is bathed in solar energy, some of which is absorbed and some of which is reflected back to space. One major class of Earth Observation instruments are termed optical because they obtain data by recording this reflected energy across various wavelengths, including visible light and invisible infrared bands [5]. Multiple satellites were and still are actively collecting data in the optical range of the spectrum. Spatial resolution varies from medium -order of tens of meters- to very high resolution -tens of centimetres. In this thesis, optical datasets are used as input of the SENSUM Earth Observation tools using both medium and high resolution imagery.

Landsat main features

The development of non-photographic remote sensing technology progressed rapidly after the first mapping satellite, Landsat 1, was put in orbit in 1972 [6]. Therefore, the Landsat mission represented the precursor of civilian Earth observation satellite and led the way to all the instruments available now [7]. Since 1972, a series of "heir satellites" continued to be placed in orbit every few years for sake of continued monitoring 1.3. Landsat popularity exploded when an open data license was launched in 2008 [8].

As demonstrated by tables 1.1 and 1.2, spectral bands of TM (Landsat 5), ETM+ (Landsat 7) and OLI (Landsat 8) sensors are quite similar and therefore can guarantee backward-compatibility to a large extent. This gives the chance to apply the same technique over multiple years and -eventually- define long-term monitoring services of urban evolution over time in an area-of-interest (AoI).

Sentinel-2 main features

The Sentinel satellite constellations are the "space segment" of the Copernicus initiative. Copernicus is the European programme for the establishment of a European capacity for Earth Observation. ESA has developed -and will maintain- a constellation of satellites designed to meet the operational needs of this programme. These missions carry a range of complementary technologies, ranging from multi-spectral and radar to

Table 1.1: *List of Landsat TM and ETM+ bands*

Band	Wavel. (μm)	Spatial res. (m)	Description
BAND 1	0.45 - 0.52	30	Blue
BAND 2	0.52 - 0.60	30	Green
BAND 3	0.63 - 0.69	30	Red
BAND 4	0.77 - 0.90	30	Near Infrared
BAND 5	1.55 - 1.75	30	Short-wave Infrared
BAND 6	10.40 - 12.50	30	Thermal Infrared
BAND 7	2.09 - 2.35	30	Short-wave Infrared
BAND 8	0.52 - 0.90	15	Panchromatic (Landsat 7 only)

Table 1.2: *List of Landsat OLI bands*

Band	Wavel. (μm)	Spatial res. (m)	Description
BAND 1	0.43 - 0.45	30	Coastal aerosol
BAND 2	0.45 - 0.51	30	Blue
BAND 3	0.53 - 0.59	30	Green
BAND 4	0.64 - 0.67	30	Red
BAND 5	0.85 - 0.88	30	Near Infrared
BAND 6	1.57 - 1.65	30	Short-wave Infrared (SWIR) 1
BAND 7	2.11 - 2.29	30	Short-wave Infrared (SWIR) 2
BAND 8	0.50 - 0.68	15	Panchromatic
BAND 9	1.36 - 1.38	30	Cirrus
BAND 10	10.60 - 11.19	100	Thermal band (TIRS) 1
BAND 11	11.50 - 12.51	100	Thermal band (TIRS) 2

Table 1.3: *List of Sentinel-2 bands*

Band	Wavel. (nm)	Spatial res. (m)	Description
BAND 1	433 - 453	60	Coastal aerosol
BAND 2	458 - 523	10	Blue
BAND 3	543 - 578	10	Green
BAND 4	650 - 680	10	Red
BAND 5	698 - 713	20	Near Infrared 1
BAND 6	733 - 748	20	Near Infrared 2
BAND 7	773 - 793	20	Near Infrared 3
BAND 8a	855 - 875	20	Near Infrared 4
BAND 8	785 - 900	10	Near Infrared 5
BAND 9	935 - 955	60	Infrared 1
BAND 10	1360 - 1390	60	Infrared 2
BAND 11	1565 - 1655	20	Short-wave Infrared 1
BAND 12	2100 - 2280	20	Short-wave Infrared 2

atmosphere and ocean monitoring. In this dissertation, focus is on Sentinel-2A satellite, launched in June 2015 as the first component of a two-satellite constellation with the aim to produce high resolution optical imagery and ensure continuity for the former ENVISAT mission. The revisit time at the moment is 10 days, reduced to 5 when the twin satellite will be in orbit. Table 1.3 illustrates the spectral features of the Sentinel-2 sensor.

CBERS data

CBERS-2B satellite (result of the Brazil-China cooperation) is an optical satellite equipped with different sensors: HRC (2.5 meters spatial resolution), CCD (20 meters), IRMSS

(80 meters) and WFI (260 meters). An important advantage of CBERS is the open access policy for the archive and the continuous monitoring of cities in quick-growing economies like Brazil and China. High resolution HRC imagery is of particular interest for the definition of built-up areas.

VHR data

Commercial satellites offer a vast spectrum of collected bands along with a very high spatial resolution. The first commercial satellites, EarlyBird-1 and Ikonos, were launched in the late 90's. Since then, launched satellites provided ever increasing spatial resolution and collected bands. For example, Digital Globe's World View-3 sensor can reach 0.3 meters. VHR images can be used in the pre-event phase for the extraction of footprints and buildings-related information thanks to the high spatial resolution. They also proved to be useful in a post-disaster scenario as a tool to highlight the most affected areas.

1.4.2 SAR data

Spaceborne radar is an active sensor based on the emission and successive measurement of the electromagnetic response of targets. In particular, Earth Observation radar satellites usually work in the C, L and X bands, therefore with frequencies ranging from 1 to 10 GHz. In order to avoid ambiguity in locating backscattering sources, the acquisition cannot be nadiral. Synthetic Aperture Radar (SAR) is an evolution of the original radar principle based on the Doppler effect and is capable to overcome the inherent resolution limitations. SAR sensors are mostly used in interferometry to estimate crustal shift and deformation. In the context of damage assessment, SAR imagery is typically used to determine changes between pre- and post-event data. The main advantages in respect of optical remote sensing are the all-weather all-day acquisition capability and the oblique view, able to detect damage on walls and not only on roofs.

COSMO-SkyMed

COSMO-SkyMed is a constellation of radar Earth Observation satellites designed by the Italian Space Agency with the capability to serve at the same time both civil and military users. The system is made of 4 satellites operating in X-band. Different acquisition modes are available, ranging from 100 meters (ScanSAR) to 1 meter (SpotLight) spatial resolution.

TerraSAR-X

A public-private joint venture between the German Aerospace Center (DLR) and Astrium managed to set up an X-band radar Earth Observation satellite called TerraSAR-X. The system is composed by two satellites: one dedicated to radar data acquisition while the twin satellite is used to collect data synchronously and produce an accurate Digital Elevation Model. Main imaging modes are SpotLight (up to 1 meters), StripMap (up to 3 meters) and ScanSAR (up to 16 meters).

Figure 1.4: *Example of picture taken from a smartphone.*

ALOS-2 PALSAR-2

ALOS-2 is the successor of ALOS, designed and launched by the Japanese Space Agency (JAXA). Unlike COSMO-SkyMed and TerraSAR-X, this system is equipped with 3 different sensors: a panchromatic visible and near infrared instrument, a stereo mapping sensor and a Phased Array L-band SAR sensor, called PALSAR-2. Like similar missions, spatial resolution can vary from 3 to 100 meters depending on the acquisition mode.

1.5 Crowdsourcing and citizen science

Until recent years, all geographic information was produced by specialized agencies and corporations, and distributed in form of paper maps or atlas. New technologies from the early nineties gave the possibility to citizens to determine their own position accurately using a GPS receiver or Internet services. The combination of this fundamental capability with the diffusion of smartphones and tablets, paved the way to what can be pictured as a "dense network of observers" (see figure 1.4), which has demonstrated its effectiveness in supporting and complementing Earth Observation.

According to a seminal paper published in 1995 [10], the term "Citizen Science" conveys both senses of the relationship between *science* and *citizens*. In particular, the definition evokes a science which assists the needs and concerns of citizens rather than "doing science" in a strict sense; at the same time, citizen science implies a form of science developed and put into effect by citizens themselves at large. It should be noted that the expression of the former concept implies no criticism on "doing science for sake of science" as opposed to "steering science towards people's needs", nor does it intend to be a contribution to the evergreen debate of whether "basic science" or rather "applied science" should receive more public resources on the grounds that they will produce greater return to the public good. It is instead a mere categorization of an increasingly observed phenomenon; it is still to be recognized that its effects may be constructive [11], and generally achieved without consuming public resources.

Even if this original concept was initially proposed in the sociology and psychology ambit, it is now commonly referred in several scientific branches; for example,

citizen data scientists represent a very hot topic in the big data world [12] [13] [14]. A COST (European Cooperation in Science and Technology) Action was initiated in 2012 entitled "Mapping and the Citizen Sensor" with the aim to enhance the role of citizen sensing in mapping [15]. A European Citizen Science Association was also established in June 2013 during the EU Green Week [16]; its ten principles, listed in the referenced website, clearly specify that both the professional and the citizen scientists should benefit from taking part to the project or task, for example considering the learning opportunities as a reward.

Looking at Earth Observation, the term "Citizen Science" has started circulating with increasing frequency more or less at the same time, until it triggered the organization of sessions in international conferences such as the invited session of citizen science for Earth Observation [17] at IGARSS 2015 (IEEE International Geoscience and Remote Sensing Symposium 2015); or even dedicated conference and workshops like the Earth Observation Open Science conference [18] organized by the European Space Agency (ESA); a wealth of research proposals and several funded projects [19] [20] [21] [22]. The term "citizen science" also applies to the Earth observation context, declined to two possible meanings (see figure 1.5) . The first one is related to "general citizen science", i.e. similarly to other disciplines, using and fiddling with data to learn and/or to provide one's own contribution. Examples are e.g. the GalaxyZoo Project [23] [24], where users can download and classify Deep Space Observation data, and the GeoWiki project [25], where users can provide intelligent feedback on previous EO-based land cover classifications. The second declination of Citizen Science in Earth Observation is declined into the *Citizen Sensor*, meaning that people can contribute to tackling a specific challenge by providing in-situ data. This may sound weird for a discipline which has always been -by its own nature- constantly centred on measurements from remote locations, but this combination of in-situ and remote sensing has a great potential indeed.

Different types of acquisition media and pro-activeness of the contributor determine the sub-type of "citizen sensing". In "active citizen sensing", a specific action by the contributor generates the data which is then submitted to the system. In "passive citizen sensing" the data is harvested from public repositories of data such as social networks. In the next subchapters we will discuss these different paradigms of Citizen Science in the context of Earth Observation.

1.5.1 General: "Playing" with data (raising awareness, organizing and disseminating knowledge)

The idea comes from the original approach of Public Participation Geographic Information Systems (PPGIS) [26] where the practice of mapping, including through GIS systems, is delegated to the local level. Participating people will donate their local knowledge and time to generate maps, which thus will incorporate information that would not be otherwise available. In exchange they learn new skills in mapping and geographic information systems. Volunteered Geographic Information, known as VGI, takes advantage of the Web to create and disseminate geographic information provided by individuals [27] [28] and represents another concept onto which collaborative mapping is based. The *Did You Feel It?* (DYFI) service from the United States Geological Survey (USGS) [29] collects perceptive information from people who felt an earth-

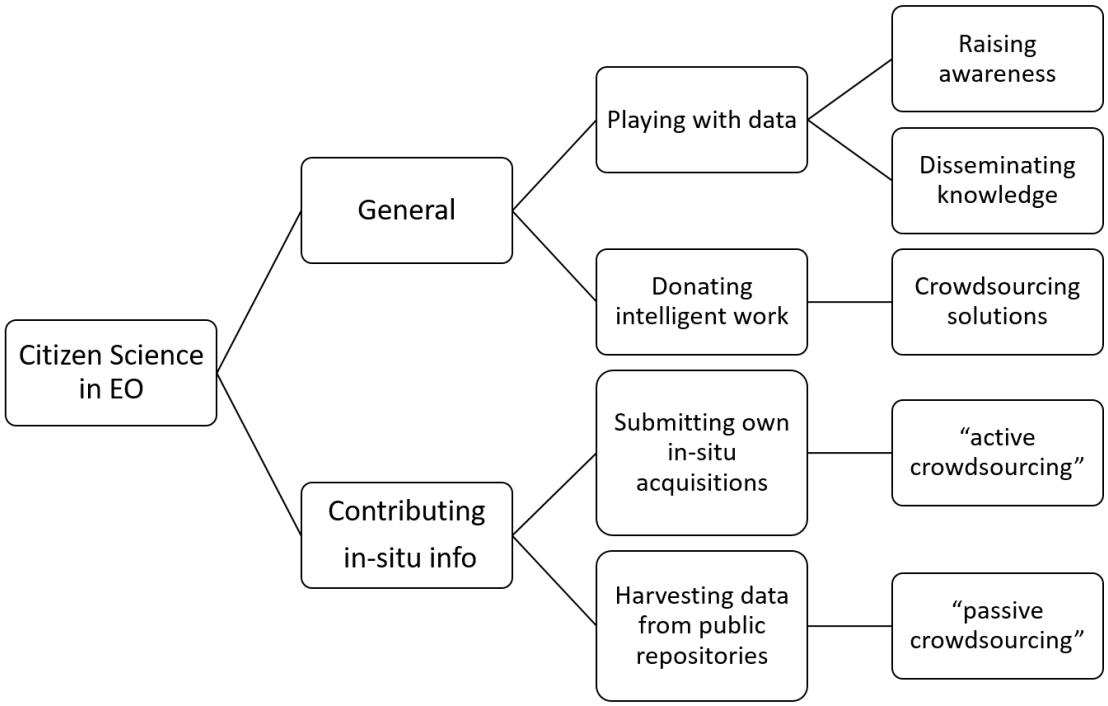


Figure 1.5: *Citizen Science in EO.*

quake; as reported in [30], since the launch of the service in 1997, until June 2015, more than 2.8 million have been collected. Although the resolution of the generated map was not very high, the project contributed to raising awareness in citizens about seismic hazard and its mechanisms. Several other examples may be cited, like the Citizens and Remote Sensing Observation Network (CARSON) initiative of the U.S. National Aeronautics and Space Administration [31] or the "Earth" section of the "ESA for Kids" [32] initiative from the European Space Agency. The former encourages citizens to explore satellite data to make local-to-wide-scale environmental observations, and to connect and compare them with official environmental data from authoritative sources. The latter is offered to younger citizens and presents a series of animations and conceptual experiments which usher in the newcomers to the world of Earth observation from space.

1.5.2 General: Donating intelligent work (crowdsourcing solutions)

The term geospatial crowdsourcing [33] is itself quite general and broad, as very different actions may fit under this term. On the one side, the scientific community refers uses it to mean outsourcing micro-tasks to several expert or non-expert users; the degree of knowledge necessary to accomplish these micro-tasks differs according to the nature of the work itself. In case of recent major disasters, for example, several initiatives were started to assess the damage location and extent from Earth observation data relying on the on-line community at large. The affected area is usually split into smaller pieces submitted to registered volunteers, whose role is to detect and outline damage, commonly in the form of points [34] [35] [36]; one of the most famous crowdsourcing

campaigns of this genre was the search of the missing Malaysia plane MH370 [37], suddenly disappeared from the radar screens while flying over the ocean, and whose remnants have only been found in part and long later.

In cases like these, no highly specific knowledge is required to fulfil the assigned tasks, and an operational manual with some practical examples is usually sufficient to make a broad audience capable of contributing. This type of activity does fit the broad definition of citizen science but it is not directly linked with the *citizen sensor* definition.

Moving closer to the focus of this paper, in the following subsection we consider another meaning of crowdsourcing, looking at people as a "dense network of observers" and therefore matching the concept of citizen sensor.

1.5.3 Crowdsourcing for generating in-situ data

Crowdsourcing -referring to its meaning of *dense network of observers* from now on- can be divided into different subcategories according to the tools employed and to the degree of targeted efforts that lie behind the collected data. In terms of tools, the distinction hinges on whether the application requires additional sensors not commonly available on *smartphones*; there exist projects and initiatives which require participating volunteers to install, for example, small meteorological stations or pollution sensors. In terms of specific efforts, two independent approaches can be followed. In the first one, called *participatory sensing* [38], the data is actively generated and submitted, meaning that a specific mobile app or service is built to collect user-generated reports. In the second approach, defined *opportunistic sensing* [39] public data can be harvested from social networks and repositories, such as *Twitter* and *Picasa*, with end-users not directly involved in the process, and possibly not even aware of the data collection in progress. After the term "crowdsourcing" made its first official appearance on a magazine article [40] in 2006, it started circulating with increasing frequency also on scientific publications. It is interesting to note that the same trend can be highlighted for the terms "crowdsourcing", "citizen sensor" and "participatory sensing", as visible in figure 1.6, where yearly occurrences of the three terms are plotted from a search on the Scopus engine [41], which claims to be "the largest abstract and citation database of peer-reviewed literature". This may suggest that the implementation of crowdsourcing went hand-in-hand with distributed, volunteer-based data collection, which is now one of the most compelling declinations of it. In the context of Earth Observation, there exist two main declinations of crowdsourcing for the generation of in-situ data, using different strategies, according to whether the user is actively involved in the process or not, i.e. "active" vs. "passive" crowdsourcing. These two declinations are treated in the following sub-subsections.

Contributing in-situ information: Submitting own in-situ acquisitions ("active crowdsourcing")

Ferster et al. in [42] reviewed different implementations of EO using mobile personal communication devices with topics ranging from ornithology to identification of tree species and phenology of plants. The value of participatory sensing combined with Earth Observation is also demonstrated by a more recent paper by Kotovirta et al. in 2015 [43]. Topics range from biomass analysis to emergency management and water

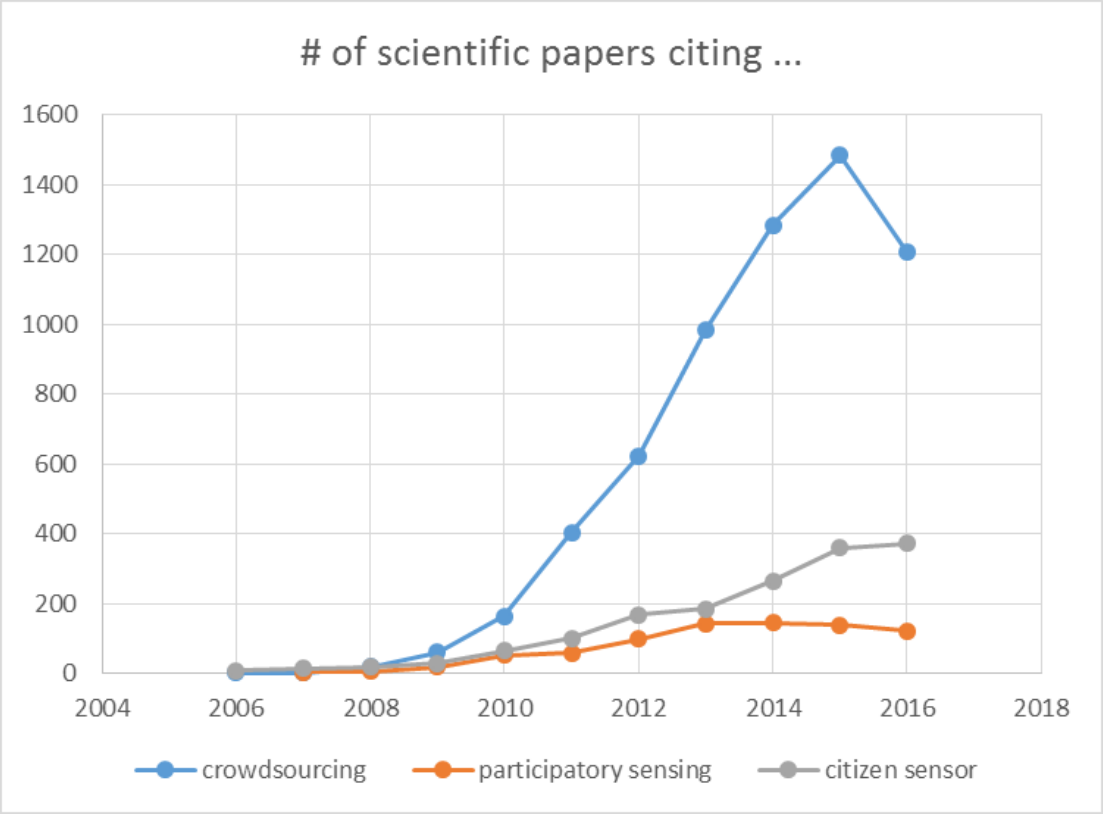


Figure 1.6: *Yearly occurrences of keywords "crowdsourcing", "citizen sensor" and "participatory sensing" in scientific literature as reported by the Scopus™ search engine. Please note that the sudden drop in the current year is expected, and not a sign of plunging interest. The search engine, indeed, will not predict the final datum for the current year but rather report the actual number of papers published until the search date.*

quality monitoring; for example, volunteers can contribute by supplying algae observations in the Baltic sea used in combination with Earth Observation data. Mobile and participative in-situ forest biomass measurements have also been proposed by Molinier et al. in [44] where the idea is to combine data gathered from a mobile app with GeoEye VHR optical data on the area of interest.

Several projects extensively involved volunteers as a key source of data. One example is COBWEB (Citizen Observatory Web) [21], funded under the European Union's Seventh Framework Programme (FP 7), theme Environment. Its main aim is to create a testbed environment which will enable citizens living within Biosphere Reserves to collect environmental data using mobile devices. Citclops [20], as quoted on its website, "aims to develop systems to retrieve and use data on seawater colour, transparency and fluorescence, using low-cost sensors combined with people acting as data carriers, contextual information (e.g. georeferencing) and a community-based Internet platform, taking into account existing experiences (e.g. Secchi Dip-In, Coastwatch Europe and Oil Reporter). [...] People will be able to acquire data taking photographs of the sea surface on ferries or other vessels, on the open sea or from the beach". The objective of EducEO [22] is to better understand how the power of citizen scientists can be channelled in order to get the most from Earth Observation products. There are naturally concerns over the quality and reliability of data collected through crowdsourcing [45], even though -apparently- experts and non-experts can perform comparably well when faced with a task not requiring highly specialized knowledge [46]. Still, malicious behaviour is a concrete threat since the beginnings of crowdsourcing [47], and obviously requires control mechanisms to be put in place besides simple compulsory registration, or dumb filtering of crowdsourced data.

Contributing in-situ information: Harvesting data from public repositories ("passive crowdsourcing")

Crowdsourcing may even be activated without voluntary contributors knowledgeably "pushing" reports to a centralized system for a specific purpose. A different paradigm of geospatial crowdsourcing consists of "pulling" (i.e. fetching and ingesting) georeferenced data made available on public repositories of e.g. georeferenced pictures. There are several possible sources to download from; examples are Instagram, which at the end of 2015 celebrated 400 million users reached, sharing 80 million photos per day [48]; Flickr, which on the same year could boast a 10-billion photos repository [49], growing at a rate of 1 million per day.

Twitter operating in Exact Location mode may also represent a source of georeferenced data, although only a few percents of the tweets are sent with precise geolocation attached [50]. Due to security concerns, the geolocation feature is deactivated by default, and needs to be explicitly activated by the user, which normally will not bother to do so. Tweets, however, hinge on textual contents rather than on pictures, calling for semantic analysis if useful information is to be extracted from them. While semantic analysis is out of the scope of this paper, tweets with pictures attached can also be considered a source in our context, given the free access to the tweet repository granted by Twitter.

In our perspective, these examples can be placed under a single umbrella as derived from the scientific literature.

Stefanidis et al. in [51] mentioned for the first time the concept of Ambient Geographic Information (AGI), arguing that social media represent a potential source of geographic information, although users may not disseminate or upload data for a specific goal other than amusement. Twitter feeds, for example, demonstrated their usefulness in responding to an earthquake, providing a new way for the identification and localization of the impact area of the event [52]. An equivalent concept is expressed by the term Social Geographic Data (SGD) introduced by Spyrators et al. in [53]. Passive crowdsourcing has also been defined as iVGI (involuntary voluteered geographic information) [54] because the users' actual intention is to share an opinion on social media and not to feed a geographic dataset. Estima and Painho in [55] propose a preliminary study using pictures harvested from Flickr and combined with CORINE Land Cover (CLC). All in all, the impression from our literature survey is passive crowdsourcing has just started to be addressed. While results are still limited, it appears to have a great potential especially considering the trend towards deep learning, which enables complex interpretation tasks to be carried out on e.g. harvested shared pictures [56].

1.6 Motivation and novelties

This thesis looks at the combination of different sources of information in order to obtain pre-event exposure-oriented data -taking advantage of remote sensing along with crowdsourcing- and combine it with post-event change detection measures computed after an earthquake. The SENSUM project, started in 2013, gave the chance to fill the gap in geospatial risk assessment, by stimulating the development of algorithms and workflows for the automatic and semi-automatic extraction of exposure and vulnerability indicators by using a combination of satellite and in-situ data (involving a StreetView-like camera). My main contribution is related to the SENSUM Earth Observation Tools QGIS plugin, you'll find more information following. Further development was performed in the framework of the RASOR and MARSITE projects, enhancing the capabilities of the proposed solutions and testing in different locations. To further enhance the visibility of the tools, a fruitful collaboration with the ESA Research and Service support led to the integration of part of the tools into the Grid Processing On Demand architecture, therefore giving the chance to simplify the know-how required to run the algorithms. However, the central idea is to include different sources of information, thus relying on volunteers, in order to fill the gaps left by spaceborne remote sensing. In particular, the aim was to build a mobile app to collect geotagged pictured seasoned with easy-to-recognize features related to buildings (marking a difference with expert-oriented tables); the SEGUICI and CLOOPSy projects gave the chance to build a general framework for geotagged data collection but the target was more on vegetation and general land cover. Implied concept, which I did not have the time to explicitly develop for a lack of time, is to modify a few parts of the more stable CLOOPSy framework (officially released with an EUPL software license in the end of September 2016) in order to collect the desired buildings-related features. Final task was then to combine pre-event exposure information -derived from space and ground- with post-event seismic damage assessment, computed from satellite imagery. The purpose was to determine the beneficial contribution of the proposed solutions for both risk assessment (not directly tackled in this dissertation) and damage estimation.

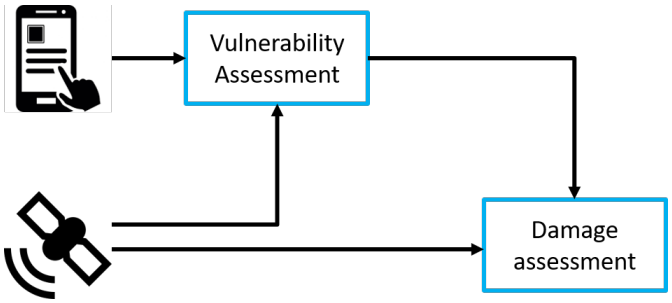


Figure 1.7: *Data flow from pre- and post-event modules.*

The idea is inspired by a paper published by Dell'Acqua et al. in 2013 [57] with two main additional steps: the use of participative sensing from the vulnerability side and the different methods used to extract damage from satellite. Possible end-users of the pre-event exposure assessment are mainly represented by stakeholders and decision-makers, whose theoretical role is to plan actions with the goal to reduce exposure and vulnerability in hazardous areas. On the other side, damage assessment is always useful for search and rescue activities and the civil protection departments. The schema displayed in figure 1.7 further clarifies the mentioned data flow and how different data sources are integrated within this dissertation.

CHAPTER 2

Pre-event exposure information: the Earth Observation Tools

Remote sensing is a very important source of information in risk and damage assessment. Regarding the former goal, a set of exposure and vulnerability proxies was identified and discussed within the EU FP7 project SENSUM [58]. The objective was to provide algorithms designed to ingest Earth Observation imagery and extract those indicators reducing the user input to the least possible amount. A stable version of all the tools was delivered, as scheduled, in December 2014; however, a few improvements and maintenance were performed in the framework of two other EU FP7 projects, RASOR [59] and MARSITE [60]. The 2015 IGARSS (International Geoscience and Remote Sensing Symposium) summer school on data fusion for risk mapping was held in Pavia and the SENSUM Earth Observation tools were explained and demonstrated.
This chapter will illustrate the mechanism behind each developed algorithm. Source code is available at the following repositories [61]. Several papers have been published related to the tools developed for the mentioned projects: [62], [63], [64], [65], [66], [67], [68], [69], [70]

2.1 Vulnerability indicators

The physical vulnerability of a building is defined as the probability of structural failure in the extreme situation of natural hazards like quakes. The concept of vulnerability has been changing since it started being considered as a risk element; in particular, it can be expressed in different forms depending on the desired typology of assessment:

- vulnerability indices: vulnerability is estimated with indicators that are not directly related to the hazard intensity and are generally useful to support a qualitative assessment

Table 2.1: *List of indicators for physical vulnerability assessment*

Indicator	Remote Sensing	Ancillary Data
Age of built-up area	Med. and high res.	Supplementary
Built-up area	Med. and high res.	Supplementary
Land use	Med. and high res.	Supplementary
Building density	Med. and high res.	
Building footprint	High res.	
Neighbourhood	High res.	
Building age	High res.	Supplementary
Roof type	High res.	Supplementary
Footprint regularity	High res.	
Building height	High res.	Supplementary
Building alignment	High res.	
Street network	Med. and high res.	Supplementary
Road width	High res.	Supplementary
Construction type	Supplementary	Yes
Public transportation network	Supplementary	Yes
Communication towers	Supplementary	Yes
Surface pipelines	Supplementary	Yes
Power Supply units	Supplementary	Yes
Bridges	Supplementary	Yes
Dams	Supplementary	Yes
Water supply units	Supplementary	Yes
Accessibility	High res.	Supplementary
Open space	High res.	Supplementary

- vulnerability curves: they specify a quantitative relationship between the intensity of the input generated by the considered hazard and the level of damage. They reflect the monotonic increasing relationship between the damage state and the intensity level of the "disaster" input. Such curves may imply an acceptable degree of uncertainty on the expected damage level.

Physical vulnerability can be presented from a wide spectrum of indicators, presented in literature. A draft list of exposure indicators related to earthquake and landslides hazards was established as a start-up contribution within the SENSUM project. This draft list was a collection of individual indicators extracted from separate studies in the literature. In a second stage, indicators were prioritized according to their possibility of being extracted from remotely sensed data, and to their inter-dependence. The adopted indicators for physical vulnerability should include two main features: they should reflect the difference in seismic performance of buildings and infrastructure and should consider other affecting factors that enhance or decrease the direct physical losses (e.g. the structure location). Mostapha Harb, former PhD student from the University of Pavia, originally took care of the proposed list, used as base to define the set of algorithms called SENSUM Earth Observation tools [70]. All the proposed indicators are listed in table 2.1, including the foreseen data input.

Table 2.2 shows the algorithms included in the SENSUM Earth Observation tools QGIS plugin [71]. This list of tools can basically be subdivided into two main groups: multi-purpose tools -basic bricks used for single and simple tasks- and indicators-related workflows -built by combining simple tools.

Table 2.2: *List of algorithms*

Icon	Name	Description
Pan	Pansharpening	Pansharpening algorithm from OrfeoToolbox
Cla	Classification	Unsupervised/Supervised classification from OrfeoToolbox
Seg	Segmentation	Segmentation algorithms from OrfeoToolbox, TerraAIDA (InterImage) and Skimage (python library)
Fea	Features	Computation of spectral and textural features from segments
Cor	Co-Registration	Co-registration algorithm designed for medium resolution. SURF and FFT alternatives are included. While services of OpenCV library are used in our approach, the FFT algorithm comes from Numpy (python library)
Stck	Stack Satellite	Stack satellite workflow including co-registration and built-up extraction with 5 different methodologies
Chg	Unsupervised Change Detection	Automatic analysis of the outcome of the object-based built-up area extraction algorithms
Extr	Footprints extraction	Supervised extraction of building footprints (Landsat 7 only)
Hgt	Building height	Combination of shadows and footprints with acquisition date for height extraction
Den	Building density	Calculation of the density of building in an area of interest
Reg	Building regularity and alignment	Computation of alignment and regularity of buildings

2.2 Multi-purpose tools

Different basic operations can be accomplished using these tools. They have been integrated in order to provide end-users with a set of accessible and familiar pieces.

2.2.1 Pansharpening

This well-known operation fuses together multi-spectral and panchromatic images in order to combine high spatial resolution with multiple spectral bands. At first, the multi-spectral image is coregistered and resampled so as to match the panchromatic sample; next, a pixel-by-pixel fusion is computed. Many different pan sharpening algorithm are available in literature; the proposed solution makes use of the equation 2.1 implemented in the OrfeoToolbox (OTB) library [72].

$$PXS(i,j) = \frac{PAN(i,j)}{PAN_{smooth}(i,j)} \cdot XS(i,j), \tag{2.1}$$

where:

- i and j are pixels indices
- PAN is the panchromatic image
- XS is the multi-spectral image
- PAN_{smooth} is the panchromatic image smoothed with a kernel that fits the multi-spectral image scale.

2.2.2 Classification

Classification is a common procedure applied to generic datasets with the aim to produce clusters based on different features. Unsupervised classification is an automated method to cluster remote sensing data. Theoretically, it does not require any experience in satellite images interpretation for end-users. OTB implements the K-Means algorithm [73]. Required parameters are the number of classes and the number of iterations.

Supervised classification is better suited for advanced users, with background in remote sensing and visual image interpretation. The main input of this type of classification is a training set, basically a set of samples for each class; based on this collection, the classifier generates a thematic map. Several different algorithms are selectable from users:

- Normal Bayes [74] assumes that the features of the training set, for each class, are normally distributed. This implies that the whole distribution is assumed to be a Gaussian mixture. Using the training data, the algorithm estimates mean vectors and covariances matrices for every class and then it uses them to do the prediction.
- Support Vector Machine [75] maps feature vectors into a higher-dimensional space using a kernel function. An optimal linear discriminating function in this space is built or an optimal hyper-plane that fits into the training data. The feature vectors that are closest to the hyper-plane are called support vectors. The other vectors do not affect the hyper-plane (the decision function)

- Decision Tree [76] is a binary tree (tree where each non-leaf node has two child nodes). The tree is built recursively, starting from the root node. All training data is used to split the root node. In each node the optimum decision rule is found based on some criteria. To reach a leaf node and to obtain a response for the input, the prediction procedure starts with the root node.
- K-Nearest-Neighbours [77] takes the decision based on the nearest neighbours. The K-Nearest-Neighbours looks for the k-near elements to classify the element
- Random Forest [78] is a collection of tree predictors that is called forest. All the trees are trained with the same parameters but on different training sets and the input is classified with every tree in the forest. Each result is taken into account and the final decision is based on majority voting.
- Gradient Boosted Trees [79] improves the decision tree classifier. Like other boosting methods, gradient boosting combines weak learners into a single strong learner.

2.2.3 Segmentation

Segmentation is a fundamental step for object-based analysis. The basic idea is to group similar and neighbouring pixels. Several solutions are available, with algorithms collected from different sources:

- Felzenszwalb [80] makes simple *greedy* decision. The *scale* parameter coarsely defines the size of the segments. The *sigma* parameter is the diameter of a Gaussian kernel, used for smoothing the image prior to segmentation.
- Watershed [81] processes the image considering it as a topographic relief. In a gray-scale image the peaks can be seen as *hills* (watersheds) while low intensity values are seen as valleys.
- Meanshift [82] segmentation consists in two steps which are filtering and clustering. The filtering step seeks modes or local maxima of density in the feature space. For each data point, the algorithm associates it with the nearby peak of the dataset probability density function. In the clustering step, each data point in the feature space has been replaced by its corresponding mode.
- Edison-meanshift [83] differs from the previous one for the clustering step that is achieved using a Region Adjacency Graph (RAG) to hierarchically cluster the modes. Furthermore, an edge detector is also used to better guide the clustering step.
- Baatz-Schäpe [84] doesn't limit to the digital number of each pixel but, it also take advantage of other details such as color, size, shape and texture.
- Morphological profiles [85] uses morphological operators such as erosion, dilation, opening, closing, top-hat transforms, etc.. A multi-scale approach is obtained using a range of Structured Element (SE) in a so-called Derivative Morphological Profile (DMP).

- Region growing [86] is the simplest segmentation algorithm. It requires initial seeds points as input. The neighbouring pixel of each seed are analysed and then, the decision to merge them is based on user-defined rules.

2.2.4 Features

Every segment can be used to compute features from the original input image. Two large groups are provided: spectral and textural features. For the former case, each pixel found within the segment is used to compute the features; for the latter circumstance, the smallest rectangular window around the segment is used to calculate the GLCM. Calculation is performed for each band. A schematic workflow of the algorithm is displayed in figure 2.1.

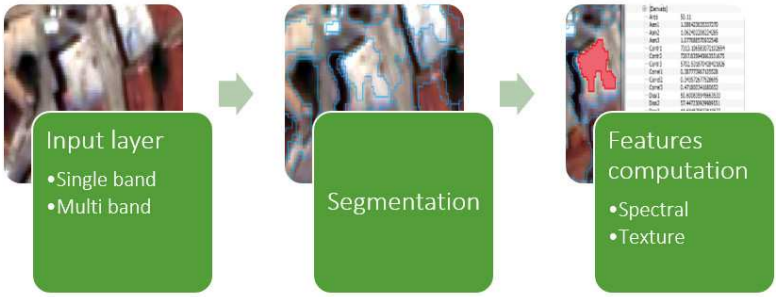


Figure 2.1: *Workflow of the feature extraction algorithm.*

- Mean

$$\mu = \frac{\sum_{i=1}^{N}}{N}, \tag{2.2}$$

- Standard Deviation

$$\sigma = \sqrt{\frac{1}{N} \cdot \sum_{i=1}^{N} (x_i - \mu)^2}, \tag{2.3}$$

- Mode
- Maximum, Minimum and Weighted Brightness

$$wB = \frac{1}{I \cdot J} \cdot \sum_{i=1}^{I} \sum_{i=1}^{J} p_{i,j}, \tag{2.4}$$

- NDVI Mean and Standard Deviation

$$NDVI = \frac{NIR - Red}{NIR + Red}, \tag{2.5}$$

Textural features are derived from the GLCM, considering i and j row and column number of the co-occurrence matrix. $P_{i,j}$ is defined as follows:

$$P_{i,j} = \frac{V_{i,j}}{\sum_{i,j=0}^{N-1} V_{i,j}}, \tag{2.6}$$

- Angular Second Moment

$$ASM = \sum_{i,j=0}^{N-1} P_{i,j}^2, \tag{2.7}$$

- Contrast

$$Contrast = \sum_{i,j=0}^{N-1} P_{i,j}(i-j)^2, \tag{2.8}$$

- Correlation

$$Correlation = \sum_{i,j=0}^{N-1} P_{i,j}[\frac{(i-\mu_i)\cdot(j-\mu_j)}{\sqrt{(\sigma_i^2)\cdot(\sigma_j^2)}}], \tag{2.9}$$

- Dissimilarity

$$Dissimilarity = \sum_{i,j=0}^{N-1} P_{i,j}\cdot|i-j|, \tag{2.10}$$

- Energy

$$Energy = \sqrt{\sum_{i,j=0}^{N-1} P_{i,j}\cdot(-\ln P_{i,j})}, \tag{2.11}$$

- Homogeneity

$$Homogeneity = \sum_{i,j=0}^{N-1} \cdot\frac{P_{i,j}}{1+(i-j)^2}, \tag{2.12}$$

2.2.5 Co-Registration

In order to build a reliable dataset, repeatability and consistency must be achieved and guaranteed for the images of interest. Misregistration, presence of clouds and differences in the illumination represent the biggest issues. Images of the same scene sometimes turned out to be shifted by meters, tens of meters and sometimes even hundreds of meters, which is unacceptable for some applications. The co-registration solution was initially developed within the SENSUM project for Landsat data because several considered scenes were affected by this issue; major improvements (e.g. registration of CBERS) have been added in the framework of the TOLOMEO project during the exchange period at the Pontificia Universidad Catolidade (PUC) Rio de Janeiro. Goward et al. in [87] presented an inclusive outlook on the desired geometric corrections for a proper extraction of information from remote sensing images. He categorized the registration into three sets: band-to-band, image-to-image and image-to-map. Focusing on the image-to-image approach, registration is performed assuming one image as reference and correcting the other images accordingly. The most common method used for correction is manual, meaning that a set of tie points is determined by hand

and used to compute the equation coefficients. As the accuracy of the manual registration relies essentially on the correctness of the selected tie points, road intersections form reliable point-candidates for registering built-up areas. However, when the set of images to be aligned is large, manual registration would be cumbersome and time-consuming therefore impractical. Moreover, the visual selection of a set of tie points is challenging particularly from low and medium resolution images. The difficulties are largely attributed to the changes of the components contributing to the pixel intensity value, and the possible indistinct appearance of building clusters where urban areas and moderate resolution images are concerned. Automatic registration approaches can be divided into two major categories: area-based and feature-based techniques. On one hand, area-based techniques use pixel values either in correlation approaches (e.g. normalized cross correlation and mutual information) or in the frequency domain (e.g Fourier-based). On the other hand, feature-based methods can also be separated into two categories according to the specified features: low-level features (e.g. edges and corners) or high-level features (objects e.g. water bodies contours).

Two methods have therefore been implemented in the plugin: a FFT-based approach and a feature-based process. The former is based on a scientific paper by Tzimiropoulos et al. [88]. The Fast Fourier Transform is applied directly to reference and target images; next, the amount of shift is derived from the translation property of the Fourier transform. Assuming I_1 and I_2 as the input images with the same size and related by an unknown translation.

$$I_2 = I_1(x + t) \Rightarrow \hat{t} = \arg_u max\{C(u)\}, \tag{2.13}$$

where:

$$C(u) = I_1(u) * I_2(u) = \int I_1(x) I_2(x + u) dx \Rightarrow C(u) = F^{-1}\{\hat{I_1}(k)\hat{I_2*}(k)\}, \tag{2.14}$$

The latter, instead, was designed within the project and is built upon a feature-based technique -Speeded Up Robust Features (SURF) [89]- used to determine common points and consequently the amount of shift between the considered images. This algorithm is capable to extract points-of-interest between the reference and floating image (image to be fixed) in the form of blobs; the process is based on the combination of integral images and the approximation of second order Gaussian derivatives with box filters: the result is a very fast and reliable algorithm. Every extracted feature is labelled using the BRIEF descriptor (Binary Robust Independent Elementary Features): pixel intensities are compared on location pairs around the feature of interest leading to a binary string output [90]. The matching step is performed using the Brute-Force matcher: every descriptor is compared to all the others and the one with the minimum Hamming distance (because of the binary nature of the descriptor) is chosen. According to the hypothesis of linear shift, a slope filter was designed and combined with a threshold on the Hamming distance. The latter is based on the selection of the three points with lower hamming distance. The former instead is based on the deviation of the values coming from the first distance-based selection. Goal of filtering is to remove unreliable points and to avoid the error propagation to the next steps. Output before and after filtering is shown in figures 2.2 and 2.3.

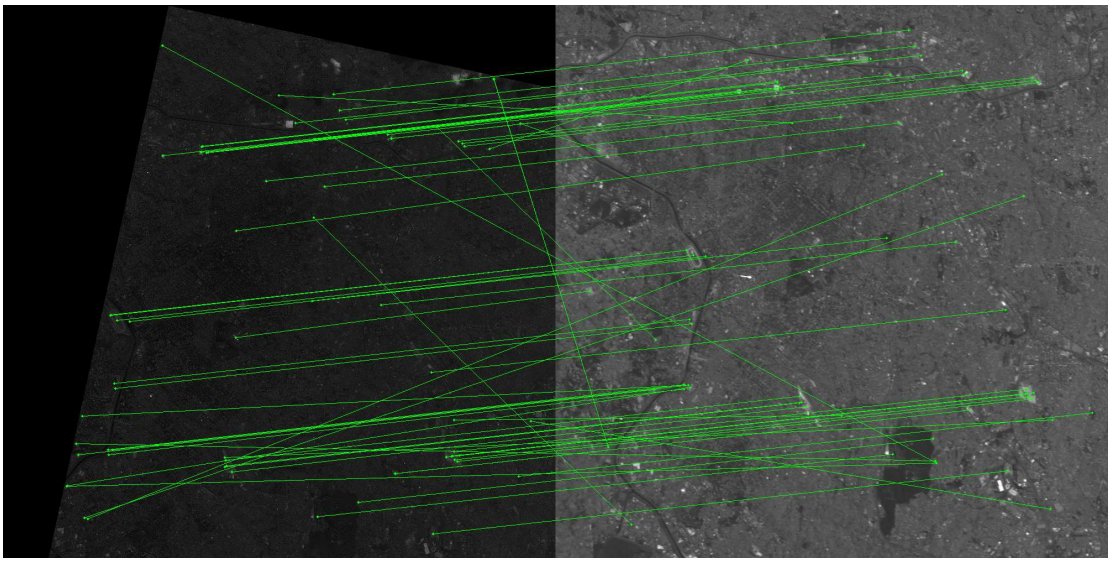


Figure 2.2: *Output of detection, description and matching steps.*

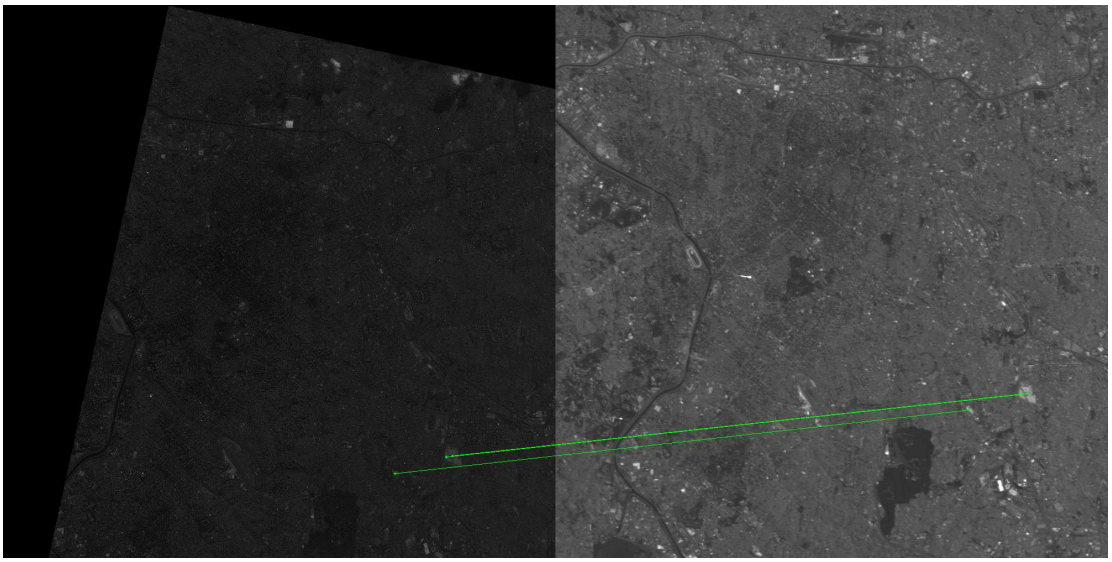


Figure 2.3: *Matching points output of the filtering process.*

The first proposed solution (2.4) involved the use of user-defined area-of-interests, mainly delineated around urban areas. This simple approach could be justified by the interest in monitoring urban areas, one of the vulnerability indicators. Moving to a more general approach -considering a full Landsat or CBERS scene- the proposed solution showed inherent limitations in determining matching points and also in processing time; moreover, the FFT-based method demonstrated its disadvantages in working on full scenes.

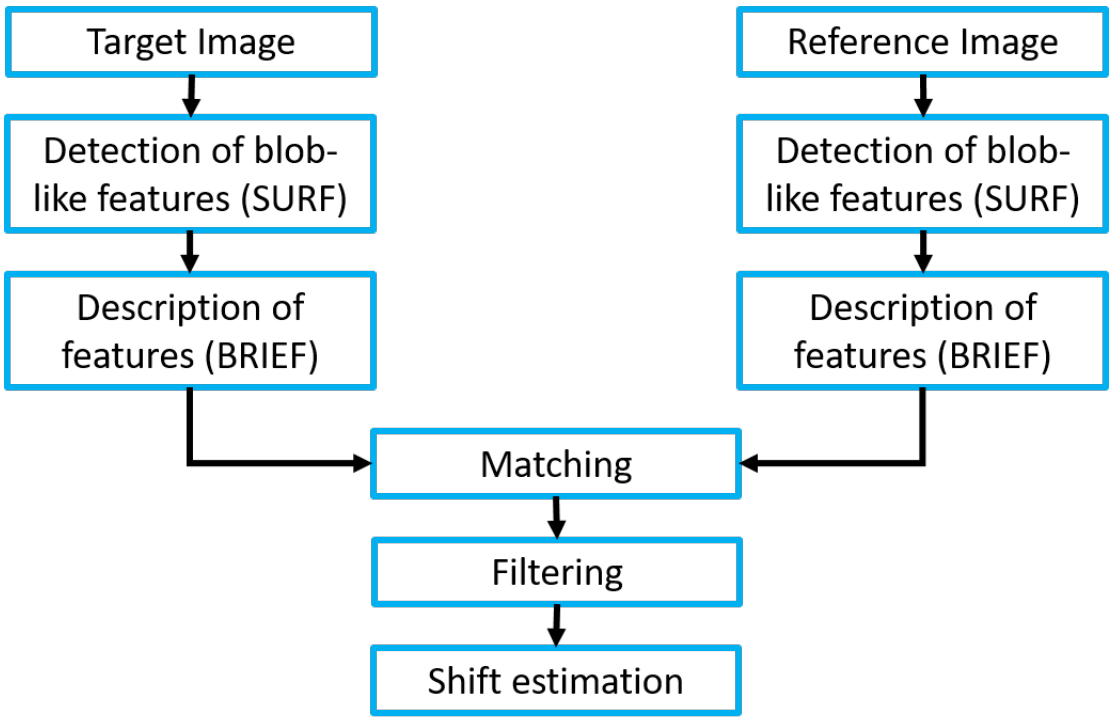


Figure 2.4: *Original SURF-based co-registration workflow.*

A modified version of the co-registration algorithm was therefore developed within the TOLOMEO project, including major improvements.

Automatic tile selection

First, a pre-processing step able to automatically select the best tiles to feed the area- or feature-based methods. The aim is to automatically extract subsets of the input image trying to avoid clouds -a major issue in registration-, reduce processing time and look for high variability in order to improve the consequent feature selection. Two different solutions have been designed and integrated: grid-based and unsupervised-classification-based. The former performs a division in regular tiles and computes statistics for each tile (e.g. frequency of maximum and minimum values, distance between max and min and the frequency standard deviation). For example, figures 2.5a and 2.5b show a selected tile and relative histogram while figures 2.5c and 2.5d display an unfavourable case.

The unsupervised-classification-based method clusters multispectral data in 5 classes 2.6a (defined empirically according to what can usually be seen from medium resolution data) and computes statistics like the maximum and minimum values, the standard deviation, etc. The class showing the highest variation is extracted 2.6b and processed with a morphological filter 2.6c in order to remove salt&pepper noise produced by the classification step. In the end, the polygon with the biggest area is selected 2.6d.

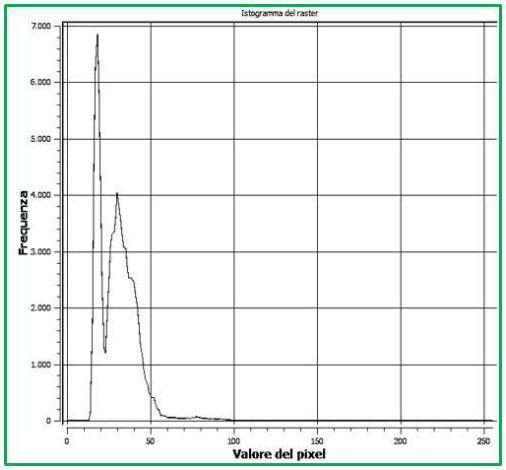
CCD-to-CCD co-registration

A shift is recorded among scenes acquired from different sensors at the same time; of particular interest, the goal is to match the panchromatic and the multispectral data.

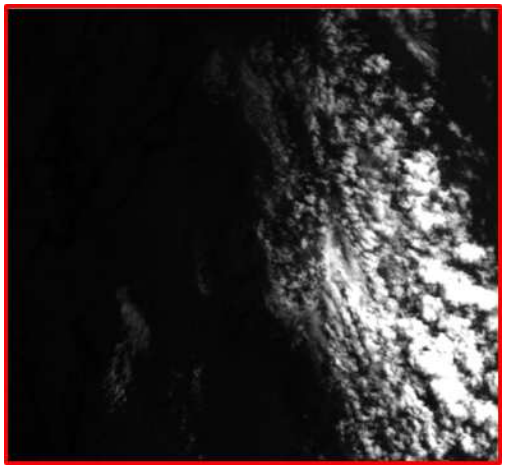
Figure 2.5: *Example of different tiles with relative histogram.*

(a) *Example of selected tile*

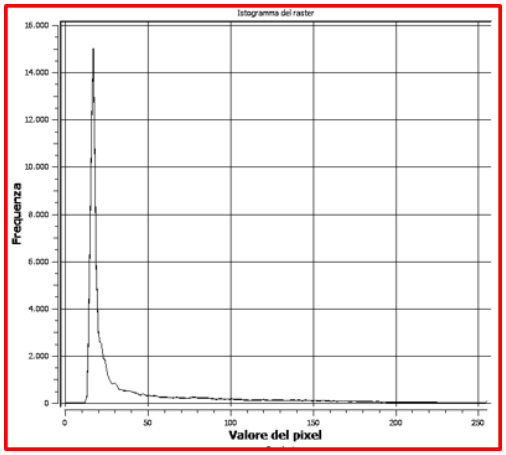
(b) *Built-Up index*



(c) *PCA unsupervised classification*



(d) *Dissimilarity hybrid-based*



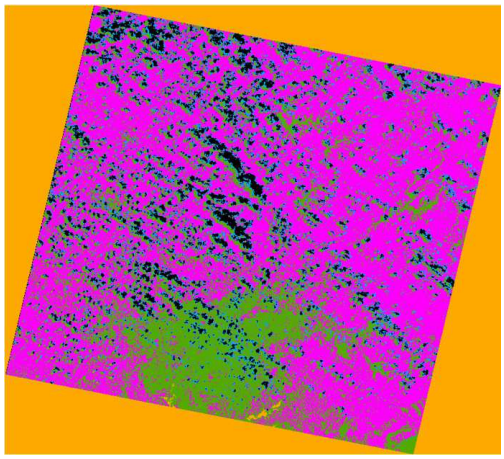
Based on the assumption of linear shift, the proposed SURF-based method has been used to register the input images. The obtained results are listed in table 2.3. Accuracy was determined by computing the shift among manually determined points before and after the registration procedure. The Root Mean Square Error (RMSE) reached 16.44 meters (0.82 pixels) while the Average Euclidean (AEE) scored 15.99 meters (0.80 pixels).

Table 2.3: *Shift between sample points computed before and after the registration process*

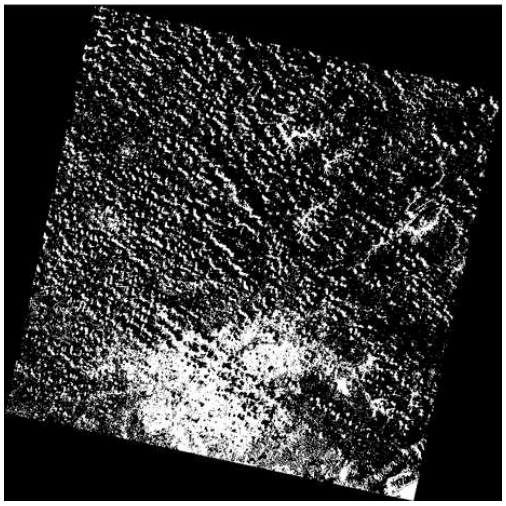
Point N	Dist. before (m)	Dist. after (m)	Dist. before (px)	Dist. after (px)
1	127.58	17.98	6.38	0.90
2	126.41	12.54	6.32	0.63
3	135.47	12.22	6.77	0.61
4	122.71	13.59	6.13	0.68
5	142.03	20.84	7.10	1.04
6	128.01	12.89	6.40	0.64
7	134.10	21.84	6.70	1.09

Figure 2.6: *Example related to all the unsupervised-classification-based method steps.*

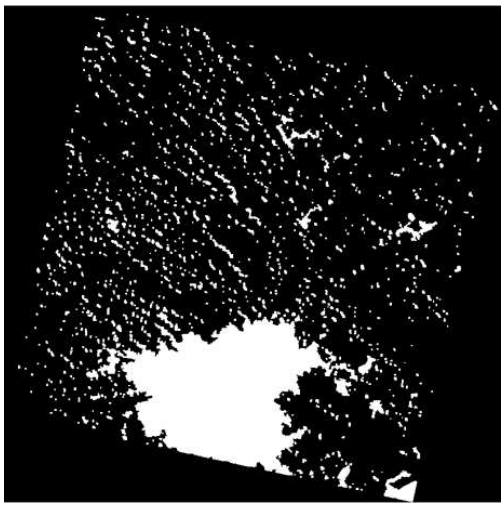
(a) *Outcome of the classification*



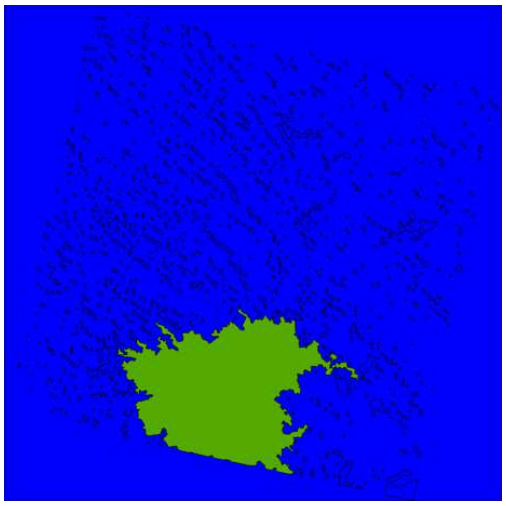
(b) *Mask built from the selected class*



(c) *Output of the morphological filter*



(d) *Maximum area selection*



CCD-to-HRC co-registration

High resolution HRC imagery is of particular interest for the definition of built-up areas. Unfortunately, CBERS images are affected by geographic misregistrations, presumably due to wrong attitude information [91]: for example, HRC data are misplaced among each other and do not overlay with CCD images covering the same area of interest. The type of shift was initially supposed linear; further measures (presented in tables 2.4 and 2.5) demonstrated that the shift can be roughly considered linear at the CCD resolution (20 meters). The values were determined by drawing polygons around noticeable areas and computing the difference of centroids locations. The approach starts by considering the high resolution image and down-sampling it to 20 meters in order to match the medium resolution.

Figure 2.7: *Example showing the deviation between CCD and HRC imagery.*

(a) *CCD images with a river highlighted in red*

(b) *HRC image of the same area (river highlighted in blue) as a comparison with the CCD-drawn line*

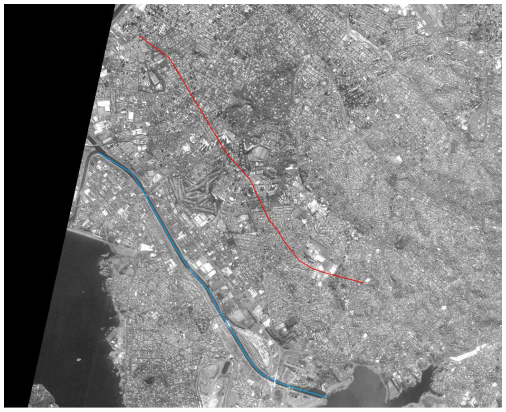


Table 2.4: *Manual estimation of shift - Rio de Janeiro case (HRC 2010-03-09, CCD 2010-03-09)*

Point ID	Δ X (m)	Δ Y (m)	Δ X deviation (m)	Δ Y deviation (m)
1	1389.25	415.34	-15.91	-9.55
2	1441.77	420.12	36.90	-4.77
3	1394.03	448.76	-11.14	23.87
4	1384.48	410.57	-20.69	-14.32
5	1408.35	405.80	3.18	-19.09
6	1413.12	448.76	7.96	23.87
AVG	1405.17	424.89		

Table 2.5: *Manual estimation of shift - Sao Paulo case (HRC 2008-07-10, CCD 2008-03-06)*

Point N	Δ X (m)	Δ Y (m)	Δ X deviation (m)	Δ Y deviation (m)
1	-5439.46	3920.38	-34.68	-32.81
2	-5431.62	3902.10	-26.85	-51.09
3	-5387.25	3983.02	17.52	29.83
4	-5408.13	3967.36	-3.36	14.17
5	-5371.59	3951.69	33.19	-1.49
6	-5408.13	3946.48	-3.35	-6.71
7	-5387.25	4001.29	17.53	48.10
AVG	-5408.78	3953.19		

Once the detection, description and matching steps are performed, a set of reliable points is derived. Once the collection of reliable matching points is complete, the misplacement is computed as average of the difference among the extracted points (average of the shift). The floating image is corrected according to the compute values: the matrix with geographic coordinates associated to the tiff image is edited according to the results. This technique is capable to preserve the content of the image avoiding distortions and data loss produced by warping procedures; this correction is for linear-shift only.

The CCD-to-HRC registration algorithm was tested on two different datasets over the cities of Rio de Janeiro (RJ, Brazil) (see table 2.6) and Sao Paulo (SP, Brazil) (listed in table 2.7). The RMSE and AEE values for the Rio de Janeiro case reached 5.47

meters and 29.95 meters respectively. For the Sao Paulo case, the RMSE stopped at 13.80 meters while the AEE at 190.45 meters.

Table 2.6: *Residual shift - Rio de Janeiro case*

Point ID	Δ X (m)	Δ Y (m)	Δ X deviation (m)	Δ Y deviation (m)
1	-14.32	-4.78	-7.95	0.39
2	28.64	-38.19	35.01	-33.02
3	-19.10	19.10	-12.73	24.27
4	-14.32	-23.87	-7.95	-18.70
5	-33.42	-7.16	-27.05	-1.99
6	14.32	23.87	20.69	29.04
AVG	-6.37	-5.17		

Table 2.7: *Residual of shift - Sao Paulo case*

Point ID	Δ X (m)	Δ Y (m)	Δ X deviation (m)	Δ Y deviation (m)
1	-120.67	-134.74	-4.49	15.66
2	-84.40	-98.16	31.78	52.24
3	-129.42	-169.62	-13.24	-19.22
4	-113.73	-144.64	2.45	5.76
5	-142.20	-229.38	-26.02	-78.98
6	-101.54	-108.64	14.64	41.76
7	-121.29	-167.60	-5.11	-17.20
AVG	-116.18	-150.40		

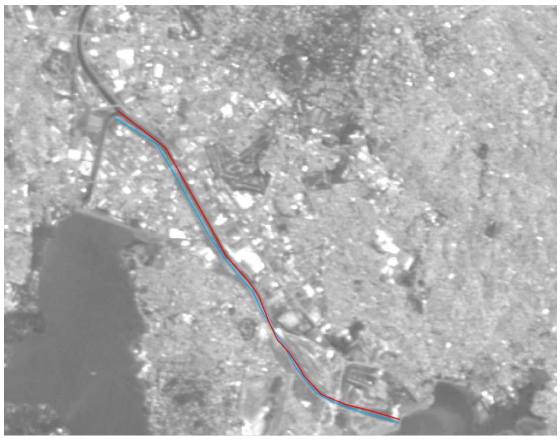


Figure 2.8: *CCD image after the correction. The blue and red lines are almost overlapping now.*

The obtained results show that a residual and not negligible shift is still noticeable, in particular for the Sao Paulo case. An additional improvement of the algorithm takes into account the availability of multiple HRC data for the same area of interest: the aim is to use more images for the correction in order to reduce variability of the estimated shift. For example, as demonstrated in figure 2.9, 5 HRC images could be picked randomly for the process.

Figure 2.9: *Set of all the HRC data superimposed on a CCD image over the same area.*

Next, the idea is to extract matching points for each intersection and filter those points according to their slope. A weighted average is computed in the end taking into account the frequency of detected points. The results are listed in table 2.8; the RMSE value reached 9.88 meters while the AEE amounted to 97.61 meters.

Table 2.8: *Residual of shift of the multiple HRC method - Sao Paulo case*

Point ID	Δ X (m)	Δ Y (m)	Δ X deviation (m)	Δ Y deviation (m)
1	-51.58	-50.84	1.15	5.79
2	-75.13	-75.85	-22.40	-19.22
3	-15.31	-29.07	37.42	27.56
4	-68.17	-145.49	-15.44	-88.86
5	-15.03	-85.42	37.70	-28.79
6	-91.67	78.89	-38.94	135.52
7	-52.20	-88.64	0.53	-32.01
AVG	-52.73	-56.63		

2.3 Built-Up Area extraction (medium resolution)

Earth Observation satellites, thanks to their recurrent revisiting and their capability to cover large areas with a sufficient spatial resolution, represent a useful tool in monitoring cities around the globe. Urban growth is one of the biggest challenges in terms of resource consumption, impact on natural and human systems and sustainability [92]. However, a global automatic extraction of built-up areas from different datasets still represents a major research challenge due to different environments surrounding urban areas and material of which buildings are made. Different solutions can be found in literature; for example, a low resolution global layer is described in [93] and [94]; Trianni et al. in [95] and Patel et al. [96] described the implementation of Landsat-based urban mapping based on the definition of a training set and running on Google Earth Engine. Finally, Pesaresi et al. in [97] proposed a solution for global human settlement mapping using high-res data but unfortunately the outcome is not released to the public yet.

A boost on the development of novel methods for EO-based urban area monitoring was triggered by the decision to release Landsat data with an open license. This decision suddenly threw open to Earth observation researchers an unprecedented opportunity to access a catalogue spanning over 40 years. Moreover, the recent launch of Sentinel satellites opened a new stream of open EO data, which further contributed to promoting the use of medium resolution satellite data for monitoring and assessing changes related to urbanization, climate change and pollution. Spectral and spatial similarity between those two open data sources, as illustrated in depth in the next section, offer a chance to assemble a complete and actionable data archive.

Due to the limited swath extension, though, AoIs may easily span across multiple swaths. As a consequence, different atmospheric conditions can appear within a single AoI, thus implying an even more challenging extraction of the desired information. Digital Numbers (DN) can be converted to Top-of-Atmosphere (ToA) radiance to compensate illumination issues using the following formula:

$$BANDn_{rad} = BANDn * M_n + A_n, \tag{2.15}$$

where M_n and A_n are respectively the band-specific multiplicative and additive rescaling factor. These parameters are included as metadata by the USGS [98] in its data distribution. The provided code automatically retrieves the necessary parameters for calibration.

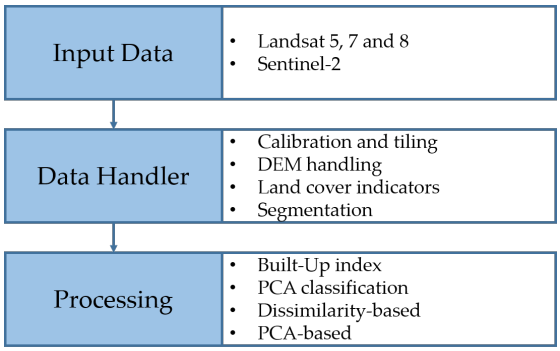


Figure 2.10: *Schematic view of the general implemented framework for built-up extraction.*

Similarities can be found in spectral band widths and in the number and spectral location of collected bands. A few differences can also be highlighted between Sentinel-2 and Landsat sensors, for example in the number of bands in the infrared range. The swath size of the European satellite is bigger and data are available from the repository as a set of tiles. Therefore, memory constraints could become a limiting factor in case of simultaneous processing.

For the following paragraphs, an ad-hoc, common set of band names among satellites is established; despite the possible confusion generated, this solution offers the possibility to discuss the process and data properties without being tied to one specific source.

Table 2.9: *List of bands used by the algorithm for each satellite*

Ref. name	Landsat 5	Landsat 7	Landsat 8	Sentinel-2
BAND1	B1	B1	B2	B02
BAND2	B2	B2	B3	B03
BAND3	B3	B3	B4	B04
BAND4	B4	B4	B5	B08
BAND5	B5	B5	B6	B11
BAND6	B7	B7	B7	B12

2.3.1 Data Handler module - Preparing data

A description of every step of the process follows. The aim of this module is to prepare the data and the complementary sources of information for being input into the core of the algorithm, i.e. extraction of built-up areas using different methodologies.

Area-of-Interest (manual)

The end-user is requested to outline a rectangular AoI, i.e. the area to be processed by the urban mapping algorithm, defined by the upper left and lower right corners. In case no specific outline is provided, the entire extension of the input data is considered.

Pre-processing (automatic)

A variety of pre-processing tasks are carried out here, for example checking from metadata whether cloud coverage is below the established ceiling, performing radiometric calibration (only for Landsat to compensate illumination issues) and merging all tiles into one single image for the subsequent steps. Landsat boundary effects are also removed using tiles provided by the USGS [99]. A distinction is in order here: due to the different distributions adopted, the approach of calibration and merging was designed for Landsat images. Sentinel-2 should not suffer from the same issue.

DEM mask generation (semi-automatic)

An optional feature included in the algorithm is the possibility of filtering the output obtained from the automatic methods with the SRTM DEM of the region of interest. This is helpful in removing false positives extracted by the algorithm because of the spectral confusion between bare soil and built-up areas. The assumption is that cities are commonly built in less irregular terrains, where irregularity is determined using roughness computed using the DEM as input. At this stage, the user can provide as input the SRTM 1-arcsec file downloadable from the USGS repository [100]. Further improvements are planned in the near future, which will activate automatic download of the data, in the local-machine version of the algorithm. Elevation files are automatically composed for the area-of-interest and roughness is computed. A default threshold value is provided to differentiate between mountains and flat land. User intervention is necessary in case of errors in mask definition. The decision-level is not based on the altitude itself in order to take into account high cities. The DEM mask is an optional step but it is highly recommended by the authors to improve the false-positive performance in the extraction. The next action is the calculation of land cover indicators, used as a

way to enhance the spectral difference between the different classes and facilitate the segmentation task.

Land Cover indicators (automatic)

Following different examples found in the literature [101], a few indexes can be computed as the ratio among different bands. The goal is to increase the separability among the different classes and facilitate the extraction of built-up areas. Most of the proposed indexes are already listed in literature. For example, the Normalized Difference Vegetation Index (NDVI) (equation 2.16) is the most widely used and known. Other useful indices are: Soil Adjusted Vegetation Index (SAVI), Normalized Difference Built-Up Index (NDBI) [102], Modified Normalized Difference Water Index (MNDWI) [103] (equations 2.17, 2.18 and 2.19).

$$NDVI = \frac{BAND4 - BAND3}{BAND3 + BAND4}, \tag{2.16}$$

$$SAVI = \frac{(BAND4 - BAND3) * 8.5}{BAND3 + BAND4}, \tag{2.17}$$

$$NDBI = \frac{BAND5 - BAND4}{BAND5 + BAND4}, \tag{2.18}$$

$$MNDWI = \frac{BAND2 - BAND4}{BAND2 + BAND4}. \tag{2.19}$$

Several attempts to enhance the spectral difference between urban areas and other classes (vegetation, bare soil and water) are reported in literature [104] [105] [106] [107] [108]. Unfortunately, most of them strongly depend on the surrounding environment, limiting the application to a few test cases. The choice made of offering different methods is consistent with the non-homogeneous spectral and textural characterization of urban areas across the globe. An additional index called built-up index was designed in order to further highlight built-up areas (equation 2.20).

$$BUILT - UP = \frac{BAND6 + BAND2 - 1.5 * BAND4}{BAND2 + BAND5 + BAND6}. \tag{2.20}$$

Segmentation (automatic)

Different methods are offered by the tool to achieve the final goal of extracting built-up areas. Two of the proposed methods are the so-called hybrid-based, taking advantage of the combination of segments with pixel-based measures in order to increase the homogeneity of the output and reduce errors. The algorithms used within the proposed workflow are the Meanshift and the Edison-Meanshift, both made available by the open-source Orfeo-Toolbox library. Default parameters are suggested; however, expert users can specify a customized set according to their needs. For sake of continuity and consistency, segments should be extracted from the most recent dataset (namely the most recent acquisition date) and applied to all the others. Pixel-based measures are aggregated by averaging all the values inside each segment.

2.3.2 Processing module - Core methods for built-up area extraction

Different methods are provided to end-users; the underlying assumption is that, due to different environmental and spectral conditions, it is really difficult to define a one-size-fits-all method capable of considering all the possible variations involved. The authors decided to provide the end-users with a set of different algorithms and let them decide which method fits best the AoI.

Built-Up index (semi-automatic)

Based on the built-up index mentioned earlier (see equation 2.20), the outline of an urban area can be extracted by simply applying a suitable threshold to a suitably computed index. The block diagram is displayed in figure 2.11. The method is pixel-based and requires user experience in the definition of parameters. It was proposed as a quick way for end users to get a grasp of where built-up areas are located (see figure 2.16b). It is not always easy to find a proper threshold, therefore the method is to be considered no more than just a support to the more structured and promising hybrid-based methods.

Figure 2.11: *Workflow of the built-up index method.*

PCA unsupervised classification (semi-automatic)

Principal Component Analysis (PCA) is a well-known mathematical technique often used for data fusion and noise removal. The input is linearly converted to a new coordinate system where the greatest variance lies on the first coordinate. Previous studies in literature illustrated its potential in combining multi-temporal Landsat scenes and also fusing different datasets [109] [110] [111] [112]. In the proposed technique, PCA is applied directly to a single-date set of bands and the outcome is passed on to a K-Means classifier (workflow in figure 2.12). User intervention is limited to the selection of class(es) related to built-up areas. Due to the different spatial distribution of the response recorded by the satellite, PCA can actively separate built-up areas from more-homogeneous desertic zones, which spectrally resemble urban areas. Thanks to all the PCA implementations available, the entire process is quite fast and results are acceptable. As for the built-up index, this solution is a way to get an idea of where the urban areas are located (see figure 2.16c).

Figure 2.12: *Workflow of the PCA unsupervised classification method.*

Dissimilarity hybrid-based method (semi-automatic)

The Haralick "dissimilarity" [113] texture is computed on bands 1, 2 and 6 in order to highlight built-up areas with respect to other classes [65] (see figure 2.14). The texture

is computed using a 7x7 window, a step of 3 and averaging over all the possible pixel-based displacement angles (0, 45, 90 and 135 degrees). The Gray Level Co-occurrence Matrix (GLCM) is quantized to 64 levels using a linear quantization algorithm removing the tails of the total distribution (cutting at 2% and 98% of the cumulative distribution). Parameters like window-size and step have been selected empirically by visual comparison of different results. The term hybrid comes from the combination of pixel-based and object-based techniques, also highlighted in the workflow presented in figure 2.13.

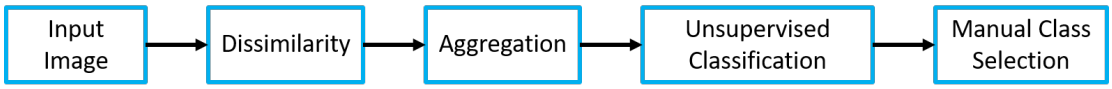


Figure 2.13: *Workflow of the dissimilarity hybrid-based method.*

From figure 2.14, it can be seen that the built-up area is characterized by a different set of values in respect of other land cover classes. However, a direct classification of the output would not help in directly extracting built-up areas due to the high level of noise. An example of the output is displayed in figure 2.16d.

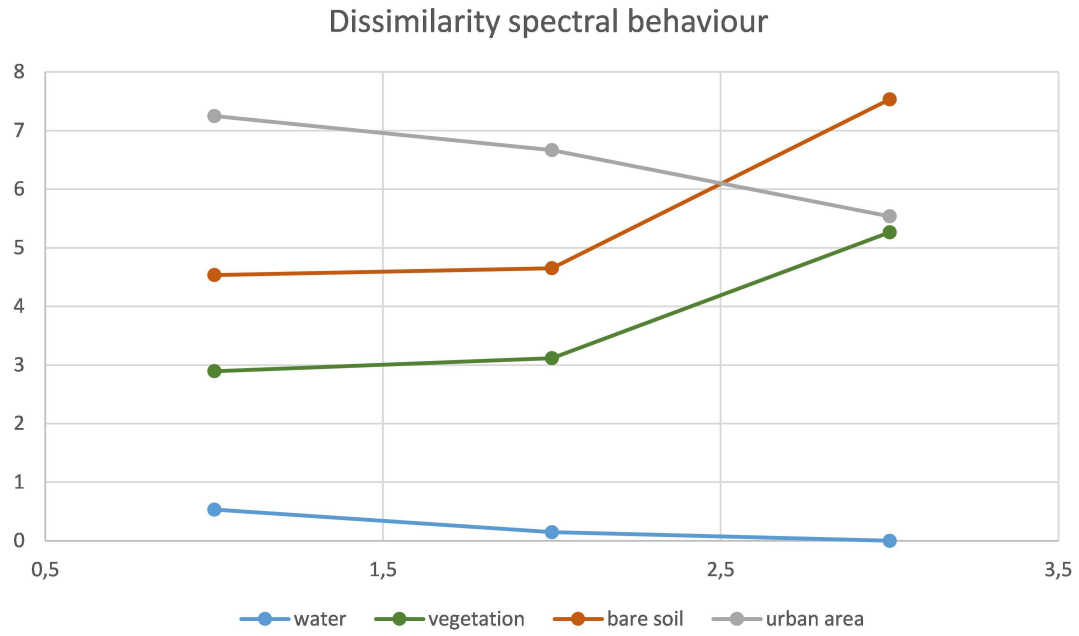


Figure 2.14: *Dissimilarity behaviour for different spectral classes.*

PCA hybrid-based method (semi-automatic)

This method builds on a PCA transformation; values are averaged within each segment and then classified using a K-Means classifier [69]. The aim is to reduce noise deriving from a standard pixel-based classification and to create more homogeneous areas taking advantage of segments. The approach is a variation of the dissimilarity-based one. The block diagram is shown in figure 2.15. Figure 2.16e shows a sample.

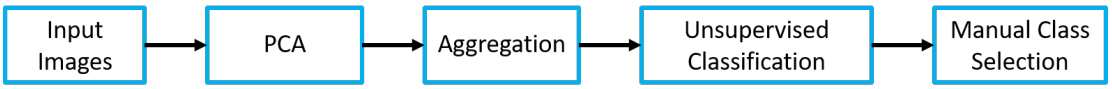


Figure 2.15: *Workflow of the PCA hybrid-based method.*

2.3.3 GPOD implementation

The potentially broad interest of the basic tools, together with the heavy computational load they sometimes require, make them ideal candidates for remote operation on computer grids. This was indeed done during my thesis, and this section reports about integration of the tools on the European Space Agency servers. The Research and Service Support provided by the European Space Agency [114] [115] has the mission to provide tools and services that support the user community in EO data exploitation. A fruitful collaboration with the ESA RSS team led to software adaptation and testing by exploiting the RSS CloudToolbox service for subsequent integration as GPOD service. Cloud computing has become increasingly popular in recent years; the main advantages are related to the easy access to end-users and powerful processing facilities. Part of the SENSUM Earth Observation Tools have therefore been integrated into the ESA RSS facilities. Thanks to the direct access to the data repository of the GPOD system, it won't be necessary to manually download the required imagery.

RSS CloudToolbox

The RSS CloudToolbox service consists of an on-demand provisioning of customised virtual machines equipped with pre-installed software according to the user requirements. The virtual machine is created upon request with flexible hardware resources depending on the user needs. This resource is hosted on a cloud infrastructure and it is accessible via a secure shell connection. This solution was adopted in order to avoid endless installation and compatibility issues while developing and adapting the algorithm in preparation of the GPOD integration. The SENSUM Earth Observation Tools have been successfully tested into this service and therefore they're ready for the integration into the GPOD environment.

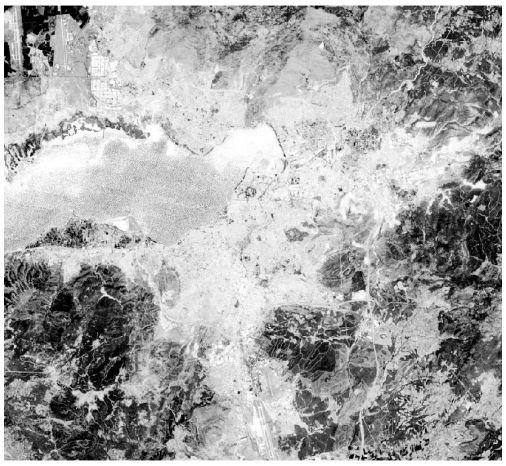
GPOD service

Following the objectives of the Big Data from Space challenge, a new parallel-driven structure has been designed to run on the ESA GPOD system [116] provided by the agency to support Earth Observation research, development and data exploitation [**?**] [117]. The algorithm, known as Stack Satellite [65] [69] has been re-arranged and configured in order to retrieve data from the ESA repository and directly process it. The parallel architecture and the time saved in configuring the machine and downloading the data are among the main advantages. The processing workflow "backbone" is unchanged, meaning that the algorithms designed to extract built-up areas are formally equivalent to the ones included in the QGIS plugin. Major adjustments, like re-writing the most demanding operations in Cython, have been carried out to further reduce the amount of time needed while keeping the same algorithm. The data flow has been redesigned, instead.

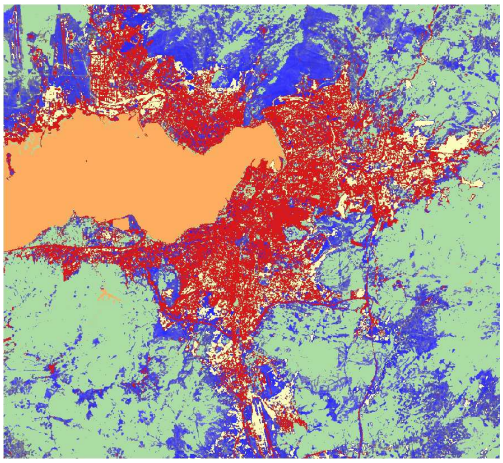
Figure 2.16: *Example of output for each method on the case study of Izmir, Turkey. Accuracy figures will be presented later.*

(a) *True colour rendition of original Landsat data*

(b) *Built-Up index*

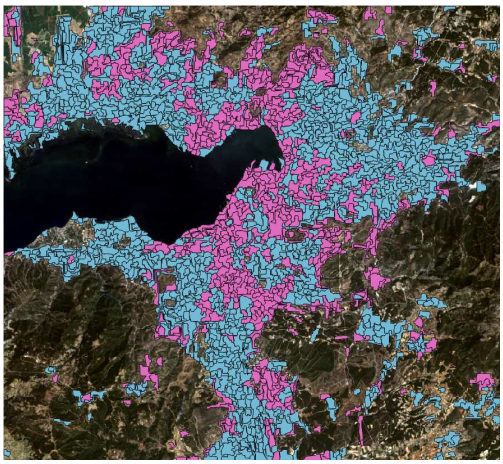


(c) *PCA unsupervised classification*



(d) *Dissimilarity hybrid-based*

(e) *PCA hybrid-based*



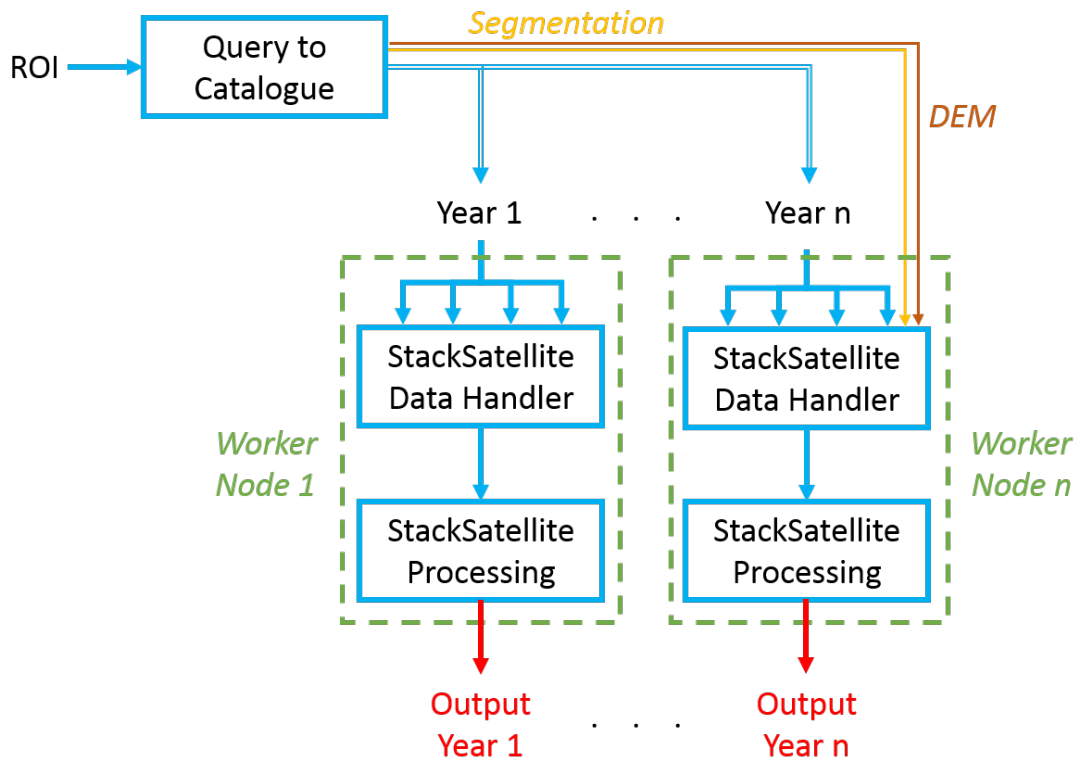


Figure 2.17: *New parallelized structure.*

The new structure, shown in figure 2.17, takes advantage of multiple worker nodes depending on the amount of data to process. The *data handler* module (DH) converts the set of input imagery to a stack of bands over the area of interest. If requested, DEM mask and segmentation are computed using the most recent year, aiming at a stable set for all the other inputs; these intermediate products are then propagated to all the worker nodes involved. The *processing* module (PR) includes the different built-up extraction algorithms, including 3 pixel-based and 2 hybrid-based modules.

The user experience is quite straightforward (see figure 2.18):

1. Define the region of interest (a rectangle), the desired range of dates and the data set; a query to the catalogue is submitted in order to retrieve a list of matching items.

2. Choose from the list according to what is the objective of the extraction (e.g. the year gap) and define the methods to apply.

Once the inputs are defined, processing tasks are distributed within the system according to the available resources. The user is notified as soon as the process is completed with results available from the "Results visualization" tab.

2.3.4 Results

Different case studies and truth layers were used to test the algorithm. Izmir and Van were selected by the SENSUM project as resident test cases therefore the proposed so-

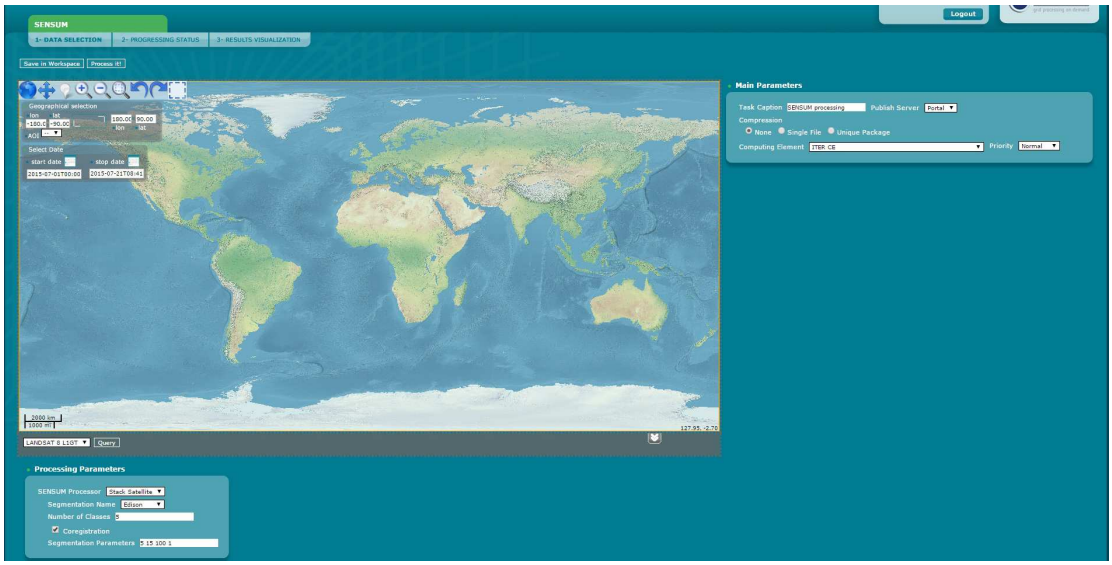


Figure 2.18: *User interface of the implemented tool.*

lutions included a comparison over different years of the most promising techniques, the hybrid-based methods. The extension of these two cities was manually delineated using the same Landsat inputs and layers were compared pixel-by-pixel. The output values were computed using the first version of the algorithm. Further changes, developed for the GPOD integration, were included and tested over Port-au-Prince. Due to the unavailability of a real ground truth layer, the output of an accurate supervised classification was used as benchmark.

Izmir

Overall, for the Izmir case, the PCA-based extraction performed better, mostly in terms of the Cohen's kappa response. In particular, values of producer and user accuracy were quite satisfying while the K value ranked low only for the 2014 case. Further investigation on the issue revealed the presence of clouds and the dis-homogeneity of values in respect of the original Landsat 5 and 7. This feedback was used to tune the second version of the algorithm, including a DEM filter and the radiometric calibration suggested by USGS.

Table 2.10: *Accuracy on the Izmir case study using the dissimilarity based method.*

Year	Producer Acc.	User Acc.	Omiss. Err.	Commiss. Err.	K
1990	80.36	32.53	19.64	67.47	0.29
2000	94.06	55.77	5.93	44.23	0.57
2006	98.45	45.65	1.54	54.35	0.38
2014	78.29	83.87	21.70	16.13	0.73

Table 2.11: *Accuracy on the Izmir case study using the PCA based method.*

Year	Producer Acc.	User Acc.	Omiss. Err.	Commiss. Err.	K
1990	81.12	60.85	18.80	39.15	0.62
2000	78.44	79.80	21.55	20.19	0.72
2006	92.83	65.23	7.16	34.77	0.64
2014	89.28	50.41	10.72	49.59	0.43

Van

Regarding the Van test case, different conclusions can be derived. Overall, the dissimilarity-based method returned better results with a K value over 0.7 in three out of four cases. The PCA-based method performance was quite stable along the years but recorded lower accuracy values in respect of the other method. It must be reminded that performances vary depending on the environment surrounding urban areas.

Table 2.12: *Accuracy on the Van case study using the dissimilarity based method.*

Year	Producer Acc.	User Acc.	Omiss. Err.	Commiss. Err.	K
1984	68.05	22.97	31.95	77.03	0.23
2000	88.75	75.60	11.25	24.39	0.79
2009	59.04	75.86	40.96	24.14	0.78
2013	75.84	88.98	24.15	11.02	0.77

Table 2.13: *Accuracy on the Van case study using the PCA based method.*

Year	Producer Acc.	User Acc.	Omiss. Err.	Commiss. Err.	K
1984	54.34	43.49	45.66	56.51	0.42
2000	88.33	67.14	11.67	32.86	0.72
2009	98.61	52.09	1.39	47.91	0.58
2013	59.08	70.01	40.92	29.98	0.55

Port-au-Prince

The version 2 of the Stack Satellite algorithm has been compared on Landsat and Sentinel-2 data over Port-au-Prince, Haiti. This city is situated in a risk-prone area, and a major earthquake took place indeed in January 2010. Performances were evaluated against the output of a supervised classification. The variety of environmental conditions in the study area makes it challenging and therefore appropriate to test the algorithms. A SRTM DEM layer was used as input of the process to automatically exclude mountainous areas.

The classification accuracy has been computed using the result of supervised classification as reference and comparing the results pixel by pixel. Tables 2.14 and 2.16 show a comparison of the different methods; accuracy values are reported in tables 2.15 and 2.17 instead. report the accuracy figures divided in built-up and non-built-up areas for the Landsat and Sentinel cases. The same reference layer has been used for both sources.

Different considerations can be derived from the obtained results. Focusing on the Landsat data performance, all the pixel-based methods scored quite high overall accuracies. However, an in-depth analysis demonstrated that either omission or commission

Table 2.14: *Comparison of Landsat extraction with supervised classification over Port-au-Prince*

Sup. class. compar.	Urban GT	Non-urban GT
Urban - Built-Up index	4.81	1.92
Non-urban - Built-Up index	11.96	81.31
Urban - PCA class.	7.06	2.88
Non-urban - PCA class.	9.71	80.35
Urban - Dissimil. hyb.-bas.	9.08	8.75
Non-urban - Dissimil. hyb.-bas.	7.69	74.48
Urban - PCA hyb.-bas.	10.53	12.69
Non-urban - PCA hyb.-bas.	6.23	70.55

Table 2.15: *Accuracy over Port-au-Prince*

Superv. Class. Accur.	Built-Up index	PCA class.	Dissimil. hyb.-bas.	PCA hyb.-bas.
Prod. Acc. (%)	28.7	42.1	54.2	62.8
User Acc. (%)	71.4	71.0	50.9	45.4
Omis. Err. (%)	71.3	57.9	45.8	37.2
Comm. Err. (%)	28.6	29.0	49.1	54.6
Over. Acc. (%)	86.1	87.4	83.6	81.1
Cohen's K	0.86	0.87	0.83	0.81

Table 2.16: *Comparison of Sentinel-2 extraction with supervised classification over Port-au-Prince*

Sup. class. compar.	Urban GT	Non-urban GT
Urban - Built-Up index	9.45	7.52
Non-urban - Built-Up index	7.33	75.70
Urban - PCA class.	7.73	4.93
Non-urban - PCA class.	9.05	78.29
Urban - Dissimil. hyb.-bas.	12.40	9.78
Non-urban - Dissimil. hyb.-bas.	4.39	73.43
Urban - PCA hyb.-bas.	12.49	17.02
Non-urban - PCA hyb.-bas.	4.30	66.19

Table 2.17: *Accuracy of Sentinel-2 extraction over Port-au-Prince*

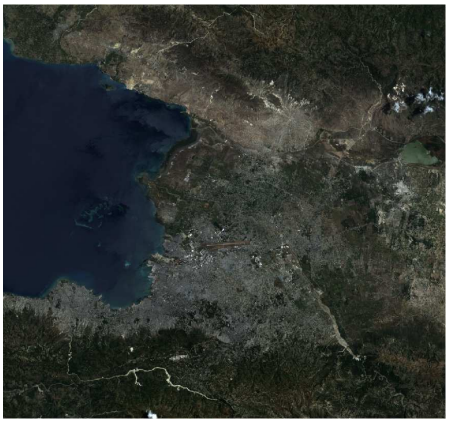
Superv. Class. Accur.	Built-Up index	PCA class.	Dissimil. hyb.-bas.	PCA hyb.-bas.
Prod. Acc. (%)	56.3	46.1	73.9	74.4
User Acc. (%)	55.7	61.1	55.9	42.3
Omis. Err. (%)	43.7	53.9	26.1	25.6
Comm. Err. (%)	44.3	38.9	44.1	57.7
Over. Acc. (%)	85.1	86.0	85.8	78.7
Cohen's K	0.85	0.86	0.85	0.78

errors tends to be substantial. The dissimilarity hybrid-based method is the one showing overall the best performance in all the indicators, including the Cohen's kappa value.

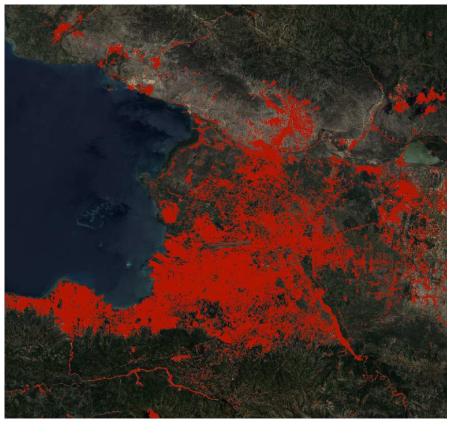
Regarding Sentinel-2 data performance, recorded results are in general agreement with the Landsat-derived ones; the dissimilarity hybrid-based turned out to be the most promising one even in this case. In both cases, the high commission error can be explained through the different origin of the results: while the ground truth is derived from supervised classification -and, therefore, pixel-based- the outputs of the hybrid methods are based on segments, inherently more homogeneous and closer to the actual aspect of a urban area. The Cohen's kappa values are quite high in case of both

Figure 2.19: *Output of all the offered methods in the Port-au-Prince case study for Landsat-8 input.*

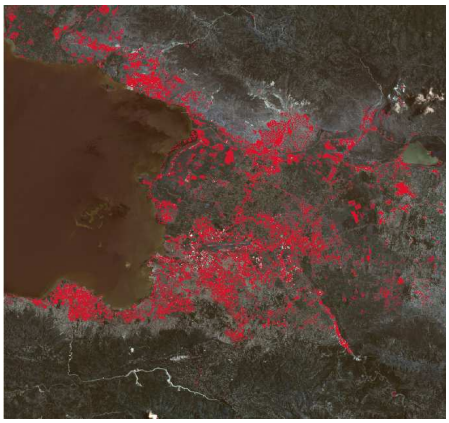
(a) *True colour rendition of original Landsat-8 data*



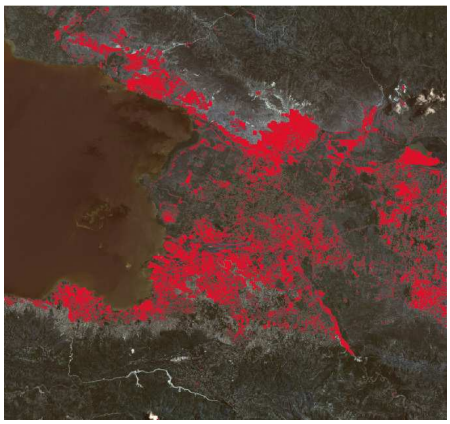
(b) *Supervised classification*



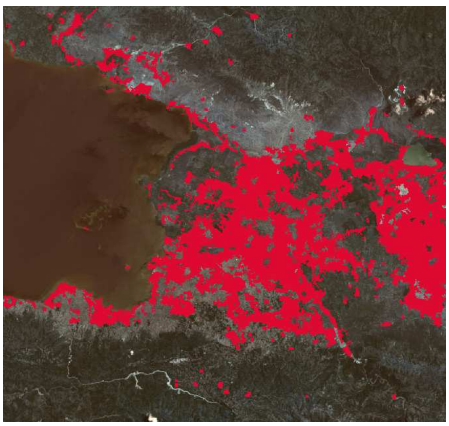
(c) *Built-Up index*



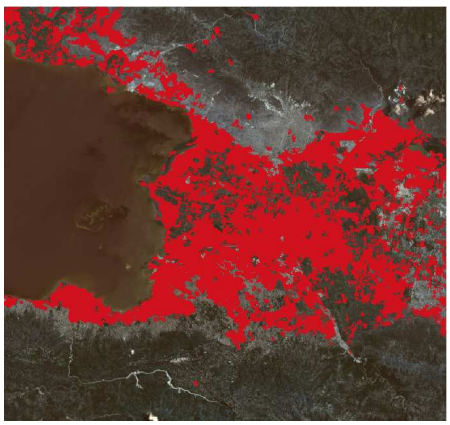
(d) *PCA unsupervised classification*



(e) *Dissimilarity hybrid-based*



(f) *PCA hybrid-based*

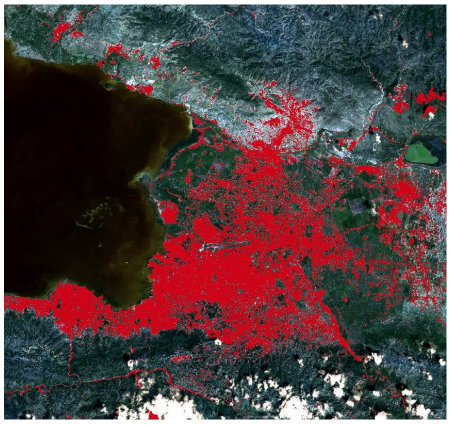


satellites.

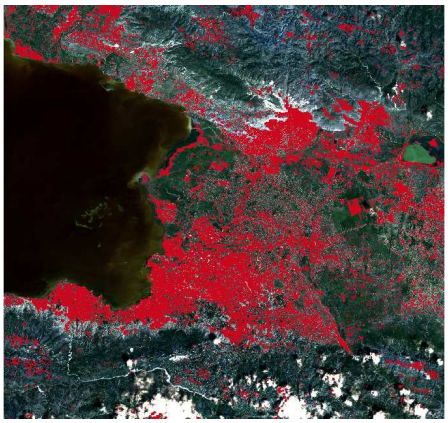
Figure 2.20: *Output of all the provided solutions for Sentinel-2 input.*

(a) *True colour rendition of original Sentinel-2 data*

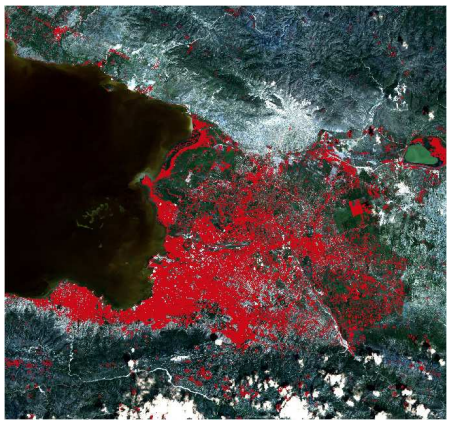
(b) *Supervised classification*



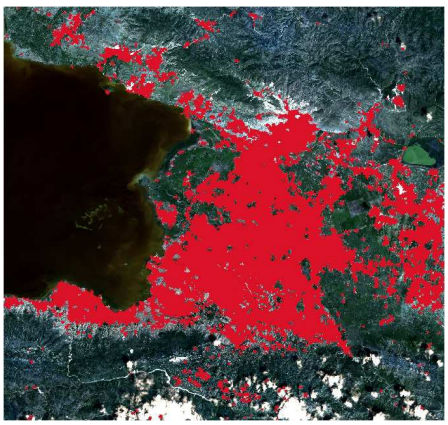
(c) *Built-Up index*



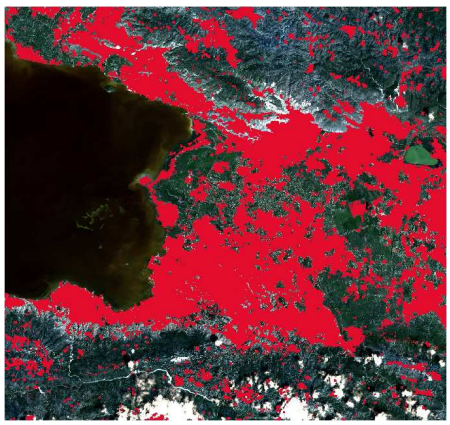
(d) *PCA unsupervised classification*



(e) *Dissimilarity hybrid-based*



(f) *PCA hybrid-based*



2.4 Age of built-up (medium resolution)

Urban expansion is a very hot topic because of the natural resources requirements and land consumption; therefore, it is fundamental to track the evolution in time in order to plan a sustainable growth. Change detection is by definition the capability to detect and

highlight changes occurring in space and time. Earth Observation satellites represent a fundamental source of information thanks to the repeatability of acquisitions in time and the guaranteed spatial resolution; in fact, they are considered an important asset for change detection and monitoring [118] [119]. The unsupervised change detection technique is based on the capability to process a series of Landsat or Sentinel-2 images available for an area-of-interest over a given time span. The suggested time step to consider for the process ranges from 3 to 5 years according to the literature [120]. The algorithm was developed considering inputs produced by the hybrid approaches presented in [65] and [69]; the role of segments is twofold: they are less affected than pixels by classification noise and they are used as fundamental units through the change detection process resulting in an improvement on processing times. However, the implemented change detection does not strictly rely on the layer generated by [65] and [69]; for example, a supervised object-based classification could also be used as input. The method can be defined as unsupervised because the user input is limited to the selection of the class (or classes) related to built-up areas in case of input generated by the previously mentioned hybrid approaches. Extraction errors can be fixed thanks to logic assumptions and continuity in time. Three different filters are included with the aim to correct apparently wrong extractions, following the common hypothesis that cities grow (and do not shrink) over time [121].

Single-change filter

The purpose of this filter is to recognize the segments showing a "regular" behaviour through the years. This label is assigned to segments behaving as in figure 2.21, i.e. showing no other changes than switching once from non-urban to urban. The segment follows the expected growth in time of a built-up area.

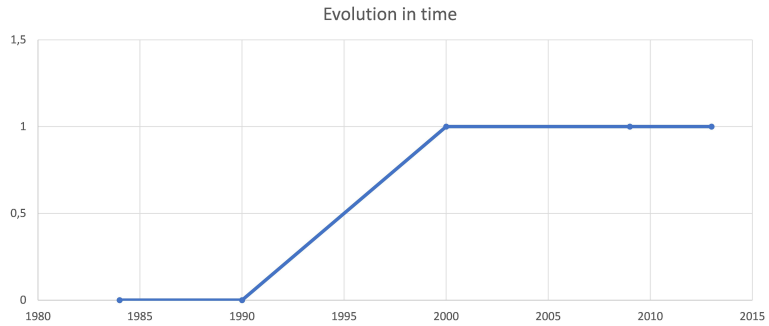


Figure 2.21: *Example of stable segment, meaning a segment following the expected growth in time of a built-up area.*

Double-change filter

Every segment not classified as "regular" from the first filter is further analysed taking into account continuity over time. The idea is to consider a 4-observations-long window (experimentally tuned) and evaluate changes in time: class continuity is the metric used to decide whether the segment was incorrectly classified at some point. An example of "non-regular" behaviour is shown in figure 2.22. Segments showing more than two changes are instead labelled as "uncertain" (figure 2.23).

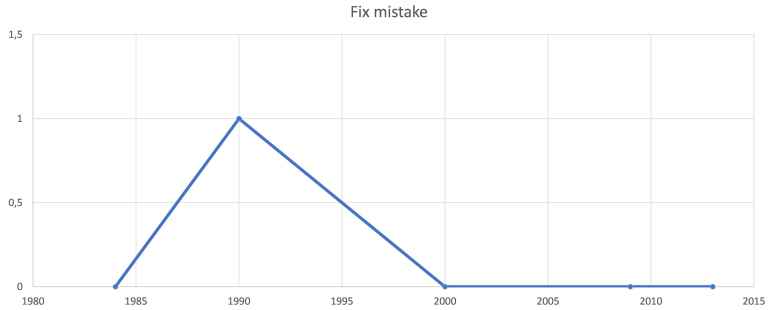


Figure 2.22: *Example of how the method can fix unstable results. The segment shows a "non-regular" behavior that can be interpreted as a misclassification in year 1990 and consequently corrected.*

Spatial filter

The third implemented filter takes care of the segments labelled as "uncertain" by the second filter. Segments around the "uncertain" segment are analysed and a decision is made according to the presence or absence of "regular" segments around the object of interest.

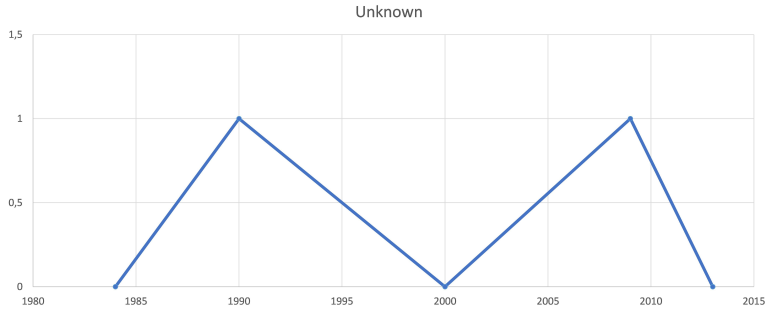


Figure 2.23: *Example of case where the spatial filter will be applied because it is difficult to determine a precise pattern.*

An example of the outcome of the proposed algorithm is shown in figure 2.24.

2.5 Footprints extraction (high resolution)

The OpenStreetMap initiative can provide useful data related to footprints, usage and land use. Unfortunately, layers still show gaps outside of major cities. The proposed tool represents a way to fulfil these gaps thanks to the combination of very high resolution imagery (up to 50 cm) and the Support Vector Machine (SVM) [75] supervised classification. A schematic workflow is shown in figure 2.25.

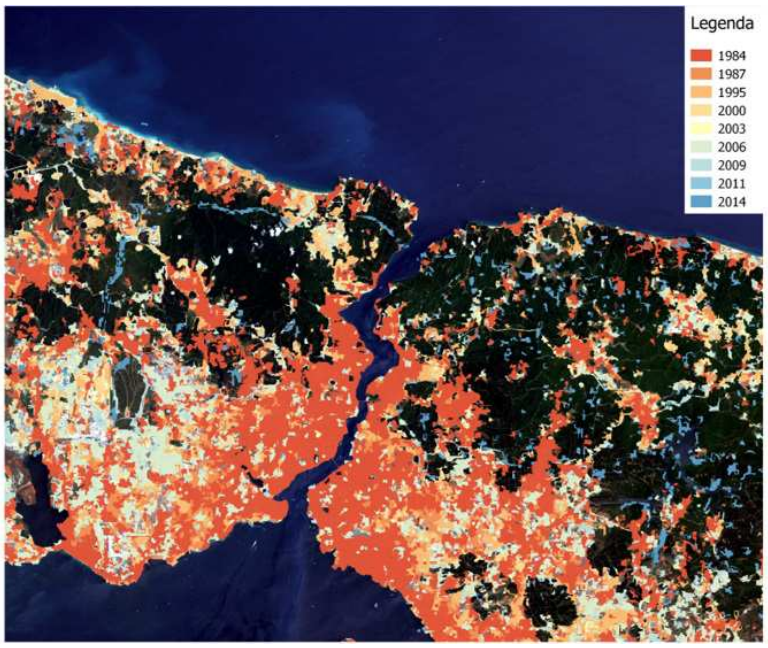


Figure 2.24: *Age of built-up applied over Istanbul (Turkey) from 1984 to 2014.*

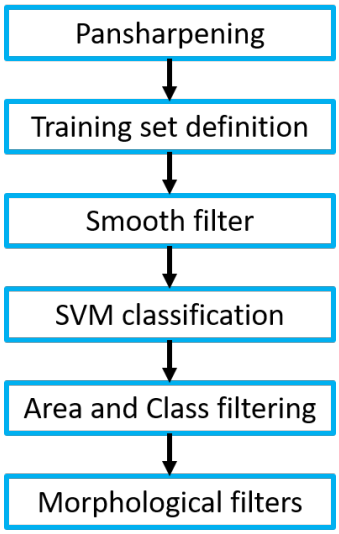


Figure 2.25: *Workflow of the proposed footprint extraction algorithm.*

Pansharpening

A very popular technique for combining multi-spectral and panchromatic data is pan-sharpening. For the case of interest, the algorithm described in the *Multi-purpose tools* paragraph has been used.

Definition of the training set

A set of samples, made of polygons, is necessary to train the classifier and generate clusters. The training set is defined by the analyst, with samples covering all the possi-

ble set of classes.

Smooth filter

In most cases, classification results suffer from salt-and-pepper noise, meaning non-continuous outputs and therefore reduced footprints accuracy. Precious help comes from the meanshift filter, provided by OTB as a way to segment big images with reduced distortion. In the buildings case, it helps in reducing the variability of pixels and therefore extracting more stable results. An example of the input image before and after this processing step is displayed in figures 2.26 and 2.27.

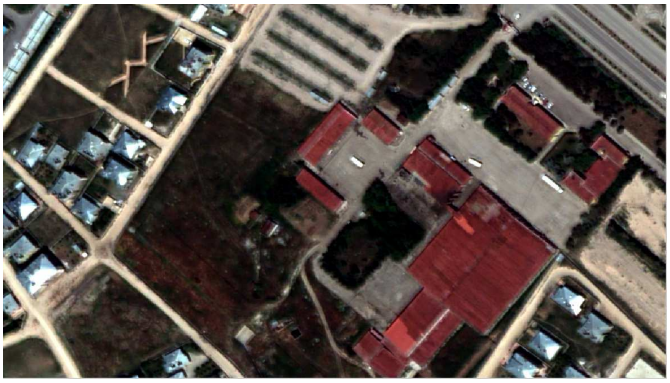


Figure 2.26: *VHR optical input image before the smooth filter.*

Figure 2.27: *VHR optical input image after the smooth filter.*

Supervised classification

The Support Vector Machine algorithm is to used to classify the input scenes. This choice was made based on a visual comparison with other classifiers, including random forest, maximum likelihood and Bayes. Parameters were left as default.

Area and class filtering

The outcome of the classification is further processed, excluding classes unrelated to buildings and applying area-based filters, thus excluding too small polygons.

Figure 2.28: *Results over a high density area in Izmir, Turkey. Buildings are separated by roof colour.*

Morphological filters

Results are further refined using morphological filters, namely closing and opening filters (controllare) with the aim to regularize the final shapes.

2.6 Height extraction (high resolution)

The response of buildings, and therefore their fragility, is definitely dependent on the number of storeys [122]. The height can be determined by combining the shadows length with the sun elevation and azimuth (see figure 2.29).

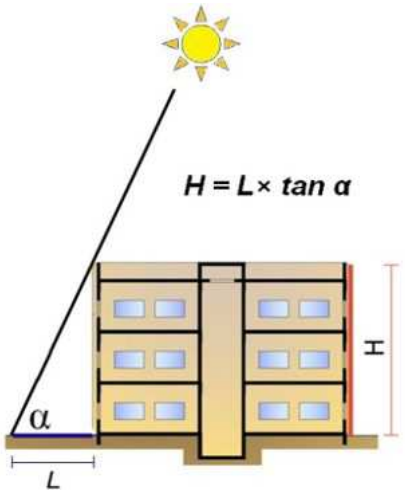


Figure 2.29: *Drawing of the height calculation.*

The process is quite straightforward:

1. extract the shadows using a supervised classification; for example, they can be retrieved as a by-product of the footprints extraction

2. provide the acquisition date of the satellite image; this will be used to automatically determine azimuth and elevation of the sun

3. based on the sun position, the length of each shadow is computed

4. using trigonometry, the length is converted into the building height

5. every value obtained is automatically transferred from shadows to input buildings using the sun direction and the distance between the entities.

The proposed approach naturally suffers from some limitations. In detail, results strictly depend on the density of buildings and image ortho-rectification; moreover, tall and tightly packed buildings complicate the extraction due to non-fully-unfolded shadows. Crowdsourcing could in this case help in filling the gap and provide more accurate information.

2.7 Building density (high resolution)

Density is expressed as in equation 2.21; it is the result of a ratio over two areas therefore it is without unit of measure and associated to each building. A schematic drawing is also provided in figure 2.30. Due to the way the density is computed, to an increased value corresponds a lower distance between buildings. Taubenböck et al. [123], in a paper published two years after the tools release, report how the term density is generally understandable and self-explanatory but the definition assumes a multidimensional nature according to the field of study. The proposed solution does not take into account the dimension of blocks but only the distance between structures.

$$d = \frac{\sum_{i=1}^{N} A_i}{\pi \cdot r^2}, \tag{2.21}$$

where:

- A_i is the intersection area between the circle defined from the polygon centroid and a building around

- N is the total number of buildings intersected by the circle

- r is the radius of the circle.

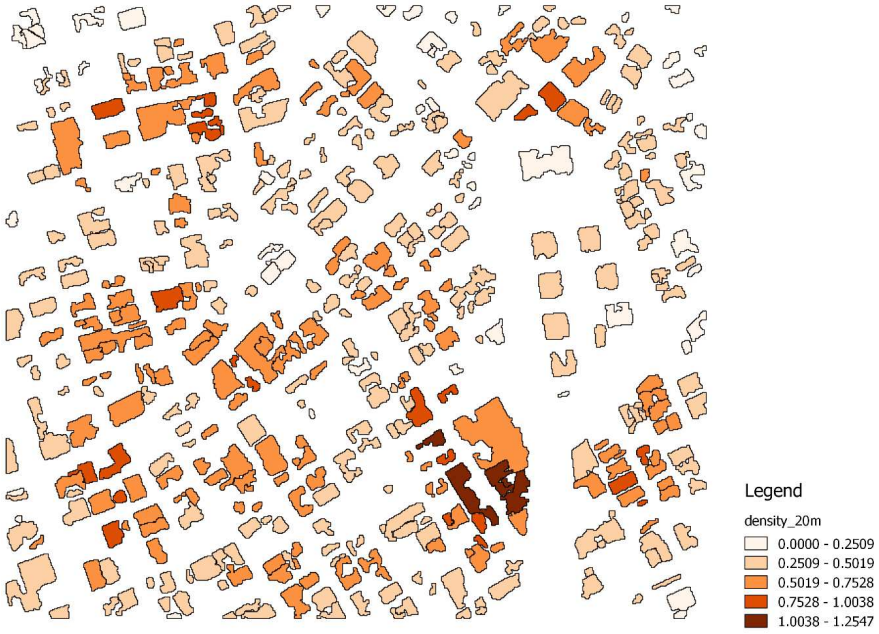


Figure 2.31: *Outcome of the density computation process over Izmir, Turkey.*

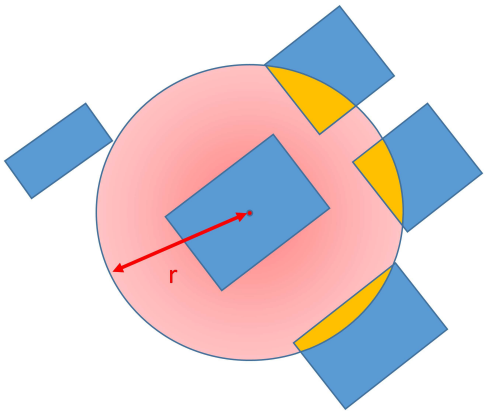


Figure 2.30: *Schematic representation of the density calculation.*

The proposed indicator can be calculated over datasets from different sources. For example, in figure 2.31 the footprints were extracted using the tool previously proposed. On the other hand, figure 2.32 shows the results of the same algorithm using OpenStreetMap footprints as input.

2.8 Building regularity (high resolution)

Regularity of buildings is another important feature to be considered for vulnerability assessment: basically, irregular layouts are more subject to heavier damage. Euro code 8 [124] defined a building plan regularity using the footprint length (a) and width ratio (b):

$$\frac{a}{b} < 4, \tag{2.22}$$

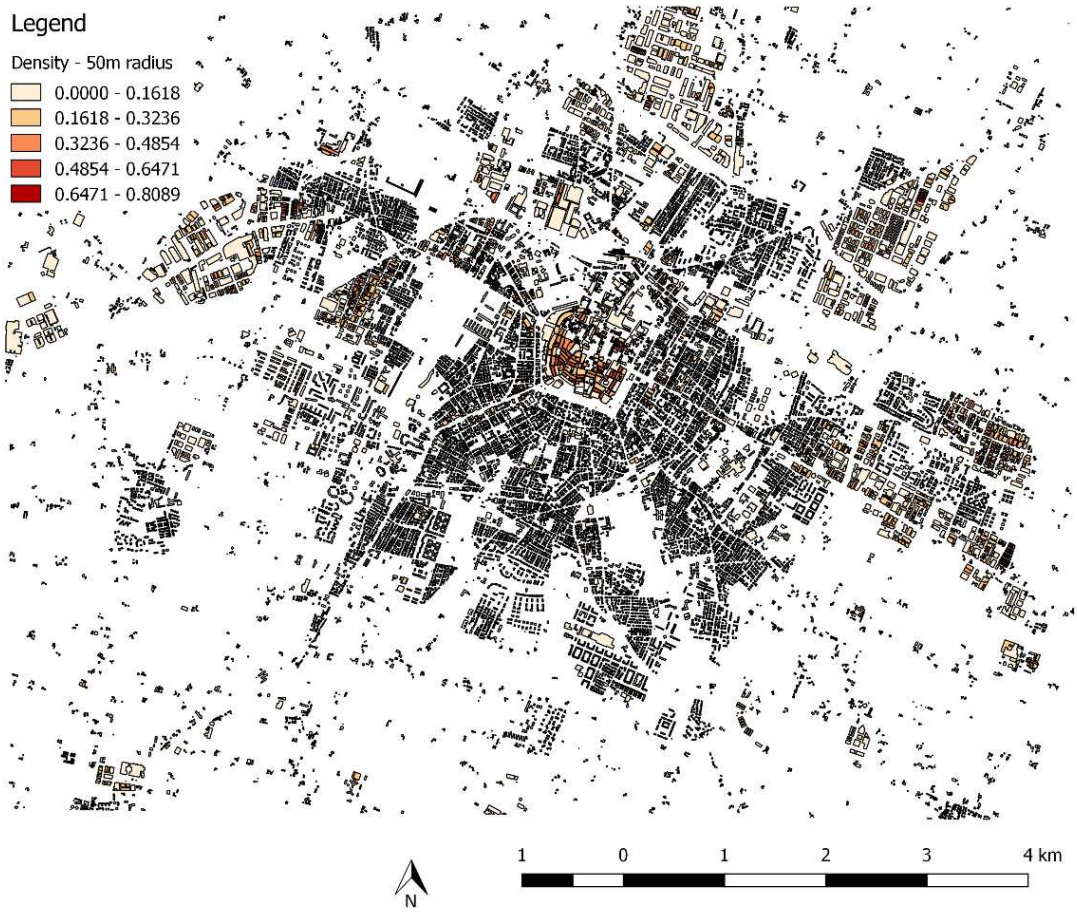


Figure 2.32: *Result of the density computation process over Modena, Italy.*

stands for a regular footprint building.

CHAPTER 3

Pre-event exposure information: Crowdsourcing

3.1 Implemented framework

Based on the idea of volunteers ("citizen sensors") contributing geocoded information using their smartphones, a generic and multi-purpose framework for geo-localized data collection has been designed [125]. The system is based upon two main components:

- a mobile app for the distributed collection of data;
- a Django-SQL server, common recipient for the generated reports.

A schematic representation of the architecture is shown in figure 3.1. The client-server-based architecture enables distributed users to upload and save georeferenced images into a database or filesystem. Requests are made through the HTTP protocol and processed thanks to an ORM (Object-Relational Mapping) handler. The API structure of iOS and Android is fully integrated in the HTTP protocol. The database and server structure are specifically designed for the presented application but they can be easily expanded and modified according to user needs. The *crowd* is the data provider; the input method can be a customized mobile app or web platform but the core remains the same without time-consuming modifications. Every submitted report is stored in a database accessible to selected agents; available data can then be downloaded as jsons or shapefiles. All the code used for the implementation of the core architecture is released under an open-source license. The architecture described above is an evolution of previously defined frameworks for citizen-sensors purposes.

The proposed framework is quite generic, meaning that it can be used to collect different types of geotagged data; for example, users could be asked to upload data related to the number of floors in specific locations thus filling the gap left by satellite imagery.

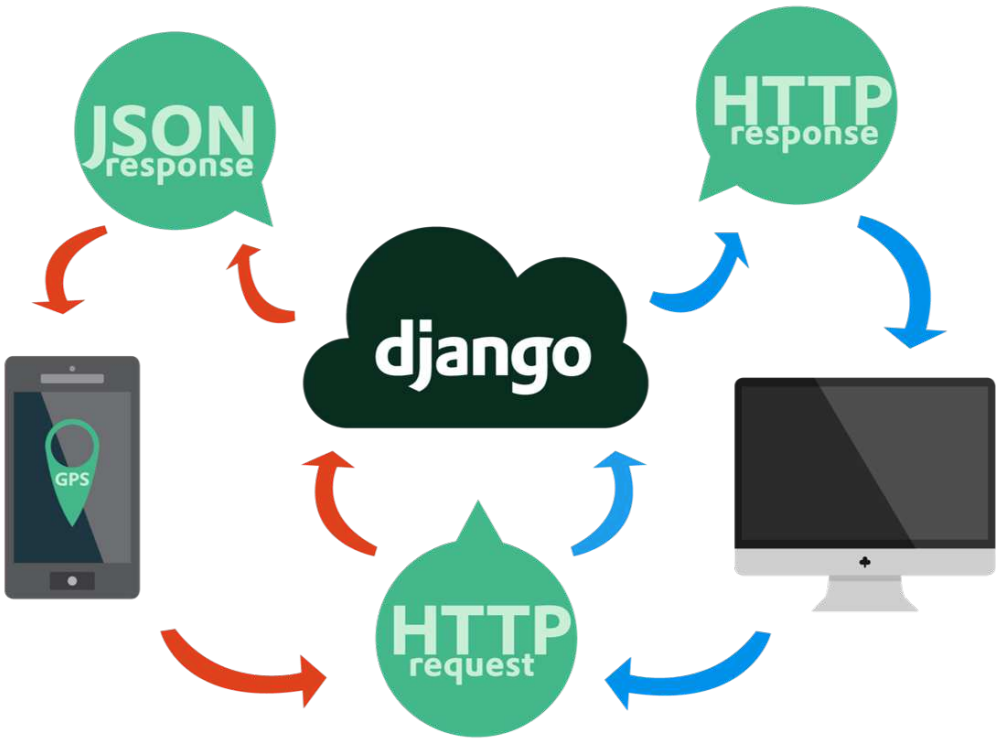


Figure 3.1: *Implemented architecture for the collection of reports.*

3.2 Applications

Different projects offered us some opportunities to implement the proposed framework in two different contexts. The first one was the SEGUICI project, focused on the optimization of water resources by feeding hydrogeological models with satellite-based Earth Observation and crowdsourced data. In particular, the goal of the mobile app was to collect data related to the growth status of crops (rice, maize, soy, etc.), in order to feed the water consumption model and to determine crop vulnerability to drought and flooding.

The second project, called CLOOPSy (Copernicus Land-cOver crOwdsourcing Platform for Sentinel-based mapping), resulted among the winners of the MyGEOSS contest [126]The aim is to collect data related to land cover, following the CORINE Land Cover taxonomy, providing fresh updates to the service; at the same time, the collected reports will be used to validate automatic or semi-automatic algorithms working on the extraction of land cover information from satellite.

3.2.1 SEGUICI Vegetation Report

The idea is to use volunteers to collect pictures of crops and ask them to indicate the recognized type from a pool of most commonly found vegetation species. Other questions asked are the height of vegetation -directly related to the stage of growth- and if the field is flooded or not [127]. Every collected report, which includes a geocoded and

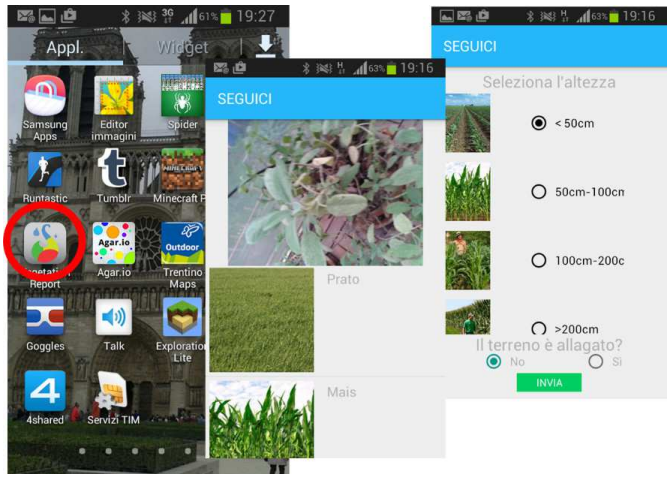


Figure 3.2: *SEGUICI Mobile app graphic interface.*

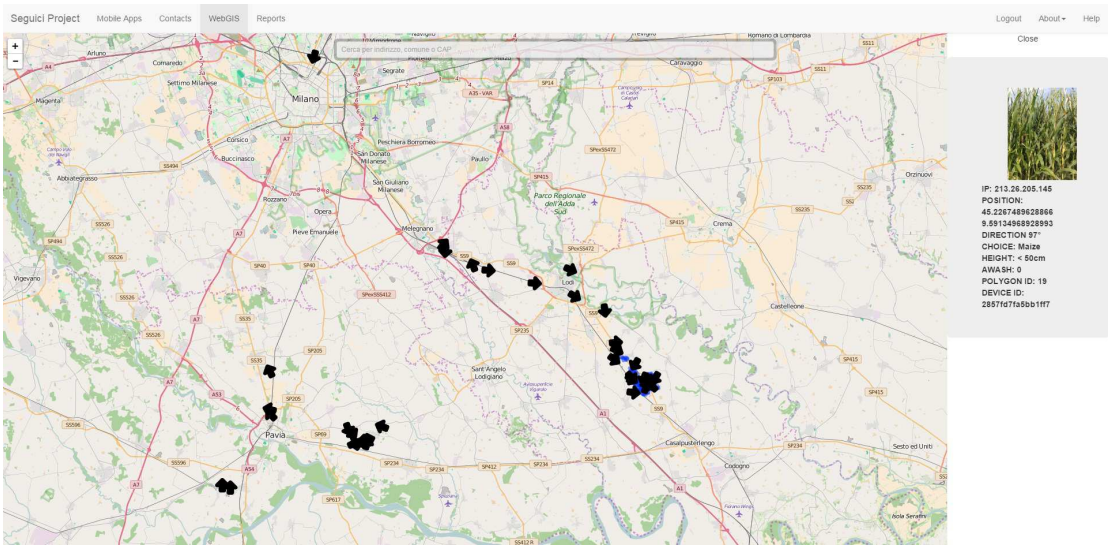


Figure 3.3: *SEGUICI WebGIS interface.*

oriented picture plus answers to the above mentioned group of questions, is automatically uploaded and stored in a remote database. To avoid weighing on the contributor's mobile data quota, an option can be activated to store reports on the hand-held device and upload them only when a WiFi connection becomes available. A webgis interface is used to display data on a OpenStreetMap-based map. Within the server, an algorithm can automatically associate the geo-localized reports with polygons related to each single field using GPS position and compass direction. The app has been released on the Google Play Store [128] and was written in java using Android Studio. A demo of the mobile app is available on YouTube [129]. The ESA RSS provided server space to host the webgis and the database, besides guidelines for use and assistance in activating and running the hosting service. The source code has been released as open-source and is available to public from a BitBucket repository [130].

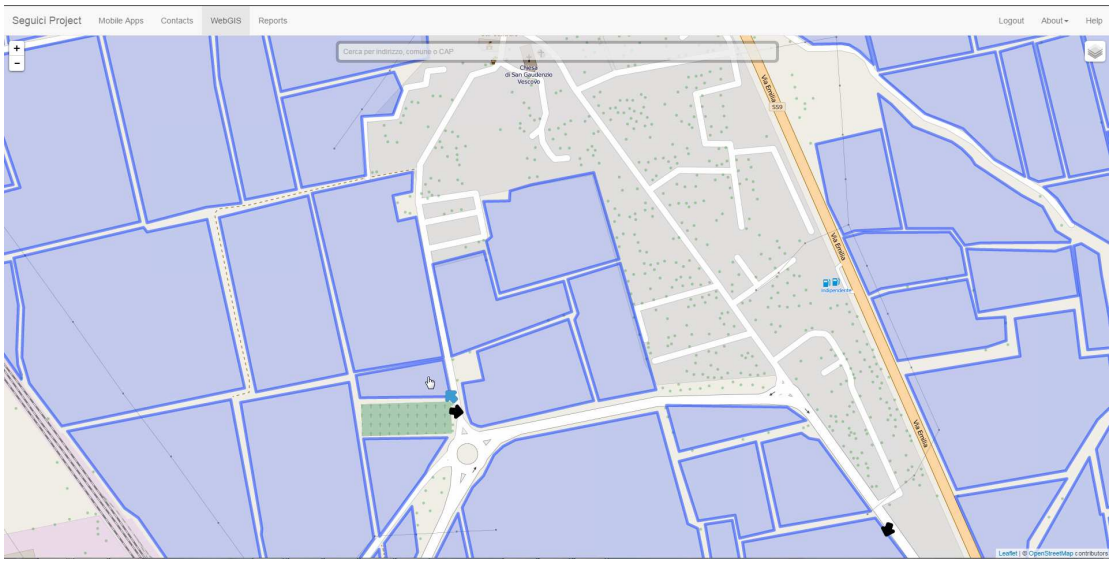


Figure 3.4: *Example of report automatically associated with one GIS polygon out of many defining homogeneous cultivation areas in a rural environment in Northern Italy.*

3.2.2 CLOOPSy

I was also involved in the development of a *mobile app* (mobile-based application) aimed at distributed collection of data related to land cover. *CLOOPSy* (Copernicus Land cOver crOwdsourcing Platform for Sentinel-based mapping) [131] has been selected as one of the winning applications of the MyGEOSS contest [132], an H2020 project with the goal to stimulate the use of geospatial data in mobile apps at the European level [126]. The proposal envisaged an *app* designed to generate ad-hoc in-situ georeferenced data to be fused with Sentinel data in order to improve the accuracy of products from Copernicus Land mapping services and to validate them by providing ground reference data. The rationale is to contribute to the integration of space and in-situ components within the Copernicus framework -a concept expressed by the red arrow in figure 3.5- by collecting reference points for validation of the algorithms and procedures using in processing satellite images or for training of machine learning algorithms.

The implemented procedure asks contributors to submit a picture of what is around them, enriched with geographic location and compass direction. At the same time, the user is requested to select a category label for the depicted land parcel, based on the CORINE Land Cover taxonomy. In his/her choice, the user is supported via a selection of example pictures (screenshots of the mobile app are included in figure **??**). Categories are automatically synchronized with the server; from the moment when enough reports will be collected, an algorithm will take care of limiting user input to the most common categories around him/her.

A tutorial is dispensed to beginners in order to familiarize with the app. Viewing the tutorial is required to a new registrant before he/she can activate the app and submit reports. This is imposed in order to try and pursue a minimum quality level for the input data.

Submitted reports are stored in a private section, accessible only by the sender. Once

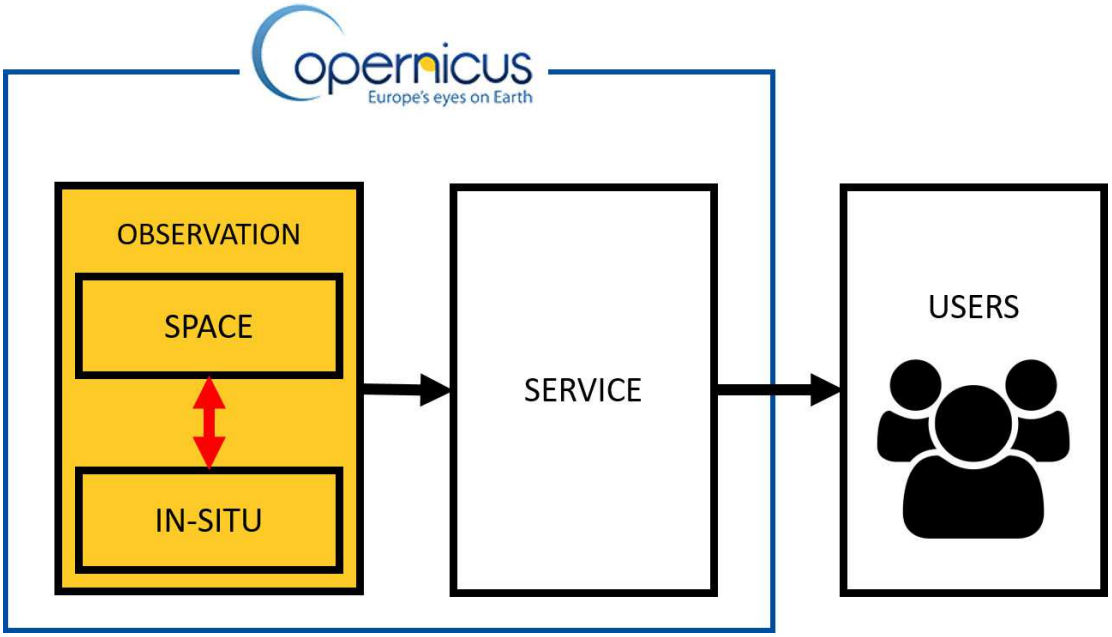


Figure 3.5: *Components and sub-components of Copernicus. The additional red arrow represents the envisaged exchange and fusion of information between the space and in-situ components.*

approved by the administrators, data is released to public and therefore is available to anyone who wants to download it (*public* reports). The aim is to avoid direct publication of reports to filter out the wrong or inappropriate ones. Automated procedures are being developed for filtering and quality checking of submitted reports, leaving to the human controller only those items which do not produce a definite quality / appropriateness score.

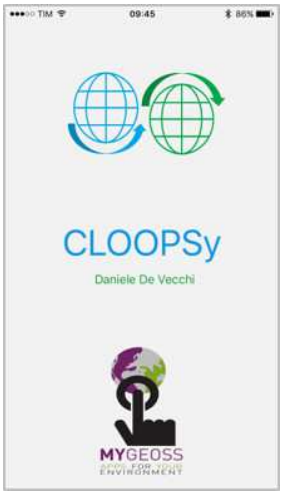
Once published, geotagged data is available for download to anyone thanks as JSON thanks to integrated APIs [133] [125]. The server integrates an algorithm for automatic matching with parcel GIS layers, where available. The web service is hosted by the ESA Research and Service Support (RSS) [134] in view of easier integration with the Sentinel-2 data stream. The mobile apps are available for iOS and Android systems on their respective app stores and were built using Xamarin. Code is available here [135].

3.3 Future development

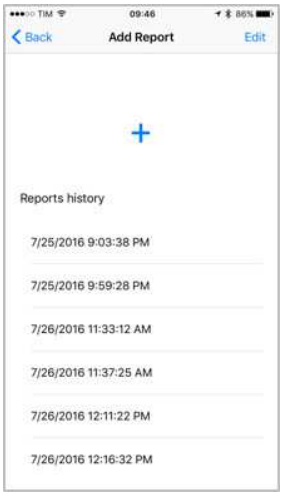
The previously mentioned "SEGUICI Vegetation Report" and "CLOOPSy", in which I was involved as developer, imposed different objectives rather than looking at building features and exposure. The developed multi-purpose framework was expressly designed in order to be flexible and collect different types of data. Moreover, the experience and know-how in mobile app development acquired thanks to these two projects will be very useful in the future. Unfortunately, there was no time left to adapt the proposed framework to collect records on buildings. In the near future, the idea is to refine the app, hand it to non experts and ask them to provide information regarding the number of floors and the material of which the buildings are made of. These two values, as discussed in the next chapter, are fundamental in order to have a rough as-

Figure 3.6: *Screenshots of CLOOPSy iOS app.*

(a) *Splash screen*



(b) *Add Report*



(c) *Take Picture*

(d) *Choose category*

(e) *Submit report*

sessment of the fragility of buildings and are therefore of key relevance. The number of floors is quite easy to determine even by non experts, it usually corresponds to the number of rows of windows in a building. On the other side, the material is not that easy to recognize and classify; the idea is to provide tutorials to users and to restrict the selection to the most common construction materials in the area (depending also on the history of the area of interest).

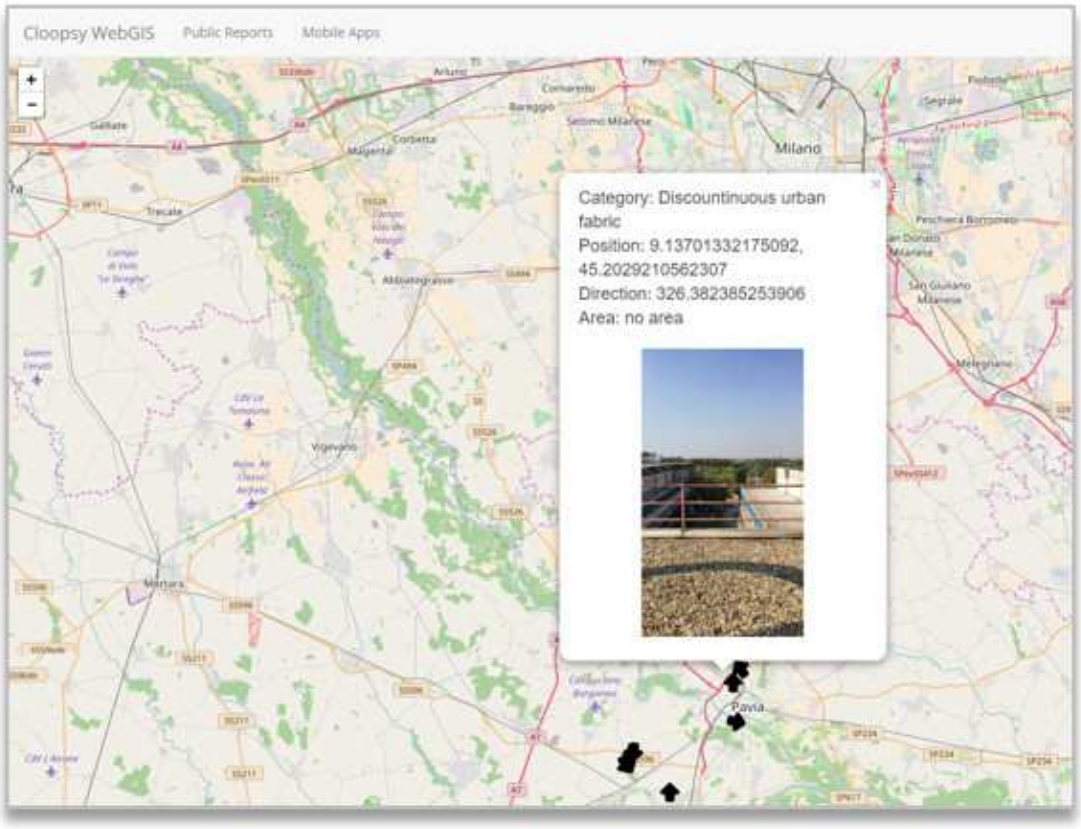


Figure 3.7: *Screenshots of the CLOOPSy webGIS. Non-registered users can only check* public *reports while logged users can also retrieve their own private reports.*

Figure 3.8: *Example of automatic intersection of a generated report with available GIS parcels data.*

CHAPTER 4

Post-event: seismic damage assessment

Earthquakes are among the deadliest and costliest natural disasters, with more than 1 million estimated casualties and more than 700 billions dollars damage since 1970 [136]. Earth Observation demonstrated its usefulness in estimating seismic damage assessment in different cases, spanning from the early nineties until now [137].

Focussing on the response phase, my thesis work focused on rapid seismic damage assessment; the objective was to provide useful information to first responders by highlighting the most affected areas. This chapter describes the different attempts made to extract damage working with different SAR sensors, techniques and combination when possible with pre-event data; the first solution focussed on the use of post-event only imagery, based on previous experience from the Pavia group. The other approaches are focused on the use of change detection, arranged during a 3 months exchange period at the Chiba University, Japan, under the supervision of Prof. Fumio Yamazaki.

4.1 Literature review

Several contributions are available in literature, describing usage of different sensors and combinations of data [138]. Manual delineation of damage is the most common solution [139]; collaborative efforts based on volunteers highlighting affected areas [140] are also a common procedure for major events [34]. Focusing on the semi- or fully-automatic approaches based on optical imagery, Chini et al. in [141] used change detection trying to delineate damaged areas. Different approaches were also proposed combining optical and SAR data with the aim to increase the capability of detecting damage [142] [143]. Considering SAR data-only methods, Matsuoka et al. in [144] demonstrated how change detection of radar data can be used to extract damage; Balz and Liao in [145] modelled the different building response in case of damage. Brett and

Guida in [146] proposed an automatic method to evaluate corner reflection changes. Dell'Acqua et al. in [147] investigated a possible correlation between texture measures extracted from post-event-only VHR SAR data and damage at block level. Limitations of seismic damage assessment depend mostly on the spatial resolution and the type of sensors. Optical imagery benefits of increasingly finer spatial resolution but its natively nadiral vision limits the detection of damage to the fraction which is visible from above, while leaving largely ignored serious damage phenomena such as pancake collapse. On the other side, SAR data, thanks to its inherent side-looking view, is capable of detecting damage inflicted to walls; but its highest resolution is currently limited to approximately 1 meter (COSMO-SkyMed and TerraSAR-X spotlight mode). In addition to the limit on highest resolution itself, spotlight data is generally considered low priority with respect to other lower-resolution modes that are employed to feed more consolidated techniques such as interferometric displacement mapping.

4.2 Test cases

4.2.1 L'Aquila

The medieval city of L'Aquila was hit by a 6.3 moment magnitude scale earthquake on April 6th 2009 (as shown in figure 4.1). Official data report 308 casualties and around 1500 injured. Pictures of from the ground show major devastation in the city centre, mostly built in the XVIII century.

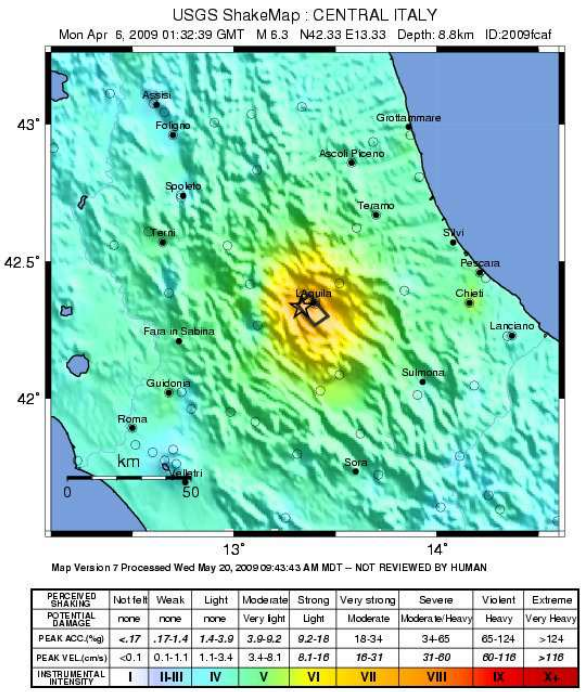
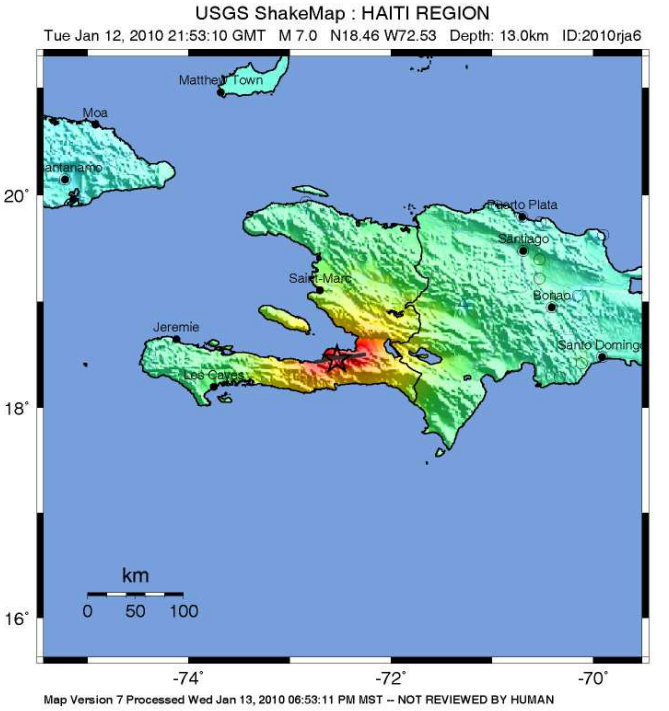


Figure 4.1: *USGS ShakeMap related to the L'Aquila earthquake.*

4.2.2 Port-au-Prince

On January 12th 2010 a catastrophic magnitude 7.0 earthquake struck the island of Haiti, with an epicentre 25 kilometres west of Port-au-Prince 4.2. An official number of casualties is not available but the number ranges from 100,000 to 316,000 deaths according to different sources. Consequences were aggravated because Haiti is one of the poorest countries in the World [148].



PERCEIVED SHAKING	Not felt	Weak	Light	Moderate	Strong	Very strong	Severe	Violent	Extreme
POTENTIAL DAMAGE	none	none	none	Very light	Light	Moderate	Moderate/Heavy	Heavy	Very Heavy
PEAK ACC.(%g)	*<.17*	*.17-1.4*	*1.4-3.9*	*3.9-9.2*	*9.2-18*	18-34	34-65	65-124	>124
PEAK VEL.(cm/s)	<0.1	0.1-1.1	1.1-3.4	3.4-8.1	*8.1-16*	*16-31*	*31-60*	*60-116*	*>116*
INSTRUMENTAL INTENSITY	I	II-III	IV	V	VI	VII	VIII	IX	X+

Figure 4.2: *USGS ShakeMap related to the Haiti earthquake.*

4.2.3 Bhaktapur

The Gorkha earthquake struck Nepal on April 25th 2015, killing nearly 9,000 people and injured around 22,000. The epicentre was located east of Gorkha district at Barpak (see the USGS ShakeMap 4.3) with a recorded magnitude of 7.8. Katmandu, the capital city, suffered minor damage. Bhaktapur experienced major damage instead and was chosen as a test case for the algorithms.

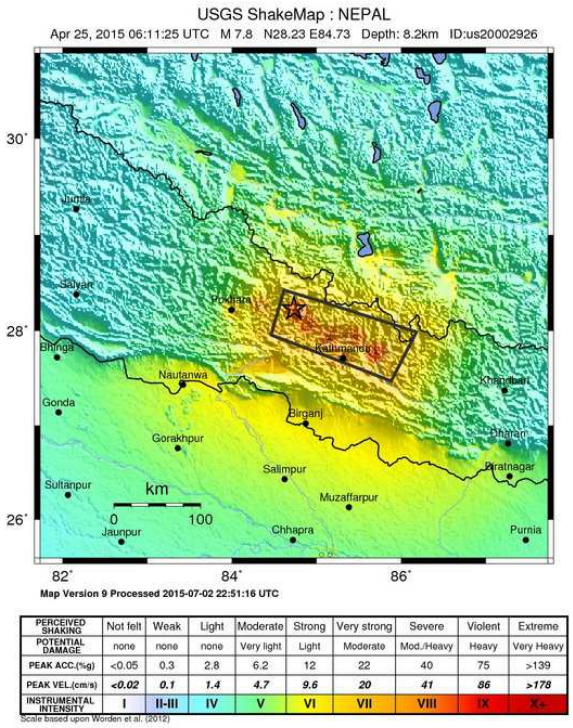


Figure 4.3: *USGS ShakeMap related to the Gorkha earthquake.*

4.2.4 Mashiki

A series of earthquakes affected the Kumamoto prefecture in April 2016. A foreshock of magnitude 6.2 occurred on April 14 while the mainshock reached magnitude 7.0 on April 16 (see figure 4.4). Casualties stopped at 49 people while a total of 3,000 people was registered as injured. Most of the damage was concentrated in Mashiki, a suburb east of Kumamoto city.

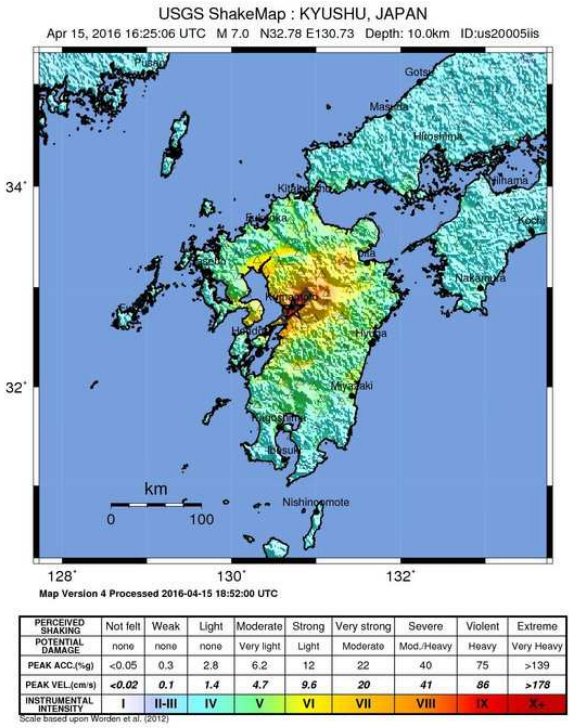


Figure 4.4: *USGS ShakeMap related to the Kumamoto earthquake.*

4.3 Post-event only method

Our research group gave a modest contribution to this investigation by analysing 1-m resolution spotlight images produced by COSMO-SkyMed and TerraSAR-X systems across various seismic events to discover a weak link between some selected texture measures and local damage level. Texture-to-damage correlation was strengthened by integrating pre-event building fragility information estimated from optical EO data [57]. Building density impacts both on radar backscatter patterns and on combined fragility of shaken buildings, and it is thus a good candidate to steer the data analysis. The basic idea is still to use GLCM measures computed over post-event images as the starting point, just like for the previously developed method. The novelty consists of adding independent information in the form of local building density extracted from optical satellite data through suitable open-source tools. The method was tested on L'Aquila because of the availability of spotlight COSMO-SkyMed imagery and pre-event information.

4.3.1 Preprocessing

The considered datasets consist of VHR SAR data, specifically for our test case 1-m spotlight acquisitions by COSMO-SkyMed on L'Aquila, Italy, an urban area stricken by the 6th April 2009 seismic event. Images were acquired before the event (February 16th, 2009) and after the event (April 13th, 2009) with an incidence angle of 50.57 degrees. Pre-processing involved conversion from complex to amplitude and automatic

co-registration between pre- and post-event imagery.

4.3.2 Texture computation

In accordance with previous research results [137], a set of GLCM textures has been computed using a smaller window (11 instead of 21 and more used in the cited paper) and evaluating the different results with more steps values (1, 3 and 5) and angles (0, 45, 90 and 135 degrees). This solution allowed considering multiple damage possibilities and directions. A complete set of textures has been extracted using the Sentinel-1 toolbox, available online [149].

4.3.3 Merging with blocks

Every texture has been averaged within polygons grouping buildings believed to have similar fragility according to visual interpretation of images by structural engineers. All the combinations of steps and angles indicated above (e.g. step 1 angle 0, step 1 angle 45 and so on) are averaged as well; the idea is to avoid the definition of a precise combination, mostly dependent on the specific case of damage considered. The set of polygons is presented in figure **??** and it is the same used in [57]. Every block was also tagged with a block-averaged level of damage known as DAR (or Damage Area Ratio) (4.1).

$$DAR_j = \frac{\sum_i d_{ij} \cdot A_{ij}^B}{A_j^P} \tag{4.1}$$

where:

- DAR_j is the DAR value on *j*-th GIS polygon
- d_{ij} is the “damage flag" (with values 0 or 1) indicating whether building *i* in polygon *j* was damaged by the earthquake
- A_{ij}^B is the footprint area of the *i*-th building in *j*-th polygon
- A_{ij}^P is the total area of the *j*-th polygon.

This measure is based on visual interpretation of aerial images carried out by experts on very-high-resolution pictures taken in the immediate aftermath of the event.

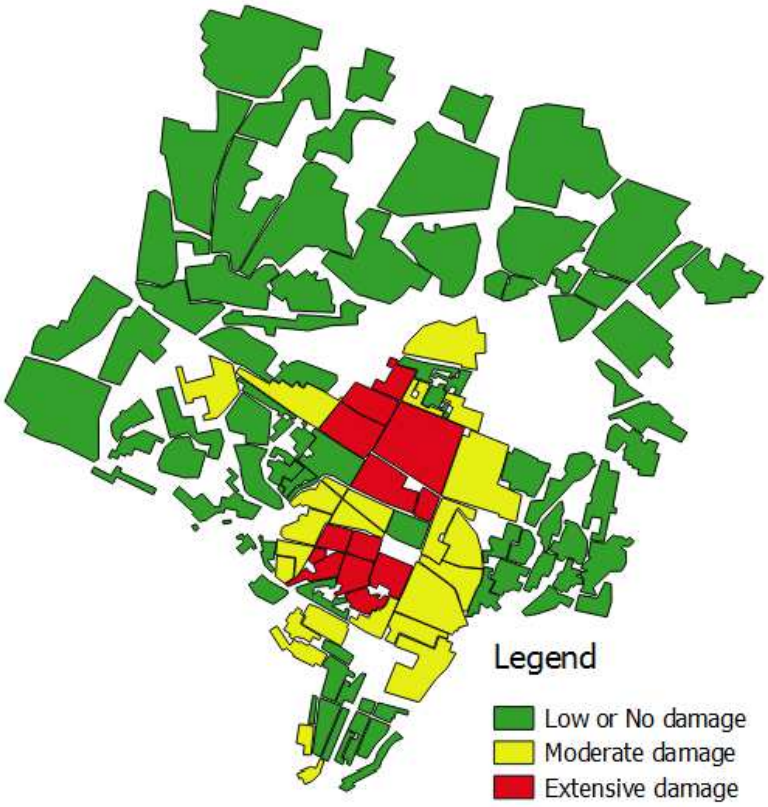


Figure 4.5: *Set of blocks used. The colour is associated to the level of damage.*

4.3.4 Density information

The additional pre-event information we propose to include is represented by the density of buildings, meaning the aggregated area of building footprints in a given block divided by the area of the block. This measure is proven to have an impact on damage [70].

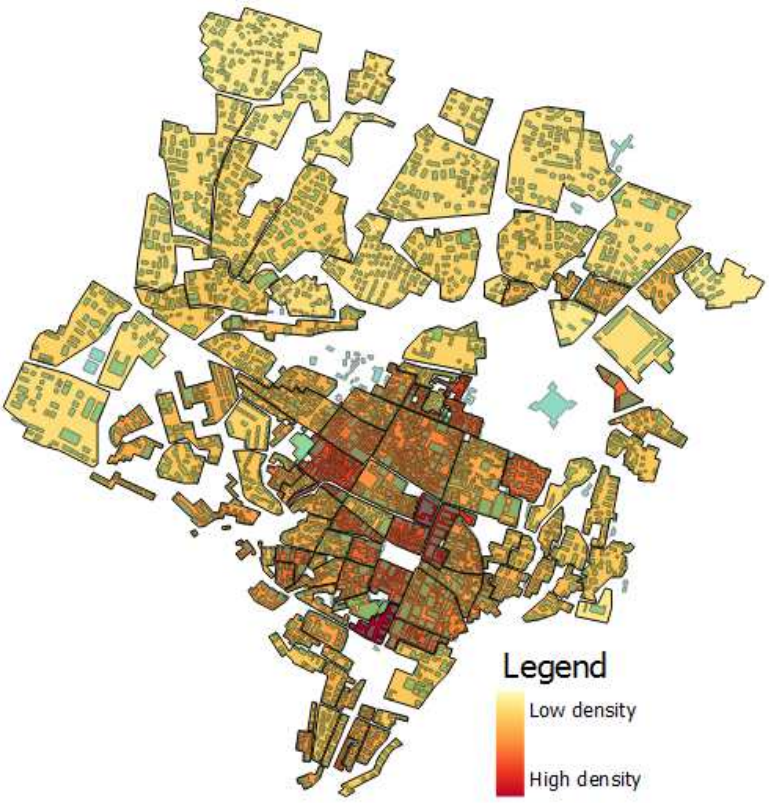


Figure 4.6: *Blocks displayed with the associated density measures.*

4.3.5 Texture and density vs. damage

All the textures computed considering the different variables have been compared to the damage value thresholded according to the distribution of DAR values. Two different layers of damage have been included, one extracted from aerial imagery, and a more comprehensive one from a ground surveying mission by experts of the Italian national institute of geophysics (INGV).

4.3.6 Results

A subset of the images available for L'Aquila has been processed following the workflow explained before. The DAR values have been thresholded according to the distribution in order to recall what was already shown in [57]. The scatter plot in figure 4.7 displays the distribution of the entropy texture, averaged over three different windows and four angles and compared with the density measure. Figure 4.8 shows the contrast measure instead. The numbers confirmed the first impression: correlation between contrast and damage reached 0.45 while the comparison with entropy reached only 0.15. Correlation of contrast computed on the pre-event image (16^{th} February 2009) reached only 0.28, therefore suggesting that the proposed measures are indication of damage, as no significant pre-event correlation was present. The centroids for each class of damage are also included in the scatter plots using triangles: in the entropy case the three values are close to each other while in the contrast case they are more separated.

Unfortunately, the proposed fusion of the pre-event density information with post-event texture measures did not reach the expected results. These findings can be justified by the complexity of the considered scenes and the pattern of damage caused by earthquake. In fact, most of the buildings show a grade 3 damage while the DAR values used for the comparison were determined by considering only heavily damaged or collapsed buildings. Therefore, the idea was to steer from post-event-only damage assessment and to build on existing methods based on change detection.

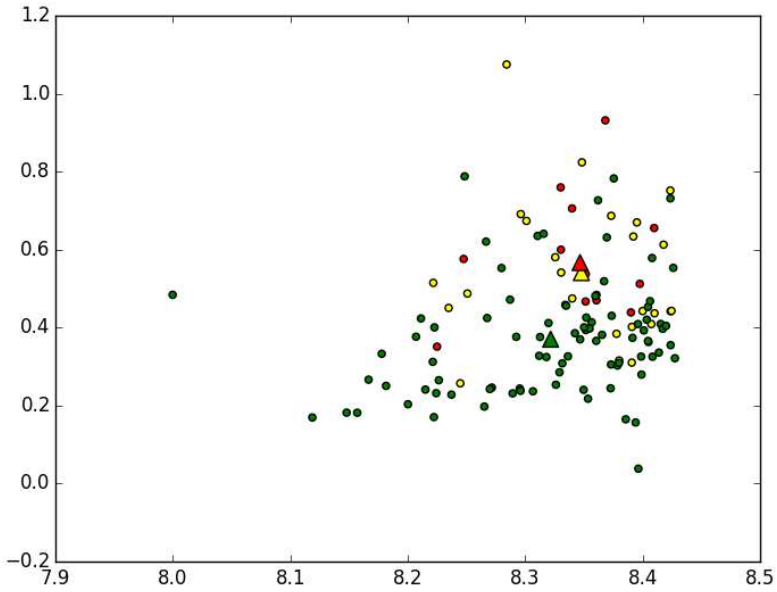


Figure 4.7: *Scatter plot with entropy (x-axis) and density (y-axis). Red=high damage, yellow=medium damage, green=low or no damage. Triangles represent averages over each colour.*

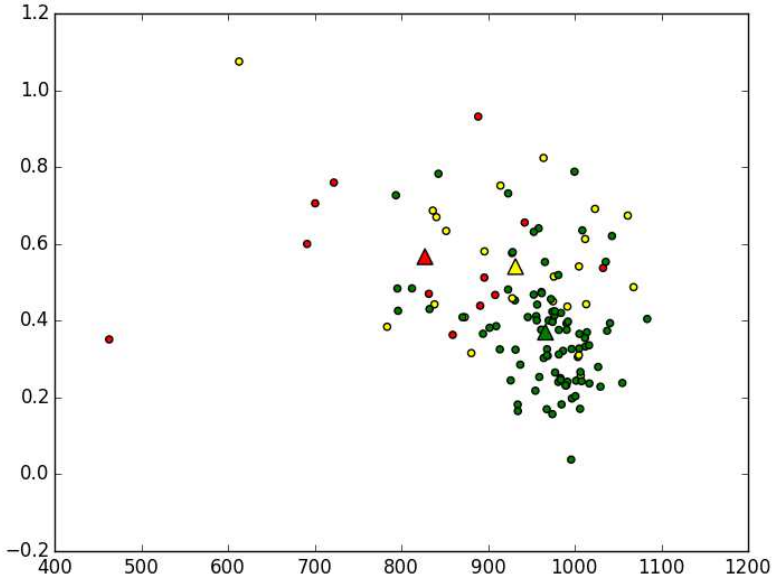


Figure 4.8: *Scatter plot with contrast (x-axis) and density (y-axis). Same visual coding as above.*

The results obtained show that, not only different fragility levels, but also differ-

ent density values seem actually to correlate with different levels of damage and thus suggest a further correction to be applied to the radar-based estimate.

4.4 Change detection approach

A different and more classical approach was also considered. Starting from a well-known change detection procedure, a few modifications have been proposed in order to move towards a fully automatic procedure. This method has been tested on TerraSAR-X stripmap and ALOS-2 PALSAR-2 data, with a spatial resolution ranging from 1.25 to 6 meters.

4.4.1 Preprocessing

Standard preprocessing involved geocoding and radiometric calibration -carried out using SARScape- and speckle filtering.

4.4.2 Difference and Correlation

Using a combination of images taken before and after the event it is possible to compute two measures, difference (see eq. 4.4) and correlation (eq. 4.3); previous papers have already demonstrated the effectiveness of this approach [144] [150] [151].

$$d = |\overline{Ib_i} - \overline{Ia_i}|, \tag{4.2}$$

where:

- *d* is the absolute difference of pre- and post-event images
- $\overline{Ib_i}$ is the *i*-th pixel value of the pre-event image in dB and averaged using a 9x9 window
- $\overline{Ia_i}$ is the *i*-th pixel value of the post-event image in dB and averaged using a 9x9 window.

$$r = \frac{N\sum_{i=1}^{N} Ia_i \cdot Ib_i - \sum_{i=1}^{N} Ia_i \cdot \sum_{i=1}^{N} Ib_i}{\sqrt{(N\sum_{i=1}^{N} Ia_i^2 - (\sum_{i=1}^{N} Ia_i)^2) \cdot (N\sum_{i=1}^{N} Ib_i^2 - (\sum_{i=1}^{N} Ib_i)^2)}}, \tag{4.3}$$

where:

- *r* is the correlation of pre- and post-event images computed over a 9x9 window
- *N* is the number of pixels determined by the window size
- Ia_i is the *i*-th pixel value of the post-event image in dB
- Ib_i is the *i*-th pixel value of the pre-event image in dB.

However, an inherent limitation is the need to define thresholds in order to associate values to damage levels; additionally, threshold values are not stable across different test cases and the visual definition relies on the damage validation dataset. Due to all these problems, an improvement is proposed with the aim to avoid the manual selection of the boundary value. A first attempt was based on the calculation of correlation and difference, the aggregation within polygons (footprints, blocks, grids) and the successive classification of the results. The underlying idea was that the aggregation could help smoothing out the false positives and delineate only the real damage. However, after a few experiments over the Port-au-Prince test case, the original hypothesis turned out not to be as good as expected. More details related to the completed experiments are included in the results paragraph. Modifications to the original procedure led to move the classification step before the aggregation; in this way, the damage mask is determined by a direct selection of the class showing the highest difference, an hint that those values could be related to damage. The aggregation is used to remove salt-and-pepper noise generated by the classification and reduce the false positive values.

An evolution of the method has also been proposed to handle multiple pre-event acquisitions. The idea is to calculate difference and correlation on the pre-pre combination, obtain a mask for vegetated and soil areas subject to changes not due to earthquake damage and, in the end, use the mask to highlight only affected areas. A similar workflow was already described in literature by Matsuoka et al. [152]: the idea was to compute the correlation between the pre-event datasets, extract the stable areas by applying a threshold and then compute the difference of correlations using this formula:

$$r_{diff} = r_{ab} - r_{bb} \tag{4.4}$$

where:

- r_{diff} is the difference of correlations
- r_{ab} is the correlation between post- and pre-event data
- r_{bb} is the correlation between pre- and pre-event data.

However, a direct application of the same technique did not produce the expected results, thus calling for a modified version. The proposed approach is built upon the following steps:

1. compute the difference of pre-pre used to discard areas with changes (related for example to soil)
2. use the difference of correlations to remove zones showing a positive correlation difference
3. classify the difference of differences and pick the two classes with the highest positive and highest negative changes (most probably related to damage)

4.4.3 Results

Port-au-Prince

One pre- and one post-earthquake TerraSAR-X stripmap data are available for the AoI. Regarding the ground truth, two layers are publicly downloadable:

- PDNA assessment: a set of points obtained using crowdsourcing to manually interpret satellite and aerial images; it covers both damaged and undamaged buildings.

- EEFIT mission: a sample of footprints with validation provided by earthquake engineers

Different attempts were made in order to compare the truth layers with the outcome of the change detection process. Aggregation of values has two main advantages: creating a base to compare the results and reducing the false positive errors due to salt-and-pepper noise. In the first attempt an average within the grid was computed but the output was uncorrelated with damage. The explanation is that averaging causes a shrink of the values into a normal distribution, therefore flattening the differences caused by damage. A second attempt was made trying to average within the EEFIT footprints; the drawback is represented by the reduced size of these footprints compared with the spatial resolution of the input image. Therefore, it was impossible to properly distinguish between damaged and undamaged buildings. Also the number of buildings available was not sufficient from a statistical point of view. In the last attempt, classes were automatically selected and then aggregated using the artificially created 50x50m grid. For sake of comparison, also the PDNA points were counted within each rectangle producing 3 different "classes" (see figure 4.9):

- No buildings in case of 0 points within the polygon (cyan)

- Undamaged if the majority of points has a value ranging from 0 to 2 (green)

- Damaged if the PDNA damage ranges from 3 to 5 (red).

Because of the automatic nature of the proposed change detection method, highlighted changes can be related to vegetation and not only to buildings. Grid elements without buildings were therefore excluded from the accuracy calculation.

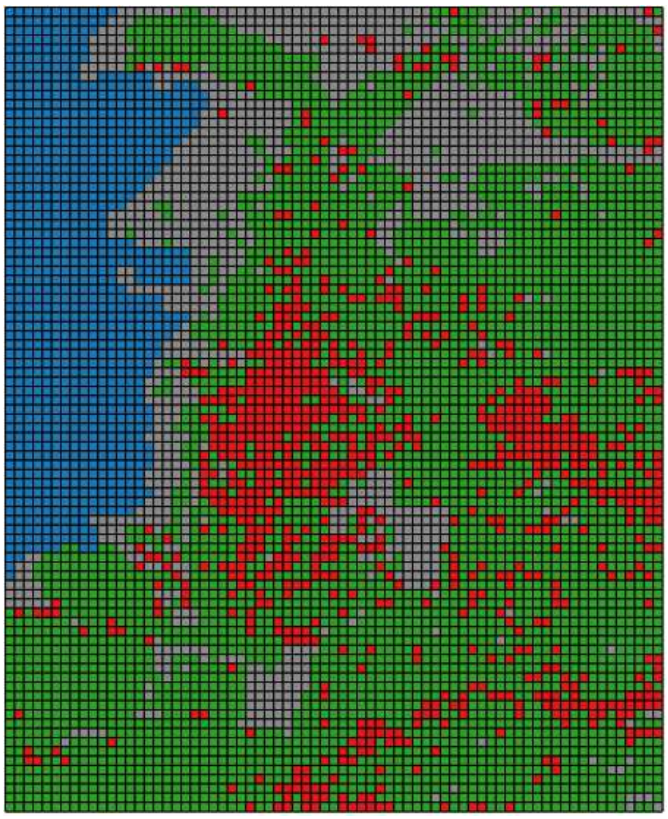


Figure 4.9: *Result of the PDNA aggregated values. In blue sea-related units, in gray blocks without buildings, in green areas with limited damage while in red areas with extensive damage.*

Table 4.1: *Accuracy of the change detection approach over Port-au-Prince. In blue sea-related units, in gray blocks without buildings, in green areas with limited damage while in red areas with extensive damage.*

	True Positive	True Negative	
Change det. Positive	431	767	35.98%
Change det. Negative	406	2605	86.51%
	51.49%	77.25%	72.13%

Table 4.2: *Available datasets*

Acquisition Date	Incidence angle (deg)	Pre/Post
2014/04/10	36.3	pre
2015/02/21	36.3	pre
2015/05/02	36.3	post

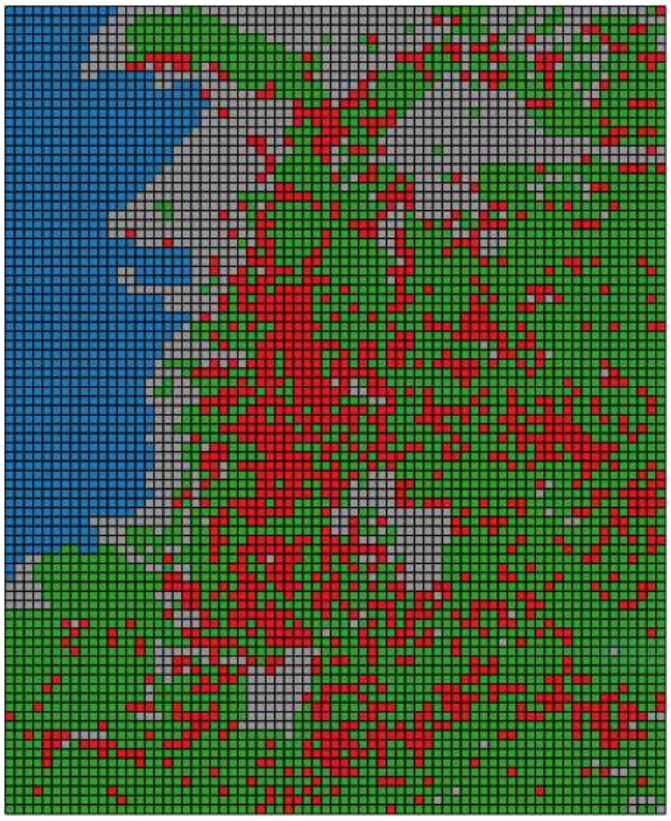


Figure 4.10: *Aggregated damage mask.*

Bhaktapur

The modified approach was applied because of the availability of multiple pre-event data. A different behaviour is also expected because of the different satellite used, ALOS-2 PALSAR-2, which offers a lower spatial resolution. This gave us a chance to evaluate the stability of the method at different spatial resolutions. Three datasets were available for the event (see table 4.2).

The availability of two pre-event images gave the chance to check the effectiveness of the change detection approach. Ground truth data is based on a collaborative effort coordinated by UNOSAT (United Nations Institute for Training and Research) (see figure 4.11); damage was assessed from satellite imagery and only damaged buildings were marked. This represents a disadvantage in respect of Port-au-Prince because it is not possible to follow the same aggregation principle. Moreover, these two layers are not directly comparable because of the different origin: the crowdsourced layer was generated based on very high resolution images while the proposed result was calculated on a lower resolution base.

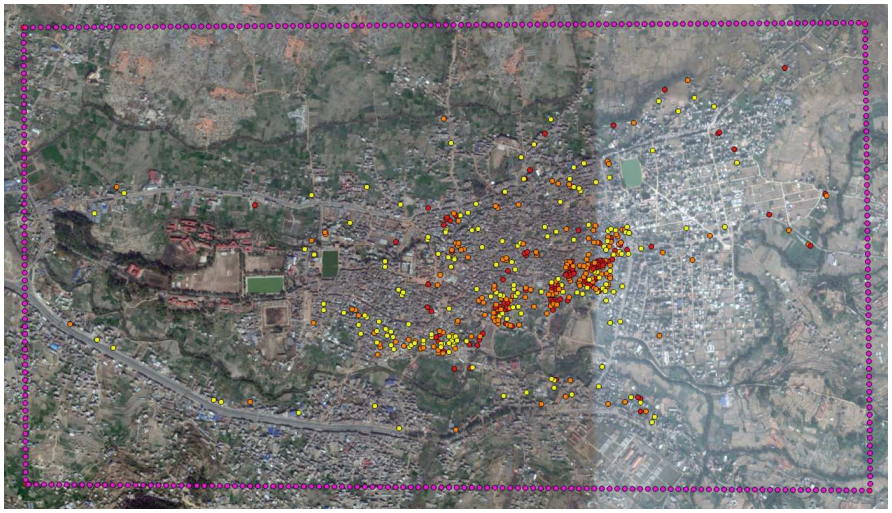


Figure 4.11: *Bhaktapur with damage assessed by UNOSAT. Yellow points are related to EMS grade 3 damage, orange points to level 4 and red points represent grade 5 buildings.*

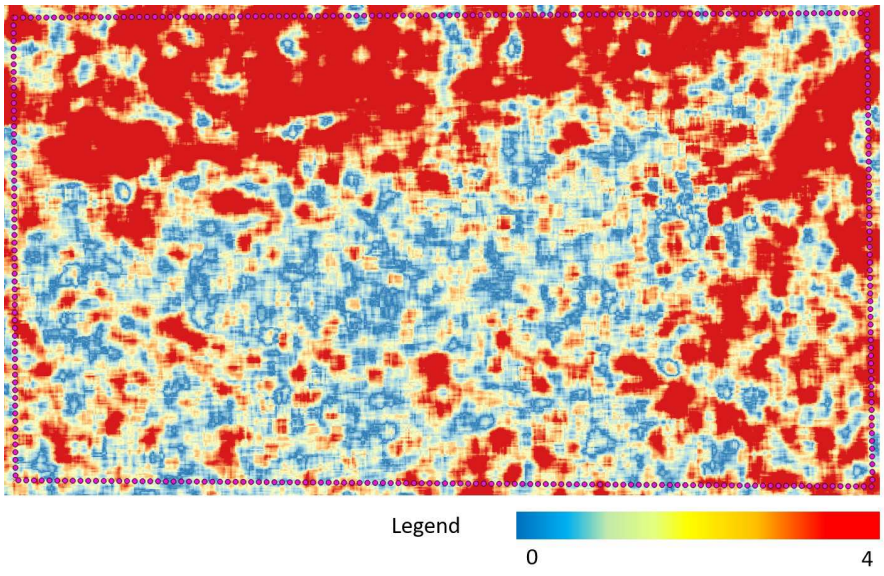


Figure 4.12: *Bhaktapur difference computed between the two pre-event images.*

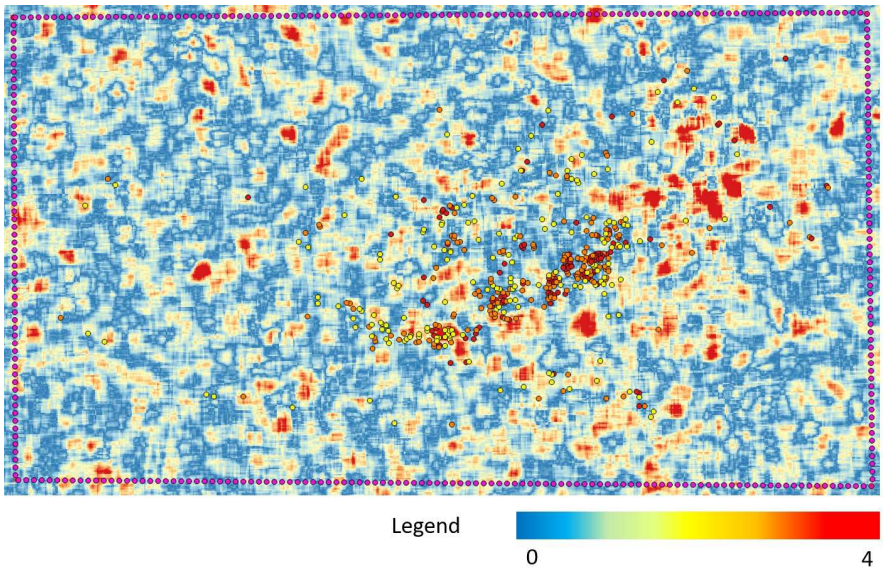


Figure 4.13: *Bhaktapur difference computed between pre- and post-event images.*

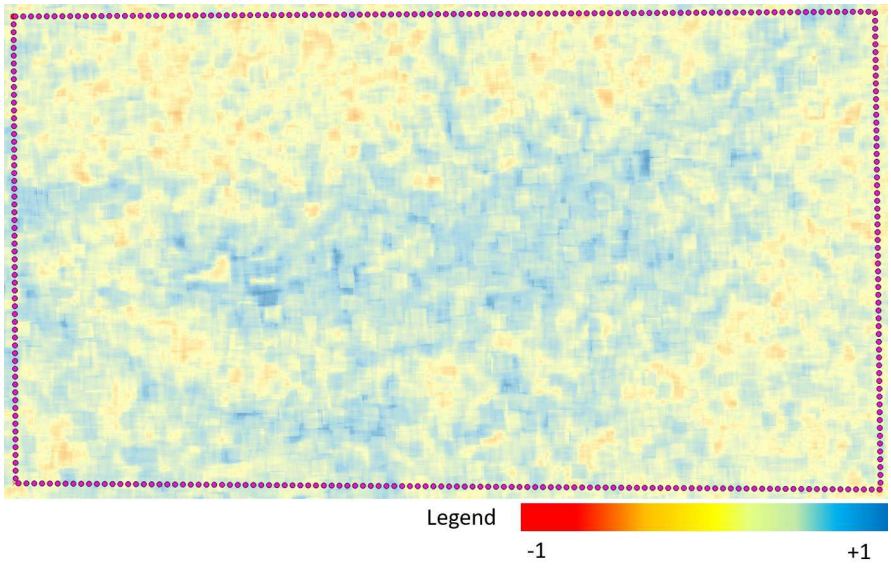


Figure 4.14: *Bhaktapur correlation computed between the two pre-event images.*

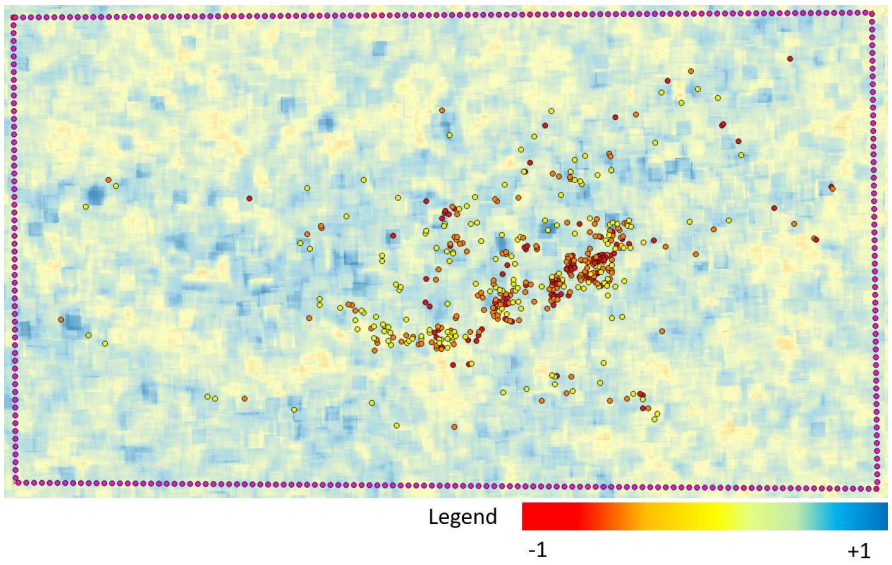


Figure 4.15: *Bhaktapur correlation computed between pre- and post-event images.*

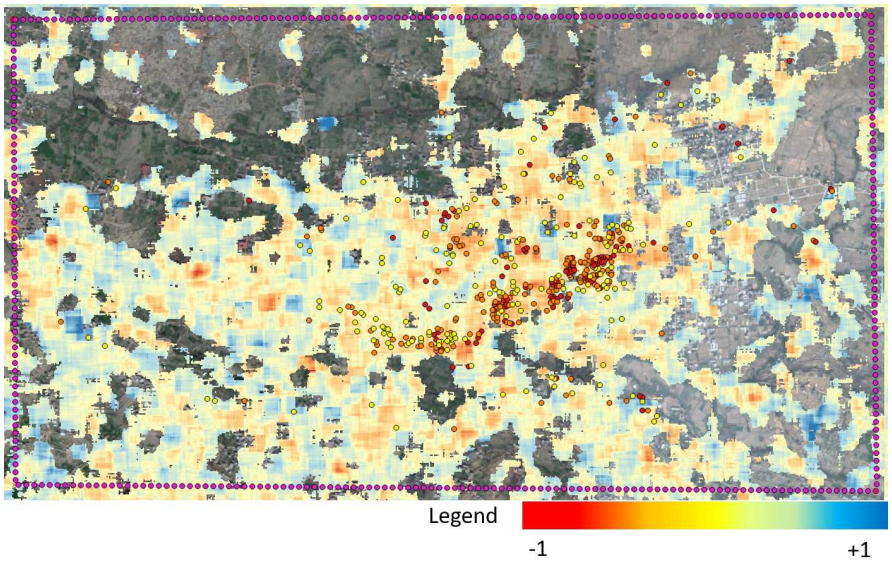


Figure 4.16: *Bhaktapur difference of correlation.*

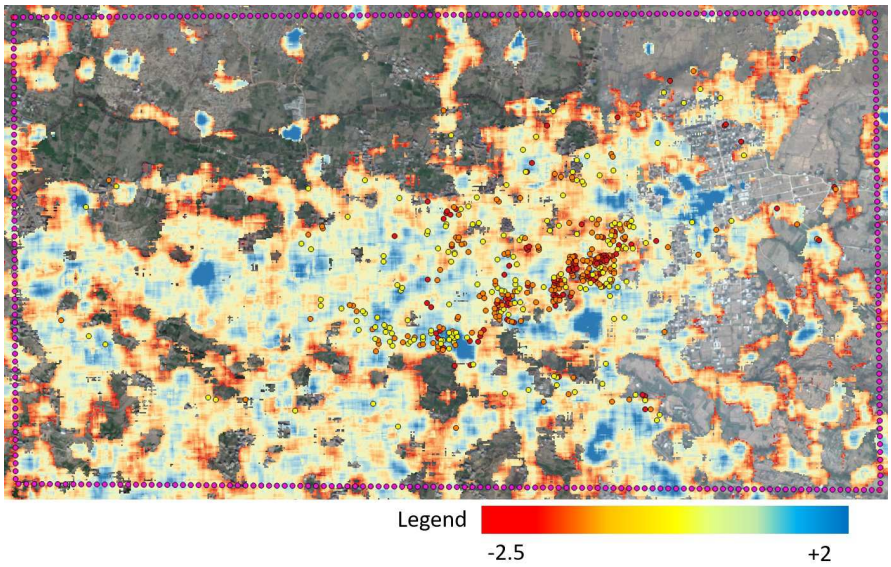


Figure 4.17: *Bhaktapur difference of difference.*

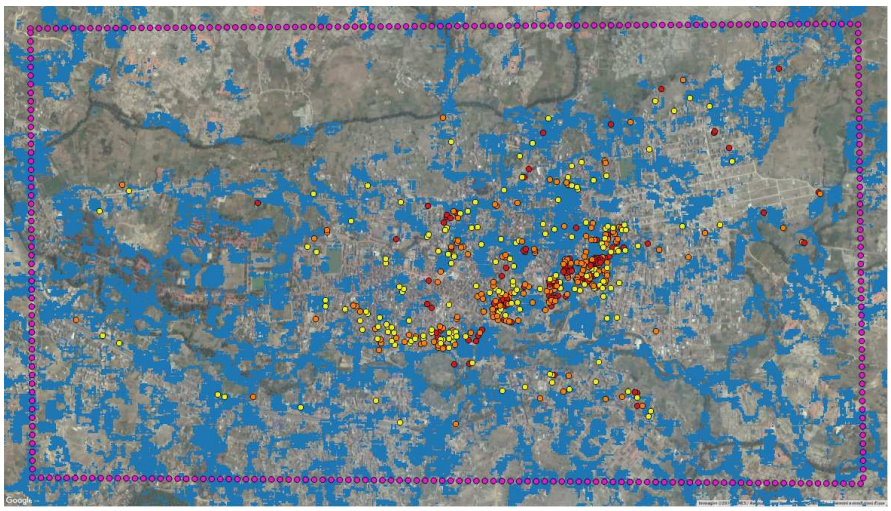


Figure 4.18: *Bhaktapur output mask. Damage extracted from change detection is represent by the blue colour.*

A visual inspection of the results led to the conclusion that the output mask is over-estimating damage. A few possible explanations are:

- unstable classification of the difference layer (by considering one single band as input -therefore one dimension- the KMeans algorithm becomes a quantization process)
- insufficient spatial resolution
- high density of buildings limiting the wave penetration
- a challenging type of damage, inherently difficult to estimate from satellite

It must be reminded that ground truth was generated using optical images with a limiting nadiral view. The assessment will most probably underestimate the real damage.

Mashiki

The method based on two pre-event images was also applied to the Kumamoto earthquake, with particular focus on Mashiki. Unfortunately, no comprehensive ground truth layer was available at the moment on the area of interest. Therefore, the damage layer was manually delineated by visual interpretation combining pre- and post-event data suddenly made available in Google Earth and GSI aerial imagery. Three different types of damage could be retrieved: damaged roof, partial and total collapse. Buildings footprints were also made available by the GSI organization.

Table 4.3: *Available datasets on Mashiki*

Acquisition Date	Incidence angle (deg)	Pre/Post
2015/11/30	32.8	pre
2016/03/07	32.8	pre
2016/04/18	32.8	post

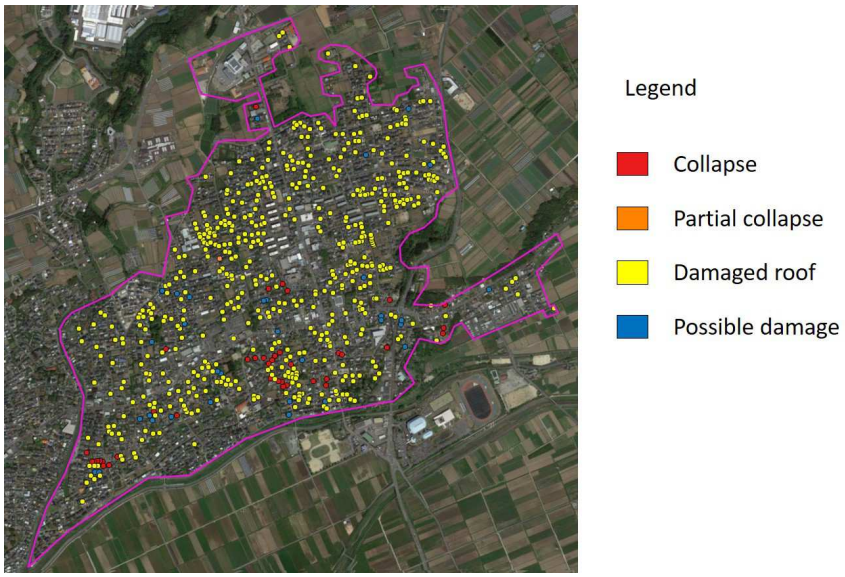


Figure 4.19: *Mashiki with manual damage assessment made of points.*

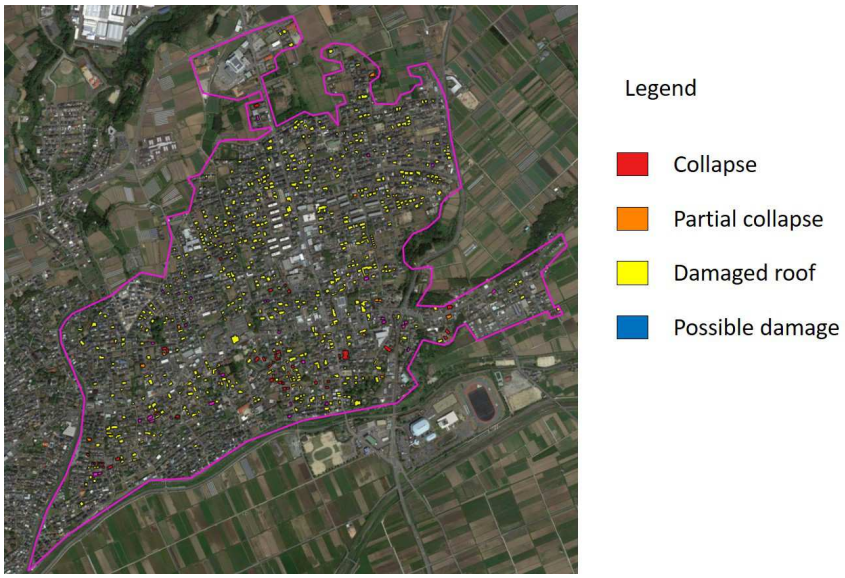


Figure 4.20: *Mashiki damage assessment aggregated to footprints.*

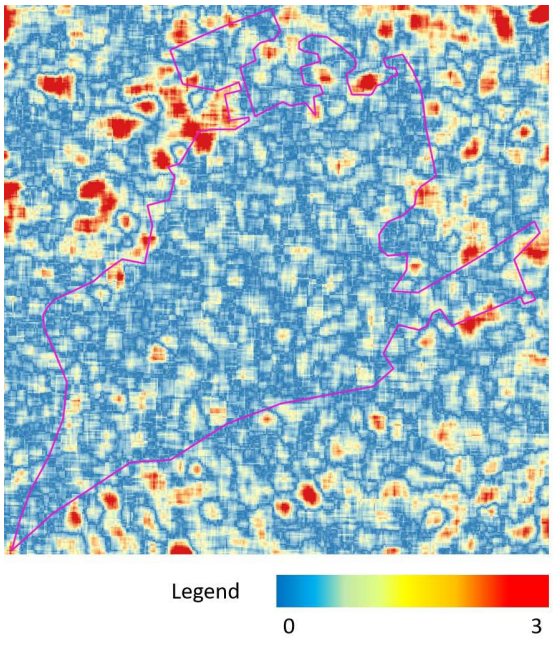


Figure 4.21: *Mashiki difference computed between the two pre-event images.*

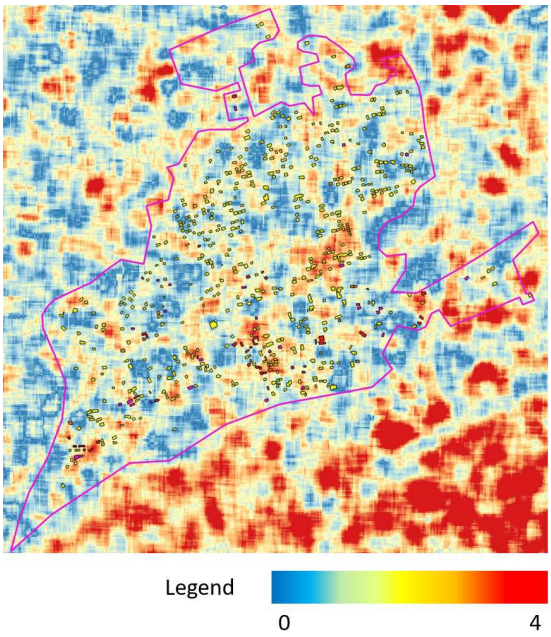


Figure 4.22: *Mashiki difference computed between pre- and post-event images.*

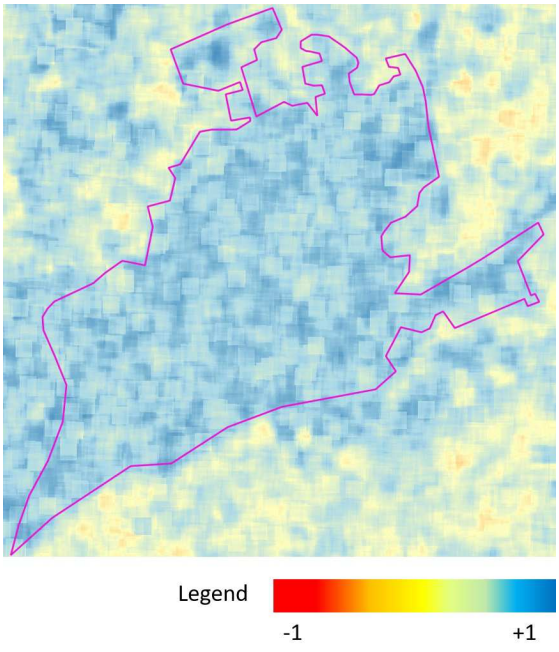


Figure 4.23: *Mashiki correlation computed between the two pre-event images.*

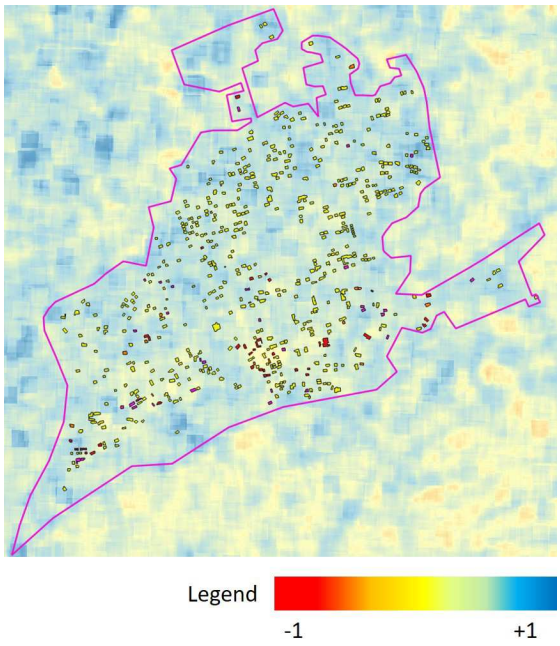


Figure 4.24: *Mashiki correlation computed between pre- and post-event images.*

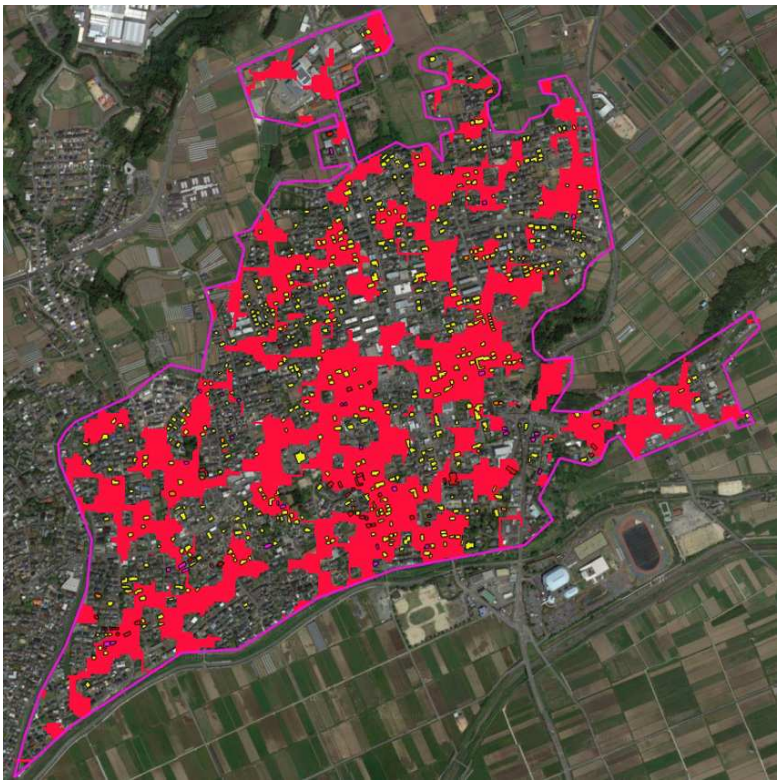


Figure 4.25: *Mashiki output mask.*

In this case, the output mask tends to underestimate damage. This can be explained by the majority of damage due to collapsed roofs, therefore not clearly recorded at

the ALOS-2 PALSAR-2 spatial resolution. The combination of different masks can influence the output; for example, areas facing changes before the earthquake could be masked out.

4.5 Combination with pre-event information

The proposed workflow is shown in figure 4.26. The idea is to apply a common procedure for SAR data by computing the difference and correlation between pre- and post-event images. At the same time, additional data gathered before the earthquake are integrated into the assessment, specifically the density of footprints and the output of the fragility curves are used in the damage assessment procedure. A detailed description of each step follows.

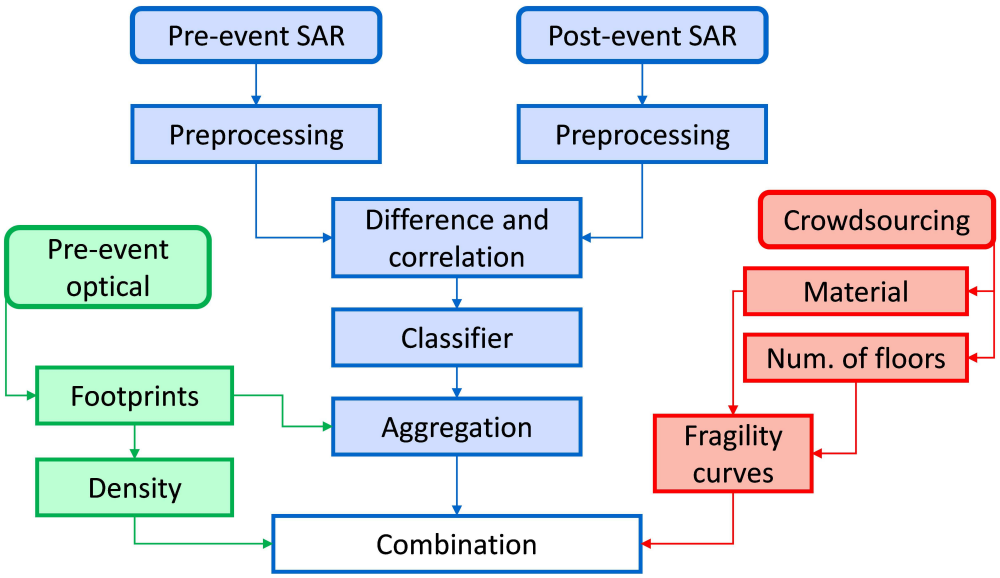


Figure 4.26: *Workflow of the proposed method.*

4.5.1 Preprocessing

Input images were already geocoded as a result of previous research work [147]. As reported in this latter publication, perfect radiometric calibration was impossible to achieve due to software-related issues. Once geocoded and converted to dB, they were ready for the next step. No speckle filter was applied due to the unexpected introduction of artefacts, mainly blobs of no data, reducing its benefit. Unfortunately, the problem came out even when using different software tools.

4.5.2 SAR change detection

Starting from the mask produced using previously proposed change detection method, the results are aggregated using footprints, meaning that a percentage of intersection between the damage layer and polygons is computed for each building. The output is a set of footprints with an attribute that can be interpreted as the likelihood of damage for each building.

4.5.3 Pre-event data

Data related to the test case of interest can be retrieved before the event, focusing on the exposure and vulnerability parameters of the risk equation. Several international initiatives (e.g. FP7 projects SENSUM [58] and RASOR [59], the Global Earthquake Model foundation [153]) contributed in the definition of platforms and tools for the extraction of exposure parameters from different sources of data, mainly Earth Observation and expert judgement. A set of exposure indicators derivable from remote sensing is included in [70]. Moreover, satellite imagery is not the only possible source of information: the circulation of smartphones and tablets with built-in GPS receivers and cameras led to what can be pictured as a dense network of observers spread all over the world; through crowdsourcing one can take advantage of this network in order to collect data and integrate spaceborne remote sensing.

For the specific topic of seismic damage assessment, the authors considered three different attributes: density of buildings, number of storeys and material. The former measure can be derived from remote sensing using the "density tool" of the SENSUM Earth Observation Tools while the others can be collected from mobile apps and are fundamental for the selection of the appropriate fragility curves.

Density

The tool is the same reported in chapter 2.

Material and number of storeys

These pieces of information are difficult to retrieve from remote sensing and also not commonly available around the world. Crowdsourcing could contribute in filling by the gap, asking volunteers to provide basic information through their smartphones. The GEM-IDCT (Inventory Data Capture Tool) Direct Observation tool was designed with the aim to collect very complex information necessary to build a solid and comprehensive model; however, for the scope of interest, material and number of storeys proved to be enough for the selection of the appropriate fragility curve. Unfortunately, for the test case of interest this simple app for collection was not ready; therefore, the extraction was accomplished by manual interpretation of Google Earth 3D models, available for free from the Google Earth Pro platform. An example is showed in figure 4.27. The number of floors can be derived by counting horizontal rows of windows while material information is guessed based on the exterior aspect and a priori knowledge of local history.

Figure 4.27: *Example of 3D model used to estimate number of floors and material. Google ©.*

The collected information is used, along with the PGA (Peak Ground Acceleration) recorded for the event, to determine the appropriate fragility curve. Given a level of expected damage and the PGA value, the curve returns the probability of exceeding a given "damage threshold" for the specific building.

4.5.4 Combination and damage assessment

Fusion of the obtained information from pre/post SAR data change detection, pre-event fragility and density can be obtained by a linear combination of the likelihoods adjusted using suitable weights (equation 4.5). The assumption is that remote sensing can provide clues but it has inherent limitations, mostly related to wave penetration and spatial resolution of the data; at the same time, fragility curves are usually generic for a certain area of interest and only return a probability of exceedance:

$$d_{l,i} = c_p \cdot I_{p,i} + c_d \cdot D_{b,i} + c_f \cdot F_{b,i}, \tag{4.5}$$

where:

- $d_{l,i}$ is the damage likelihood for the *i*-th polygon
- c_p is the coefficient related to the damage mask extracted by change detection
- $I_{p,i}$ is the percentage of intersection with the *i*-th footprint
- c_d is the density coefficient
- $D_{b,i}$ is the density for the *i*-th building
- c_f is the fragility curve coefficient
- $F_{b,i}$ is the fragility value given by the curve for the *i*-th footprint.

The aim of the proposed methodology is to provide a quick and rough estimate of damage after the event, thus supporting the search&rescue efforts by prioritizing heavy damaged areas.

4.5.5 Results

In order to compare the same entities, the output mask generated by the algorithm is aggregated into the previously mentioned footprints; the first experiment regarded the change-detection-only damage assessment. Based on the mask shown in figure 11, footprints were marked as damaged in case of area intersection ratio exceeding 20%; the threshold value was defined based on the assumption that such changes would in most cases correspond to non-negligible damage. In the comparison with the reference layer, every building reporting a grade greater than or equal to 3 was considered as damaged. A summary of the results can be found in table 4.4 where the number of buildings for each category is reported. Another test was made by modifying the threshold value, considering as damaged only those buildings categorized with levels 4 and 5, basically looking at partial or total collapse. The underlying hypothesis is the supposed higher visibility of collapse with respect to non-structural damage to walls. The results of this experiment are included in table 4.5.

Table 4.4: *Accuracy of the change detection-only approach. A building is considered damage in case of EMS value greater than or equal to 3.*

	Positive GT	Negative GT	
Change det. Positive	452	176	71.97%
Change det. Negative	857	309	26.50%
	34.53%	63.71%	42.42%

Table 4.5: *Accuracy of the change detection-only approach with values of damage equal or greater than 4.*

	True Positive	True Negative	
Change det. Positive	147	481	23.41%
Change det. Negative	318	848	72.72%
	31.61%	63.81%	55.46%

In both cases the approach based on pure change detection offered poor values of overall accuracy. The difficulty in correctly estimating damage in this case can be explained by the type of damage suffered by buildings and the conformation of the city centre; in particular, the value of 42.42% obtained from the first attempt can be explained by the challenging extraction of grade 3 damage from remote sensing. The third test regarded the proposed combined approach of fragility, density and change detection. As per the ground truth layer, every building labelled with a level greater or equal to 3 was considered damaged (see table 4.6). Footprints were considered damaged with a damage likelihood value greater than 0.25. The set of coefficients was based on visual inspection:

- density coefficient $c_d = 0.2$
- change detection coefficient $c_p = 0.35$
- fragility curve coefficient $c_f = 0.45$.

Table 4.6: *Accuracy of the combined approach with values of damage equal or greater than 3.*

	Positive GT	Negative GT	
Change det. Positive	743	284	72.35%
Change det. Negative	139	102	42.32%
	84.24%	26.42%	66.64%

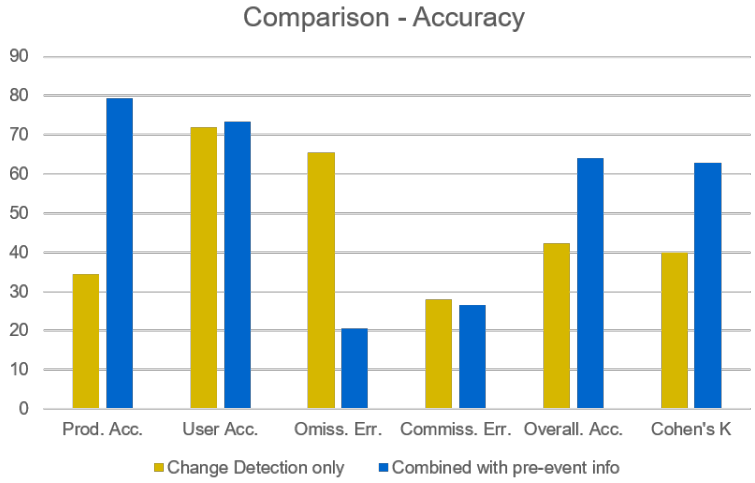


Figure 4.28: *Comparison of the different methods tested. The Cohen's K value has been multiplied by 100 for a better comparison.*

Table 4.7: *Comparison of the different methods*

Accuracies	CD grade 3	CD grade 4	Combined
Prod. Acc. (%)	34.53	31.61	79.37
User Acc. (%)	71.97	23.41	73.48
Omis. Err. (%)	65.47	68.39	20.63
Comm. Err. (%)	28.02	76.59	26.52
Over. Acc. (%)	42.42	55.46	64.05
Cohen's K	0.40	0.53	0.63

The results demonstrated that the combination of fragility and change detection led to a substantial improvement of the detected damage. The recorded accuracies are still not very high but could support a quick-and-rough estimation, as expected (see figure 4.28 and table 4.7). The K value trend is positive compared with the change detection-only methodologies. A limitation of the proposed approach lies in the scarce availability of pre- and post-event spotlight data, an acquisition mode not frequently activated in case of earthquakes. The resolution of other types of data like stripmap, more frequently available, does not permit the aggregation by footprint due to the limited spatial resolution.

CHAPTER 5

Conclusions

This thesis highlights the importance of the combination of remote sensing and crowdsourcing in the disaster cycle, in particular looking at mitigation, preparedness and response phases of the disaster cycle.

In the framework of the SENSUM project, a list of exposure proxies has been discussed and settled by Mostapha Harb, former PhD student from the same research group at the University of Pavia. Based on the proposed table, a set of python tools has been designed with the aim to push automation as much as possible. The aim is to reduce the experience needed in order to process medium and high resolution EO data, in particular in the risk assessment field. All the code produced has been released with an open source license and it is free to use and modify for non-commercial purposes. Compared to public layers only recently released, the SENSUM Earth Observation Tools can be executed on user-defined data and not on predefined time periods. This represents an advantage due to the differences in urban development experienced in different areas of the world. Limitations in the user friendliness and usability of the tools have been reported and only partially tackled. For example, the medium resolution-oriented algorithms have been implemented within the ESA GPOD system, where input from users is limited to the region-of-interest and the selection of the acquisition dates. Processing is automatically and simultaneously performed within the ESA machines, therefore avoiding a cumbersome installation and data preparation.

However, remote sensing is not the only "theme" in this dissertation, as discussed in chapter 3. The extensive diffusion of smartphones defined a dense network of sensors, capable to provide complementary data. Two projects led to the development of a general and multi-purpose framework for data collection and validation. The first - SEGUICI Vegetation Report- was developed by Daniel Aurelio Galeazzo while my contribution was limited to the definition of the user experience and the algorithm,

hosted directly on the server, for the automatic matching of reports with polygons. The second -named CLOOPSy- was designed and developed by myself as part of the MyGEOSS contest. Unfortunately, there was no time left to have an operational mobile service for the collection of exposure-related data by non-expert users. The idea is to ask volunteers to provide a picture, to count the number of floors and choose the construction material among the most common ones in the considered area. The mobile apps produced within the mentioned projects will require just a few modifications to follow the proposed new specifications.

Moving to the response phase, different approaches have been considered depending on the input datasets. First, a post-event only method has been tested building on previous works of Pavia group. Next, as part of the experience in Japan with Prof. Yamazaki group, change detection approaches -built upon their experience- and further enhancements have been implemented and tested using different datasets and test cases. The results obtained so far suggest that there is actually some correlation between certain features of pre- and post-event SAR data and the level and distribution of seismic damage in urban areas. Evidence also suggest that damage assessment based on SAR data may be substantially strengthened by the injection of pre-event vulnerability information supported by the collection of crowdsourced data from mobile phones of volunteer contributors. Currently, though, the levels of accuracy that could be obtained are still unsatisfactory except for purposes of a quick-and-rough assessment aimed at prioritizing emergency intervention. In order for the methods to become useful, much more research effort has to be invested including possibly also a deeper integration of fragility information about the mapped buildings. Moreover, it is still hard to find a set of fragility curves for each earthquake-prone area. Central Italy and L'Aquila are well studied and data is available; however, this assumption is not valid everywhere. To be fully operational, the proposed approach should adapt to the amount of usable data in the affected area and consider different SAR inputs according to the availability.

Acknowledgements

First of all, I'd like to thank my supervisor Prof. Fabio Dell'Acqua, for his guidance, support, and unlimited patience demonstrated in these years. He managed to transfer his passion and to stimulate me to go always a step further.

Many thanks also to Prof. Paolo Gamba for his precious feedback and helpful suggestions. I would also like to thank Prof. Raul Feitosa and Prof. Gilson Costa from Pontificia Universidade Catolica do Rio de Janeiro, Brazil. I'm also grateful to Prof. Fumio Yamazaki and Prof. Wen Liu from Chiba University, Japan. I would like to thank the reviewers of this manuscript for their precious feedback.

Special thanks to Mostapha Harb, with whom I shared part of my work, lots of ideas and tonnes of challenges. Furthermore, I'm very grateful to all my lab colleagues (Gianni Cristian Iannelli, Niccoló Ricardi, Amparito Morales Figueroa, Andreas Salentinig, Andrea Marinoni, Gianni Lisini, Anna Vizziello, Taqi Eddine, Giovanni Rogolino, Luca Pasca) which made working fun and stimulating at the same time.

Particular thanks goes to my friend and colleague Daniel Aurelio Galeazzo, who collaborated with me in several projects and whose help has been fundamental.

I would also like to thank my friends, who helped me in understanding that there's a lot more than just work.

The greatest thanks goes to Camilla, for her fundamental support, her advices, her never ending belief in me and most of all for her love, which enlightened the darkest and the brightest moments.

Last but not least, I wish to thank my family. Without their never ending support and patience I would not have achieved this important goal.

Bibliography

[1] "Sendai Framework for Disaster Risk Reduction - UNISDR." [Online]. Available: http://www.unisdr.org/we/coordinate/sendai-framework

[2] UNISDR, "2009 UNISDR Terminology on Disaster Risk Reduction."

[3] "The Disaster Management Cycle." [Online]. Available: http://www.gdrc.org/uem/disasters/1-dm_cycle.html

[4] "Disaster Management Cycle." [Online]. Available: http://mjcetenvsci.blogspot.it/2014/11/disaster-management-cycle.html

[5] ESA, "How does Earth Observation work?" [Online]. Available: http://www.esa.int/Our_Activities/Observing_the_Earth/How_does_Earth_Observation_work

[6] "ESA - Eduspace EN - Home - History of Earth observation," Nov. 2009. [Online]. Available: https://www.esa.int/SPECIALS/Eduspace_EN/SEM1NP3Z2OF_0.html

[7] "Landsat Missions Timeline." [Online]. Available: http://landsat.usgs.gov/about_mission_history.php

[8] M. A. Wulder and N. C. Coops, "Make Earth observations open access," *Nature*, vol. 513, pp. 30–31, Sep. 2014. [Online]. Available: http://www.nature.com/news/satellites-make-earth-observations-open-access-1.15804

[9] "Landsat Timeline." [Online]. Available: http://landsat.gsfc.nasa.gov/a-landsat-timeline/

[10] A. Irwin, *Citizen Science: A Study of People, Expertise and Sustainable Development*. Psychology Press, 1995.

[11] Y. Bhattacharjee, "Citizen scientists supplement work of cornell researchers," *Science*, vol. 308, no. 5727, pp. 1402–1403, 2005. [Online]. Available: http://science.sciencemag.org/content/308/5727/1402

[12] J. Silvertown, "A new dawn for citizen science," *Trends in ecology & evolution*, vol. 24, no. 9, pp. 467–471, 2009. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S016953470900175X

[13] ""Citizen Data Scientist" Revolution." [Online]. Available: http://www.kdnuggets.com/2016/03/citizen-data-scientist-revolution.html

[14] K. Noyes, "The rise of the citizen data scientist," Apr. 2016. [Online]. Available: http://www.computerworld.com/article/3051605/big-data/the-rise-of-the-citizen-data-scientist.html

[15] "Cost action - mapping and the citizen sensor." [Online]. Available: http://www.cost.eu/COST_Actions/ict/TD1202

[16] "European Citizen Science Association (ECSA)." [Online]. Available: http://ecsa.citizen-science.net/

[17] "IGARSS 2015 | 2015 IEEE International Geoscience and Remote Sensing Symposium | 13-18 July 2015 | Milan, Italy." [Online]. Available: http://www.igarss2015.org/RegularProgram.asp

[18] "eoscience20.org." [Online]. Available: http://www.eoscience20.org/

[19] B. Fishbain, U. Lerner, N. Castell, T. Cole-Hunter, O. Popoola, D. M. Broday, T. M. Iniguez, M. Nieuwenhuijsen, M. Jovasevic-Stojanovic, D. Topalovic, R. L. Jones, K. S. Galea, Y. Etzion, F. Kizel, Y. N. Golumbic, A. Baram-Tsabari, T. Yacobi, D. Drahler, J. A. Robinson, D. Kocman, M. Horvat, V. Svecova, A. Arpaci, and A. Bartonova, "An evaluation tool kit of air quality micro-sensing units," *Science of The Total Environment*, Sep. 2016. [Online]. Available: http://linkinghub.elsevier.com/retrieve/pii/S0048969716319799

[20] "Citclops." [Online]. Available: http://www.citclops.eu/

[21] "Home | Cobweb." [Online]. Available: https://cobwebproject.eu/

[22] "ESA EDUCEO | Education for Earth Observation capitalizing on a Citizen Science approach." [Online]. Available: http://educeo.info/

[23] K. W. Willett, C. J. Lintott, S. P. Bamford, K. L. Masters, B. D. Simmons, K. R. Casteels, E. M. Edmondson, L. F. Fortson, S. Kaviraj, W. C. Keel, and others, "Galaxy Zoo 2: detailed morphological classifications for 304 122 galaxies from the Sloan Digital Sky Survey," *Monthly Notices of the Royal Astronomical Society*, p. stt1458, 2013. [Online]. Available: http://mnras.oxfordjournals.org/content/early/2013/09/15/mnras.stt1458.short

[24] "GalaxyZoo." [Online]. Available: https://www.galaxyzoo.org/

[25] "Geo-Wiki." [Online]. Available: http://geo-wiki.org/

[26] R. Sieber, "Public participation geographic information systems: A literature review and framework," *Annals of the Association of American Geographers*, vol. 96, no. 3, pp. 491–507, 2006. [Online]. Available: http://www.tandfonline.com/doi/abs/10.1111/j.1467-8306.2006.00702.x

[27] M. F. Goodchild, "Citizens as sensors: the world of volunteered geography," *GeoJournal*, vol. 69, no. 4, pp. 211–221, Nov. 2007. [Online]. Available: http://link.springer.com/10.1007/s10708-007-9111-y

[28] S. Elwood, M. F. Goodchild, and D. Z. Sui, "Researching Volunteered Geographic Information: Spatial Data, Geographic Research, and New Social Practice," *Annals of the Association of American Geographers*, vol. 102, no. 3, pp. 571–590, May 2012. [Online]. Available: http://www.tandfonline.com/doi/abs/10.1080/00045608.2011.595657

[29] "USGS - Did You Feel It?" [Online]. Available: https://earthquake.usgs.gov/data/dyfi/

[30] "Did You Feel It?: Crowdsourcing Earthquake Maps." [Online]. Available: https://www.citizenscience.gov/2015/06/25/did-you-feel-it/

[31] "Citizens and Remote Sensing Observation Network (CARSON)." [Online]. Available: https://terra.nasa.gov/citizen-science

[32] "ESA for Kids." [Online]. Available: https://www.esa.int/esaKIDSen/Earth.html

[33] C. Heipke, "Crowdsourcing geospatial data," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 65, no. 6, pp. 550–557, Nov. 2010. [Online]. Available: http://linkinghub.elsevier.com/retrieve/pii/S0924271610000602

[34] S. Ghosh, C. K. Huyck, M. Greene, S. P. Gill, J. Bevington, W. Svekla, R. DesRoches, and R. T. Eguchi, "Crowdsourcing for rapid damage assessment: The global earth observation catastrophe assessment network (GEO-CAN)," *Earthquake Spectra*, vol. 27, no. S1, pp. S179–S198, 2011. [Online]. Available: http://www.earthquakespectra.org/doi/abs/10.1193/1.3636416

[35] L. Barrington, S. Ghosh, M. Greene, S. Har-Noy, J. Berger, S. Gill, A. Y.-M. Lin, and C. Huyck, "Crowdsourcing earthquake damage assessment using remote sensing imagery," *Annals of Geophysics*, vol. 54, no. 6, 2012. [Online]. Available: http://www.annalsofgeophysics.eu/index.php/annals/article/view/5324

[36] DigitalGlobe, "Open data to support hurricane response in Haiti | Seeing a better worldâ." [Online]. Available: http://blog.digitalglobe.com/2016/10/07/open-data-volunteer-mapping-to-support-hurricane-response-in-haiti/

[37] L. See, P. Mooney, G. Foody, L. Bastin, A. Comber, J. Estima, S. Fritz, N. Kerle, B. Jiang, M. Laakso, H.-Y. Liu, G. Milcinski, M. Niksic, M. Painho, A. Podör, A.-M. Olteanu-Raimond, and M. Rutzinger, "Crowdsourcing, Citizen Science or Volunteered Geographic Information? The Current State of Crowdsourced Geographic Information," *ISPRS International Journal of Geo-Information*, vol. 5, no. 5, p. 55, Apr. 2016. [Online]. Available: http://www.mdpi.com/2220-9964/5/5/55

[38] J. Burke, D. Estrin, M. Hansen, A. Parker, N. Ramanathan, S. Reddy, and M. B. Srivastava, "Participatory sensing," in *In: Workshop on World-Sensor-Web*

(WSWâ06): Mobile Device Centric Sensor Networks and Applications, 2006, pp. 117–134.

[39] W. Z. Khan, Y. Xiang, M. Y. Aalsalem, and Q. Arshad, "Mobile Phone Sensing Systems: A Survey," *IEEE Communications Surveys & Tutorials*, vol. 15, no. 1, pp. 402–427, 2013. [Online]. Available: http://ieeexplore.ieee.org/document/6177188/

[40] J. Howe, "The Rise of Crowdsourcing," *Wired Magazine*, Jun. 2006. [Online]. Available: https://www.wired.com/2006/06/crowds/

[41] "Scopus." [Online]. Available: https://www.scopus.com/home.uri

[42] C. J. Ferster and N. C. Coops, "A review of earth observation using mobile personal communication devices," *Computers & Geosciences*, vol. 51, pp. 339–349, Feb. 2013. [Online]. Available: http://linkinghub.elsevier.com/retrieve/pii/S0098300412003184

[43] V. Kotovirta, T. Toivanen, R. Tergujeff, T. HÃme, and M. Molinier, "Citizen Science for Earth Observation: Applications in Environmental Monitoring and Disaster Response," *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. XL-7/W3, pp. 1221–1226, Apr. 2015. [Online]. Available: http://www.int-arch-photogramm-remote-sens-spatial-inf-sci.net/XL-7-W3/1221/2015/

[44] M. Molinier, C. López-Sánchez, T. Toivanen, I. Korpela, J. Corral-Rivas, R. Tergujeff, and T. HÃme, "RelasphoneâMobile and Participative In Situ Forest Biomass Measurements Supporting Satellite Image Mapping," *Remote Sensing*, vol. 8, no. 10, p. 869, Oct. 2016. [Online]. Available: http://www.mdpi.com/2072-4292/8/10/869

[45] M. F. Goodchild and L. Li, "Assuring the quality of volunteered geographic information," *Spatial Statistics*, vol. 1, pp. 110–120, May 2012. [Online]. Available: http://linkinghub.elsevier.com/retrieve/pii/S2211675312000097

[46] L. See, A. Comber, C. Salk, S. Fritz, M. van der Velde, C. Perger, C. Schill, I. McCallum, F. Kraxner, and M. Obersteiner, "Comparing the Quality of Crowdsourced Data Contributed by Expert and Non-Experts," *PLoS ONE*, vol. 8, no. 7, p. e69958, Jul. 2013. [Online]. Available: http://dx.plos.org/10.1371/journal.pone.0069958

[47] U. Gadiraju, R. Kawase, S. Dietze, and G. Demartini, "Understanding malicious behavior in crowdsourcing platforms: The case of online surveys," in *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, ser. CHI '15. New York, NY, USA: ACM, 2015, pp. 1631–1640. [Online]. Available: http://doi.acm.org/10.1145/2702123.2702443

[48] "Celebrating a community of 400 million." [Online]. Available: http://blog.instagram.com/post/129662501137/150922-400million

[49] "Find every photo with flickrâs new unified search experience." [Online]. Available: http://blog.flickr.net/en/2015/05/07/flickr-unified-search/

[50] K. H. Leetaru, S. Wang, G. Cao, A. Padmanabhan, and E. Shook, "Mapping the global twitter heartbeat: The geography of twitter," *First Monday*, May 2013. [Online]. Available: http://firstmonday.org/article/view/4366/3654

[51] A. Stefanidis, A. Crooks, and J. Radzikowski, "Harvesting ambient geospatial information from social media feeds," *GeoJournal*, vol. 78, no. 2, pp. 319–338, Apr. 2013. [Online]. Available: http://link.springer.com/10.1007/s10708-011-9438-2

[52] A. Crooks, A. Croitoru, A. Stefanidis, and J. Radzikowski, "#Earthquake: Twitter as a distributed sensor system," *Transactions in GIS*, vol. 17, no. 1, pp. 124–147, 2013. [Online]. Available: http://dx.doi.org/10.1111/j.1467-9671.2012.01359.x

[53] S. Spyratos, M. Lutz, and F. Pantisano, "Characteristics of citizen-contributed geographic information," in *Proceedings of the AGILE'2014 International Conference on Geographic Information Science*, ser. AGILE'2014, 2014. [Online]. Available: https://agile-online.org/Conference_Paper/cds/agile_2014/agile2014_109.pdf

[54] F. Fischer, "VGI as Big Data," *GeoInformatics*, vol. 3, pp. 46–47, 2012.

[55] J. Estima and M. Painho, "Flickr geotagged and publicly available photos: Preliminary study of its adequacy for helping quality control of corine land cover," in *Computational Science and Its Applications - ICCSA 2013*, 2013, ch. 3, pp. 205–220.

[56] S. Lee, H. Zhang, and D. J. Crandall, "Predicting geo-informative attributes in large-scale image collections using convolutional neural networks," in *2015 IEEE Winter Conference on Applications of Computer Vision*, Jan 2015, pp. 550–557.

[57] F. Dell'Acqua, I. Lanese, and D. A. Polli, "Integration of EO-based vulnerability estimation into EO-based seismic damage assessment: a case study on l'aquila, Italy, 2009 earthquake," *Natural Hazards*, vol. 68, no. 1, pp. 165–180, Aug. 2013. [Online]. Available: http://link.springer.com/10.1007/s11069-012-0490-0

[58] "SENSUM project." [Online]. Available: http://www.sensum-project.eu/

[59] "Rasor Project." [Online]. Available: http://www.rasor-project.eu/

[60] "MARsite - New Directions in Seismic Hazard Assessment through Focused Earth Observation in the Marmara Supersite." [Online]. Available: http://marsite.eu/

[61] "SENSUM-project/sensum_rs_qgis." [Online]. Available: https://github.com/SENSUM-project/sensum_rs_qgis

[62] M. Harb, F. Dell'Acqua, and D. De Vecchi, "Multi-risk buildings exposure and physical vulnerability mapping from optical satellite images: Developing an integrated toolset," in *2014 IEEE Geoscience and Remote Sensing Symposium*. IEEE, 2014, pp. 1164–1166. [Online]. Available: http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=6946637

[63] D. De Vecchi, M. Harb, and F. Dell'Acqua, "Refining registration of Large, Multi-Temporal Stacks of Medium-Resolution images: a Novel, Automated approach for Big Heritage Data," *Proc. of BiDS*, 2014. [Online]. Available: https://www.researchgate.net/profile/Fabio_DellAcqua/publication/270756248_REFINING_REGISTRATION_OF_LARGE_MULTI-TEMPORAL_STACKS_OF_MEDIUM-RESOLUTION_IMAGES_A_NOVEL_AUTOMATED_APPROACH_FOR_BIG_HERITAGE_DATA/links/54b3c04f0cf2318f0f957259.pdf

[64] D. De Vecchi, M. Harb, G. C. Iannelli, P. Gamba, F. Dell'Acqua, and R. Q. Feitosa, "A feature-based approach to register CBERS CCD and HRC imagery for built-up area extraction purposes," in *2015 Joint Urban Remote Sensing Event (JURSE)*. IEEE, 2015, pp. 1–4. [Online]. Available: http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=7120350

[65] M. Harb, D. De Vecchi, and F. Dell'Acqua, "Automatic hybrid-based built-up area extraction from Landsat 5, 7, and 8 data sets," in *2015 Joint Urban Remote Sensing Event (JURSE)*. IEEE, 2015, pp. 1–4. [Online]. Available: http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=7120475

[66] D. De Vecchi, M. Harb, F. Dell'Acqua, and D. Aurelio Galeazzo, "An integrated, open-source set of tools for urban vulnerability monitoring from Earth observation data," in *EGU General Assembly Conference Abstracts*, vol. 17, 2015, p. 7434. [Online]. Available: http://adsabs.harvard.edu/abs/2015EGUGA..17.7434D

[67] M. Harb, D. De Vecchi, P. Gamba, F. Dell'Acqua, and R. Feitosa, "Automatic clouds/shadows extraction method from CBERS-2 CCD and LANDSAT data," in *2015 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*. IEEE, 2015, pp. 4594–4597. [Online]. Available: http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=7326851

[68] D. De Vecchi, D. A. Galeazzo, M. Harb, and F. Dell'Acqua, "Unsupervised change detection for urban expansion monitoring: An object-based approach," in *2015 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*. IEEE, 2015, pp. 350–352. [Online]. Available: http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=7325772

[69] D. De Vecchi, M. Harb, and F. Dell'Acqua, "A PCA-based hybrid approach for built-up area extraction from Landsat 5, 7 and 8 datasets," in *2015 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*. IEEE, 2015, pp. 1152–1154. [Online]. Available: http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=7325975

[70] M. M. Harb, D. De Vecchi, and F. Dell'Acqua, "Phisical Vulnerability Proxies from Remotes Sensing: Reviewing, Implementing and Disseminating Selected Techniques," *IEEE Geoscience and Remote Sensing Magazine*, vol. 3, no. 1, pp. 20–33, Mar. 2015. [Online]. Available: http://ieeexplore.ieee.org/document/7067063/

[71] "QGIS Python Plugins Repository." [Online]. Available: https://plugins.qgis.org/plugins/sensum_eo_tools/

[72] "Orfeo ToolBox - Orfeo ToolBox is not a black box." [Online]. Available: https://www.orfeo-toolbox.org/

[73] S. Lloyd, "Least squares quantization in PCM," *IEEE transactions on information theory*, vol. 28, no. 2, pp. 129–137, 1982. [Online]. Available: http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=1056489

[74] K. Fukunaga, *Introduction to statistical pattern recognition*, 2nd ed., ser. Computer science and scientific computing. Boston: Academic Press, 1990.

[75] A. Ben-Hur, D. Horn, H. T. Siegelmann, and V. Vapnik, "Support vector clustering," *Journal of machine learning research*, vol. 2, no. Dec, pp. 125–137, 2001. [Online]. Available: http://www.jmlr.org/papers/v2/horn01a.html

[76] J. R. Quinlan, "Induction of decision trees," *Machine learning*, vol. 1, no. 1, pp. 81–106, 1986. [Online]. Available: http://link.springer.com/article/10.1023/A:1022643204877

[77] N. S. Altman, "An Introduction to Kernel and Nearest-Neighbor Nonparametric Regression," *The American Statistician*, vol. 46, no. 3, p. 175, Aug. 1992. [Online]. Available: http://www.jstor.org/stable/2685209?origin=crossref

[78] T. K. Ho, "Random decision forests," in *Document Analysis and Recognition, 1995., Proceedings of the Third International Conference on*, vol. 1. IEEE, 1995, pp. 278–282. [Online]. Available: http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=598994

[79] J. H. Friedman, "Greedy function approximation: a gradient boosting machine," *Annals of statistics*, pp. 1189–1232, 2001. [Online]. Available: http://www.jstor.org/stable/2699986

[80] P. F. Felzenszwalb and D. P. Huttenlocher, "Efficient graph-based image segmentation," *International Journal of Computer Vision*, vol. 59, no. 2, pp. 167–181, 2004. [Online]. Available: http://link.springer.com/article/10.1023/B:VISI.0000022288.19776.77

[81] F. Meyer, "Un algorithme optimal pour la ligne de partage des eaux," vol. 2, Lyon, France, 1991, pp. 847–857.

[82] D. Comaniciu and P. Meer, "Mean shift analysis and applications," in *Computer Vision, 1999. The Proceedings of the Seventh IEEE International Conference on*, vol. 2. IEEE, 1999, pp. 1197–1203. [Online]. Available: http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=790416

[83] C. M. Christoudias, B. Georgescu, and P. Meer, "Synergism in low level vision," in *Pattern Recognition, 2002. Proceedings. 16th International Conference on*, vol. 4. IEEE, 2002, pp. 150–155. [Online]. Available: http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=1047421

[84] M. Baatz and A. Schape, "Multiresolution Segmentation: an optimization approach for high quality multi-scale image segmentation," 2000, oCLC: 611240115.

[85] M. Pesaresi and J. A. Benediktsson, "A new approach for the morphological segmentation of high-resolution satellite imagery," *IEEE transactions on Geoscience and Remote Sensing*, vol. 39, no. 2, pp. 309–320, 2001. [Online]. Available: http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=905239

[86] L. S. Bins, L. M. G. Fonseca, G. J. Erthal, and F. M. Ii, "Satellite imagery segmentation: a region growing approach," *Simposio Brasileiro de Sensoriamento Remoto*, vol. 8, no. 1996, pp. 677–680, 1996. [Online]. Available: http://www.academia.edu/download/6503536/10.1.1.93.4555.pdf

[87] S. N. Goward, J. G. Masek, D. L. Williams, J. R. Irons, and R. J. Thompson, "The Landsat 7 mission: Terrestrial research and applications for the 21st century," *Remote Sensing of Environment*, vol. 78, no. 1, pp. 3–12, 2001. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0034425701002620

[88] G. Tzimiropoulos, V. Argyriou, S. Zafeiriou, and T. Stathaki, "Robust FFT-Based Scale-Invariant Image Registration with Image Gradients," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 10, pp. 1899–1906, Oct. 2010. [Online]. Available: http://ieeexplore.ieee.org/document/5467089/

[89] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool, "Speeded-Up Robust Features (SURF)," *Computer Vision and Image Understanding*, vol. 110, no. 3, pp. 346–359, Jun. 2008. [Online]. Available: http://linkinghub.elsevier.com/retrieve/pii/S1077314207001555

[90] M. Calonder, V. Lepetit, C. Strecha, and P. Fua, "Brief: Binary robust independent elementary features," in *European conference on computer vision*. Springer, 2010, pp. 778–792. [Online]. Available: http://link.springer.com/chapter/10.1007/978-3-642-15561-1_56

[91] F. Liporace, L. Fonseca, and J. D. de Matos, "Brazilian EO Satellite Program Update," AMS Kepler, Tech. Rep., Apr. 2012.

[92] P. Gamba and M. Herold, *Global Mapping of Human Settlement*, ser. Taylor&Francis Series in Remote Sensing Applications. CRC Press, Taylor&Francis Group, 2009.

[93] A. Schneider, M. A. Friedl, and D. Potere, "A new map of global urban extent from MODIS satellite data," *Environmental Research Letters*, vol. 4, no. 4, p. 044003, Dec. 2009. [Online]. Available: http://stacks.iop.org/1748-9326/4/i=4/a=044003?key=crossref.6bde1987a5db9cd29ac85e4c73503481

[94] ——, "Mapping global urban areas using MODIS 500-m data: New methods and datasets based on 'urban ecoregions'," *Remote Sensing of Environment*, vol. 114, no. 8, pp. 1733–1746, Aug. 2010. [Online]. Available: http://linkinghub.elsevier.com/retrieve/pii/S003442571000091X

[95] G. Trianni, G. Lisini, E. Angiuli, E. A. Moreno, P. Dondi, A. Gaggia, and P. Gamba, "Scaling up to National/Regional Urban Extent Mapping Using Landsat Data," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 8, no. 7, pp. 3710–3719, Jul. 2015. [Online]. Available: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=7052336

[96] N. N. Patel, E. Angiuli, P. Gamba, A. Gaughan, G. Lisini, F. R. Stevens, A. J. Tatem, and G. Trianni, "Multitemporal settlement and population mapping from Landsat using Google Earth Engine," *International Journal of Applied Earth Observation and Geoinformation*, vol. 35, pp. 199–208, 2015. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0303243414001998

[97] M. Pesaresi, Guo Huadong, X. Blaes, D. Ehrlich, S. Ferri, L. Gueguen, M. Halkia, M. Kauffmann, T. Kemper, Linlin Lu, M. A. Marin-Herrera, G. K. Ouzounis, M. Scavazzon, P. Soille, V. Syrris, and L. Zanchetta, "A Global Human Settlement Layer From Optical HR/VHR RS Data: Concept and First Results," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 6, no. 5, pp. 2102–2131, Oct. 2013. [Online]. Available: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6578177

[98] "Using the USGS Landsat 8 Product." [Online]. Available: http://landsat.usgs.gov/Landsat8_Using_Product.php

[99] "WRS-1 Path/Row Shapefile." [Online]. Available: https://landsat.usgs.gov/tools_wrs-2_shapefile.php

[100] "EarthExplorer." [Online]. Available: http://earthexplorer.usgs.gov/

[101] J. W. Rouse, R. H. Haas, J. A. Schell, and D. W. Deering, "Monitoring vegetation systems in the great plains with ERTS," vol. SP-351, 1973, pp. 309–317. [Online]. Available: http://ntrs.nasa.gov/search.jsp?R=19740005838

[102] Y. Zha, J. Gao, and S. Ni, "Use of normalized difference built-up index in automatically mapping urban areas from TM imagery," *International Journal of Remote Sensing*, vol. 24, no. 3, pp. 583–594, Jan. 2003. [Online]. Available: http://www.tandfonline.com/doi/abs/10.1080/01431160304987

[103] H. Xu, "Modification of normalised difference water index (NDWI) to enhance open water features in remotely sensed imagery," *International Journal of Remote Sensing*, vol. 27, no. 14, pp. 3025–3033, Jul. 2006. [Online]. Available: http://www.tandfonline.com/doi/abs/10.1080/01431160600589179

[104] A. R. As-syakur, I. W. S. Adnyana, I. W. Arthana, and I. W. Nuarsa, "Enhanced Built-Up and Bareness Index (EBBI) for Mapping Built-Up and Bare Land in an Urban Area," *Remote Sensing*, vol. 4, no. 12, pp. 2957–2970, Oct. 2012. [Online]. Available: http://www.mdpi.com/2072-4292/4/10/2957/

[105] H. Xu, "Extraction of urban built-up land features from Landsat imagery using a thematicoriented index combination technique," *Photogrammetric Engineering & Remote Sensing*, vol. 73, no. 12, pp. 1381–1391, 2007. [Online].

Available: http://www.ingentaconnect.com/content/asprs/pers/2007/00000073/00000012/art00006

[106] ——, "Analysis of Impervious Surface and its Impact on Urban Heat Environment using the Normalized Difference Impervious Surface Index (NDISI)," *Photogrammetric Engineering & Remote Sensing*, vol. 76, no. 5, pp. 557–565, May 2010. [Online]. Available: http://openurl.ingenta.com/content/xref?genre=article&issn=0099-1112&volume=76&issue=5&spage=557

[107] C. Jieli, L. I. Manchun, L. I. U. Yongxue, S. Chenglei, and H. U. Wei, "Extract residential areas automatically by new built-up index," in *Geoinformatics, 2010 18th International Conference on*. IEEE, 2010, pp. 1–5. [Online]. Available: http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=5567823

[108] A. Varshney, "Improved NDBI differencing algorithm for built-up regions change detection from remote-sensing data: an automated approach," *Remote Sensing Letters*, vol. 4, no. 5, pp. 504–512, May 2013. [Online]. Available: http://www.tandfonline.com/doi/abs/10.1080/2150704X.2013.763297

[109] P. S. J. Chavez and A. Y. Kwarteng, "Extracting spectral contrast in Landsat Thematic Mapper image data using selective principal component analysis," *Photogrammetric Engineering & Remote Sensing*, vol. 55, no. 3, pp. 339–348, Mar. 1989.

[110] G. F. Byrne, P. F. Crapper, and K. K. Mayo, "Monitoring land-cover change by principal component analysis of multitemporal landsat data," *Remote Sensing of Environment*, vol. 10, no. 3, pp. 175–184, 1980.

[111] X. Li and A. G. O. Yeh, "Principal component analysis of stacked multi-temporal images for the monitoring of rapid urban expansion in the Pearl River Delta," *International Journal of Remote Sensing*, vol. 19, no. 8, pp. 1501–1518, 1998.

[112] J. S. Deng, K. Wang, Y. H. Deng, and G. J. Qi, "PCA-based land-use change detection and analysis using multitemporal and multisensor satellite data," *International Journal of Remote Sensing*, vol. 29, no. 16, pp. 4823–4838, Aug. 2008. [Online]. Available: http://www.tandfonline.com/doi/abs/10.1080/01431160801950162

[113] R. M. Haralick, K. Shanmugam, and I. H. Dinstein, "Textural features for image classification," *Systems, Man and Cybernetics, IEEE Transactions on*, no. 6, pp. 610–621, 1973. [Online]. Available: http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=4309314

[114] "RSS Join&Share." [Online]. Available: http://wiki.services.eoportal.org/tiki-custom_home.php

[115] P. G. Marchetti, G. Rivolta, S. D'Elia, J. Farres, G. Mason, and N. Gobron, "A Model for the Scientific Exploitation of Earth Observation Missions: the ESA Research and Service Support," *IEEE Geoscience and Remote Sensing Society Newsletter*, vol. 162, pp. 10–12, Mar. 2012.

[116] "Grid Processing On Demand - Home." [Online]. Available: http://gpod.eo.esa.int/

[117] R. Cuccu, G. Sabatino, J. M. Delgado, and G. Rivolta, "Enabling SAR data exploitation by processing on-demand," in *2015 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*. IEEE, 2015, pp. 1476–1479. [Online]. Available: http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=7326058

[118] K. Green, D. Kempka, and L. Lackey, "Using remote sensing to detect and monitor land cover and land use change," *Photogrammetric Engineering & Remote Sensing*, vol. 60, pp. 331–337, 1994.

[119] G. Iannelli, G. Lisini, F. Dell'Acqua, R. Feitosa, G. Costa, and P. Gamba, "Urban Area Extent Extraction in Spaceborne HR and VHR Data Using Multi-Resolution Features," *Sensors*, vol. 14, no. 10, pp. 18 337–18 352, Sep. 2014. [Online]. Available: http://www.mdpi.com/1424-8220/14/10/18337/

[120] H. Taubenböck, T. Esch, M. Wurm, M. Thiel, T. Ullmann, A. Roth, M. Schmidt, H. Mehl, and S. Dech, "Urban structure analysis of mega city Mexico City using multisensoral remote sensing data," U. Michel, D. L. Civco, M. Ehlers, and H. J. Kaufmann, Eds., Oct. 2008, p. 71100E. [Online]. Available: http://proceedings.spiedigitallibrary.org/proceeding.aspx?doi=10.1117/12.800272

[121] H. Taubenböck, T. Esch, A. Felbier, M. Wiesner, A. Roth, and S. Dech, "Monitoring urbanization in mega cities from space," *Remote Sensing of Environment*, vol. 117, pp. 162–176, Feb. 2012. [Online]. Available: http://linkinghub.elsevier.com/retrieve/pii/S0034425711003427

[122] M. Rota, A. Penna, and C. Strobbia, "Processing Italian damage data to derive typological fragility curves," *Soil Dynamics and Earthquake Engineering*, vol. 28, no. 10-11, pp. 933–947, Oct. 2008. [Online]. Available: http://linkinghub.elsevier.com/retrieve/pii/S0267726107001339

[123] H. Taubenböck, I. StandfuÃ, M. Klotz, and M. Wurm, "The physical density of the cityâdeconstruction of the delusive density measure with evidence from two european megacities," *ISPRS International Journal of Geo-Information*, vol. 5, no. 11, p. 206, 2016. [Online]. Available: http://www.mdpi.com/2220-9964/5/11/206

[124] P. Bisch, E. Carvalho, H. Degee, P. Fajfar, M. Fardis, P. Franchin, M. Kreslin, A. Pecker, P. Pinto, A. Plumier, H. Somja, and G. Tsionis, "Eurocode 8: Seismic Design of Buildings Worked examples," Joint Research Centre - European Commission, Tech. Rep., 2012.

[125] D. A. Galeazzo, D. De Vecchi, F. Dell'Acqua, and P. Demattei, "A small step towards the citizen sensor: A multi-purpose framework for mobile apps," in *Geoscience and Remote Sensing Symposium (IGARSS), 2015 IEEE International*. IEEE, 2015, pp. 1348–1350. [Online]. Available: http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=7326025

[126] "MYGEOSS." [Online]. Available: http://digitalearthlab.jrc.ec.europa.eu/mygeoss/index.cfm

[127] G. Ravazzani, C. Corbari, A. Ceppi, M. Feki, M. Mancini, F. Ferrari, R. Gianfreda, R. Colombo, M. Ginocchi, S. Meucci, D. De Vecchi, F. Dell'Acqua, and G. Ober, "From (cyber-)space to ground: new technologies for smart farming," *Hydrology Research*.

[128] "Google Play." [Online]. Available: https://play.google.com/store

[129] D. De Vecchi and F. DellAcqua, "Citizens as Source of "fresh" information: a mobile app for updated vegetation status," Apr. 2016. [Online]. Available: https://www.youtube.com/watch?v=_xr6Lr9RWxg

[130] "Seguici vegetation report android - Bitbucket." [Online]. Available: https://bitbucket.org/unsoteam/vegreport_android/src/d47c5b98bc68?at=master

[131] "Cloopsy - Public Reports." [Online]. Available: https://cloopsy.unipv.it/

[132] E. Roglia and M. Craglia, "MYGEOSS project," *GeoInformatics*, vol. 18, pp. 38–39, Sep 2015.

[133] "CLOOPSy REST API." [Online]. Available: https://cloopsy.unipv.it/rest-api/reports-public/

[134] "ESA Researh and Service Support." [Online]. Available: https://wiki.services.eoportal.org/tiki-custom_home.php

[135] "cloopsy / android." [Online]. Available: https://bitbucket.org/cloopsy/android

[136] J. Karaian, "The deadliest, and costliest, earthquakes in recent history." [Online]. Available: http://qz.com/392064/the-deadliest-and-costliest-earthquakes-in-recent-history/

[137] F. Dell'Acqua and P. Gamba, "Remote Sensing and Earthquake Damage Assessment: Experiences, Limits, and Perspectives," *Proceedings of the IEEE*, vol. 100, no. 10, pp. 2876–2890, Oct. 2012. [Online]. Available: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6235981

[138] L. Dong and J. Shan, "A comprehensive review of earthquake-induced building damage detection with remote sensing techniques," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 84, pp. 85–99, Oct. 2013. [Online]. Available: http://linkinghub.elsevier.com/retrieve/pii/S0924271613001627

[139] F. Yamazaki, Y. Yano, and M. Matsuoka, "Visual Damage Interpretation of Buildings in Bam City Using QuickBird Images Following the 2003 Bam, Iran, Earthquake," *Earthquake Spectra*, vol. 21, no. S1, pp. 329–336, Dec. 2005. [Online]. Available: http://earthquakespectra.org/doi/abs/10.1193/1.2101807

[140] M. F. Goodchild and J. A. Glennon, "Crowdsourcing geographic information for disaster response: a research frontier," *International Journal of Digital Earth*, vol. 3, no. 3, pp. 231–241, Sep. 2010. [Online]. Available: http://www.tandfonline.com/doi/abs/10.1080/17538941003759255

[141] M. Chini, C. Bignami, S. Stramondo, W. J. Emery, and N. Pierdicca, "Quickbird panchromatic images for mapping damage at building scale caused by the 2003 bam earthquake," in *Geoscience and Remote Sensing Symposium, 2008. IGARSS 2008. IEEE International*, vol. 2. IEEE, 2008, pp. II–1029. [Online]. Available: http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=4779173

[142] S. Stramondo, C. Bignami, M. Chini, N. Pierdicca, and A. Tertulliani, "Satellite radar and optical remote sensing for earthquake damage detection: results from different case studies," *International Journal of Remote Sensing*, vol. 27, no. 20, pp. 4433–4447, Oct. 2006. [Online]. Available: http://www.tandfonline.com/doi/abs/10.1080/01431160600675895

[143] M. Chini, N. Pierdicca, and W. Emery, "Exploiting SAR and VHR Optical Images to Quantify Damage Caused by the 2003 Bam Earthquake," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 47, no. 1, pp. 145–152, Jan. 2009. [Online]. Available: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=4674629

[144] M. Matsuoka and F. Yamazaki, "Use of satellite SAR intensity imagery for detecting building areas damaged due to earthquakes," *Earthquake Spectra*, vol. 20, no. 3, pp. 975–994, 2004. [Online]. Available: http://www.earthquakespectra.org/doi/abs/10.1193/1.1774182

[145] T. Balz and M. Liao, "Building-damage detection using post-seismic high-resolution SAR satellite data," *International Journal of Remote Sensing*, vol. 31, no. 13, pp. 3369–3391, Jul. 2010. [Online]. Available: http://www.tandfonline.com/doi/abs/10.1080/01431161003727671

[146] P. T. Brett and R. Guida, "Geometry-based SAR curvilinear feature selection for damage detection," in *Synthetic Aperture Radar, 2012. EUSAR. 9th European Conference on*. VDE, 2012, pp. 62–65. [Online]. Available: http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=6216946

[147] F. Dell'Acqua, P. Gamba, and D. A. Polli, "Earthquake damage assessment from post-event VHR radar data: From Sichuan, 2008 to Haiti, 2010," in *Urban Remote Sensing Event (JURSE), 2011 Joint*. IEEE, 2011, pp. 201–204. [Online]. Available: http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=5764755

[148] V. Pasquali, "Global Finance Magazine - The World's Richest and Poorest Countries." [Online]. Available: https://www.gfmag.com/global-data/economic-data/worlds-richest-and-poorest-countries

[149] "Sentinel-1 Toolbox." [Online]. Available: https://sentinel.esa.int/web/sentinel/toolboxes/sentinel-1

[150] P. Uprety, F. Yamazaki, and F. Dell'Acqua, "Damage Detection Using High-Resolution SAR Imagery in the 2009 L'Aquila, Italy, Earthquake," *Earthquake Spectra*, vol. 29, no. 4, pp. 1521–1535, 2013. [Online]. Available: http://www.earthquakespectra.org/doi/abs/10.1193/060211EQS126M

[151] G. Deodatis, P. D. Spanos, and M. Shinozuka, Eds., *Computational Stochastic Mechanics: [the Sixth International Conference on Computational Stochastic Mechanics (CSM6) was held on the island of Rhodes (Rodos), Greece on June 13 - 16, 2010 ... the conference also celebrated the 80th Anniversary of the birthday of Dr. Masanobu Shinozuka ...]*. Singapore: Research Publishing Services, 2011, oCLC: 931338607.

[152] M. Matsuoka, "Use of ALOS/PALSAR imagery for monitoring areas damaged due to recent natural disasters," vol. 6412, Dec. 2006.

[153] "GEM Foundation | GEM - Global Earthquake Model." [Online]. Available: https://www.globalquakemodel.org/