SAPIENZA
UNIVERSITÀ DI ROMA

Sapienza Università di Roma

Dottorato di Ricerca in Scienze della Terra, XXXVI Ciclo

Ph.D. thesis in Engineering Geology (GEO/05)

DATA-DRIVEN GEOLOGICAL MULTI-HAZARD RISK ANALYSIS AT URBAN SCALE

Giandomenico Mastrantoni

Academic Year 2023 – 2024

Ph.D. Supervisors:

Prof. Paolo Mazzanti

Dipartimento di Scienze della Terra

Sapienza Università di Roma

P.le Aldo Moro 5, 00185 Roma, Italy

Prof. Carlo Esposito

Dipartimento di Scienze della Terra

Sapienza Università di Roma

P.le Aldo Moro 5, 00185 Roma, Italy

Revised by:

Prof. Luigi Lombardo

Faculty of Geo-Information Science and, Earth Observation (ITC),

University of Twente,

Enschede, Netherlands

Prof. Mario Floris

Dipartimento di Geoscienze

Università di Padova

Padova, Italy

"The Scientist must set in order.

Science is built up with facts, as a house is with stones.

But a collection of facts is no more a science than a heap of stones is a house."

Henri Poincare, Science and Hypothesis.

TABLE OF CONTENTS

ABSTRACT

This research work explores the intricate domain of geological multi-hazard risk assessment in urban environments, with a particular focus on the Municipality of Rome, Italy. By leveraging data-driven methodologies, advanced Machine Learning techniques, and open-access data, this study addresses the challenges posed by ground instability hazards, including landslides, subsidence, and sinkholes. The research offers valuable insights, innovative models, and practical applications, contributing to urban resilience enhancement and risk mitigation.

As the global trend of urbanization continues to transform landscapes and increase population concentration in cities, the understanding and management of geological hazards in urban areas have never been more critical. This research encompasses three key articles published within the thesis, to delve into geological multi-hazard risk analysis in the urban context.

In Chapter 2, we tackle the challenge of integrating fragmented, incomplete, and often unreliable data sources for landslide hazard assessment. Recognizing the crucial role of hazard evaluation in risk mitigation policies, we focused on making the most of available data. The methods employed include landslide inventory review, boosting the training dataset, and optimizing Machine Learning-based susceptibility analysis. The research explores temporal recurrence by analysing multi-temporal landslide inventories and historical rainfall databases, offering valuable insights into hazard probability. The integration of spatial and temporal attributes led to the creation of Rome's first large-scale landslide hazard product. This product, designed for statutory purposes, represents a significant milestone in improving the resilience of urban areas.

Chapter 3 delves into the world of Machine Learning techniques for landslide susceptibility mapping, emphasizing the significance of data quality. The study compared various models, with the ExtraTreesClassifier emerging as the most reliable choice. However, recognizing the rarity of high-quality, site-specific landslide inventories, the research examines the reliability of open-source landslide susceptibility maps at different spatial scales. By conducting a thorough statistical and spatial analysis, the research highlights the impact of mapping unit and analysis scale on prediction accuracy. Furthermore, the fusion of multiple low-resolution susceptibility maps is introduced, offering a promising approach to overcome data limitations. This milestone not only improves our understanding of landslide susceptibility mapping but also provides practical tools for large-scale assessments in urban environments.

The development of a novel model for multi-risk assessment takes centre stage in Chapter 4. The model's goal is to support decision-makers and risk managers in prioritizing buildings exposed to

ground instability hazards, including landslides, subsidence, and sinkholes. By integrating spatial hazard assessments, satellite interferometric data, building characteristics, and asset market values, the model calculates a multi-risk score, enabling the ranking of urban buildings. The research offers a detailed explanation of how each dataset contributes to the overall risk assessment, from spatial hazard to potential economic damage and activity rates. The data harmonization process leads to a practical, semi-quantitative approach for multi-risk assessment, fostering a more proactive stance in mitigating ground instability hazards.

This Ph.D. work could bring relevant insights for geological multi-hazard risk assessment and urban resilience. Its findings lay the foundation for further research, fostering safer, more resilient cities worldwide.

Keywords: Landslide Hazard, Multi–Risk, Urban Environment, Spatial Analysis, Machine Learning, Open Data, Satellite Interferometry, Data fusion.

RIASSUNTO

Questo lavoro di ricerca esplora l'intricato campo della valutazione dei rischi geologici in contesti urbani, con particolare attenzione al Comune di Roma, Italia. Utilizzando metodologie guidate dai dati, avanzate tecniche di apprendimento automatico (Machine Learning) e dati in libero accesso, questo studio affronta le sfide poste dai pericoli legati all'instabilità del terreno, tra cui frane, subsidenza e sinkhole. La ricerca offre preziose intuizioni, modelli innovativi e applicazioni pratiche, contribuendo al potenziamento della resilienza urbana e alla mitigazione del rischio.

Poiché la tendenza globale dell'urbanizzazione continua a trasformare i paesaggi e ad aumentare la concentrazione della popolazione nelle città, la comprensione e la gestione dei rischi geologici in aree urbane non sono mai state così critiche. Questa ricerca comprende tre articoli chiave pubblicati all'interno della tesi, per approfondire la valutazione dei rischi geologici multipli nel contesto urbano.

Nel Capitolo 2 affrontiamo la sfida di integrare fonti di dati frammentate, incomplete e spesso non affidabili per la valutazione del rischio da frana. Riconoscendo il ruolo cruciale della valutazione della pericolosità nelle politiche di mitigazione del rischio, ci siamo concentrati sull'ottimizzazione dei dati disponibili. I metodi impiegati includono la revisione degli inventari di frana, il potenziamento del dataset di addestramento dei modelli e l'ottimizzazione dell'analisi di suscettibilità basata sul Machine Learning supervisionato. La ricerca esplora la ricorrenza temporale attraverso l'analisi degli inventari di frana multi-temporali ed archivi storici delle precipitazioni, offrendo preziose informazioni sulla probabilità di innesco delle frane. L'integrazione degli attributi spaziali e temporali ha portato alla creazione del primo prodotto locale della pericolosità da frana di Roma. Questo prodotto, progettato per scopi statutari, rappresenta una tappa significativa nel migliorare la resilienza delle aree urbane.

Il Capitolo 3 si addentra nel mondo delle tecniche di Machine Learning supervisionate per la mappatura della suscettibilità di frana, sottolineando l'importanza della qualità dei dati. Lo studio ha confrontato vari modelli, con l'ExtraTreesClassifier che emerge come la scelta più affidabile. Tuttavia, riconoscendo la rarità degli inventari di frana di alta qualità sito-specifici, la ricerca esamina l'affidabilità delle mappe di suscettibilità ad accesso libero esistenti a diverse scale spaziali. Attraverso una dettagliata analisi statistica e spaziale, la ricerca evidenzia l'impatto dell'unità di mappatura e della scala di analisi sulla precisione delle previsioni. Inoltre, viene introdotta la fusione di diverse mappe di suscettibilità a bassa risoluzione, offrendo un approccio promettente per superare le limitazioni dei singoli prodotti. Ciò non solo migliora la comprensione della mappatura della suscettibilità di frana, ma fornisce anche strumenti pratici per valutazioni su larga scala in contesti urbani.

Lo sviluppo di un nuovo modello per la valutazione del rischio multiplo è al centro del Capitolo 4. L'obiettivo del modello è quello di supportare i decisori e i gestori del rischio nella priorizzazione degli edifici esposti ai rischi legati all'instabilità del terreno, tra cui frane, subsidenza e sinkhole. Integrando valutazioni della pericolosità spaziale, dati interferometrici satellitari, caratteristiche degli edifici e valori di mercato degli asset, il modello calcola un punteggio di rischio multiplo, consentendo la classificazione degli edifici urbani. La ricerca offre una spiegazione dettagliata di come ciascun dataset contribuisca alla valutazione complessiva del rischio, dalla valutazione della pericolosità spaziale al danno economico potenziale e allo stato di attività dei processi. Il processo di integrazione dei dati porta a un approccio semiquantitativo pratico per la valutazione del rischio multiplo, promuovendo una posizione più proattiva nella mitigazione dei rischi legati all'instabilità del terreno.

Questo lavoro di dottorato potrebbe portare rilevanti intuizioni per la valutazione dei rischi geologici e la resilienza urbana. I suoi risultati pongono le basi per ulteriori ricerche, favorendo la creazione di città più sicure e resilienti in tutto il mondo.

Parole chiave: Pericolosità da frana; Multi-Rischio; Contesto Urbano; Analisi Spaziale; Machine Learning; Dati Aperti; Interferometria Satellitare; Fusione dei Dati.

CHAPTER 1

1. Introduction

1.1. The value of Open Science

Open science is a movement to make publicly funded scientific research transparent, inclusive, accessible, and reproducible. It is based on the principles of openness, transparency, and reproducibility, which will increase the pace and quality of scientific progress. Open science practices can be applied to all aspects of the research process, from data collection and analysis to publication and dissemination (Vicente-Saez and Martinez-Fuentes 2018). Open science is important for several reasons. First, it helps to ensure the quality and reliability of scientific research. By making research more transparent and reproducible, open science makes it easier for other scientists to identify and correct errors (Open Science Collaboration 2015). Second, it makes scientific research more accessible to the public. This can help to promote public understanding of science and build trust in the scientific community. Third, it fosters collaboration among scientists. By sharing data and code, scientists can work together to solve problems more quickly and efficiently. Among the several open science practices, some valuable examples include open access publishing, open research data, and open source software (Figure 1.1). To help build a culture of open science, NASA is championing a new initiative: the Transform to Open Science (TOPS), which is part of the Open-Source Science Initiative (OSSI), making open science a reality on a large scale.

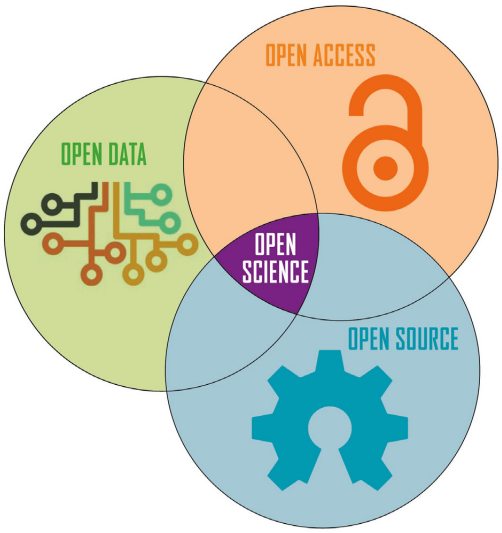


Figure 1.1: The value of open science. Open data can be defined simply as data that anyone can access, use, and share. Open access literature is digital, online, free of charge, and mostly free of copyright and licensing restrictions. Open source is software with source code that anyone can view, copy, modify, and share.

Open data is essential for conducting reliable scientific research. By making research data openly available, scientists can:

- Improve the transparency and reproducibility of research. When research data is open, other scientists can review and analyse the data to verify the findings of the original study. This can help to identify errors and biases, and to ensure that the findings are robust and reproducible.
- Foster collaboration and innovation. Open data makes it easier for scientists to collaborate on research projects and to share their findings. This can lead to new insights and discoveries that would not be possible if scientists were working in isolation (Hicks et al. 2015).
- Accelerate scientific progress. Open data can help to accelerate scientific progress by making it easier for scientists to build on the work of others. When scientists have access to a large body of open data, they can use it to develop new hypotheses and design new experiments.

In addition to the benefits listed above, open data can also help to:

- Increase public trust in science. When scientific data is open and transparent, the public can have more confidence in the findings of scientific research. This is important for building public support for science and for making science-informed policy decisions.
- Promote public education and awareness of natural hazards. Open data can be used to develop educational resources and tools that can help the public to learn about natural hazards and how to reduce risk.

In the geosciences, open data plays a crucial role in improving our understanding of geological processes and their consequences, as well as in advancing new technologies and applications. In the context of geohazard research, the prominence of open data assumes a pivotal role in the formulation and refinement of dependable hazard and risk assessment models, which constitute the pillar for identifying hazardous areas and evaluating the potential losses resulting from these natural phenomena, thus helping to protect people and property. This not only bolsters the resilience of communities but also contributes to the broader scientific understanding of geodynamics and their evolving impacts on our environment. In essence, open data transcends being merely useful; it stands as an imperative cornerstone for the safeguarding of lives and the preservation of societal assets in the face of geological hazards.

Among several strategies to share geoscientific data openly, a prominent avenue involves the publication of data in open access repositories, thereby granting unrestricted access to users. These repositories can store and share diverse datasets and geodatabases, ranging from geological maps to disaster occurrences, urban assets, and satellite imagery. In essence, they serve as dynamic hubs for the exchange of critical information. A complementary method involves the dissemination

of data and associated code through online platforms like GitHub, distinguished by their collaborative and interactive functionality, including pull-request support and user engagement features. Leveraging these and other online resources, efforts have been undertaken by the author to retrieve and categorise an extensive array of geodatabases encompassing a wide spectrum of critical domains. These endeavours, motivated by the concern to advance our understanding and preparedness in the face of natural hazards have culminated in the compilation of a comprehensive repository containing more than 110 geospatial resources (Table A1 in Appendix A). A summarisation of the repository is thoughtfully presented through a series of donut charts crafted to show the proportion between categories, covered area, data type, and access type (Figure 1.2). These plots unveil the breadth and diversity of information accessible online, as well as the accessibility and availability of the data, a fundamental consideration for researchers and stakeholders seeking to leverage open geospatial resources.

In the following sections, we delve further into the utilization of some of these resources and their impact on geohazards modelling and risk assessment.

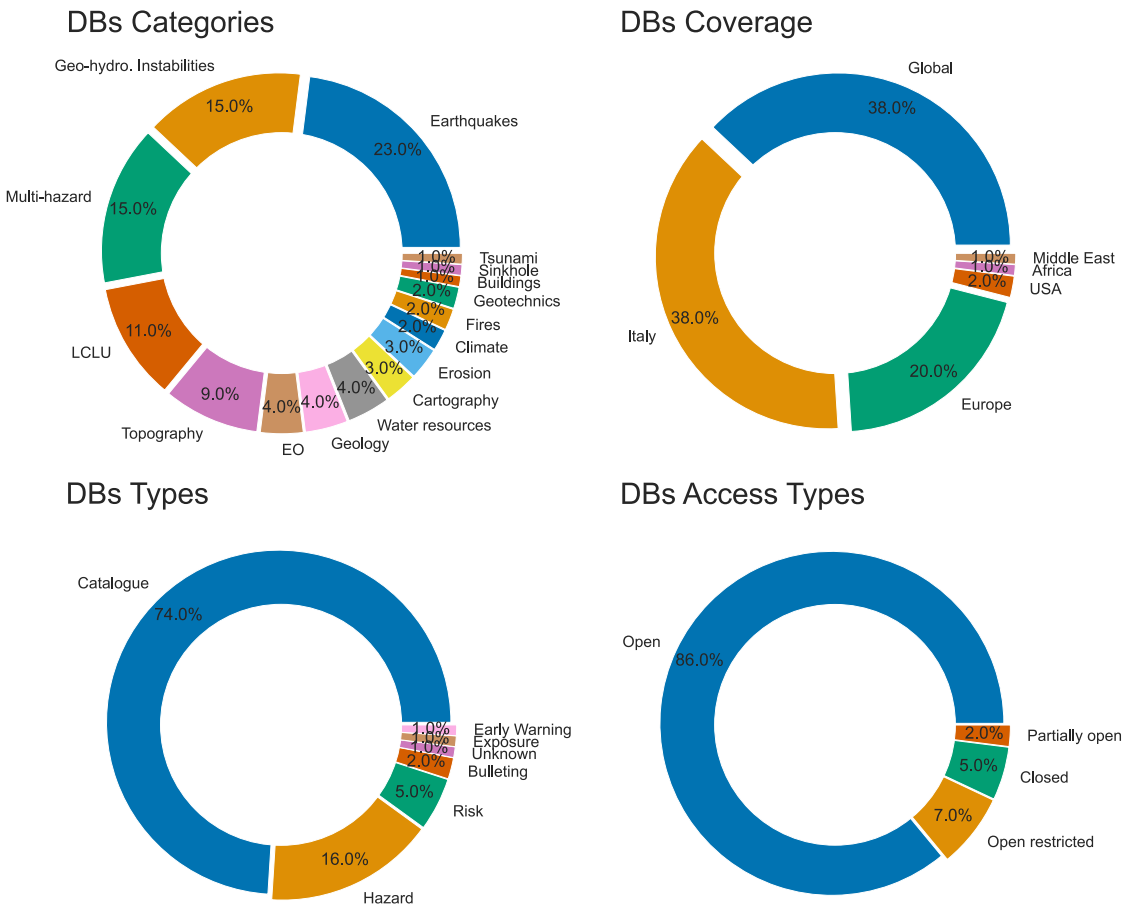


Figure 1.2: Proportion of categories, covered area, data type, and access type of the retrieved array of geodatabases.

1.2. Multi-hazard risk in urban areas

Urban areas, characterised by their concentration of infrastructure, human populations, and economic activities, face unique challenges when it comes to geohazard assessment and risk mitigation. The burgeoning urbanization trend worldwide necessitates a comprehensive understanding of geological phenomena that pose a threat to these environments. Urbanization alters natural terrain, disrupts hydrological cycles, and introduces new geotechnical challenges. Consequently, this transformation often leads to an increased susceptibility to geohazards, posing substantial risks to infrastructure, property, and population. The Sendai Framework for Disaster Risk Reduction (DRR) 2015 – 2030 (United Nations 2015) clearly highlights the need of actions in the following 4 priority areas: (1) understanding disaster risk; (2) strengthening disaster risk governance; (3) investing in disaster risk reduction for resilience; and (4) enhancing preparation for disaster and an effective response. To this purpose geohazard risk assessments are of paramount importance to defining prioritisation schemes for the design, implementation and optimisation of DRR and resilience-enhancing strategies (Sevieri et al. 2020). In geosciences, risk (R) refers to the outcome arising from potential phenomena that could impact a specific area (hazard, H), the extent, value, or nature of the elements exposed to this impact (exposure, E), and their vulnerability to damage (vulnerability, V). Consequently, when a generic event occurs, the resulting effects on the elements at risk are determined by the combination of the inherent structural attributes and the magnitude of the hazardous phenomenon.

$$R = f(H, E, V)$$

Regarding the hazard factor, it can be described as the probability of a specific adverse event, resulting in a phenomenon of a given magnitude, happening within a defined area and during a specific time frame. Hence, the hazard is described by three variables: the event's intensity, its spatial distribution, and the probability of its occurrence over time.

Among the array of geohazards that urban areas contend with, ground instabilities hold a prominent position. Landslides pose a prevalent threat in hilly or steep terrain regions. Sinkholes can emerge unexpectedly and result in substantial damage. Subsidence can lead to structural damage and infrastructure issues. The influence of climate change on European precipitation patterns is expected to bring about shifts in the occurrence and distribution of landslides and floods. In areas where heightened rainfall frequency and intensity are projected, the likelihood of landslides triggering and the exposure to landslide risk are anticipated to increase. In general, the impact of changing climate on rapid, shallow slope instabilities is forecasted to intensify and spread throughout the 21st century (Mateos et al. 2020). From 1903 to 2020, 185,753 fatalities were reported from 37,946 landslide

worldwide. Countries such as Colombia, Nepal, Costa Rica, India, USA, and China stand out, registering the highest incidence of landslides (Figure 1.3). On the European continent, the Geological Survey of Europe has aggregated data on 849,543 landslides from various databases spanning 24 countries. Remarkably, over 60% of these landslides are concentrated in Italy (Mateos et al. 2020). Recognizing the distribution and concentration of geohazards in urban areas is crucial in determining where and how preventive and corrective measures can effectively reduce vulnerability and exposure among urban populations (Johnson et al. 2016). However, despite the pronounced exposure of human assets to risk, most countries lack legislative measures within their National Land Bills to mandate the consideration of landslides and other hazards in urban planning practices (Martino et al. 2019; Mateos et al. 2020).

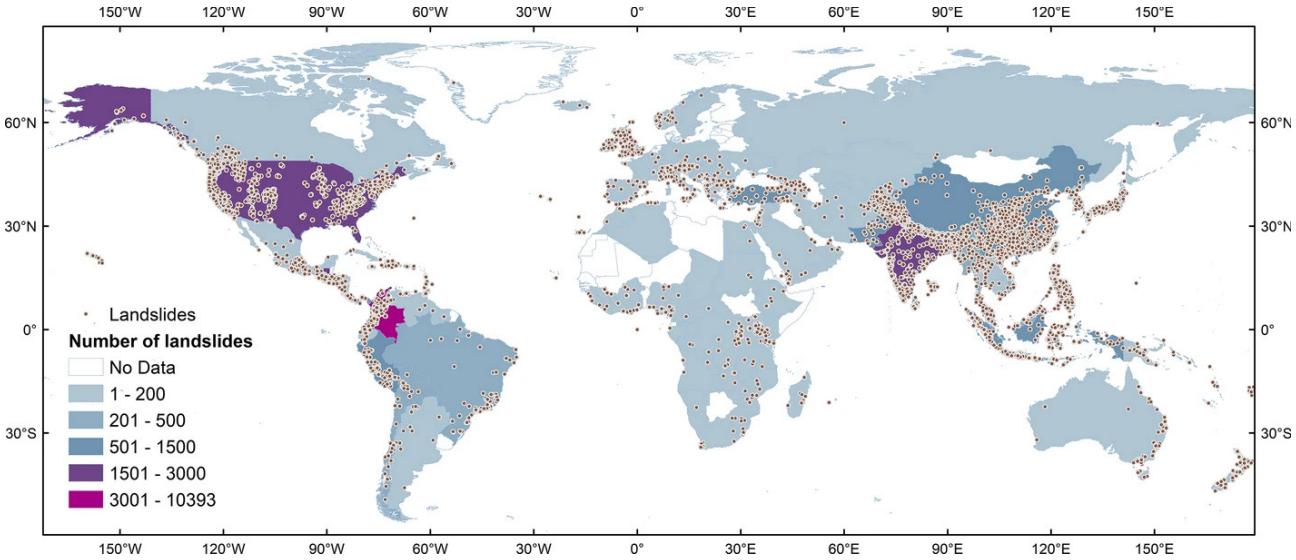


Figure 1.3: Spatial distribution of landslides worldwide. From Gómez et al. 2023.

The concept that natural hazards almost never occur individually is of great importance: this leads to the concept of multi-hazard (Carpignano et al. 2009; Gill et al. 2022). Multi-hazard analysis is the implementation of methodologies and approaches aimed at assessing and mapping the potential occurrence of different types of natural hazards in a specific area. It must consider the selection of multiple hazards that the country faces, and the specific contexts where hazardous events may occur simultaneously, cascadingly over time, and taking into account the potential interrelated effects (Gill et al. 2022; Hochrainer-Stigler et al. 2023). To date, multi-(hazard-)risk assessment still lacks a standard glossary among different communities (Gallina et al. 2016; De Angeli et al. 2022). Table 1.1 clarifies the terminology adopted in this thesis, including the definitions for damage, exposure, impact, loss, risk, vulnerability, multi-hazard, multi-hazard risk, and multi-risk.

The spatial (area that the hazard impacts) and temporal (the time duration) scales over which natural hazards impact upon the natural environment cover many orders of magnitude (Gill and Malamud

2014). Figure 1.4 presents the spatial versus temporal scales over which 16 selected hazards are reported. These hazards, are divided into five groups:

(1) *Geophysical* (earthquake, tsunami, volcanic eruption, landslide, and snow avalanche).
(2) *Hydrological* (flood and drought)
(3) *Shallow Earth Processes* (regional subsidence and uplift, local subsidence and heave, and ground collapse).
(4) *Atmospheric* (tropical cyclone, tornado, hail, snow, lightning and thunderstorm, long-term climatic change, and short-term climatic change.
(5) *Biophysical* (wildfire).

The analysis of the potential spatio-temporal overlaps of two or more natural hazards, led to the definition of various interrelations, which can be grouped into four main groups (Tilloy et al. 2019; Gill et al. 2020; Claassen et al. 2023).

(1) *Compound hazards*: two or more hazards may impact the same region and/or time period with impacts different than their sum.
(2) *Consecutive hazards*: two or more hazard that occur in succession, and whose direct impacts overlap spatially before recovery from a previous event is completed.
(3) *Triggering hazards*: one hazard causes another hazard to occur, which can result in hazard chains, networks, or cascades.
(4) *Amplifying hazards*: one hazard changes the probability or magnitude of another hazard by changing environmental conditions for the occurrence of another hazard.

Gill and Malamud (2014) identified and ranked 90 instances of linkages between 21 natural hazards (Figure 1.5). This revealed that volcanic eruptions, earthquakes, and storms are the primary hazard most linked to secondary hazards, with each having nine identified connections. In contrast, when examining secondary hazards arising from primary hazards, landslides emerged as the most linked secondary hazard, with 13 links, followed by volcanic eruptions with 11 links, and floods with 10 links. Among the 13 landslide linkages, two are associated with local subsidence and ground collapses. Consequently, landslides can either be triggered or fostered (i.e., increased probability) by ground collapse and local subsidence. These phenomena collectively fall within the category of ground instability hazards, which pose a substantial threat in urban areas. Addressing this threat necessitates a comprehensive multi-(hazard-)risk analysis and assessment to enable the implementation of preventive and corrective measures aimed at reducing future losses.

Table 1.1: Key terms and definitions related to multi-(hazard-)risk as used in this thesis. Modified from De Angeli et al. (2022); Ward et al. (2022).

Term	Definition
Damage	Negative impacts, i.e., impacts that result in negative effects on assets, people, socioeconomic and environmental systems.
Exposure	The situation of people, infrastructure, housing, production capacities and other tangible human assets located in hazard-prone areas. Exposure is described through a series of characteristics of the exposed assets, such as material, occupancy, economic value of a building or number of people.
Loss	A measure (usually in monetary terms) of a certain damage.
Vulnerability	The totality of the conditions determined by physical, social, economic, and environmental factors or processes which increase the susceptibility of an individual, or a community, assets, or systems to the impacts of hazards. In this thesis, specific reference is made to the physical vulnerability, which is the propensity of territorial elements to suffer physical damages from hazardous events. Vulnerability is represented by curves or functions, which measure the expected damage as a function of the hazard magnitude.
Risk	The potential loss of life, injury, or destroyed or damaged assets which could occur to a system, society, or a community in a specific period, determined as a function of hazard, exposure, and vulnerability.
Multi-hazard	The selection of multiple major hazards that the country faces, and the specific contexts where hazardous events may occur simultaneously, cascadingly, or cumulatively over time, and considering the potential interrelated effects.
Multi-hazard risk	Risk generated from multiple hazards and the interrelationships between these hazards (but not considering interrelationships on the vulnerability level)
Multi-risk	Risk generated from multiple hazards and the interrelationships between these hazards (and considering interrelationships on the vulnerability level).
Multi-(hazard-)risk	Used in this thesis when collectively referring to all the above.

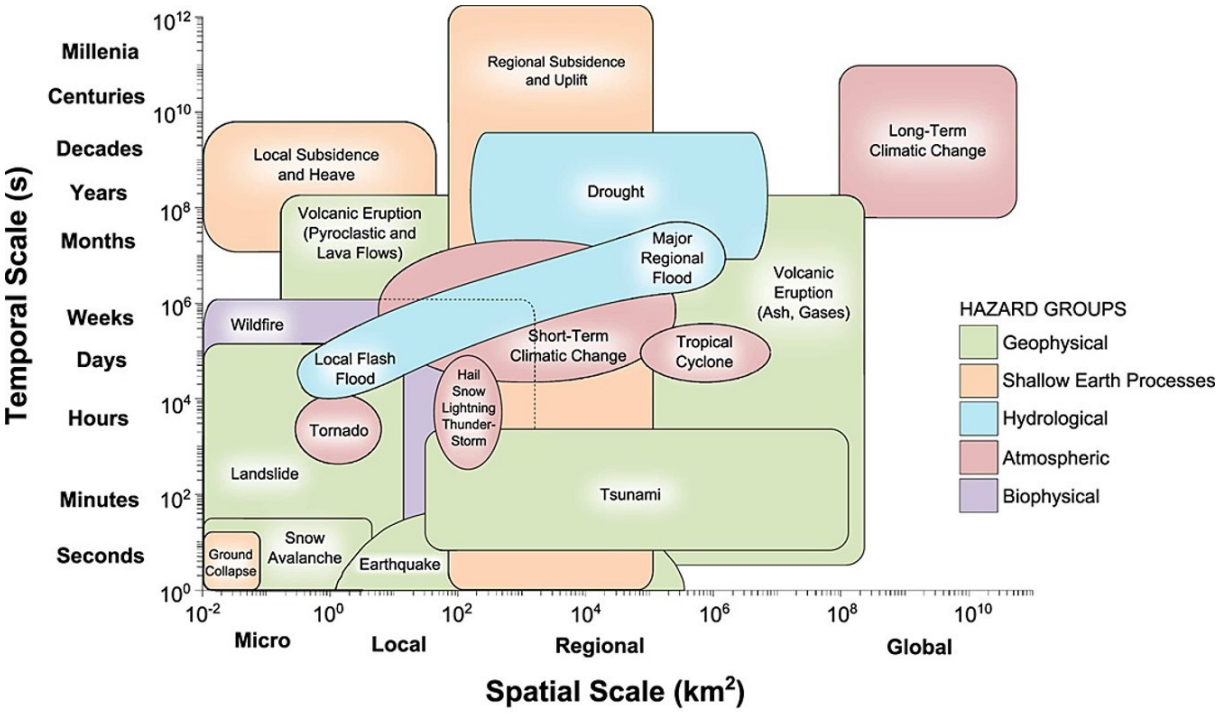


Figure 1.4: Spatial and temporal scales (on logarithmic axes) of 16 selected natural hazards and grouped into geophysical, hydrological, shallow Earth processes, atmospheric, and biophysical. From Gill and Malamud (2014).

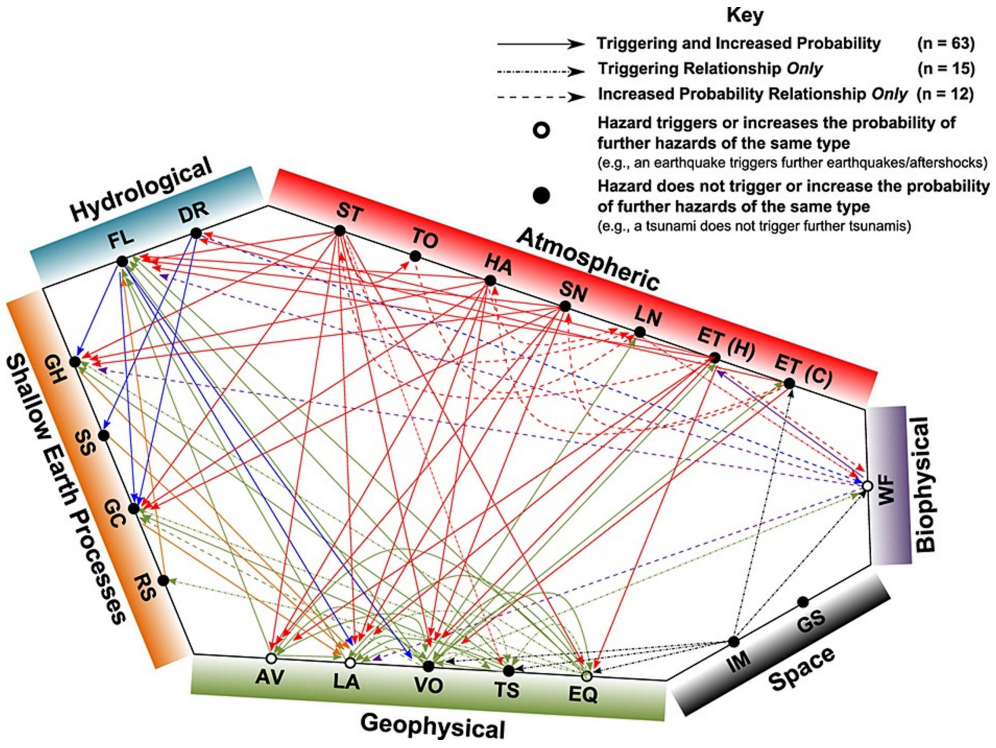


Figure 1.5: Hazard type linkages. A network diagram showing the potential hazard type linkages between 21 natural hazards. **EQ** = earthquake, **TS** = tsunami, **VO** = volcanic eruption, **LA** = landslide, **AV** = snow avalanche, **RS** = regional subsidence, **GC** = ground collapse, **SS** = soil (local) subsidence, **GH** = ground heave, **FL** = flood, **DR** = drought, **ST** = storm, **TO** = tornado, **HA** = hailstorm, **SN** = snowstorm, **LN** = lightning, **ET(H)** = extreme high temperature, **ET(C)** = extreme cold temperature, **WF** = wildfires, **GS** = geomagnetic storm, and **IM** = impact events. From Gill and Malamud (2014).

Similarly to multi-hazard, multi-risk is generated from multiple hazards that may also be correlated, affecting the same exposed elements. The interrelationships between these hazards and the vulnerability level (i.e., vulnerability changes because of different hazards affect a vulnerable element) should be considered. A multi-risk approach encompasses both multi-hazard and multi-vulnerability perspectives (Terzi et al. 2019). However, a clear framework and guidelines for multi-risk assessment and management for natural hazards is still lacking, which is currently acknowledged as the main challenge in multi risk management (Ward et al. 2022; Hochrainer-Stigler et al. 2023). Quantifying multi-risk presents further significant basic challenges, including defining the underlying concepts, acquiring extensive data, developing tailored analytical approaches, and utilizing valid modelling techniques. Moreover, the wide differences in terms of nature of the events, intensity, return periods, and potential damage to exposed assets make the comparability of natural hazards even more complicated (Carpignano et al. 2009). Kappes et al. 2012b, outline four key challenges associated with compiling a multi-(hazard-)risk assessment. These include (1) allowing different hazards to be compared, (2) interrelationships between hazards to be noted, (3) physical vulnerability assessments to be validly contrasted, and (4) the synthesis, communication, and visualization of a broad array of information from multiple disciplines and methods.

Assessing the risks and vulnerabilities of a large set of urban assets is an arduous task, demanding extensive technical expertise and financial resources that are usually not available (Kappes et al. 2012b; Julià and Ferreira 2021). However, the management of multi-risk even in highly developed countries remains a necessity, as exemplified by the 2021 flood events in Germany (Thiebes and Winkhardt-Enz 2023). A comprehensive multi-(hazard-)risk approach, which considers hazard interactions, coincidences, and dynamic vulnerabilities, would be time consuming, data and resource intensive, and necessitate the utilization and development of multiple methodologies drawn from diverse disciplines. Consequently, a prevalent practice within the natural hazards community leans toward a multi-layer single-hazard approach (Zschau 2017). While this approach may overlook interrelationships between hazards, potentially diminishing its reliability, it offers simplicity, allows for the consideration of numerous hazards, and produces easily comprehensible results. Nevertheless, conducting such multi-risk modelling can be challenging, especially on a large geographic scale and a high-resolution mapping unit, as it may not be evident how to aggregate or combine the quantification of each individual underlying hazard or risk (Tocchi et al. 2023).

1.2.1. The impact of mapping unit and scale

Geohazards analysis inherently operates across varying spatial scales and employ diverse methodologies, which range in sophistication and serve different zoning purposes. These purposes can span from providing informative or advisory guidance to serving statutory and regulatory compliance requirements, or even to engineering design (Flentje et al. 2007; Cascini 2008; Corominas et al. 2014). The reliability and effectiveness of these approaches hinge profoundly on the quality and comprehensiveness of available geospatial data and catalogues. At the core of this multifaceted process lies the pivotal decision regarding the mapping unit employed in predictive modelling. Whether constructing a susceptibility, hazard, or risk map, the choice of mapping unit is foundational. Among the mapping units routinely utilized, three prominent options have gained recognition: pixels, slope units (SU), and unique condition units. In recent years, there has been a discernible uptick in interest surrounding the implementation of SU as a mapping unit of choice for landslide hazard (Moreno et al. 2023). Understanding the implications of this choice with respect to the spatial scale and precision of hazard assessments is paramount. Mapping units exert a profound influence on the granularity of analysis and the subsequent zoning outcomes. Pixel, for instance, offer a fine-grained resolution but may become computationally intensive for large-scale assessments. On the other hand, SU strike a balance between granularity and computational efficiency, making them particularly appealing for medium to small-scale modelling.

Chapter 3 of the thesis endeavours to delve deeper into the ramifications of spatial scale of analysis and mapping unit selection in landslide susceptibility modelling. It explores how the choice of mapping unit influences the accuracy and applicability of landslide assessments across various spatial scales. By shedding light on this critical aspect, we aim to enhance our understanding of the nuanced interplay between mapping units, spatial scales, and the efficacy of landslide susceptibility modelling, ultimately contributing to more informed landslide risk management strategies.

1.2.2. The potential of Spatial Data Science

In a world where decision-making is increasingly influenced by data, it is important to understand how Spatial Data Science (SDS) can help. SDS is a subset of generic data science that focuses on the characteristics of spatial data, i.e., the importance of “where”. Data science is often referred to as the science of extracting meaningful information from data. In this context, SDS treats location, distance, and spatial interaction as core aspects of the data and employs specialized methods and software to store, retrieve, explore, analyse, visualise, and learn from geospatial data. In this sense, SDS relates to data science as spatial statistics to statistics, spatial databases to databases, and geocomputation to computation (Anselin 2020). At its core, it combines Geographic Information Systems (GIS) with advanced analytical techniques, helping us better understand the relationship between location and phenomena. In geohazards risk, this means understanding not just the

occurrence of hazards but their spatial context, evolution, and impact on our environment and society. SDS transforms how we assess natural hazards and risks. It enables us to grasp the spatial patterns of hazards by understanding “where” the risks occur. Thus, we can identify high-risk areas, improve preparedness, and develop targeted mitigation strategies. It allows us to perform the analyses at various scales, from local to global, providing a comprehensive understanding of their spatial and temporal dynamics.

One of the key strengths of SDS lies in its capacity for modelling and scripting. Spatial modelling consists of the analysis of spatial data to make inferences about the model parameters, to predict at unsampled locations, or for downscaling/upscaling applications (Banerjee 2003). It enables the development of advanced geospatial models and scripts, automating data processing, analysis, and visualization. This automation allows for timely responses and informed decision-making. It also allows continuous monitoring of dynamic conditions, early threat detection, and resource optimization, thereby bolstering resilience in the face of evolving challenges. SDS finds synergy with Machine Learning, fostering a potent partnership that enhances geohazards analysis. Machine Learning algorithms, when integrated into spatial data analysis, empower us to extract patterns, make predictions, and gain insights that may otherwise remain hidden. This integration enables the creation of predictive models that factor in complex geospatial variables, leading to more accurate hazard assessments, early warning systems, and risk mitigation strategies. By leveraging the capabilities of Machine Learning, SDS can take multi-(hazard-)risk analysis to unprecedented heights of precision and efficacy.

In the pursuit of spatial data insights, various tools have been developed (Figure 1.6). These tools empower to navigate the intricate terrain of spatial analysis. Some of the main tools include:

- Geographic Information Systems (GIS): GIS software forms the foundation of spatial data analysis. It allows for the creation, visualization, and analysis of geographic data, enabling the generation of informative maps and spatial models.
- Remote Sensing: Remote sensing technology, including satellite and aerial imagery, provides a wealth of data for monitoring and assessing changes in the Earth's surface over time.
- Spatial Databases: Robust spatial databases, such as PostGIS and PostgreSQL, organize and manage large geospatial datasets, facilitating efficient data retrieval and analysis.
- Geospatial Libraries and Frameworks: Open-source geospatial libraries and frameworks, such as GDAL, GeoPandas, and PySAL, offer powerful tools for data manipulation, analysis, and modelling.

- ❖ Machine Learning Libraries: Libraries like scikit-learn and TensorFlow integrate seamlessly with spatial data analysis, allowing for the application of machine learning algorithms to geospatial challenges.

In essence, SDS represents a pivotal paradigm shift in geospatial data analyses, such as geohazard risk assessment. It equips us with the tools to navigate our dynamic world, uncover hidden patterns, and make informed decisions that safeguard lives, assets, and the environment. As we delve further into this thesis, we will explore the practical applications, methodologies, and real-world impact of SDS applied to geological multi-(hazard-)risk assessment.

These concepts of multi-(hazard-)risk analysis relying on open source spatial data science will be the core focus of the research work, as they are applied to assess ground instability hazards and risk affecting the Municipality of Rome (Italy), with a specific emphasis on landslide hazard.

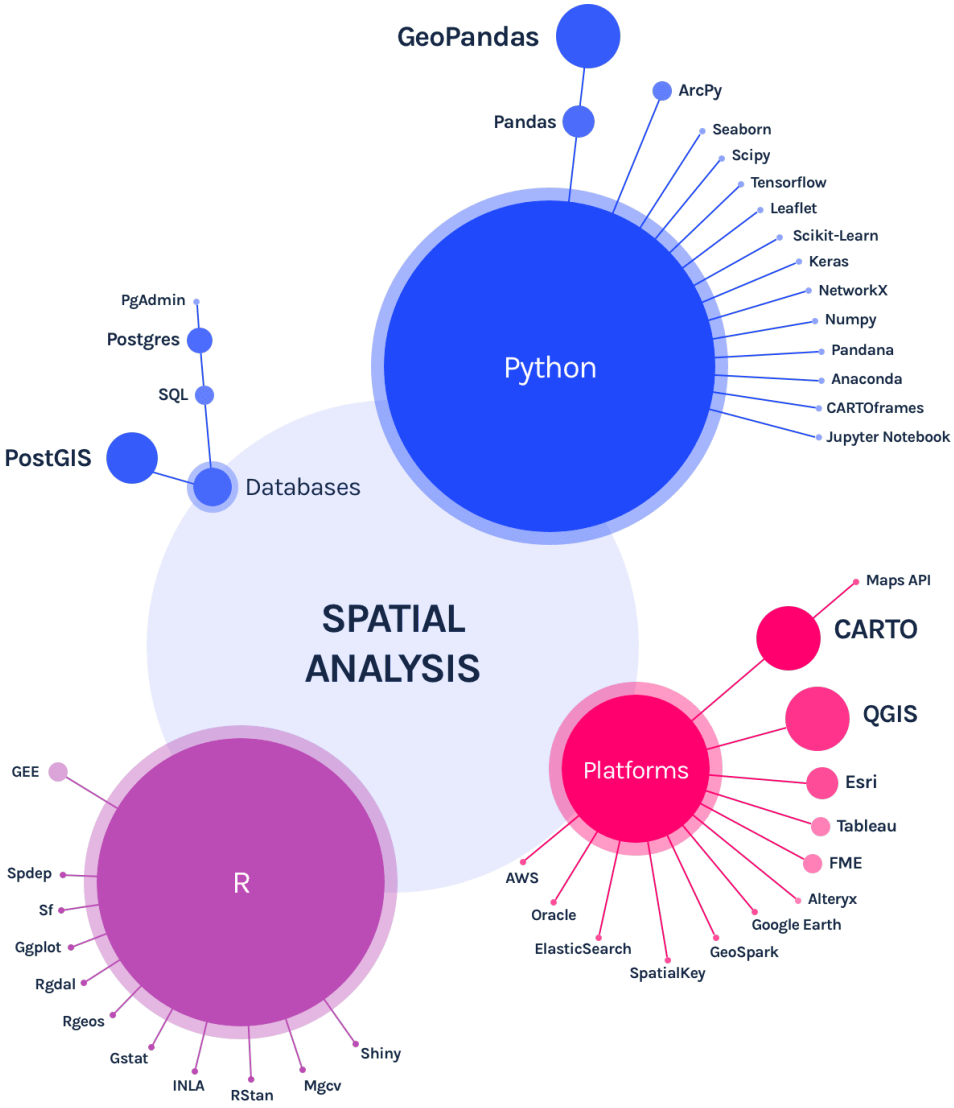


Figure 1.6: Existing tools for the pursuit of Spatial Data Science.

1.3. Objectives of the thesis

In the context of enhancing urban resilience, it is of paramount significance to comprehend and assess the array of threats that major cities confront. The evaluation of hazards plays a pivotal role in the implementation of policies aimed at mitigating risks. While numerous hazard and risk maps exist on national, European, and global scales, their precision may not suffice for effectively pinpointing potential hazard sources at the urban scale. This underscores the necessity of developing solutions tailored to urban settings. To achieve this objective, the gathering of as much information as possible regarding historical events, asset characteristics, environmental factors, and other relevant ancillary data is essential. Consequently, it becomes imperative to explore the physical risks posed by multiple natural hazards, that may overlap in space and time, by aggregating available data and integrating them with advanced modelling techniques, such as Spatial Data Science and Machine Learning. This modelling facilitates the generation of precise predictions regarding future hazard occurrences and physical risk of existing assets.

In the present doctoral thesis, we delve into the topics of shallow landslides hazard and multi-risk assessment related to ground instabilities at the urban scale. Our primary objective is to turn data into information that can guide risk managers and infrastructure stakeholders in implementing preventive mitigation measures to enhance resilience, thus contributing to the priority areas highlighted by the Sendai Framework for Disaster Risk Reduction. We focus on the Municipality of Rome as our case study, examining both the spatial and temporal components of shallow landslide hazard in large urban regions. Our study explores the potential of open-source catalogues, e.g., landslide inventories, to be cross-validated and optimised for training accurate Machine Learning models. These models aim to predict high-probability landslide zones with statutory purposes. To assess the impact of analysis scale and mapping units on prediction accuracy, we compared susceptibility levels for shallow landslides with available open-source multi-scale maps. We also explored the benefits of merging these maps to harness their full potential. Furthermore, we developed a robust, replicable, and scalable model which combines spatial hazard assessments, satellite-based interferometric data, and physical features of exposed assets. This model computes a multi-risk score to rank buildings associated with ground instability hazards, including landslides, subsidence, and sinkholes.

In this doctoral work, the following research questions will be explored:

(1) How can official open data be exploited to develop reliable landslide hazard models valuable for statutory purposes?
(2) What is the impact of mapping unit and analysis scale on the accuracy of Landslide Susceptibility Zoning (LSZ)?

(3) Can the fusion of multiple low-resolution LSZ overcome the lack of site-specific studies?
(4) How can susceptibility maps, census data, and satellite interferometry contribute to multi-risk assessment in the urban environment?
(5) How can single- and multi-risk analyses inform strategies for enhancing urban resilience and minimizing economic losses?

1.4. Outline of the thesis

The thesis includes published and unpublished works by the Author, and they are incorporated in five different chapters. The main chapters (Chapters 2 – 4) must be considered as separate manuscripts. I collected, analysed, and interpreted the data and wrote the manuscripts under the supervision of Prof. Paolo Mazzanti and Prof. Carlo Esposito (Dep. of Earth Sciences, Sapienza University of Rome, Italy). Prof. Carmine Galasso (Dep. of Civil Engineering, University College of London, UK) introduced me to multi-hazard vulnerability studies during my period abroad.

Chapter 2 focuses on research question 1 through the statistical analysis of fragmented and incomplete data related to landslides and rainfall records. This chapter has been published as Original Paper in *Landslides* journal in June 2023 (DOI: 10.1007/s10346-023-02095-7), with the following citation:

- Esposito C., **Mastrantoni G**., Marmoni G.M., Antonielli B., Caprari P., Pica A., Schilirò L., Mazzanti P., and Bozzano F. (2023). From theory to practice: optimisation of available information for landslide hazard assessment in Rome relying on official, fragmented data sources. *Landslides* 20, 2055-2073 (2023). *https://doi.org/10.1007/s10346-023-02095-7.*

Chapter 3 focuses on research questions 1, 2, and 3, through an extensive examination of open-source multiscale landslide susceptibility maps with spatial data science methods. This chapter has been published as Original Paper in *Georisk: Assessment and Management of Risk for Engineered Systems and Geohazards* in August 2023 (DOI: 10.1080/17499518.2023.2251139), with the following citation:

- **Mastrantoni G**., Marmoni G.M., Esposito C., Bozzano F., Scarascia Mugnozza G., and Mazzanti P. Reliability assessment of open-source multiscale landslide susceptibility maps and effects of their fusion. *Georisk: Assessment and Management of Risk for Engineered Systems and Geohazards* (2023). *https://doi.org/10.1080/17499518.2023.2251139*.

Chapter 4 focuses on research questions 4 and 5, through the development of a novel multi-risk analysis method aimed at addressing and prioritising buildings exposed to ground instability hazards.

This chapter has been published as Original Paper in *Geomatics, Natural Hazards and Risk* in November 2023 (DOI: 10.1080/19475705.2023.2275541), with the following citation:

- **Mastrantoni G.**, Masciulli C., Marini R., Esposito C., Scarascia Mugnozza G., and Mazzanti P. A novel model for multi-risk ranking of buildings at city level based on open data: the test site of Rome, Italy. *Geomatics, Natural Hazards and Risk* (2023). *https://doi.org/10.1080/19475705.2023.2275541*

Chapter 5 corresponds to the Conclusions of the thesis. This chapter summarizes the main findings, answering the research questions of this thesis and presents possibilities for future development of the work presented.

References

Anselin L. 2020. Spatial Data Science. In: International Encyclopedia of Geography [Internet]. [place unknown]: John Wiley & Sons, Ltd; [accessed 2023 Oct 2]; p. 1–6. https://doi.org/10.1002/9781118786352.wbieg2015

Banerjee SB Bradley P Carlin, Alan E Gelfand, Sudipto. 2003. Hierarchical Modeling and Analysis for Spatial Data. New York: Chapman and Hall/CRC. https://doi.org/10.1201/9780203487808

Carpignano A, Golia E, Di Mauro C, Bouchon S, Nordvik J. 2009. A methodological approach for the definition of multi-risk maps at regional level: first application. Journal of Risk Research. 12(3–4):513–534. https://doi.org/10.1080/13669870903050269

Cascini L. 2008. Applicability of landslide susceptibility and hazard zoning at different scales. Engineering Geology. 102(3–4):164–177. https://doi.org/10.1016/j.enggeo.2008.03.016

Claassen JN, Ward PJ, Daniell J, Koks EE, Tiggeloven T, de Ruiter MC. 2023. A new method to compile global multi-hazard event sets. Sci Rep. 13(1):13808. https://doi.org/10.1038/s41598-023-40400-5

Corominas J, van Westen C, Frattini P, Cascini L, Malet J-P, Fotopoulou S, Catani F, Van Den Eeckhaut M, Mavrouli O, Agliardi F, et al. 2014. Recommendations for the quantitative analysis of landslide risk. Bull Eng Geol Environ. 73(2):209–263. https://doi.org/10.1007/s10064-013-0538-8

De Angeli S, Malamud BD, Rossi L, Taylor FE, Trasforini E, Rudari R. 2022. A multi-hazard framework for spatial-temporal impact analysis. International Journal of Disaster Risk Reduction. 73:102829. https://doi.org/10.1016/j.ijdrr.2022.102829

Flentje PN, Miner A, Whitt G, Fell R. 2007. Guidelines for landslide susceptibility, hazard and risk zoning for land use planning. 42(1):28.

Gallina V, Torresan S, Critto A, Sperotto A, Glade T, Marcomini A. 2016. A review of multi-risk methodologies for natural hazards: Consequences and challenges for a climate change impact assessment. Journal of Environmental Management. 168:123–132. https://doi.org/10.1016/j.jenvman.2015.11.011

Gill JC, Duncan M, Ciurean R, Smale L, Stuparu D, Schlumberger J, de Ruiter M, Tiggeloven T, Torresan S, Gottardo S, et al. 2022. D1.2 Handbook of multi-hazard, multi-risk definitions and concepts [Internet]. [accessed 2023 Jun 18]. https://nora.nerc.ac.uk/id/eprint/533237/

Gill JC, Malamud BD. 2014. Reviewing and visualizing the interactions of natural hazards. Reviews of Geophysics. 52(4):680–722. https://doi.org/10.1002/2013RG000445

Gill JC, Malamud BD, Barillas EM, Guerra Noriega A. 2020. Construction of regional multi-hazard interaction frameworks, with an application to Guatemala. Natural Hazards and Earth System Sciences. 20(1):149–180. https://doi.org/10.5194/nhess-20-149-2020

GitHub. [accessed 2023 Sep 19]. https://github.com

Gómez D, García EF, Aristizábal E. 2023. Spatial and temporal landslide distributions using global and open landslide databases. Nat Hazards. 117(1):25–55. https://doi.org/10.1007/s11069-023-05848-8

Hicks D, Wouters P, Waltman L, de Rijcke S, Rafols I. 2015. Bibliometrics: The Leiden Manifesto for research metrics. Nature. 520(7548):429–431. https://doi.org/10.1038/520429a

Hochrainer-Stigler S, Šakić Trogrlić R, Reiter K, Ward PJ, de Ruiter MC, Duncan MJ, Torresan S, Ciurean R, Mysiak J, Stuparu D, Gottardo S. 2023. Toward a framework for systemic multi-hazard and multi-risk assessment and management. iScience. 26(5):106736. https://doi.org/10.1016/j.isci.2023.106736

Johnson K, Depietri Y, Breil M. 2016. Multi-hazard risk assessment of two Hong Kong districts. International Journal of Disaster Risk

Reduction. 19:311–323. https://doi.org/10.1016/j.ijdrr.2016.08.023

Julià PB, Ferreira TM. 2021. From single- to multi-hazard vulnerability and risk in Historic Urban Areas: a literature review. Nat Hazards. 108(1):93–128. https://doi.org/10.1007/s11069-021-04734-5

Kappes MS, Keiler M, von Elverfeldt K, Glade T. 2012b. Challenges of analyzing multi-hazard risk: a review. Nat Hazards. 64(2):1925–1958. https://doi.org/10.1007/s11069-012-0294-2

Martino S, Bozzano F, Caporossi P, D'Angiò D, Della Seta M, Esposito C, Fantini A, Fiorucci M, Giannini LM, Iannucci R, et al. 2019. Impact of landslides on transportation routes during the 2016–2017 Central Italy seismic sequence. Landslides. 16(6):1221–1241. https://doi.org/10.1007/s10346-019-01162-2

Mateos RM, López-Vinielles J, Poyiadji E, Tsagkas D, Sheehy M, Hadjicharalambous K, Liscák P, Podolski L, Laskowicz I, Iadanza C, et al. 2020. Integration of landslide hazard into urban planning across Europe. Landscape and Urban Planning. 196:103740. https://doi.org/10.1016/j.landurbplan.2019.103740

Moreno M, Lombardo L, Crespi A, Zellner P, Mair V, Pittore M, Westen CJ van, Steger S. 2023. Space-time data-driven modeling of precipitation-induced shallow landslides in South Tyrol, Italy [Internet]. [accessed 2023 Sep 29]. https://eartharxiv.org/repository/view/5987/

Open Science Collaboration. 2015. Estimating the reproducibility of psychological science. Science. 349(6251):aac4716. https://doi.org/10.1126/science.aac4716

Sevieri G, Galasso C, D'Ayala D, De Jesus R, Oreta A, Grio MEDA, Ibabao R. 2020. A multi-hazard risk prioritisation framework for cultural heritage assets. Natural Hazards and Earth System Sciences. 20(5):1391–1414. https://doi.org/10.5194/nhess-20-1391-2020

Terzi S, Torresan S, Schneiderbauer S, Critto A, Zebisch M, Marcomini A. 2019. Multi-risk assessment in mountain regions: A review of modelling approaches for climate change adaptation. Journal of Environmental Management. 232:759–771. https://doi.org/10.1016/j.jenvman.2018.11.100

Thiebes B, Winkhardt-Enz R. 2023. Challenges and opportunities using new modalities and technologies for multi-risk management. Nat Hazards. 119(2):1137–1140. https://doi.org/10.1007/s11069-022-05516-3

Tilloy A, Malamud BD, Winter H, Joly-Laugel A. 2019. A review of quantification methodologies for multi-hazard interrelationships. Earth-Science Reviews. 196:102881. https://doi.org/10.1016/j.earscirev.2019.102881

Tocchi G, Ottonelli D, Rebora N, Polese M. 2023. Multi-Risk Assessment in the Veneto Region: An Approach to Rank Seismic and Flood Risk. Sustainability. 15(16):12458. https://doi.org/10.3390/su151612458

TOPS. NASA TOPS [Internet]. [accessed 2023 Sep 19]. https://nasa.github.io/Transform-to-Open-Science/

United Nations O for DRRG Switzerland. 2015. Sendai framework for disaster risk reduction 2015–2030. Center, Asian Disaster Reduction [Internet]. [accessed 2023 Oct 5]. https://www.apec-epwg.org/media/2584/e1a8e2e1c1125430bcf585c521ca6bcb.pdf

Vicente-Saez R, Martinez-Fuentes C. 2018. Open Science now: A systematic literature review for an integrated definition. Journal of Business Research. 88:428–436. https://doi.org/10.1016/j.jbusres.2017.12.043

Ward PJ, Daniell J, Duncan M, Dunne A, Hananel C, Hochrainer-Stigler S, Tijssen A, Torresan S, Ciurean R, Gill JC, et al. 2022. Invited perspectives: A research agenda towards disaster risk management pathways in multi-(hazard-)risk assessment. Natural Hazards and Earth System Sciences. 22(4):1487–1497. https://doi.org/10.5194/nhess-22-1487-2022

Zschau J. 2017. Where are we with multihazards, multirisks assessment capacities? [Internet]. [accessed 2023 Oct 6]. drmkc.jrc.ec.europa.eu

CHAPTER 2

2. From Theory to Practice: Optimisation of Available Information for Landslide Hazard Assessment in Rome Relying on Official, Fragmented Data Sources

Authors

Carlo Esposito[1], Giandomenico Mastrantoni[1], Gian Marco Marmoni[1], Benedetta Antonielli[1], Patrizia Caprari[1], Alessia Pica[1], Luca Schilirò[2], Paolo Mazzanti[1,3], Francesca Bozzano[1,3].

[1] Dipartimento di Scienze della Terra e Centro di ricerca per i rischi geologici CERI, Sapienza Università di Roma, P.le Aldo Moro 5, 00185, Roma, Italy.

[2] Istituto di geologia ambientale e geoingegneria (IGAG), Consiglio Nazionale delle Ricerche (CNR), Roma, Italia.

[3] NHAZCA S.r.l. start-up di Sapienza Università di Roma, P.le Aldo Moro 5, 00185, Roma, Italy.

Published in Landslides (2023)

Abstract

The definition of landslide hazard is a step-like procedure that encompasses the quantification of its spatial and temporal attributes, i.e., a reliable definition of landslide susceptibility and a detailed analysis of landslide recurrence. However, available information is often incomplete, fragmented, and unsuitable for reliable quantitative analysis. Nevertheless, landslide hazard evaluation has a key role in the implementation of risk mitigation policies and an effort should be done to retrieve information and make it useful for this purpose. In this research, we go through this topic optimising the information available in catalogues, starting from landslide inventory review and constitution of a boosted training dataset, propaedeutic for susceptibility analysis based on machine learning methods. The temporal recurrence of landslide events has been approached here either through the definitions of large-scale quantitative hazard descriptors or by analysis of historical rainfall (i.e., the main triggering factor for the considered shallow earth slope failures) databases through the definition of rainfall probability curves. Spatial and temporal attributes were integrated, selecting potential landslide source areas ranked in terms of hazard. Data integration was also pursued through Persistent Scatterer Interferometry analysis which pointed out areas of interest within potential landslide source areas featured by

ongoing ground movement. The consequential approach led to the definition of the first hazard product of the city of Rome at a local scale functional for advisory purposes or the statutory level, representing a thematic layer able to orient the risk managers and infrastructure stakeholders.

Keywords: Susceptibility; Machine Learning; Rainfall Probability; Landslide Hazard; Landslide Inventories; Interferometry.

2.1. Introduction

A shift from the ideal scientific framework to its actual applicability is required when dealing with risk-oriented hazard analyses related to land-use-planning purposes (Thiery et al., 2020). Such a task can be particularly challenging, as the application of rigorous and reliable scientific approaches can be data demanding, in terms of quality and quantity of information needed to fully exploit their potential (Corominas et al., 2014; Glade, 2001). On the contrary, in some cases the quality and quantity of available data do not match the optimal (and sometimes even fair) requirements to ensure reliable results for regulatory purposes (Cascini, 2008). Furthermore, the possibility of acquiring additional or higher resolution data can be limited at times, especially for large-scale prevention and mitigation activities (i.e., land-use planning and the propaedeutic hazard zoning), being related to available financial resources as well as time constraints. Nevertheless, local or national administrations in charge of risk prevention and mitigation could optimize the available information to fit as much as possible scientific standards. For this purpose, data-driven and knowledge-based methods can be implemented to integrate and boost available data sources.

In this perspective, we detailly report and discuss an integrated approach devoted to a reliable assessment of landslide hazards in one of the more challenging contexts exposed to risk: the urban area of Rome. The scenario of gravitational instabilities and related damage resulting from an intense rainfall event occurred in 2014 (Alessi et al. 2014) highlighted how the landslide risk is not negligible in the city of Rome. This statement is justified by the occurrence of recurring, usually small-size landslide events (both as single slope failures and multiple phenomena over large areas), and the high exposure as regards the number and value of the exposed elements. The landslide conditions are known (Amanti et al., 2013; Del Monte et al., 2016) and taken into due consideration in the official planning tools, such as the Hydro-geological Structure Plan (https://www.autoritadistrettoac.it/planning/hydrographic-basin-planning/documentation-of-the-tiber-basin-plan) and the municipal land-use plan (http://www.urbanistica.comune.roma.it/prg-2008-vigente/). Collaborations between research and government institutions (Amanti et al., 2014, Esposito et al., 2019) show a growing awareness and interest on this issue. However, a proper hazard analysis and the relative zoning of the territory has not been carried out so far. The knowledge of the hazard, from the temporal point

of view, is instead a fundamental requirement for a correct and exhaustive definition of the risk and subsequent actions for risk mitigation (Corominas et al., 2014). Although this task is pursued by the municipal administration, at least as Civil Protection plans (https://www.comune.roma.it/web-resources/cms/documents/Fasc3_RischioFrane_2021.pdf),the fragmentariness of landslide inventories available for Rome, together with the frequent lack of information regarding the date of occurrence of the surveyed landslide (re)activations, severely limits the possibility of evaluating the temporal probability of occurrence. On the other hand, a hazard assessment as detailed as possible can be useful at least for advisory purposes.

In this study we present a first attempt to quantify the landslide hazard by exploiting the existing databases reporting relevant information for the investigated phenomenon, such as landslide inventories and other predisposing, preparatory and triggering factors. First, we performed a full review of the landslide databases by means of a geomorphological review of the known and mapped landslides and a check of the available dated landslide events. We then produced an integrated and edited landslide inventory; on this basis it has been possible to perform a spatial hazard assessment that required additional efforts (i.e., “inventory boosting”) to make such an information fit the requirements needed for a reliable analysis. Furthermore, a statistical analysis of historical rainfall data was performed through the Generalised Extreme Values (GEV; Jenkinson, 1955) method for both daily and hourly rainfall. Such an analysis allowed us to infer quantitative rainfall thresholds and assess the return period of rainfall events capable of inducing landslides.

In addition, a preliminary overview of the slopes potentially susceptible to landslides was derived through the Persistent Scatterer Interferometry (PSI) based on the Sentinel-1 SAR images. The presence of movement or changes in the state of activity of already-catalogued phenomena were detected by analysing InSAR data, and the results have been integrated with the susceptibility analysis, to add information on the ongoing slope deformations to the intrinsic characteristics of the slope, prone to landslide. For the same level of susceptibility, this integration allowed to rank the actual critical situations by highlighting areas with ongoing deformation, then providing a cognitive dynamic information of practical use for local administrators involved in the management of urban areas.

2.2. A Brief Introduction to Geohazards in Rome

The city of Rome lies in a hilly area where the present geological and morphological setting is mainly related to deposition and erosion in marine (late Pliocene – lower Pleistocene) and continental (middle Pleistocene – Holocene) environments.

The geological context is featured by a succession of Pliocene marine clays (Monte Vaticano Formation), silt and silty sands (Monte Mario Formation) in Lower Pleistocene which turn into littoral, transitional and continental sediments (Ponte Galeria Formation) during lower-middle Pleistocene. These deposits largely outcrop on the hills on the right bank of the Tiber (Funiciello e Giordano, 2008). Over these terms, a succession of alternating volcanic deposits (resulting from the activity, about 600 ka ago, of the surrounding Colli Albani and Sabatini Volcanic Districts) and continental (mainly alluvial and palustrine) sediments deposited in the area during middle-upper Pleistocene. The present landscape is strongly influenced by valleys and slopes carved by the Tiber River and its tributaries, and only partially filled by alluvial deposits whose thickness can reach tens of meters (Bozzano et al., 2000; Figure 2.1).

The hydraulic circulation is governed by the superimposition of medium to high permeability volcanic deposits over the regional aquiclude made up of the blue clays of the Monte Vaticano Fm. and the silty and clayey horizons of the Monte Mario Fm. (La Vigna et al., 2015). Ephemeral and perched water tables develop in the weathered soil covers, that transiently over impose as a surficial seepage within unsaturated soils.

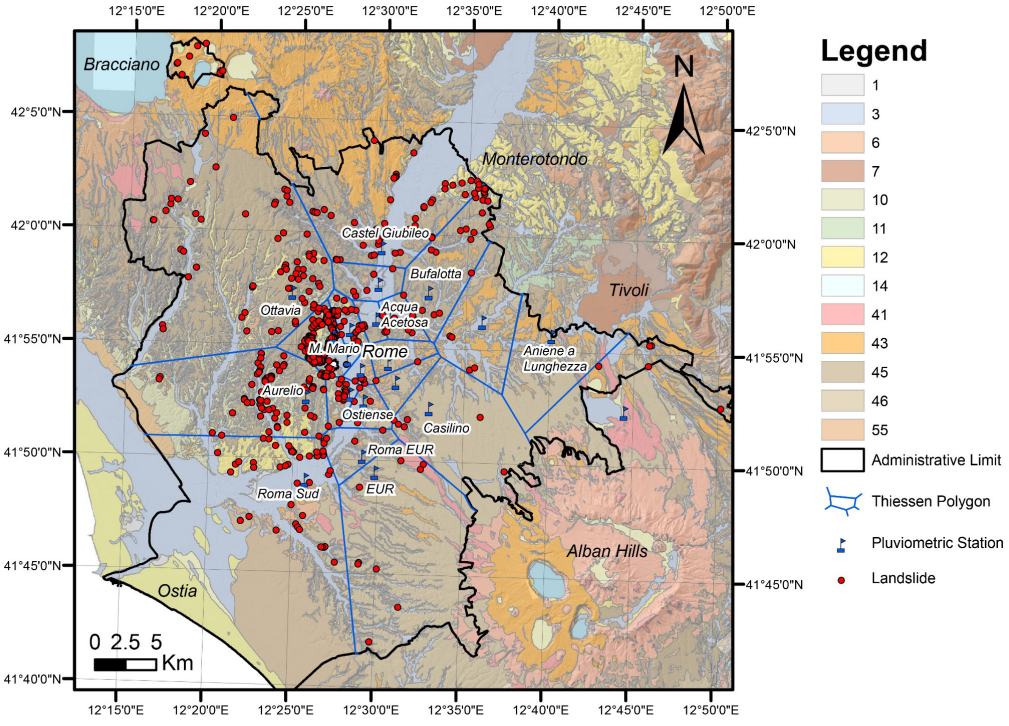


Figure 2.1: *Lithological units in the study area of the municipality of Rome with the location of the occurred landslides. Thiessen polygons adopted considering the available pluviometric stations are also reported in the map. Key to legend: 1) Anthropic deposits; 3) Recent and terraced sandy-gravelly alluvial deposits, eluvio-colluvial deposits; 6) silty-sandy alluvial deposits, fluvio-lacustrine deposits; 7) travertines; 10) Plio-pleistocene clayey and silty deposits; 11) Marine Pliocene clays; 12) Debris and talus slope deposits, Conglomerates and cemented breccias; 14) Marls, Marly limestones and calcarenites; 41) Leucititic/trachytic lavas; 43) Lithoid*

Tuffs, Pomiceous ignimbritic and phreatomagmatic facies; 45) Welded tuffs, tufites; 46) Pozzolanic sequence; 55) Alternance of loose and welded ignimbrites.

The different geological units outcropping on the left and right embankment of the Tiber River valley control the different responses to natural and anthropogenic hazard processes that affect the urban area (Funiciello et al., 2008), including subsidence, sinkholes, floods, and landslides (Table 2.1). In particular, the landslide response is closely related to the outcrop distribution of Plio-Pleistocene sedimentary units, which are frequently involved in shallow and translational landslides (Amanti et al., 2008; Alessi et al., 2014). Ground instabilities are often localised within unsaturated, shallow soil covers resulting from the chemical and physical weathering of the underlying deposits (Schilirò et al., 2019). Ephemeral hydraulic circulation develops along permeability contrasts between volcanic, or debris covers over less permeable sedimentary deposits, can influence the local hydraulic circulation and, thus, the slope stability. Furthermore, falls and topples sometimes involve rocky slopes, mainly made up of volcanic tuffs (Amanti et al., 2008; Alessi et al., 2014).

Landslides with volumes of tens up to hundreds of cubic meters (Amanti et al., 2013; Bozzano et al., 2006) as well as natural or anthropogenic sinkholes (Esposito et al., 2021,) are the most frequent and impacting processes in the urban area. Based on the 2014 event (Alessi et al., 2014) and other information collected from a subset of the known landslides (e.g., http://sgi2.isprambiente.it/franeroma/), most of the slope failures have been triggered by intense or prolonged rainfalls. Such events can cause significant damages specially to pipelines, aqueducts, and road infrastructure.

Table 2.1: Summary of the main geological hazards involving the Municipality of Rome.

	Involved areas	***Latest significant event***	***Event References***	***Hazard/ Susceptibility Maps***	***Inventories and maps***	***Intensity***	***Return Period (yrs)***
Subsidence	Alluvial plain of Tiber River and its main tributaries		Campolunghi et al., 2008; Bozzano et al., 2015; 2018; Delgado Blasco, 2019;	n.a.	-	Low	Continuous
Sinkhole	Widespread over the urban area, mainly concentrated in areas with relevant thicknesses of volcanic deposits (left bank of Tiber River);	Via Zenodossio;	Esposito et al., 2021	Ciotoli et al., 2015, 2016; Esposito et al., 2021;	Ciotoli et al., 2013	High	2-5

Flood	Alluvial plain of Tiber River and its main tributaries	Cremera Creek Flood (2014)	Remedia et al., 1998;	PAI Flood Hazard Map; Di Salvo et al., 2017	Bersani & Benciveng a 2001;	High	20-50
Landslides	Spread over the urban area, mainly concentrated in areas featured by sedimentary deposits (right bank of Tiber River)	Diffuse landslides in 2014 following intense rainfall	*This study*, Alessi et al., 2014; Amanti et al. 2012; 2013; Bozzano et al., 2006	n.a.	IFFI catalogue; Del Monte et al., 2016	Medium	5-10

2.3. Material And Methods

Beyond the general goal of providing a first comprehensive landslide hazard analysis in Rome, the main specific objectives of this research can be summarized in: i) dataset preparation; ii) susceptibility assessment, iii) in-depth check of multi-temporal information about landslide (re)activations and hydrological analysis with the purpose of evaluating a temporal probability of occurrence, iv) testing the integration of remotely-sensed displacement data and susceptibility zoning to provide refined information about critical areas (i.e., those prone to landslide and experiencing an actual deformation).

Basically, whatever the analysis approach chosen among the large variety of those available (Reichenbach et al., 2018), all the methods for susceptibility assessment are based on landslide inventories, that is the fundamental element of input and/or validation of the analysis (Corominas et al., 2014). Furthermore, the quantification of relationships among causative factors and landslide occurrence requires the acquisition of as many datasets and related information as the number of variables (usually of environmental type, such as geo-thematic and morphometric data) related to predisposing/preparatory and, sometimes, triggering factors (Reichenbach et al., 2018; Fell et al., 2008).

In order to predict the location of potential slope failures (i.e., landslide susceptibility), and to evaluate their intensity and temporal probability of occurrence, the landslide inventories must include information about the exact timing, size and mechanism of the slope movements. The definition of landslide recurrence thus requires at least a raw dating of the events (monthly and yearly precision) to get a reliable statistical assessment of the return period. In case of availability of more precise information (i.e., day and hour), for rainfall-induced landslides it could be possible to assess the landslide frequency as the recurrence interval of the causative factors which, in turn, can be assessed by hydrological analyses.

Since the precise and complete timing of a landslide in a database is the most valuable but often lacking attribute, the full hazard definition represents the most challenging task for both scientific community and local administrators (Corominas et al., 2014).

2.3.1. Inventory review and dataset preparation

Our study started from the acquisition and review of the available data sources, including landslide inventory, as well as rainfall time series available in the urban area of Rome. The available landslide inventories in the Rome municipality area include:

The Official inventories covering the national territory, among which the Aree Vulnerate Italiane (AVI) project (1996), the Italian Inventory of landslide phenomena (IFFI, 2007) now distributed as open access in the IDROGEO Platform (Iadanza et al., 2021), the Tiber River basin hydrogeological risk mitigation plan (PAI, 2012) and Rome's regulatory plan for the urban area (PRG, 2007). The IFFI project inventory is the national DB of landslides that collects information from different sources. The DB architecture is designed to provide a wide range of information for each landslide, but often lacks or has incomplete data. Mapping criteria and landslide geometries are inconsistent, with some landslides identified as points and others as polygons. Additional processing is required to extract the detachment areas, which are necessary for susceptibility assessments. In addition, key information such as the date of (re)activation, volume assessment, geotechnical parameters, landslide type, and state of activity are not always available in the area. The ISPRA's "Frane Roma" Project is a database of landslides in Rome, providing location, source of information, and description of the event based on Amanti et al., 1995, 2008, 2013; ISPRA, 2014; Ventriglia, 1990, 2002. The CERI provides a database of landslides triggered by rainfall in 2014, surveyed on the field (Alessi et al., 2014). The Geomorphological Map of Rome by Del Monte et alii (2016) covers the downtown area and integrates previous studies with remote analysis and field surveys.

During the collection and review of the available sources, database completeness and homogeneity analyses were approached, analysing the documental sources, and comparing the different elements of the ancillary catalogues to get a unique, coherent, and corrected data source for subsequent analyses.

Such an integration has been achieved by a sequence of actions:

1. Merging of available data to i) check and erase duplicates, ii) if needed, correct the location and/or attribute a type of movement, by means of DTM- and orthophoto-based interpretations. Specifically, to check for duplicates between the various DBs, buffers of 20 m were created around each landslide: their intersection, if present, could mean a duplicate of the same process and, thus, all the redundant records by a supervised procedure were

deleted. The selection was made starting from those most reliable in terms of precision and checking the quality in the geolocation of the surveyed landslide phenomena.

2. Reviewing and standardizing the classification in terms of type of movement, since different descriptions for the same typology have been adopted in the native DBs.

3. Ensuring a geometrical homogeneity by associating at each landslide the minimal informative geometry, i.e., a point (Landslide Initiation Point - LIP), which is located in the topmost part of the corresponding line (main scarp edge), or polygon (whole instability or accumulation) lends itself better to susceptibility analyses for shallow landslides.

The so-defined inventory includes (where available) both landslide polygon and LIPs, despite only points having been adopted to train the susceptibility model. Given the landslide population and the different landslide types available in the catalogues, shallow landslides and earth slide mechanisms have been considered.

2.3.2. Spatial component of the landslide hazard

From the information contained in the landslide catalogues and their related attributes, the definition of rainfall-induced landslide hazard was attempted, starting from the definition of the spatial attribute of landslide hazard (i.e., the landslide susceptibility).

To analyse the relationship between causative factors and landslide presence/absence, we relied on a data-driven approach featured by Machine Learning (ML) techniques aimed at inferring the multivariate combination of causative and predisposing factors over stable and unstable sectors of the investigated area. As Machine Learning algorithms strongly rely on quality and amount of input data, we decided to boost (i.e., modify and make it reliable for a proper hazard assessment) the landslide database. Specifically, we developed a methodology to assess the extent of the detachment areas and to maximize the amount of information for each landslide generating additional “synthetic” LIPs to better catch the variability of predisposing factors within a given detachment area. This solution adopts a bounding box enveloping landslide crowns or create half-circle buffers around point elements indicating non-mappable landslides. This workflow allows the reconstruction of detachment areas that were used to sample the synthetic LIPs randomly over areas located at elevation lower than the crown and thus within the potential landslide source area. Once obtained, the synthetic LIPs were validated by testing the similarity between the distribution of their features with that of the original LIPs. Chi-squared (CS test) and Inverted Kolmogorov-Smirnov D statistic (KS test). As regards the factors predisposing landslide initiation, the most important and adopted variables were considered based on the available products. All the DEMs available in the municipality of Rome were evaluated to extract morphometric and hydrological derived variables. A new and more reliable DTM was defined starting from the extraction of the numeric 1:5,000

topographic map of contour lines and quoted points referred to the terrain. The DTM was generated by means of the ANUDEM algorithm implemented in the TopoToRaster tool by ArcGIS. Lithological and Land use maps available in open access by the Lazio Region (https://geoportale.regione.lazio.it/) were collected to derive litotechnical and land cover classes at a scale of 1:25'000. Given the scarcity and inhomogeneity of the available geotechnical data, additional informative variables about the mechanical parameters of soil cover and bedrock were not considered.

Then, some of the most important and used predisposing variables (Reichenbach, 2018) were adopted. These variables and parameters include:

Table 3.2: Summary of the explanatory variables adopted in the analysis.

Category	***Variable***	***Acronym***	***Scale/Resolution***	***Source/Reference***
Lithological and hydrogeological factors	Lithotechnical units	*litotec*	1:25,000	Lazio Region
	Distance to boundary of permeability contrasts	*dist_lim_k*	10x10m (Map 1:100'000)	Hydrogeological Map of Lazio Region (Capelli et al. 2012)
	Distance to main river	*dist_idro*	10x10m (Map 1:100'000)	ISPRA
Primary and derived Morphometric variables	Elevation	*dtm*	5x5m (Map 1:5000)	CTR Lazio Region
	Slope	*slope*	5x5m	-
	Aspect	*aspect_8*	5x5m	-
	Relative Relief Index	*rrelief100*	5x5m	Hesse, 2010
	Total, Planar and Profile Curvature	*cur_tot; cur_prof; cur_plan*	5x5m	-
	Flow Direction	*flow_dir*	5x5m	-
	Flow Accumulation	*flow_acc*	5x5m	-
	Topographic Position Index	TPI	5x5m	Jenness, 2006

			(Neighbourhood 100m)	
	Topographic Wetness Index	TWI	5x5m	Beven & Kirkby, 1979
Anthropogenic factors	Land use	*cus*	10x10m	Lazio Region Land Use Map DGR n.953/2000
	Distance to Road	*dist_road*	10x10m	Open Street Map (OSM) consortium
	Soil consumption	*soil_cons*	10x10m	Municipality of Rome

To include preparatory factors (sensu Julian and Anthony, 1996) ed as proxies of middle-term “hydraulic stress” and weathering efficiency which, in turn, can be related to the material strength decrease. The mean annual rainfall (ord_rain) was defined for each rain gauge and then attributed to the related Thiessen polygon.

In pre-processing stages, correlations between features, or multicollinearities, were calculated and variables were systematically removed if a pairwise correlation exceed 75% (Mastrantoni et al., 2023).

Explanatory variables were sampled in landslide initiation points (LIPs) and random stable points after pre-processing. Stable areas were defined as non-landslide areas, which were used to train the statistical model. The LIPs and stable point datasets were balanced with a 35-65% ratio of unstable and stable points, respectively based on the need of maximising predictive ability on Positives but at the same time avoiding an excessive distance between the actual statistical population (i.e., actual proportion of unstable and stable areas in the whole territory) and the sampled training datasets, thus limiting an excessive False Positive Rate. The datasets were then split into 80% for training and 20% for validation. The susceptibility function was trained and tested on 10 K-fold random partitions of the dataset to estimate the accuracy of the predictive model and avoid selection bias.

Given the spatial concentration of landslide effects in the Monte Mario area caused by many landslides occurred during the 31 January - 2 February 2014 meteorological event, the landslide LIPs training dataset was split into two different portions, where the first represents the generic catalogue of instabilities that happened before and after this intense event and the second “nested” subset composed by the 2014 event-inventory. Since previous studies remark the exceptionality of this rainfall event in the NW sector of the city (Alessi et al., 2014), the above-mentioned subsets represent the result of ordinary (on an average basis) and rare rainfall events, respectively. On the generic landslide database, a single predictive susceptibility function was trained. To derive the most

appropriate predictive function, we tested several ML models, among which the Extra Trees Classifier (Guerts et al., 2006) implemented in the Scikit-learn package (Pedegrosa et al., 2011) outperformed the other ones in the study area (Mastrantoni et al., submitted). The Extra Trees (abbreviation for extremely randomised trees) is an ensemble supervised machine learning method that fits many randomised decision trees on various sub-samples of the dataset and uses averaging to improve the predictive accuracy and control over-fitting. In Extra Trees, randomness goes one step further in the way splits are computed. It does not come from bootstrapping of data but rather comes from the random splits of all observations.

The Extra Trees algorithm creates many decision trees, but the sampling for each tree is random, without replacement. This creates a dataset for each tree with unique samples. A random subset of candidate features is used for each tree. The most important and unique characteristic of the Extra Trees is the random selection of a splitting value for a feature. Instead of calculating a locally optimal value using Gini or entropy to split the data, thresholds are drawn randomly for each candidate feature, and the best of these randomly generated thresholds is picked as the splitting rule. This makes the trees diversified and uncorrelated. With this approach, both bias and variance are handled. The former is reduced by using the whole original sample instead of a bootstrap replica. The latter is dwindled by randomly choosing the split point of each node. Since splits are selected at random for each feature in the Extra Trees Classifier, it is less computationally expensive than a Random Forest (Breiman, 2001).

Validation of susceptibility product was also performed on training and test datasets expressing the predictive capability of the model by Confusion Matrixes and resulting Receiver Operating Characteristic (ROC) curves as well as other metric performance indicators.

Furthermore, considering the value that the susceptibility prediction assumes in correspondence with the LIPs, detection rate curves (DRCs) were constructed, which represent the cumulative percentage of LIPs correctly predicted for increasing susceptibility values, regarded as the separation threshold between the true and false positive condition of each LIP. This method allows setting susceptibility thresholds corresponding to predetermined detection rates, driving the discrete classification of the susceptibility values. In this paper, to avoid coarse “expert” choice on landslide susceptibility value, fixed thresholds on landslide detection rate were chosen, adopting percentile values of landslide detection equal to 50, 75, 95 and 97.5% to divide very low, low, moderate, high, and very high susceptibility classes. Specifically, two DRCs were built: one using only the LIPs of the 2014 event (recognized as extraordinary) and the other with the remaining dataset (which mainly includes events triggered by ordinary rainfall). To attribute temporal recurrence to such scenarios and transpose them into hazard, approaches of data integration have been adopted looking for the best definition

of the landslide hazard with the available data, shifting from its theoretical definition to operative practices of use for stakeholders.

2.3.3. Temporal component of the landslide hazard

Once defined the landslide susceptibility from the generic catalogue of landslide effects, further steps in the definition of the landslides hazard have been addressed, facing it through a multi-stage and multi-level approach, consistent with the quality and quantity of available data.

The analysis of landslide frequency has been addressed by means of i) a preliminary evaluation of landslide event recurrency on the few dated landslides; ii) quantification of large-scale hazard descriptors (Corominas et al., 2014); iii) detailed quantitative attribution of Return Periods (RPs) of the rainfall events that caused landslide triggering basing on a hydrological analysis, hereafter discussed.

Given the composition of the above-mentioned landslide database with both generic and event-based inventories, the here-defined resulting landslide susceptibility product can be referred to as much as common “ordinary” scenarios as to the severe and non-ordinary conditions experienced in 2014.

2.3.3.1. Preliminary estimation of landslide recurrence

Temporal hazard was preliminary approached by exploiting the available dating’s information retrieved in the catalogues. Selecting landslides of same type featured by multiple reactivations, a gross estimation of the RP was attempted, defining the temporal range between the first and the last (re)activation. Because of the reduced number of dated landslides events and the presence of replicas, the number of effects that experienced multiple reactivations was defined in 17 out of the 471 total landslides.

2.3.3.2. Large-scale quantitative hazard description

Since we face off with large-scale analysis for advisory purposes, low-resolution hazard descriptors at a regional scale can be defined according to Corominas et alii (2014) based on indices defined as landslide density and frequency or landslides/year/km^2. For this purpose, Thiessen polygons were delineated for rain gauges featured by a decades-wide time series, to spatialise the area of interest of each weather station (Figure 2.1). Hence, dated landslides of the same mechanism were considered to evaluate temporal recurrence within Thiessen polygon and to relate it to the spatial landslide density.

Landslide effects with multiple reactivations, i.e., effects were more than one date of rainfall associated, give us the opportunity to calculate the period ΔT between the first and the last rainfall-

induced landslide and the number of reactivations of landslides (n) within each Thiessen polygon, so the ratio n/ΔT provides the minimum landslides temporal frequency. From these results, a synthetic hazard descriptor (H_d) for each Thiessen polygons was calculated as the ratio of number of landslides activations over ten years.

2.3.3.3. Evaluation of RPs of landslide trigger rainfall events

The most advanced insights on landslide hazard were retrieved from the analysis of the recurrence of the landslide triggering factors by analysing the intensity and RPs of the main rainfall events that caused ground instabilities by proper hydrological analysis.

To this aim, a hydrological analysis on rainfall regime was performed by choosing reference pluviometric stations considering the time span of the historical record and the time sampling of the rainfall series (daily or hourly data). An overall time period of 70 years was considered, despite the analysis of rain gauge stations data revealing a reduced continuity in data logging and a time coverage often reduced to several years, thus not sufficient to perform rigorous hydrological analysis. For this reason, only weather stations with a dataset equal to or broader than ten years were adopted, despite at least 20 years should be considered in the analysis for reliable results (Serrano, 2010). The available pluviometric stations are reported in Appendix B.

A statistical analysis of maximum rainfall intensity data was performed on daily and hourly data, to evaluate the return period of heavy rainfalls for different durations. The hydrological–statistical analysis of the maximum values requires the cumulative rainfall at different time intervals to calculate the rainfall probability curves. Daily and hourly rainfall data were used to calculate cumulative rainfalls over the territory of the municipality of Rome for time stages of 1, 2, 5, 10, 20, 30, 60, 90,120 and 180 days, and 1, 3, 6, 12, and 24-hours, respectively. The generalised extreme value (GEV) distribution (Jenkinson, 1955) was adopted, which follows the following function:

$$F(x) = \exp\left\{-\left(1+\xi\frac{x-\mu}{\sigma}\right)^{\frac{1}{\xi}}\right\} \qquad \text{[Eq. 2.1]}$$

Where μ, σ, and ξ are referred to as the location, scale, and shape parameters, respectively. These parameters have been defined by applying the probability-weighted moments (PWM) method (Hosking et al., 1985) based on the maximum values of the above-mentioned rainfall time periods available from the dataset. First, the RPs of each considered variable were obtained by inverting the probability function. Then the obtained cumulative rainfall value was fitted by a power law distribution to build the rainfall probability curves.

For every certified landslide (i.e., with a defined date), the daily rainfall value recorded at the closest rain gauge was attributed to infer the range of potential RPs of the intense rainfall causing instabilities and, indirectly, the RPs of the landslide event.

We used the cumulative rainfall values recorded at the relevant rain gauge in correspondence of the date of landslide occurrence, and then we compared these values with both hourly and daily probability curves. Where the hourly rainfall data or timing of landslide occurrence lacks, we referred to the total cumulative daily rainfall, inferring a posteriori the possible range of admissible RPs if concentrated rainfalls (within 3 hours) or distributed precipitations (over 24 hours) are considered.

Such an inference is needed due to the absence of the time of occurrence of the landslide events, which makes the precise attribution of RP impractical. Based on the rainfall linked to the dated landslides with respect to the rainfall probability curves, it was possible to establish the order of magnitude of the RPs of each landslide available in the record.

Despite the effort to gain quantitative information about the temporal recurrence of landslides and face off with the definition of landslide hazard, this method is threatened by the scarcity of dated landslides and the intermittent recordings of several rain gauges.

2.3.4. Persistent Scatterer Interferometry

The landslide hazard quantification in Rome was completed by evaluating the areas where probability can be considered unitary, since they are experiencing slope movements to date and should be prioritised in the risk analyses. Such areas of interest have been identified using satellite remote sensing technique, and the InSAR (Interferometric Synthetic Aperture Radar) technique (Massonnet & Feigl, 1998; Hanssen, 2005). Such analyses provided information on the distribution of ground movement, giving valuable insights into assessing the state of activity of the detected land movements and additional information on possible landslide reactivations of pre-existing landslides. Furthermore, the contribution of such relevant data can integrate the static landslide susceptibility assessment, providing dynamic information on landslide activity. For these purposes, the Advanced Differential SAR Interferometry (A-DInSAR) approach (Ferretti et al., 2001; Kampes, 2006) was applied to the study area, by the processing of C-band Sentinel-1 (European Space Agency, ESA) SAR images, in ascending and descending orbital geometries, covering the Rome municipality area. The acquired images cover a time span of about five years (from October 2014 to April 2019) and have been processed through the Persistent Scatterers Interferometry analysis (Ferretti et al., 2000; 2001; Kampes, 2006; Crosetto et al., 2016), that is optimal for detecting deformations in an urban environment, where the density of permanent scatterers (PS) is generally high.

The resulting PSs have been filtered after by applying a high temporal coherence threshold (> 0.6). The results are represented by the velocity of displacement maps (mm/yr) for both ascending and

descending datasets, where the movement velocity is measured along the satellite line of sight (LOS). These results were post-processed by setting specific and reliable thresholds to identify areas experiencing slope movement and select the “landslide candidates” in the PSs velocity maps. Such thresholds were fixed on velocity values of PSs (velocity < -2.5 or > +2.5 mm/yr) and PSs location, considering only PSs in areas with slope > 5°.

2.3.5. Data Integration

In order to perform an accurate and reliable landslide hazard analysis for the city of Rome and shed lights on the most critical sectors in terms of first-generation landslides and state of activity of existing ones, approaches of data integration can be used, combining the results of spatial and temporal hazard, adding on its evidence of slope movements derived by PSI analysis.

The raw attribution of landslide recurrence in each Thyssen polygon has been combined with its landslide susceptibility, thus attributing it to the highest susceptible areas, zoning the territory by landslide hazard indicators. Specifically, from the above-described susceptibility analysis, a landslide Hazard Index (H_i) was defined combining, cell-by-cell, the susceptibility level with the classes values of the large-scale hazard descriptors (H_d) expressing the mean landslide temporal recurrence in every Thiessen area. H_i has been thus defined as:

$$H_i = \text{Susceptibility} / (1 - H_d)$$

The results of Hi has been after reclassified to rank the hazard in view of final user requirements. In this paper a 5-classes reclassification based on Natural Breaks was adopted. Then, potential landslide source areas have been extracted within the high and very high susceptibility classes defined considering the intensity-based DRC classification.

Hence, if two intensity scenarios are considered and two DRC-based classification adopted, it is possible to obtain hazard maps concerning different intensity levels, delineating potential landslide source areas reactivable under different ranges of RPs, achieving a first robust landslide hazard assessment in Rome.

2.4. Results

2.4.1. Inventory review and dataset preparation

From the collection and comparison of the available databases, we noticed a certain degree of fragmentation of information, such as lack of heterogeneity and consistency, record repetition and different attributes related to landslide effects. The differences are evident also about number of elements, publication date (from 1995 to 2018), description of the type of movement (if present),

updating, purposes and scale of observation (from national to local). Moreover, the data associated with the landslides are not homogeneous for what concerns temporal information. Certified (i.e., dated) landslides are available only in the "Frane Roma" ISPRA Project. After the data collection, the geometry of the databases also appears heterogeneous: within the same database, shapefiles can vary from points or polygons to points-lines-polygonal vectors. However, many inventories use a single geometry as a comprehensive outline for source transit– and deposit areas.

According to all the available sources, the uniformed DB is composed by as many points as the filtered landslides (566), including earth slope instabilities (earth slides and shallow i.e., soil slip) movements, rockfalls and remaining n.d. effects. Most of the landslides in the city of Rome occurred in the last decades (Figure. 2.2), with slope failures that have almost always been associated with heavy rainfall between 2008 and 2014 (67 landslides out of the total occurred during the storm of 31 January 2014; Alessi et al., 2014).

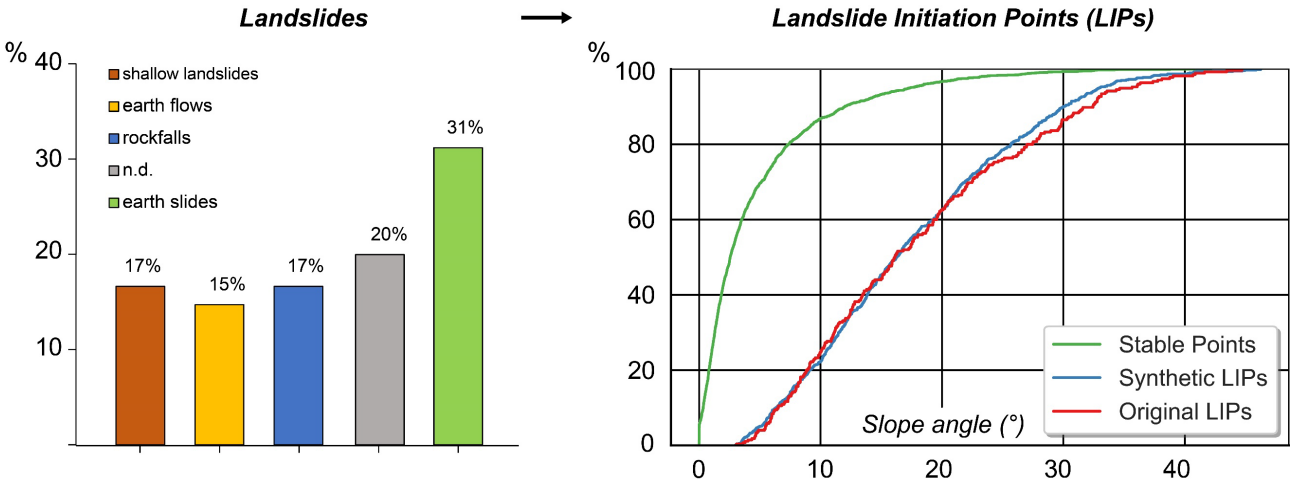


Figure 2.2: a) Distribution of landslide effects available in the catalogues over the territory of Rome. b) Distribution of LIPs and Stable Points vs Slope angle adopted in the susceptibility model.

Based on the reviewed catalogue, the dataset of LIPs to be considered for the susceptibility analysis was defined, accounting for shallow earth failure mechanisms only, hereafter considered as a unique ensemble. Test of validation of the synthetic vs original LIPs was performed through the evaluation of similarity of the slope angle distribution between synthetic and original LIPs (Figure 2.2). CS and KS tests resulted in scores of 0.666 and 0.923, respectively.

Hence, the final point-based landslide database is represented by 1099 LIPs (289 original and 810 synthetic), excluding the 67 related to the January 2014 extreme rainfalls.

2.4.2. Spatial component of the landslide hazard

The here conducted analysis allowed us to point out the landslide susceptibility to earth slide and shallow failures in the municipal area or Rome. The continuous landslide susceptibility map shown

in Figure 2.3 reveals local maxima in the right embankment of the Tiber River, in the Monte Mario and Monte Ciocci ridges, where the M. Vaticano and M. Mario sedimentary units crop out. The portion of Rome west of the Tiber River shows higher susceptibility with respect to the eastern one because of the presence of tuffs and pyroclastic deposits of Sacrofano and La Storta unit from the Sabatini Volcanic District (Sottili et al., 2004).

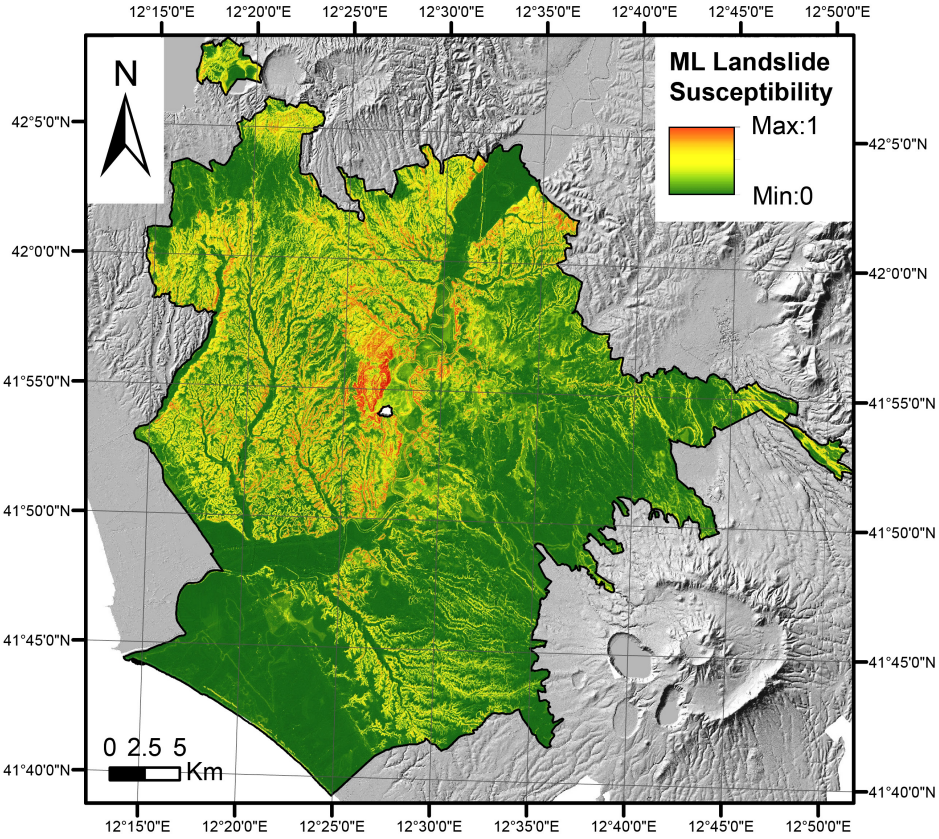


Figure 2.3: Continuous map of landslide initiation susceptibility of Rome area according to the adopted ML model.

The most important predisposing and preparatory variables can be identified in the Feature Importance permutation graph in Figure 2.4, where slope angle and relative relief (rrelief100) are the most conditioning features, which showed a Pearson correlation coefficient equal to 0.7, thus acceptable since lower than the threshold of 0.75 recommended by Kuhn and Johnson (2013). Relevant importance can be attributed to lithology and land use types as well as to the distances to rivers and permeability limits, which can be considered a proxy of hydraulic circulation and the presence of temporary water tables. A marked preparatory role was played by the predictive variable of the annual rainfall, whose distribution can be considered a factor capable of controlling soil moisture and a stressor for slope stability.

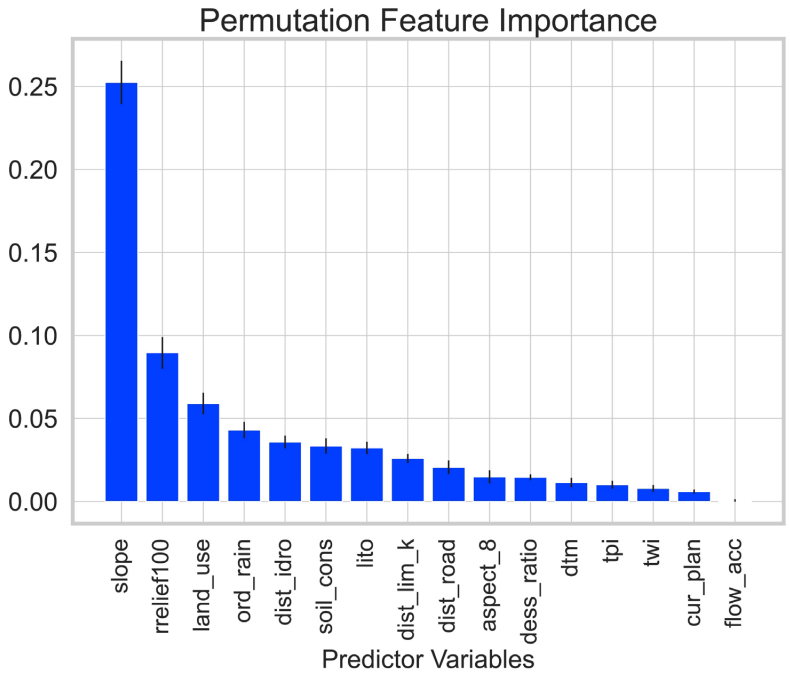


Figure 2.4: Feature importance resulting from the Extra Trees Classifier model.

The quantification of susceptibility performance revealed on the test dataset by the confusion matrix and resulting ROC curves highlighted the very high quality of the function in predicting both stability and instability (Figure 2.5), especially after the definition of the hyperparameters. An Area Under Curve (AUC) value of 0.96 was found, with excellent specificity and sensitivity.

Afterwards, the susceptibility map was the reclassified according to the DRCs extracted with respect to subsets of landslides available in the catalogue and referrable to the 2014 event and the rest of the catalogue. From this point, two different scenarios were considered adopting the two reclassification criteria, that account for the different susceptibility conditions referrable to events of different intensity.

Two different detection rate curves were thus extracted from the dataset, including LIPs of the 2014 event and the other antecedent 2014 (pre-2014 in Figure 2.6), to fix susceptibility thresholds under fixed detection rates. Such thresholds drove the discrete classification of the susceptibility in five classes. For example, moderate to very-high susceptibility classes moved from susceptibility equal to 0.27 to values greater than 0.65 if the rare scenarios are considered (Figure 2.6).

The spatialisation of the so-defined susceptibility classes points out how Monte Mario and Monte Ciocci hills face a limited increase in landslide susceptibility when moving from the ordinary to the rare scenario, that reflected in limited variations of the areal extension of the high and very high susceptibility classes (Figure 2.7). Larger differences can be found in the right embankment of the Tiber River, where wide areas fall into the moderate and high classes given the reduction of the class breaks from 0.65 to 0.27 (Figure 2.7).

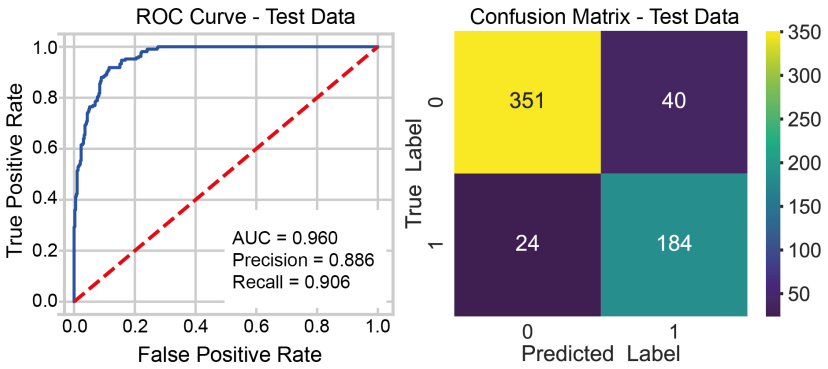


Figure 2.5: ROC curve defined by the confusion matrix on the testing dataset. AUC, Precision, and Recall are also reported.

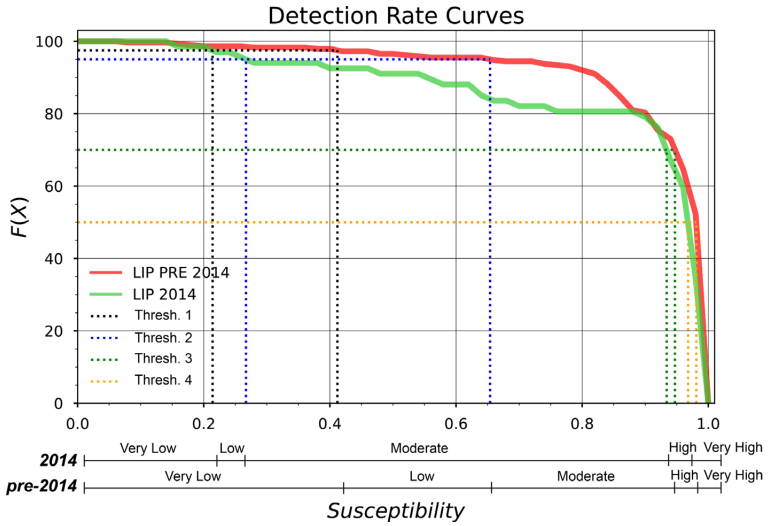


Figure 2.6: Detection rate curves obtained on the susceptibility results considering LIPs before the 2014 meteorological event and dated 31/01/2014 drive the discrete classification of the susceptibility.

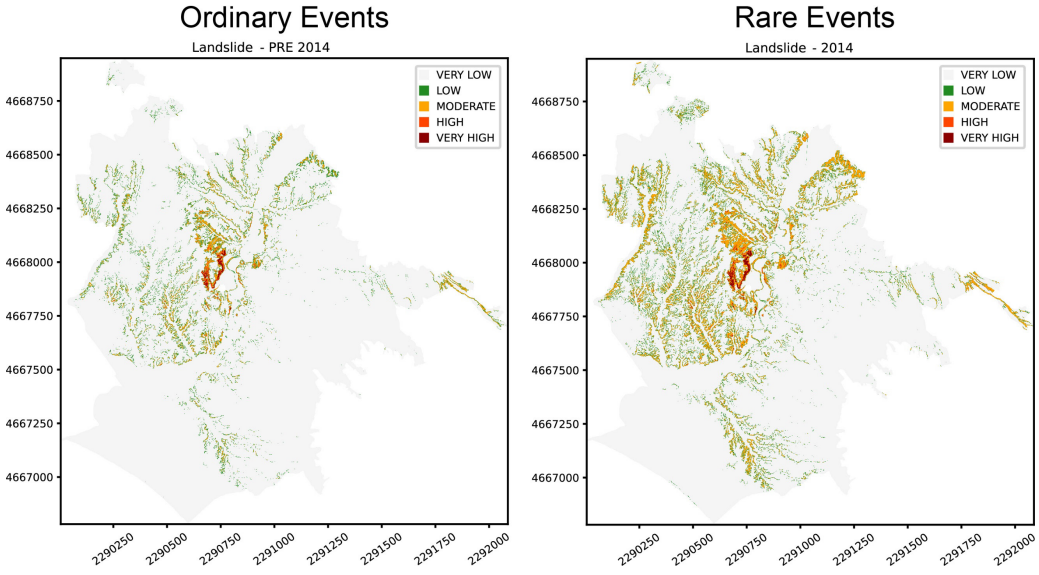


Figure 2.7: Classification of landslide susceptibility for earth failure mechanism in the municipal area of Rome according to detection rate thresholds defined for ordinary scenarios (left) and rare events (right) like the one that occurred in 2014.

The moderate, high, and very high classes in the ordinary cover a percentage of the area of 3.73%, 0.32%, and 0.10%, respectively, which increase up to 10.50%, 1.64%, and 0.37% in the rare scenarios. Adopting the two different classifications, it is possible to consider two different landslide intensities (intrinsic in the event-based inventory) and refer to the susceptibility analysis to pluviometric events with increasing RPs.

From the obtained results it is possible to state that limited differences among the two intensity scenarios can be found. This similarity relies on the common average size of the landslides inventoried in the generic and 2014-event catalogues, as well as on the similarity of landslide triggering pluviometric inputs. For these reasons, a univocal landslide scenario and a resulting map can be considered valid for depicting landslide hazard in Rome.

2.4.3. Temporal component of the landslide hazard

With the aim of associating RPs to the different zones of the territory, a preliminary definition of landslide recurrency was approached by the analysis of temporal range among subsequent occurrences of landslide-triggering meteorological events; first evidence from the few reactivated landslides came out by the preliminary attempts: on the 17 reactivated effect, the return period ranges between 1 and 25 years with a mean value of 5 years.

The large-scale analysis on dated landslides, which uses hazard descriptors, pointed out the higher landslide density in the W and NW sector of Rome for shallow and earth slide failure mechanisms. The landslide density is in a strict relationship with the mean annual and maximum daily rainfall (Figure 2.8).

The resulting landslide frequency fairly matches the one obtained by the preliminary analysis on return periods of reactivated (and multi-dated) landslide, whose values range from 2 to 10 years.

In-depth information on temporal hazard has been also derived from the hydrological analysis of daily and hourly rainfall data, resulting in a series of rainfall probability curves enclosed in the Appendix B. Some of these curves are also reported in Figure 2.9. The here-conducted analysis allows to evaluate the cumulative rainfall for every pluviometric station assuming different RPs, updating the analysis carried out by Alessi et al. (2014) to the end of March 2021, and confirming that the January 2014 storm is characterised by RPs up to 100 years.

Based on the few landslides featured by a certain date and assuming as critical rainfall the cumulative, daily rainfall fallen until the landslide date, it is possible to qualitatively evaluate the range of RPs of the triggering rainfall by comparing the daily and hourly cumulative rainfall obtained from the GEV analysis (Figure 2.10).

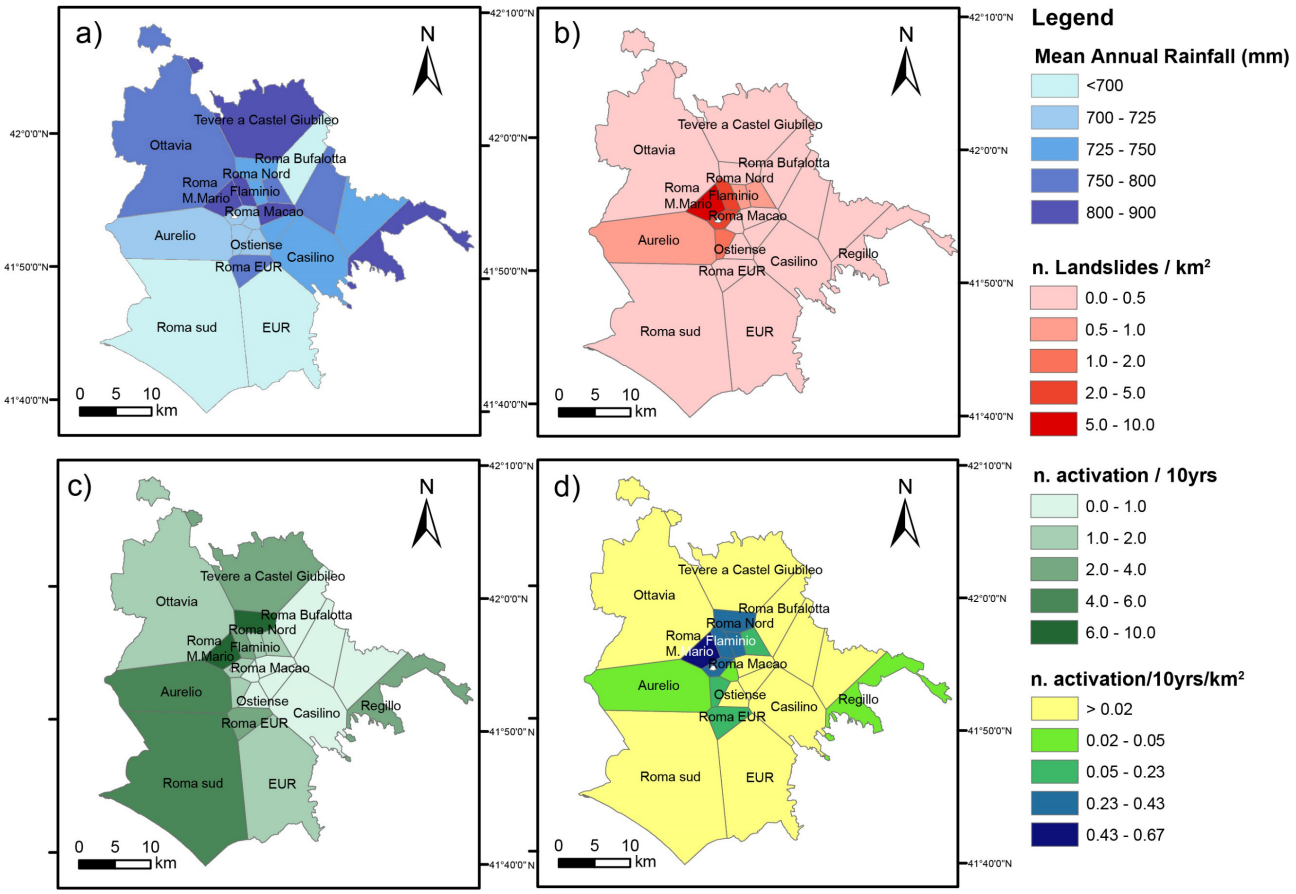


Figure 2.8: a) Mean annual rainfall in Rome according to rain gauges with a minimum time coverage of 20yrs. b) spatial density (landslide/km^2) of shallow landslides and earth slides over the total area of Rome. c) landslide temporal return period (n. of activations/10 years) of shallow landslides and earth slides over the period from first to last dated landslides. d) synthetic descriptors of hazard (H_d – n. activations/10 years/km^2).

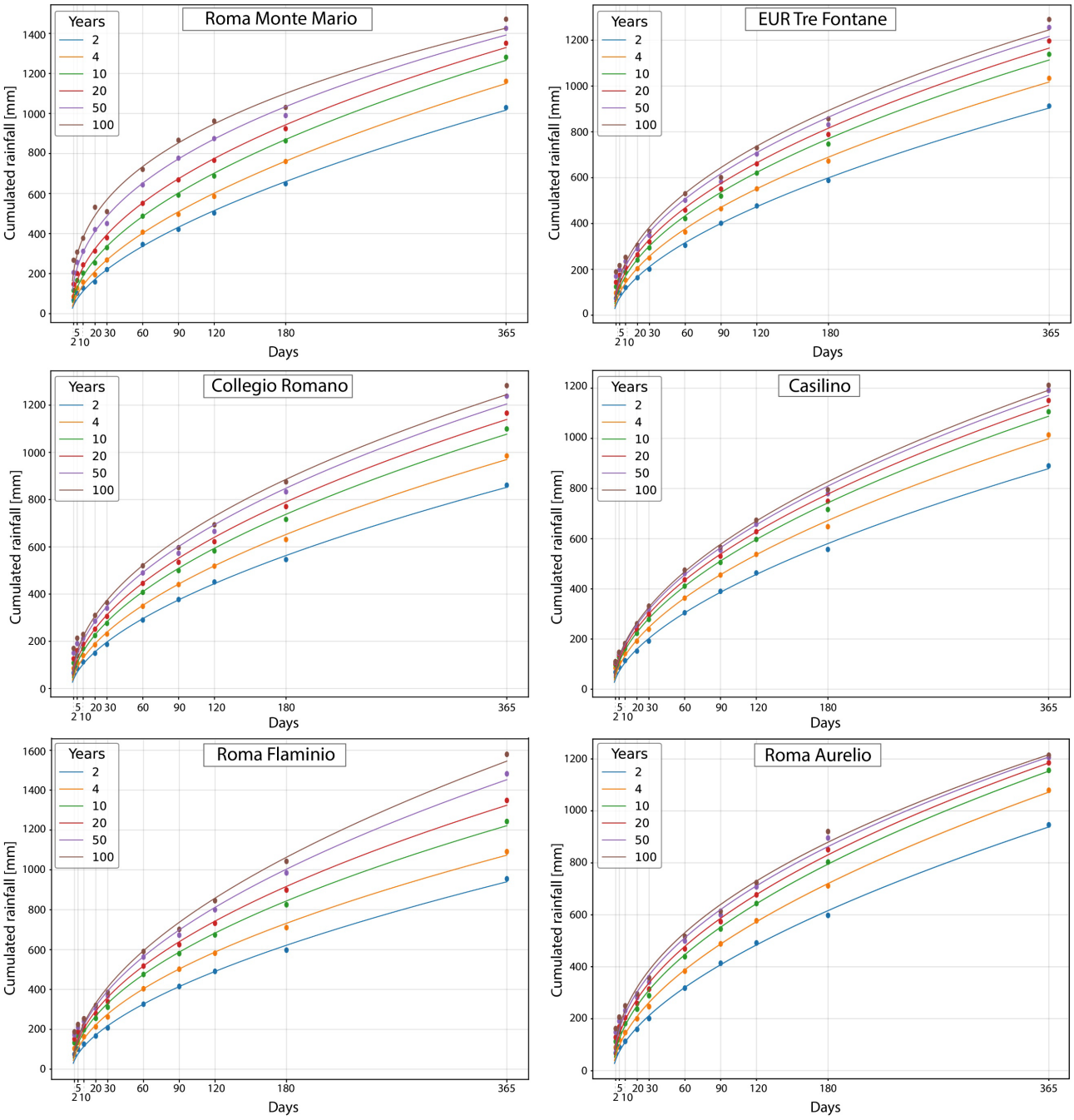


Figure 2.9: Exemplary rainfall probability curves obtained in the Municipality of Rome by the GEV analysis on daily rainfalls. The total of the processed weather station is provided in Appendix B.

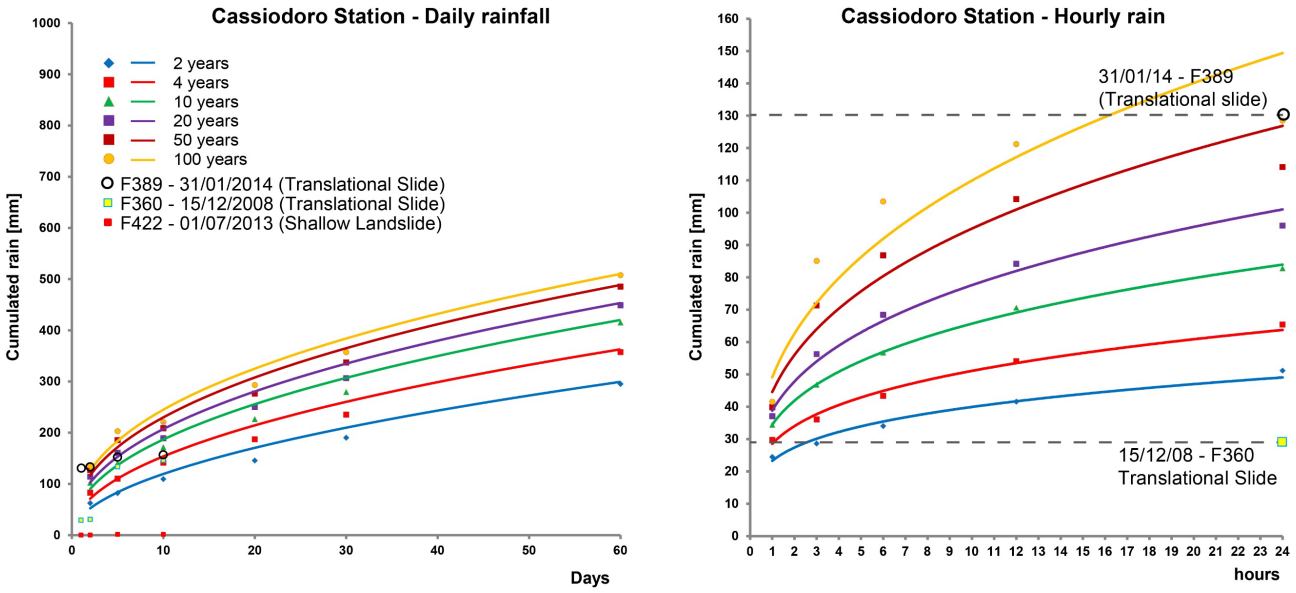


Figure 2.10: Extraction of range of RPs from the comparison between daily rainfall and results of the GEV analysis derived from daily and hourly rainfall records.

It is possible to note that in the municipality of Rome, the pluviometric events responsible for landslide triggering are featured by RPs below 2 years for most of the rainfall-induced effects. On the contrary, RPs for January 2014 and December 2008 rainfall events that, resulted in RPs always above 10 years.

In detail, the graphs in Figure 2.11 show the results of the rainfall data processing plotted on the rainfall probability curves. It is evident that TR < 2 values prevail for rainfall-induced effects that occurred before January 2014, as well as the large amount of "No Data" testifies the lack of rainfall record (Figure 2.11b). For ordinary rainfalls, instead, the comparison with rainfall probability curves obtained by GEV highlighted that more than 60% of dated landslides are associated with RPs among 2 and 10 years (Figure 2.11b). The landslide that occurred on 31/01/2014 significantly show the exceptionality of the event, with more than 25% of the landslides with RP values between 10-50 years and complementary ones ranging between 50 and 100 years (Figure 2.11c).

This confirms the exceptionality of the 2014 rainfall event, as stated by Alessi et al. (2014), that, however, reflects in average sized landslides. Furthermore, they retrieved its uneven areal distribution and the peak of rainfall in the NW sector of Rome, which is the most susceptible and where the higher landslide density and frequency exist (Figures 2.1, 2.3, 2.8).

Given the results of preliminary evaluation, hazard descriptors at large scale and detailed hydrological analysis, site specific landslide scenarios can be defined, attributing an average landslide intensity, and associating temporal attributes to the static susceptibility.

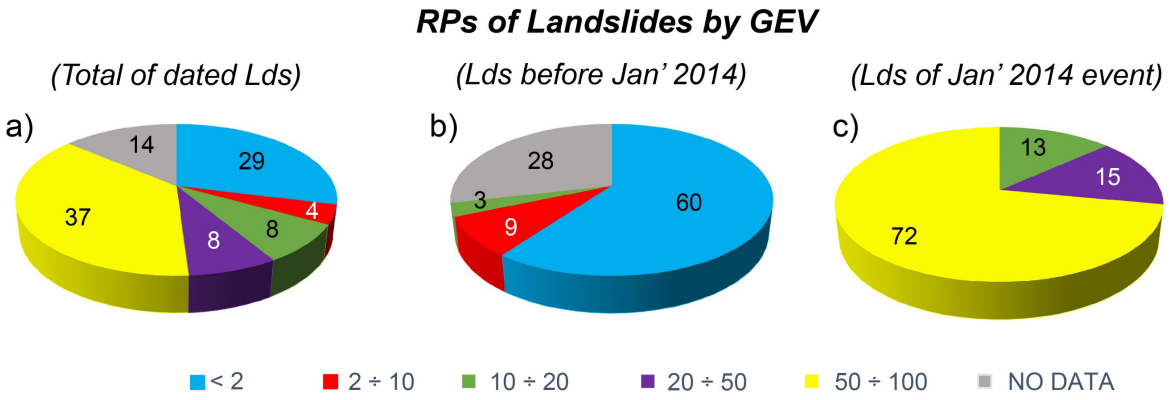


Figure 2.11: Pie chart showing the percentage of landslides with different RPs considering the entire database (left), before the 2014 event (centre), and during January 2014 (right) meteorological event.

2.4.4. Persistent Scatterer Interferometry

The complementary interferometric analysis carried out and the resulting A-DInSAR velocity maps referred to the period October 2014 - April 2019 (Figure 2.12), represents the first steps in the dynamic update of the static landslide susceptibility map.

The detected clusters of PSs affected by displacement allowed us to identify a total of 19 moving areas featured by slope angle >5° (only 3 clusters are in areas with slope < 10°). It is worth noting that the most significant part of the urban slopes, where the highest slope angles and susceptibility values were found, such as those of Monte Mario, are generally covered by vegetation that limits the possibility of retrieving PSs.

A-DInSAR analysis identifies the areas where a slope movement is taking place during the analysed time span (landslide candidates) and reveals the active deformation phenomena. The comparison of the results of the PSI analysis with the landslide susceptibility map highlights the convergence of the landslide candidates and the area more susceptible to landslides, as well as potential source areas in the very high susceptibility classes (Figure 2.13). The PSI results were also compared with the lithology outcropping in the study area (scale of 1:5000). Different clusters of PSs in motion are found in correspondence with tuffaceous and clayey lithologies or generally in high susceptibility areas (Table 2.3).

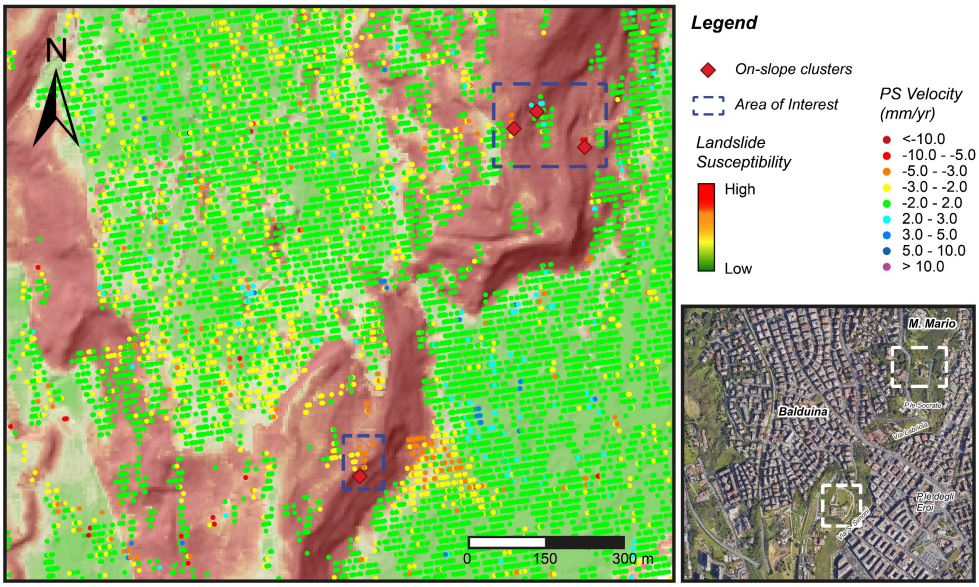


Figure 2.12: PSs map and moving on slopes clusters extracted from PSI analyses on Sentinel-1 images from October 2014 to April 2019. Ascending orbital geometry is taken as an example.

Table 2.3: List of the areas of interest resulting by combination of susceptibility analysis and Persistent Scatterer Interferometry.

Area of Interest	***PS Velocity (mm/year)***	***Susceptibility Class***
Via Simone Simoni	*3-5*	*5*
Via Fedro	*3-5*	*5*
Viale Platone	*5-10*	*5*
Passeggiata del Gianicolo	*3-5*	*5*
Via Misurina	*5-10*	*4*
Via F. Scaduto – Villa Veschi	*5-10*	*5*
Via Trionfale – Via Igea	*3-5*	*2*

2.4.5. Data Integration

Thanks to the integration of spatial susceptibility with temporal attributes resulting from the definition of hazard descriptors or hydrological analysis, a first landslide hazard quantification results from the reclassification and hierarchization of susceptibility, according to their intensity scenario and mean recurrency of the triggering input.

Landslide susceptibility results were combined and integrated with temporal attributes derived from hazard analysis on the subset of dated landslides. Hazard Index (H_i) Map pointed out the maximum hazard in the ridges of Mt. Mario and Mt. Ciocci, where the maximum value of hazard was derived, strengthening the results of spatial hazard analysis. Similar impact can be assessed in the susceptible slopes falling in the Roma Flaminio, Acqua Acetosa and Roma Nord areas, where relatively higher H_d were found. The Hazard Index classification obtained and compared with the reclassified susceptibility map was reported in Figure 2.13.

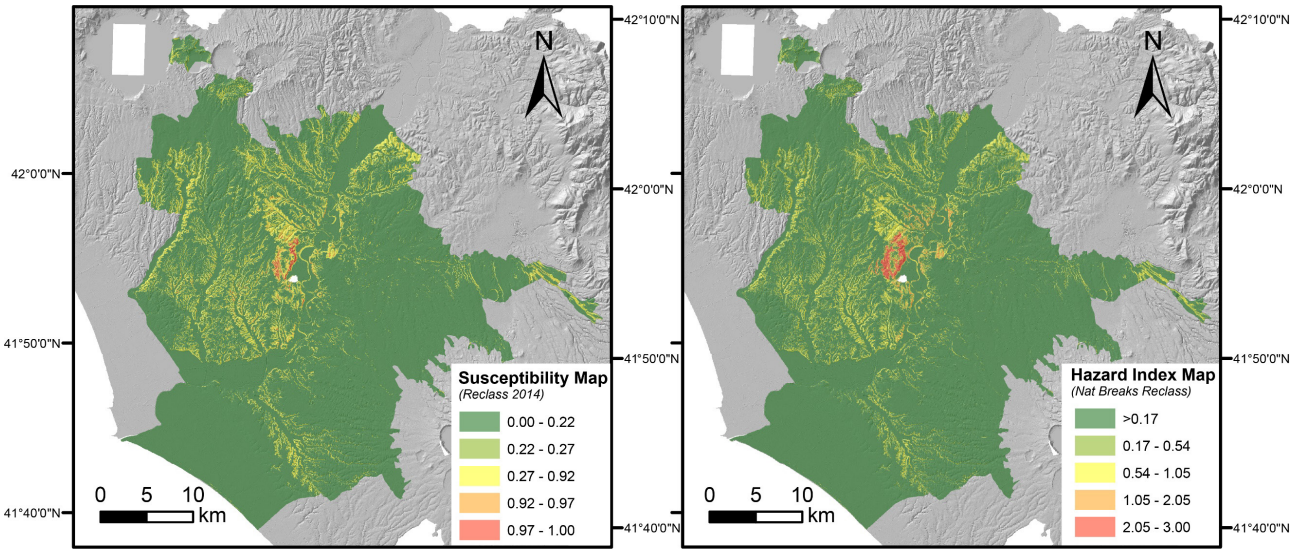


Figure 2.13: Susceptibility (left) and Hazard Index Map (right) resulting from the hazard analysis in the Municipality of Rome.

2.5. Discussions

Multivariate statistical analysis, including machine learning models, can identify landslide patterns by analysing various input data, such as an ancillary landslide database. However, data quality is often inadequate due to incomplete, heterogeneous, and erroneous data, leading to biased susceptibility maps (Steger et al., 2017). Spatial heterogeneity in DBs has proved to cause bias in landslide susceptibility maps (Loche et al., 2022). Open-source landslide inventories may have erroneous geometries and positional errors, resulting in sparse and unreliable data (Steger et al., 2016). To improve reliability, low-accuracy datasets must be integrated adequately in terms of quality and quantity (Mastrantoni et al., 2022; Titti et al., 2021).

The study involved a GIS- and ML-based combined approach to collect, check, cross-validate, and integrate open-source landslide inventories into a single database. However, the number of LIPs was too low to train reliable ML models, so we developed a methodology to derive synthetic LIPs. This helped to improve the overall consistency of the original database and allowed for more accurate ML model training.

Spatial hazard was resolved by means of ML approaches, providing the first advanced landslide susceptibility zoning useful at a statutory level. From the continuous landslide susceptibility map, a discrete reclassification was performed by means of DRCs to encompass the different intensity and account for different intensity scenarios. Results pointed out how similar is the areal extension of the two classifications within high and very high susceptibility classes. These outcomes rely on the most common type of landslides in Rome, which are mainly shallow (soil slips and translational slides) and comparable in volumes (from some cubic meters to several tens of cubic meters). On these assumptions, we can consider the obtained hazard maps as landslide scenarios of a specific intensity and a certain pluviometric event (i.e., a certain number of landslides simultaneously triggered during a rainfall event) rather than a uniform hazard analysis. Landslide scenarios associated to ordinary and rarer pluviometric input were evaluated, assuming the temporal information either from preliminary evaluation on landslide reactivation recurrency or from large scale hazard analysis or detailed hydrological analysis. These data help us to integrate the spatial hazard, by attributing to the different areas of the municipality of Rome average RPs and Hazard descriptors. Given the low number of dated landslides, more detailed analysis on landslide triggering conditions was not approached.

In this study, we constrained order of magnitudes for ordinary landslide frequency, as the estimations are based only on the RPs of the recurrent landslides in which multiple dates are available. To overcome such limitations and provide RP and critical rainfall intensity ranges for the few dated landslides, the antecedent rainfall registered before the event was compared with the standard rainfall probability curves obtained for every pluviometric station. Despite the scarcity of data, landslide frequency in Thiessen polygons and comparison between dated landslides and rainfall probability curves provided similar results.

Despite the effort to gain quantitative information about the temporal recurrence of landslides and face off with the definition of landslide hazard, this method is intrinsic limited by the scarcity of dated landslides and the intermittent recordings by some of the rain gauges. The lack of detailed information from DBs mainly affects the temporal hazard and minorly the spatial hazard.

Afterwards, potential landslide source areas were extracted from high susceptibility pixels, associating temporal attributes, and ranking the hazard zones (Figure 2.14).

For landslide risk management purposes and prevention of landslide risk in a highly urbanised area like Rome (Italy), the definition of spatial and temporal hazards and their variations over time can be crucial, however detailed multi-annual landslide inventories are a necessary and essential fact (Zieher et al., 2016; Fang et al., 2022).

The here-released product can be hence integrated either in case of availability of updated landslide inventories (e.g., insertion of new events and/or refinement of thematic maps), or additional and/or higher-resolution ancillary data become available. This updating process can be followed by adopting defined protocols, aimed at testing the validity of susceptibility models or training new ones. In this sense, cloud-based solutions, tools, or routines can be of help to both geoscientists and administrators (e.g., Titti et al., 2022).

By the proposed approach we provide a way to distinguish different landslide hazard scenarios in Rome based on the fragmented and incomplete data. The reclassification of susceptibility by DRCs also pointed out as landslides in Rome are featured by common intensity (i.e., basically given by landslide size). Furthermore, the spatial susceptibility was joined with the outcomes of the hydrological analysis, which allow linking such scenarios to RPs of 5-10 years for an ordinary rainfall, and of 20-50 years for rarer events (like the 2014 event).

In the hazard map so defined, it is possible to locate on map and extract potential source areas for first-time landslides according to their spatial probability, that increase in intensity (i.e., area) the lower is their associated hazard. On these source areas, an additional classification can be performed to differentiate with hazard attributes the proneness to landslides of specific areas of interest (Figure 2.14).

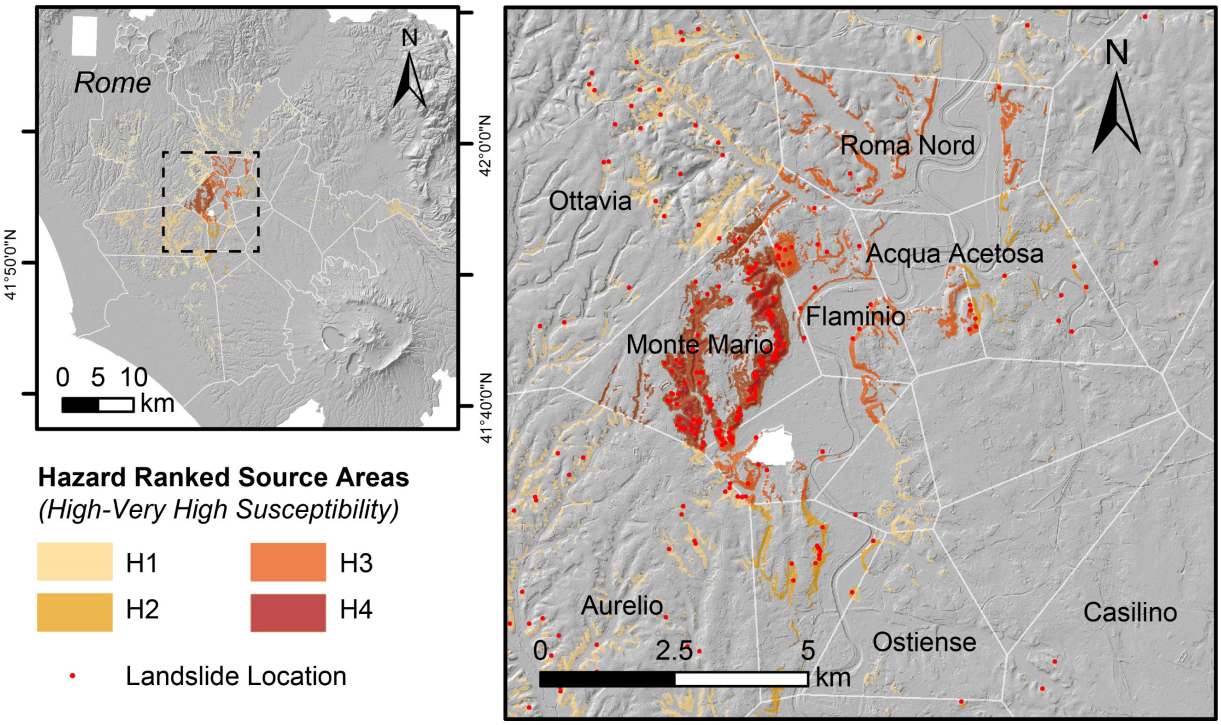


Figure 2.14: Ranking of Hazard in the potential source areas extracted by High and Very High Susceptibility classes. Locations of landslides sites are also reported.

The extraction of potential source areas can be done by an arbitrary choice or supervised procedures from high to very high susceptibility classes, or, precautionarily, considering a high rate of landslide detection (e.g., assuming as potential source areas also lower susceptibility values). Alternatively, potential source areas for shallow or translational sliding can be identified over portions of the slope not yet failed and extracted by fixing a defined percentile in the distribution of susceptibility values within class 5 (max susceptibility). For a complete overview of the hazard condition in the urban area of Rome, first time occurrence hazard map must be overlayed with location of existing landslides, where the hazard in unitary and landslide reactivations must be expected (Figure 2.14). In these hazardous areas, the static landslide susceptibility was linked to the PSI analysis: such an integration allowed to locate and update, in a dynamic way, zones of ground deformations potentially related to landslides. This analysis consists in a second order check on the current state of activity that allowed to quantify the rate of activity within potential landslide source areas. Thus, the multitemporal analysis allowed to convert the static product into dynamic thematic layers, able to steer the risk managers and infrastructure stakeholders towards the most hazardous sectors in the urban area and size the intensity of the gravitational processes (Esposito et al., 2021). According to the location, distribution and displacement characteristics of the observed PSs, some interesting evidence of slope movements were collected, at least where the PSs coverage allowed it (not in vegetated slopes). Data deriving from the European Ground Motion Service (EGMS) platform by Copernicus, will also allow a dynamic update on an annual time scale of the state of activity to rank the most critical situations (i.e., slopes with actual ongoing deformations) among the landslide-prone areas, as suggested for Sinkholes in Esposito et alii, 2021. Furthermore, advances in PSI analysis can availing of COSMO Sky-Med (ASI) images that ensure higher resolution and spatial coverage to catch in detail landslide process. For these purposes, approaches that integrate different constellations can be adopted by data fusion algorithms or by adopting advanced photo-monitoring approaches (Caporossi et al., 2018).

2.6. Conclusions

This paper presents the results of a study aimed at defining the landslide hazard in the municipal territory of Rome, with the dual objective of: i) providing a first "integrated" product useful for the prediction - at least in terms of spatial hazard - of first-generation landslides, through the systematization of all the knowledge available for the area, ii) identifying methods which, starting from an available database that is not optimal in terms of quality and quantity, would allow to obtain reliable results for implementation on an operational level, e.g. land management by policy making institutions.

As for the first objective, this study made it possible to highlight and quantify the potentially critical conditions for landslides affecting the Roman area. This geohazard factor, although considered in

the official planning documents, has always been strongly underestimated with respect to the actual risk conditions: in this sense, only 27, already existing landslides are officially "certified" as situations of potential risk.

At the same time, the work carried out allowed us to develop a methodological approach capable of exploiting the available data. The review of the DB and inventory boosting made it possible to optimize the input information, while the application of ML techniques and the comparison between the results allowed us to select high performance susceptibility functions. The integration of historical-statistical analyses on the few dated landslides and on the recurrence of triggering factors has also made it possible to define the temporal hazard, at least in terms of average frequency of occurrence. The result of the above-mentioned workflow is an integrated map which identifies areas of concern classified based on the combination of spatial hazard with high reliability and expressed in quantitative terms and an expected temporal frequency with lower reliability and measured in terms of "qualitative-quantitative" estimates. The here-defined components of the landslide hazard fulfil the requirements of advisory purposes, thus providing non-binding strategic advice to the management of the territory to address further (and focused) detailed studies in areas where the River Basin Authority could place impeding or prescriptive constraints, based on the severity of the instability. As a matter of fact, the results of this study have already been presented to and shared with local decision-makers and the River Basin Authority of the Central Apennines, starting dedicated interlocutions aimed at its operative transposition. Furthermore, in the frame of an ongoing research contract with the municipal Civil Protection this research represents the basis for the identification of sub-areas of high susceptibility where to evaluate rainfall thresholds for landslide initiation useful for Civil Protection Duties. The reported products point out the state of activity and the framework of landslide proneness for shallow and earth slide failure mechanisms, representing the operative products supporting decision-makers in the management of criticalities in the territory for an aware setup of monitoring solutions and/or prioritization of investments for prevention and mitigation.

Declarations: The authors declare no competing interests.

Data availability: The data that support the findings of this study are available from the corresponding author upon request.

Funding: Open access funding provided by Università degli Studi di Roma La Sapienza within the CRUI-CARE Agreement. This research has been partly funded by the Italian Ministry of University and Research in the frame of the PRIN project “URGENT - Urban Geology and Geohazards: Engineering geology for safer, resilieNt and smart ciTies”.

References

Alessi, D., Bozzano, F., Di Lisa, A., Esposito, C., Fantini, A., Loffredo, A. … Varone, C. (2014). Geological risks in large cities: The landslides triggered in the city of Rome (Italy) by the rainfall of 31 January–2 February 2014. Italian Journal of Engineering Geology and Environment, 1, 15–34.

Amanti M., Cesi C. & Vitale V. (2008). Le frane nel territorio di Roma in La Geologia di Roma – Dal centro storico alla periferia. Mem. Descr. Carta Geol. d'Italia, Vol. 80

Amanti M., Gisotti G. & Pecci M. (1995). I dissesti a Roma in La Geologia di Roma – II centro storico. Mem. Descr. Carta Geol. d'Italia, Vol. 50

Amanti et al 2008 in Memorie Descrittive della Carta Geologica d'Italia: La geologia di Roma. Dal centro storico alla periferia", vol. 80, 2008

Amanti, M., Cesi, C., & Vitale, V. (2008) Landslides distribution in the Roma municipality. Descr. Carta Geol. d'Italia, Vol. 80

Amanti, M., Chiessi, V., & Guarino, P. M. (2012). The 13 November 2007 rockfall at Viale Tiziano in Rome (Italy). Natural Hazards and Earth System Sciences, 12(5), 1621-1632.

Amanti M., Troccoli A. & Vitale V. (2013). Pericolosità geomorfologica nel territorio di Roma Capitale. Analisi critica di due casi di studio: la Valle dell'Inferno e la Valle dell'Almone. Mem. Descr. Carta Geol. d'Italia, Vol. 93

AVI (1996) - Aree vulnerate italiane. A cura di Coordinamento della Protezione Civile al Gruppo Nazionale per la Difesa dalle Catastrofi Idrogeologiche (GNDCI) del Consiglio Nazionale delle Ricerche (CNR) http://avi.gndci.cnr.it

Bersani, P., & Bencivenga, M. (2001). Le piene del Tevere a Roma: dal 5. secolo a. C. all'anno 2000.

Beven, K. J., & Kirkby, M. J. (1979). A physically based, variable contributing area model of basin hydrology/Un modèle à base physique de zone d'appel variable de l'hydrologie du bassin versant. *Hydrological sciences journal*, *24*(1), 43-69.

Bozzano, F., Andreucci, A., Gaeta, M., & Salucci, R. (2000). A geological model of the buried Tiber River valley beneath the historical centre of Rome. Bulletin of Engineering Geology and the Environment, 59(1), 1-21.

Bozzano, F., Martino, S., & Priori, M. (2006). Natural and man-induced stress evolution of slopes: the Monte Mario hill in Rome. Environmental Geology, 50(4), 505-524.

Bozzano, F., Esposito, C., Franchi, S., Mazzanti, P., Perissin, D., Rocca, A., & Romano, E. (2015). Understanding the subsidence process of a quaternary plain by combining geological and hydrogeological modelling with satellite InSAR data: The Acque Albule Plain case study. Remote Sensing of Environment, 168, 219-238.

Bozzano, F., Esposito, C., Mazzanti, P., Patti, M., & Scancella, S. (2018). Imaging multi-age construction settlement behaviour by advanced SAR interferometry. Remote Sensing, 10(7), 1137.

Breiman, L. (2001). Random forests. Machine learning, 45(1), 5-32.

Campolunghi, M. P., Capelli, G., Funiciello, R., Lanzini, M., Mazza, R., Casacchia, R. 2008. Un caso esemplare: la stabilità degli edifici nell'area intorno a Viale Giustiniano Imperatore (Roma, IX Municipio). In: Memorie descrittive della Carta Geologica D'Italia. Volume L. La geologia di Roma dal centro alla periferia. Istituto Poligrafico dello Stato.

Cascini, L. (2008). Applicability of landslide susceptibility and hazard zoning at different scales. *Engineering Geology*, *102*(3-4), 164-177.

Caporossi, P., Mazzanti, P., & Bozzano, F. (2018). Digital image correlation (DIC) analysis of the 3 December 2013 Montescaglioso landslide (Basilicata, southern Italy): Results from a multi-dataset investigation. *ISPRS International Journal of Geo-Information*, *7*(9), 372.

Chang, Z., Huang, F., Huang, J., Jiang, S. H., Liu, Y., Meena, S. R., & Catani, F. (2023). An updating of landslide susceptibility prediction from the perspective of space and time. *Geoscience Frontiers*, 101619.

Ciotoli, G., Corazza, A., Finoia, M. G., Nisio, S., Serafini, R., & Succhiarelli, C. (2013). Sinkholes antropogenici nel territorio di Roma Capitale. I Sinkholes: metodologie di indagine, ricerca storica, sistemi di monitoraggio e tecniche di intervento. Centri abitati e processi di instabilità naturale: valutazione, controllo e mitigazione. Mem. Descr. Carta Geol. d'It, 93, 143-181.

Ciotoli G, Nisio S, Serafini R. 2015. Analisi della suscettibilita ai sinkholes antropogenici nel centro urbano di Roma: analisi previsionale. Mem Descr Carta Geol d'It. 99:167–188.

Ciotoli, G., Di Loreto, E., Finoia, M. G., Liperi, L., Meloni, F., Nisio, S., & Sericola, A. (2016). Sinkhole susceptibility, Lazio region, central Italy. Journal of Maps, 12(2), 287-294.

Corominas, J., van Westen, C., Frattini, P., Cascini, L., Malet, J. P., Fotopoulou, S., ... & Smith, J. T. (2014). Recommendations for the quantitative analysis of landslide risk. *Bulletin of engineering geology and the environment*, *73*(2), 209-263.

Crosetto, M.; Monserrat, O.; Devanthéry, N.; Cuevas-González, M.; Barra, A.; Crippa, B. Persistent scatterer interferometry using Sentinel-1 data. Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci. - ISPRS Arch. 2016, 41, 835–839, doi:10.5194/isprsarchives-XLI-B7-835-2016.

Delgado Blasco, J. M., Foumelis, M., Stewart, C., & Hooper, A. (2019). Measuring urban subsidence in the Rome metropolitan area (Italy) with Sentinel-1 SNAP-StaMPS persistent scatterer interferometry. Remote Sensing, 11(2), 129.

Di Salvo, C., Ciotoli, G., Pennica, F., & Cavinato, G. P. (2017). Pluvial flood hazard in the city of Rome (Italy). Journal of Maps, 13(2), 545-553.

Del Monte M., D'Orefice M., Luberti G.M., Marini R., Pica A., Vergari F. (2016) Geomorphological classification of urban landscapes: the case study of Rome (Italy). Journal of Maps, DOI: 10.1080/17445647.2016.1187977

Esposito, C., Belcecchi, N., Bozzano, F., Brunetti, A., Marmoni, G. M., Mazzanti, P., ... & Spizzirri, M. (2021). Integration of satellite-based A-DInSAR and geological modeling supporting the prevention from anthropogenic sinkholes: a case study in the urban area of Rome. Geomatics, Natural Hazards and Risk, 12(1), 2835-2864.

Esposito C., Marmoni G.M., Scarascia Mugnozza G., Argentieri A., Rotella G. (2019). A preliminary large-scale assessment of landslide susceptibility in the territory of the Metropolitan City of Rome (Italy) Proceedings of 2019 IPL Symposium on Landslides, Paris, France 16 – 19 September 2019. ISBN 978-4-9903382-5-1

Fang, Z., Wang, Y., van Westen, C. J., & Lombardo, L. (2022). Space-time landslide susceptibility modelling in Taiwan.

Fell, R., Corominas, J., Bonnard, C., Cascini, L., Leroi, E., & Savage, W. Z. (2008). on behalf of the JTC-1 Joint Technical Committee on Landslides and Engineered Slopes (2008) Guidelines for landslide susceptibility, hazard and risk zoning for land use planning. *Eng Geol*, *102*(3-4), 85-98.

Ferretti, A.; Prati, C.; Rocca, F. Permanent scatterers in SAR interferometry. IEEE Trans. Geosci. Remote Sens. 2001, 39, 8–20, doi:10.1109/36.898661.

Ferretti, A.; Prati, C.; Rocca, F. Nonlinear subsidence rate estimation using permanent scatterers in differential SAR interferometry. IEEE Trans. Geosci. Remote Sens. 2000, 38, 2202–2212, doi:10.1109/36.868878.

Fowler, H. J., Kilsby, C. G., & O'Connell, P. E. (2003). Modeling the impacts of climatic change and variability on the reliability, resilience, and vulnerability of a water resource system. Water resources research, 39(8).

Funiciello, R., & Giordano, G. (2008). La nuova carta geologica di Roma: litostratigrafia e organizzazione stratigrafica. La geologia di Roma dal centro storico alla periferia. Memorie

Descrittive della Carta Geologica d'Italia, 80(1), 39-85.

Geurts, P., Ernst, D. & Wehenkel, L. Extremely randomised trees. Mach Learn 63, 3–42 (2006). https://doi.org/10.1007/s10994-006-6226-1.

Glade, T. (2001). Landslide hazard assessment and historical landslide data—an inseparable couple?. In *The use of historical data in natural hazard assessments* (pp. 153-168). Springer, Dordrecht.

Guzzetti, F., Peruccacci, S., Rossi, M., Stark, C.P., 2007. Rainfall thresholds for the initiation of landslides in central and southern Europe. Meteorol. Atmos. Phys. 98, 239–267.

Hanssen, R.F. Satellite radar interferometry for deformation monitoring: A priori assessment of feasibility and accuracy. Int. J. Appl. Earth Obs. Geoinf. 2005, 6, 253–260, doi:10.1016/j.jag.2004.10.004.

Hesse, Ralf (2010). LiDAR-derived Local Relief Models - a new tool for archaeological prospection. Archaeological Prospection 17:67-72.

Hosking JRM, Wallis JR, Wood EF (1985) Estimation of the generalized extreme value distribution by the method of probability weighted moments. Technometrics 27:251– 261. https://doi.org/10.1080/00401706.1985.10488049

Iadanza, C.; Trigila, A.; Starace, P.; Dragoni, A.; Biondo, T.; Roccisano, M. (2021) IdroGEO: A Collaborative Web Mapping Application Based on REST API Services and Open Data on Landslides and Floods in Italy. ISPRS Int. J. Geo-Inf. 2021, 10, 89. https://doi.org/10.3390/ijgi10020089

IFFI (2007) - Inventario dei fenomeni franosi in italia realizzato dall'ISPRA e dalle Regioni e Provinc Autonome. http://www.isprambiente.gov.it/it/progetti/iffi-inventario-dei-fenomeni-franosi-in-italia

ISPRA. (2014). Progetto frane Roma. Inventario dei fenomeni franosi nel territorio di Roma Capitale. Italian National Institute for Environmental Protection and Research, (ISPRA). Retrieved September 17, 2015, from:http://sgi.isprambiente.it/franeroma/default.htm

Jenness, J. (2006). Topographic Position Index (tpi_jen. avx) extension for ArcView 3. x, v. 1.3 a. Jenness Enterprises.

Jenkinson AF (1955) The frequency distribution of the annual maximum (or minimum) values of meteorological events. Q J Royal Meteorol Soc 87:158–171. https://doi.org/10.1002/qj.49708134804

Julian, M., & Anthony, E. (1996). Aspects of landslide activity in the Mercantour Massif and the French Riviera, southeastern France. Geomorphology, 15(3-4), 275-289.

Kuhn, M., & Johnson, K. (2013). Applied predictive modeling (Vol. 26, p. 13). New York: Springer.

Kampes, B.M. Radar interferometry: Persistent scatterer technique; 2006; Vol. 12; ISBN 140204576X.

La Vigna, F., Mazza, R., Amanti, M., Di Salvo, C., Petitta, M., & Pizzino, L. (2015). The synthesis of decades of groundwater knowledge: the new Hydrogeological Map of Rome. Acque Sotterranee-Italian Journal of Groundwater, 4(4).

Loche, M., Alvioli, M., Marchesini, I., Bakka, H., & Lombardo, L. (2022). Landslide susceptibility maps of Italy: Lesson learnt from dealing with multiple landslide types and the uneven spatial distribution of the national inventory. Earth-Science Reviews, 104125.

Massonnet, D.; Feigl, KL Radar interferometry and its application to changes in the earth's surface. Rev. Geophys. 1998, 36, 441–500, doi:10.1029/97RG03139.

Mastrantoni, G., Caprari, P., Esposito, C., Marmoni, G. M., Mazzanti, P., & Bozzano, F. (2022). Data requirements and scientific efforts for reliable large-scale assessment of landslide hazard in urban areas (No. EGU22-4669). Copernicus Meetings.

Mastrantoni G.*, Marmoni G.M., Esposito C., Bozzano F., Scarascia Mugnozza G., Mazzanti P. (submitted) Reliability assessment of open-source multiscale landslide susceptibility maps and effects of their fusion. Georisk

PAI (2012) - Piano di Assetto Idrogeologico a cura dell'Autorità di bacino del Tevere.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... & Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. the Journal of machine Learning research, 12, 2825-2830.

PRG Roma (2007) - Carta di pericolosità e vulnerabilità geologica del territorio comunale. Piano Regolatore Generale del Comune di Roma a cura di Mordiagliani D.

Reichenbach, P., Rossi, M., Malamud, B. D., Mihir, M., & Guzzetti, F. (2018). A review of statistically-based landslide susceptibility models. *Earth-science reviews*, *180*, 60-91.

Remedia, G., Alessandroni, M. G., & Mangianti, F. (1998). Le piene eccezionali del fiume Tevere a Roma Ripetta. Università degli Studi dell'Aquila, Dip. di Ingegneria delle Strutture, delle Acque e del Terreno (DISAT n. 3).

Schilirò, L., Poueme Djueyep, G., Esposito, C., & Scarascia Mugnozza, G. (2019). The role of initial soil conditions in shallow landslide triggering: insights from physically based approaches. Geofluids, 2019.

Segoni, S., Leoni, L., Benedetti, A. I., Catani, F., Righini, G., Falorni, G., Gabellani, S., Rudari, R., Silvestro, F., and Rebora, N.: Towards a definition of a real-time forecasting network for rainfall induced shallow landslides, Nat. Hazards Earth Syst. Sci., 9, 2119–2133, https://doi.org/10.5194/nhess-9-2119-2009, 2009.

Serrano SE (2010) Hydrology for engineers, geologists, and environmental professionals: an integrated treatment of surface, subsurface, and contaminant hydrology, Second edn. Hydroscience Inc., USA, 590 pp

Sottili, G., Palladino, D. M., & Zanon, V. (2004). Plinian activity during the early eruptive history of the Sabatini Volcanic District, Central Italy. Journal of Volcanology and Geothermal Research, 135(4), 361-379.

Steger, S., Brenning, A., Bell, R., & Glade, T. (2016). The propagation of inventory-based positional errors into statistical landslide susceptibility models. Natural Hazards and Earth System Sciences, 16(12), 2729-2745.

Steger, S., Brenning, A., Bell, R., & Glade, T. (2017). The influence of systematically incomplete shallow landslide inventories on statistical susceptibility models and suggestions for improvements. *Landslides*, 14, 1767-1781.

Thiery, Y., Terrier, M., Colas, B., Fressard, M., Maquaire, O., Grandjean, G., & Gourdier, S. (2020). Improvement of landslide hazard assessments for regulatory zoning in France: STATE–OF–THE-ART perspectives and considerations. International journal of disaster risk reduction, 47, 101562.

Titti, G., van Westen, C., Borgatti, L., Pasuto, A., & Lombardo, L. (2021). When enough is really enough? On the minimum number of landslides to build reliable susceptibility models. *Geosciences*, *11*(11), 469.

Titti, G., Napoli, G. N., Conoscenti, C., & Lombardo, L. (2022). Cloud-based interactive susceptibility modeling of gully erosion in Google Earth Engine. *International Journal of Applied Earth Observation and Geoinformation*, 115, 103089.

Van Westen CJ, Van Asch TWJ, Soeters R (2005) Landslide hazard and risk zonation; why is it still so difficult? Bull Eng Geol Environ 65(2):167–184

Ventriglia, U. (1971). La geologia della città di Roma. Amm. Prov. Roma, 417 pp.

Ventriglia, U. (2002). Geologia del territorio del Comune di Roma. Amm. Prov. Roma, 810 pp.

Zieher, T., Perzl, F., Rössel, M., Rutzinger, M., Meißl, G., Markart, G., & Geitner, C. (2016). A multi-annual landslide inventory for the assessment of shallow landslide susceptibility–Two test cases in Vorarlberg, Austria. *Geomorphology*, *259*, 40-54.

CHAPTER 3

3. Reliability Assessment of Open-Source Multiscale Landslide Susceptibility Maps and Effects of Their Fusion

Authors

Giandomenico Mastrantoni[1], Gian Marco Marmoni[1], Carlo Esposito[1], Francesca Bozzano[1, 2], Gabriele Scarascia Mugnozza[1, 2], Paolo Mazzanti[1, 2].

[1] Dipartimento di Scienze della Terra e Centro di ricerca per i rischi geologici CERI, Sapienza Università di Roma, P.le Aldo Moro 5, 00185, Roma, Italy.

[2] NHAZCA S.r.l. start-up di Sapienza Università di Roma, P.le Aldo Moro 5, 00185, Roma, Italy.

Published in Georisk: Assessment and Management of Risk for Engineered Systems and Geohazards (2023)

Abstract

Several landslide susceptibility (LS) maps at various scales of analysis have been performed with specific zoning purposes and techniques. Supervised machine learning algorithms (ML) have become one of the most diffused techniques for landslide prediction, whose reliability is firmly based on the quality of input data. Site-specific landslide inventories are often more accurate and complete than national or worldwide databases. For these reasons, detailed landslide inventory and predisposing variables must be collected to derive reliable LS products. However, high-quality data are often rare, and risk managers must consider lower-resolution available products with no more than informative purposes. In this work, we compared different ML models to select the most accurate for large-scale LS assessment within the Municipality of Rome. The ExtraTreesClassifier outperformed the others reaching an average F1-score of 0.896. Thereafter, we addressed the reliability of open-source LS maps at different scales of analysis (global to regional) by means of statistical and spatial analysis. The obtained results shed light on the difference in hazard zoning depending on the scale and mapping unit. An approach for low-resolution LS data fusion was attempted, assessing the importance of the adopted criteria, which increased the ability to detect occurred landslides while maintaining precision.

Keywords: Landslide Susceptibility; Machine Learning; Map Scale; Data Fusion; Rome.

3.1. Introduction

Shallow landslides are slope instabilities which involve the most superficial deposits, mainly colluvium, rather than bedrock formations (Baeza and Corominas 2001; Bordoni et al. 2021). They are most frequently triggered by extreme rainfall and can be densely distributed across small catchments (Hungr, Leroueil, and Picarelli 2014).In urban areas, even in case of small-sized, shallow slope failures, their occurrence as spatially distributed clusters as frequent as their triggering rainfall event is capable of tearing the physical structure as well as the network of socio-economic, cultural, material, and immaterial relations that make up the life of cities (Iadanza et al. 2009; Salvati et al. 2010; Trigila, Iadanza, and Spizzichino 2010; Iadanza et al. 2013; Trigila, Iadanza, Munafò, et al. 2015). This is due to their rapid or extremely rapid movement and intrinsic damage potential both in terms of human and economic losses (Trigila, Iadanza, Esposito, et al. 2015). The recent expansion of urban areas entails a significant soil cover consumption and demands correct and sustainable urban planning to face off with the delineation of landslide-prone areas. Human pressure has modified geological, geomorphological, and hydrological features of original terrains through centuries, acting on topography, stratigraphy, and even geotechnical properties of the subsurface, by developing networks of superficial and underground structures, modifying or obliterating water streams (Luberti et al. 2018). Therefore, the importance of the analysis and evaluation of hazard sources for risk mitigation actions in urban areas is evident. Since the 1980s, the role of urban planning in the knowledge and prevention of landslides has expanded considerably. Nowadays, landslide hazard analysis is usually mandatory to approach proper land use planning and management (Mateos et al. 2020). Nevertheless, in some cases regulatory plans lack detailed thematic mapping of geohazard-related data (Cui et al. 2019) despite cities are frequently affected by landslides and exposed to high risk (Kiersch 1964; Crosta et al. 2005; Mazzanti and Bozzano 2011; Martino et al. 2019; Tonini et al. 2022).

Landslide hazard management can be developed at different spatial scales and with different methods with increasing degrees of sophistication and different zoning purposes (Fell et al. 2008), which can be informative or advisory up to statutory and regulatory or devoted to engineering design (Flentje et al. 2007; Cascini 2008; Corominas et al. 2013), whose reliability is strongly based on quality and completeness of the available landslide catalogues.

Several open-source landslide inventories exist worldwide; however, the amount and quality of available data may be inadequate to build accurate large-scale predictive models. Open-source landslide inventories may have incomplete spatial and temporal information and heterogeneous features, which affirms the need to integrate them to gain reliability (Mastrantoni et al. 2022). Alternatively, recent developments have shown the potential of SAR satellite imagery for multitemporal mapping of landslides (Bhuyan et al. 2023).

To achieve a proper prediction of landslide susceptibility (LS), landslide inventory data must be processed together with predisposing and preparatory factors to model hidden patterns, thus suggesting new potentially unstable slopes (van Westen, Castellanos, and Kuriakose 2008; Corominas et al. 2013; Vijith et al. 2014; Basu and Pal 2018). In the past decades, various approaches to determine the degree of LS have been proposed and applied by employing heuristic, data-driven and physically based methods for various types of landslides at different observation scales (Brabb 1984; Soeters and Van Westen 1996; Guzzetti et al. 1999; Fell et al. 2008; Ngadisih et al. 2014; Günther et al. 2014; Kainthura and Sharma 2022; Luo et al. 2022). Among data-driven approaches, multivariate analysis is one of the most sophisticated techniques for LS assessment. The spatial distribution of landslides is predicted through the estimation of the relationship between several independent predisposing factors and response variables, relying on the information on previous landslide and non-landslide samples (Persichillo et al. 2017; Su et al. 2022). A recent and important statistical improvement for LS studies is the advance in machine learning algorithms (ML). During the past two decades several models have been developed (e.g., random forest, support vector machines, artificial neural networks etc.). Although some methods performed better than others, no single method proved to be superior under all conditions (Reichenbach et al. 2018). Hence, it is not recommended to select the model in advance, but as many models as possible should be built and compared. Once obtained, the best-performing one is selected and used for landslide zoning purposes.

Since several LS maps at medium to small scale are available, understanding their potentiality and that of their integration for the prediction of stable and unstable areas may become relevant when detailed products are not yet available. Therefore, two main topics were investigated in the following: i) the achievement of a reliable urban-scale assessment of landslide susceptibility in the Municipality of Rome and, ii) the comparison with each available map and with the products derived by their fusion.

The exploiting of geospatial information through Machine Learning algorithms allowed us to estimate an urban-scale landslide susceptibility zoning (LSZ) for the Municipality of Rome, weighting and validating the different models adopted, selecting the most accurate one for the area of interest. The obtained result is then compared with open-source, regional, national, European, and global scale susceptibility maps to quantitatively assess their similarity and accuracy, and therefore the overall reliability. The fusion of LSZ products based on different analysis unit and resolution was addressed to evaluate the information gain or loss of the obtained output. By merging the LSZ maps we wanted to assess whether the adopted data fusion criteria can lead to a significant increase in accuracy and reliability of small-scale maps (i.e., global to national) generally available worldwide, thus potentially rising their applicability from informative to at least advisory.

3.2. Case Study

The study area corresponds to the Municipality of Rome, extending to about 1287 km^2. The area within the city of Rome has experienced a series of geological processes over the last hundreds of thousands of years that encompass volcanic activity and relative sea-level changes that have influenced the erosional and depositional stages of the lower part of the Tiber River catchment. Sedimentary complexes deposited during the Plio-Pleistocene are constituted by both marine and continental deposits (e.g., clays, sands, and conglomerates of alluvial deposits) outcropping the hills on the right bank of the Tiber River. They are superimposed by an alternating succession of volcanic products and continental sediments. Volcanic products erupted about 600k years ago by monogenic and polygenic complexes of the Alban Hills and the Sabatini volcanic district. Recent alluvial deposits of the Tiber River and its tributaries have filled the valley reaching thicknesses of tens of meters (Bozzano et al. 2000; Funiciello and Giordano 2008; Parotto 2008). The left and right embankments of the Tiber River valley have distinct geological units, thus resulting in different responses to natural and anthropogenic hazard processes, including subsidence (Bozzano et al. 2015), sinkholes (Ciotoli et al. 2014; Esposito et al. 2021), urban and fluvial floods, and landslides (Amanti, Chiessi, and Guarino 2012; Del Monte et al. 2016), that affected the area over the years.

Landslides in Rome are mainly represented by shallow landslides (soil slips and translational slides) with reduced and comparable volumes. Recently occurred landslides in Rome are characterized by an average area of 2.1x103 m^2, that due to the shallow position of the sliding surfaces (2 to 5 meters) correspond to an average volume of 5x103 m^3 as order of magnitude (Alessi et al. 2014; Del Monte et al. 2016). Additionally, ephemeral hydraulic circulation built up along permeability contrasts between volcanic, or debris covers overlying sedimentary deposits, may induce transient seeping that reduces slope stability. The Monte Mario and Monte Ciocci hills are the areas where most of the recorded landslides have occurred (Figure 3.1). Their proneness to landsliding might be related to high slope angles formed of the Monte Vaticano and Monte Mario Formations, which consist of a thick succession of clays, silty-clays, and silty-sands highly sensitive to shallow and translational landslides. The weathering of these deposits produced unsaturated shallow soil covers, which have been proven to host ground instabilities (Schilirò et al. 2019).

According to Esposito et al. (2023), the city of Rome has experienced at least 566 landslide events over the last century, 356 of which are reported as translational, flow or shallow landslides (Figure 3.1b, c). These landslides have almost always been associated with heavy rainfall, among which sixty-seven occurred during the exceptional rainfall event registered from January 31 to February 2, 2014 (Alessi et al. 2014). This event and the landslide related damages declare Rome's susceptibility to geological hazards, which can interact with socio-economic activities, posing relevant risk conditions, and causing a significant impact, as documented by the service and road infrastructure

interruption after the event, whose consequences lasted for months. Although landslide hazard in Rome, as well as in other major cities in Italy, is widely known, it still lacks urban-scale regulatory or advisory landslide susceptibility zonation, except for national scale susceptibility or hazard indicators (Trigila et al. 2013; Iadanza et al. 2021).

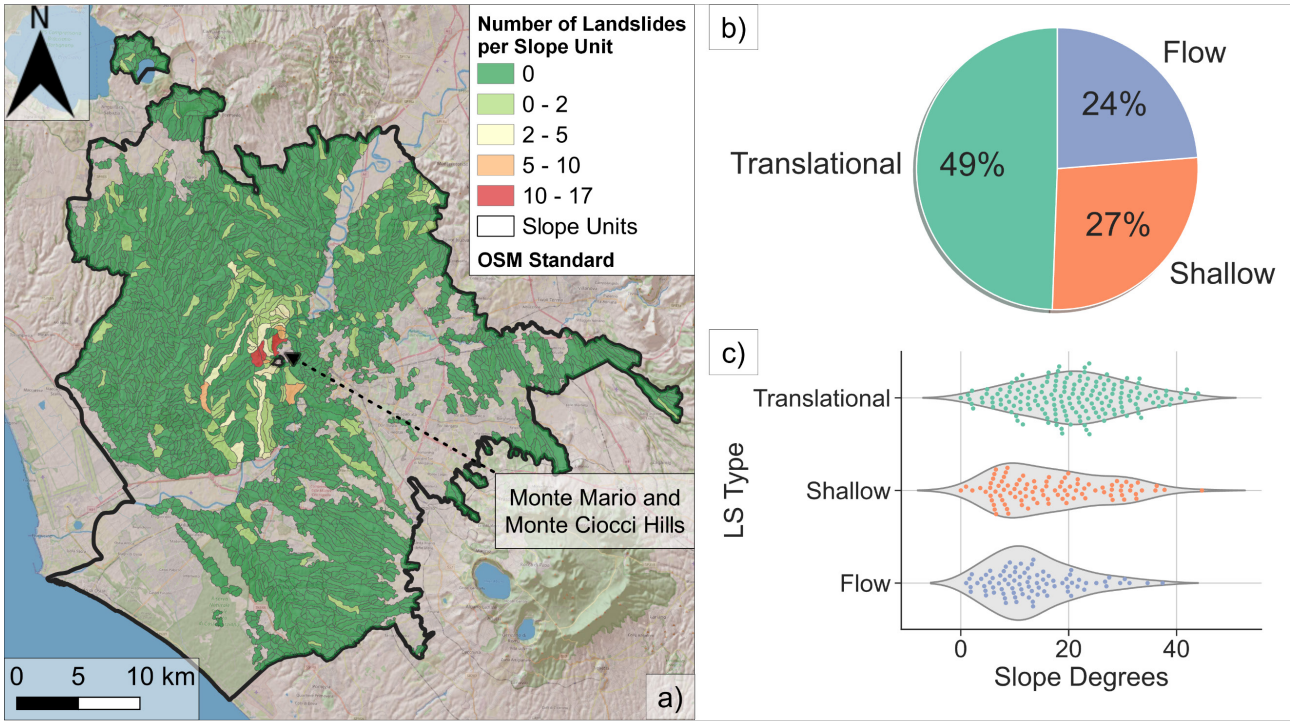


Figure 3.1: Number of landslide events per geomorphological slope unit (Alvioli, Guzzetti, and Marchesini 2020) within the Municipality of Rome (a); percentage of landslides by type of failure, as defined by Hungr, Leroueil, and Picarelli, 2014 (b); slope angle distribution of landslide events by type (c). Slope units are shown for non-flat areas only.

3.3. Material And Methods

3.3.1. Urban-scale assessment of landslide susceptibility

Analysis based on ML methods needs reliable data on which the model is based. The main prerequisite is information on the spatial occurrence of landslide events (van Westen, Castellanos, and Kuriakose 2008) together with conditioning factors picked because of their expected relationship with the occurrence of slope failures. Although landslide inventories are often incomplete in spatial and temporal terms, integration, cross-validation, and processing of open-source databases produced at different scales of observation are required to reduce bias before training the predictive models to be compared. In this study, the standardised and boosted landslide database of Rome was used to train different supervised ML algorithms.

The database contains both Landslide Initiation Points (LIP) with the related polygons representing the area involved. The best performing model was then employed to estimate the LS level for the

whole area, thus resulting in the urban-scale LSZ which in this study is referred to as “URGENT Rome”. It is developed within the PRIN project entitle “URGENT - URban Geology and geohazards: Engineering geology for safer, resilieNt and smart ciTies”.

Figure 3.2a describes the pipeline implemented in this study to achieve the urban-scale shallow landslide susceptibility map of Rome.

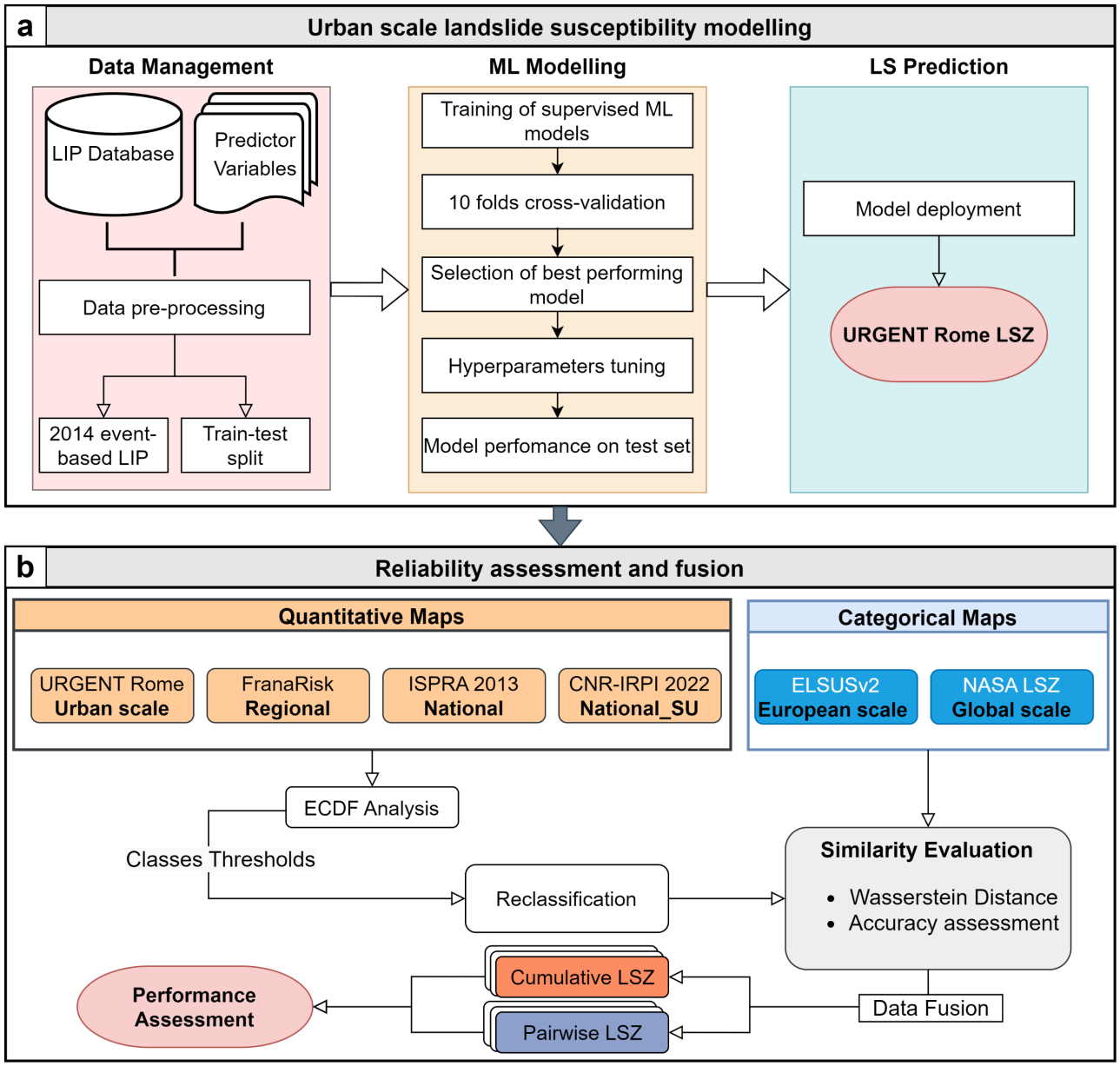


Figure 3.2: Workflow developed to obtain an urban-scale landslide susceptibility zoning of Rome territory (a), and reliability assessment of open-source LS maps at different scales of observation (b).

3.3.1.1. Data pre-processing

All ML algorithms use some input data to create outputs. Input data comprise features, which are usually in the form of structured columns. Algorithms require features with some specific characteristics to work properly. Here, the need to feature engineering arises. Feature engineering was performed with two main goals: 1) preparing the proper input dataset, compatible with the ML algorithm requirements (i.e., tidy structure), and 2) improving the performance of the models.

Data about terrain, lithological, hydrological, soil, roads and streams were selected as potential predictor variables (Table 3.1). A total of 20 variables are included, of which 14 represent terrain features derived from the 5m resolution DTM. Distances from roads, hydraulic permeability limits (i.e., hydrogeological units that cause significant contrast in hydraulic conductivity) and water streams were also computed. The remaining are categorical variables representing litho-technical units and urban land characteristics.

We derived litho-technical units from the official lithological map of Rome by grouping geological formations with similar geotechnical characteristics. We did not include the individual formations as a predictor variable due to the large number of categories, which would result in a high-cardinality feature. According to Maxwell et al. (2020), this allows embedding expert knowledge into the prediction. We also included land cover and soil consumption information to account for urban assets and activities. Thereafter we encoded categorical variables by replacing the original value of the feature with the frequency distribution, thus converting the feature into a numeric. Since we are dealing with rainfall-triggered shallow landslides, we took advantage of 33 distributed rain gauge records (from 1951 to 2021) to derive Thiessen polygons and their amount of ordinary annual rainfall.

All raster-based variables were then extracted at the mapped LIP and stable point locations using the Point sampling tool plugin (Jurgiel 2013) of QGIS 3 to generate tables from which to extract train, validation, and test data.

Table 3.1: List of variables deemed predictors of landslide susceptibility in Rome.

Variable	Abbreviation	Description
Digital Terrain Model	DTM	5x5 m DTM model derived by points and contour lines of elevation with Anudem algorithm (Hutchinson, Xu, and Stein 2011).
Slope Angle	slope	5x5 m slope angle map
Relative Relief	rrelief100	Difference between max and min altitude within a cell of 100 m (Oguchi 1997)
Topographic Position Index	tpi	TPI with a neighbourhood equal to 300 m
Dissection Ratio	dess_ratio	Topographic dissection within 100 m grid.
Drop	drop	Relative height within a radius of 50 m.
Aspect	aspect	Topographic aspect as degrees from north.
Aspect classes	aspect_8	Topographic aspect reclassified in 8 dials.
Curvatures	curv_plan, curv_prof, curv_tot	Planar, parallel, and total curvature based on 5x5 m DTM
Flow Accumulation	flow_acc	Amount of upstream area draining into a cell.

Topographic Wetness Index	twi	Steady state wetness index map.
Ordinary annual rain	ord_rain	Based on rainfall analysis of 33 distributed rain gauges records (from 1951 to 2021)
Soil thickness	soil_thickness	Soil thickness based on Saulnier's equation (Saulnier, Beven, and Obled 1997)
Land use	land_use	Regione Lazio land cover map https://dati.lazio.it/catalog/it/dataset/carta-uso-suolo-2016
Litho-technical units	litho	Based on 1:25000 lithological map of the Regione Lazio https://dati.lazio.it/catalog/it/dataset/carta-geologica-informatizzata-regione-lazio-25000
Distance to streams	dist_idro	Euclidean distance to nearest stream
Distance to roads	dist_road	Euclidean distance to nearest road.
Distance to permeability limit	dist_lim_k	Based on hydrogeological complex defined in the Hydrogeological map of Rome (La Vigna et al. 2015).

Input data consist of LIPs and stable points with a ratio of 35-65%, for an overall amount of 2992 records stored as geo-dataframe. The few dated landslide records (i.e., those triggered by the heavy rainfall of January 31, 2014) were formerly split from the database and treated as event-based test samples to assess the capability of the six LSZs to correctly detect landslide occurrences.

The importance of features in ML algorithms plays a very crucial role in prediction analysis in any field. First, we explored the data to select a subset of input features that would be relevant to the prediction, thus avoiding the issue of dimensionality, as some algorithms perform badly when in high dimensionality. Correlations between features, or multicollinearities, were calculated and variables were systematically removed if a pairwise correlation exceed 75%, as recommended by Kuhn and Johnson (2013). The algorithm used to filter out multicollinearity calculated a correlation matrix and the highest pairwise correlation was found (Figure 3.3). The variable within this pair with the lower correlation with the target variable was removed from the dataset. This was repeated until no pairwise correlation exceeded 75%. Sixteen out of 20 variables remained to be used in model training after verifying that there was no missing or redundant information.

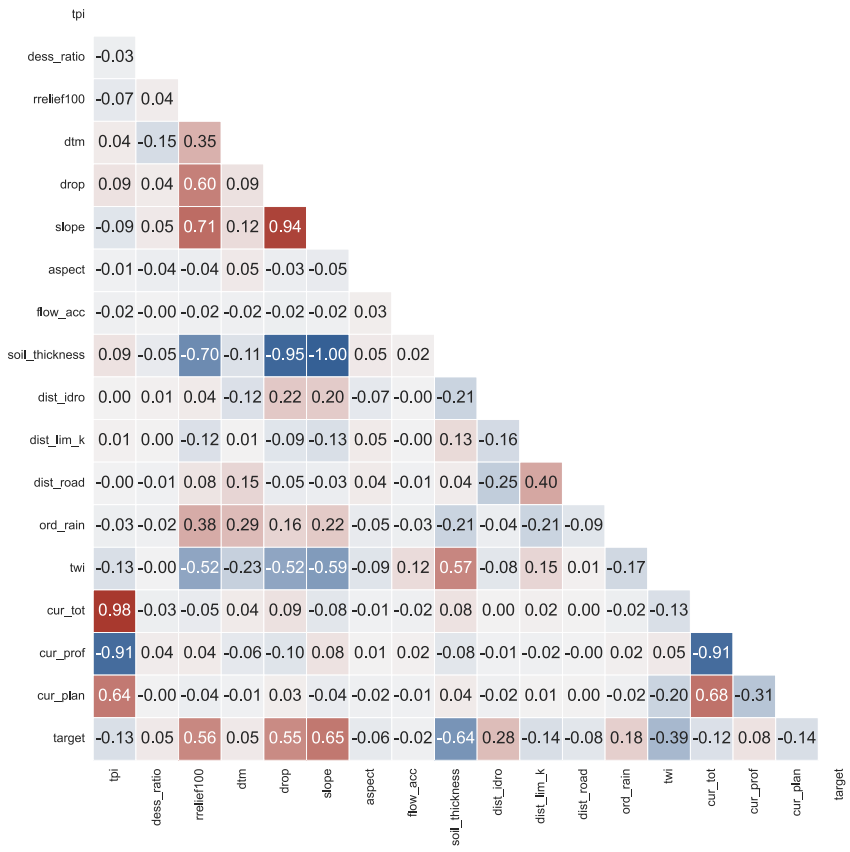


Figure 3.3: Pairwise correlation of all continuous numerical features originally included in the study.

3.3.1.2. Machine Learning modelling

Candidate models include Naïve Bayes, logistic regression, k-nearest neighbours, and decision trees-based methods. The complete list of trained models can be found in Table 3.2. The data were split into two subsets. The former being the training set containing 80% of the data, whereas the latter being the hold-out set containing the remaining 20%. Stratified random sampling was used to account for the imbalanced binary outcome (Kotsiantis, Kanellopoulos, and Pintelas 2006). K-fold cross validation was performed on the training set to train and validate the model with ten subsets validation data.

Cross-validation results were then used to compute ten confusion matrices for each model. Given a confusion matrix (Figure 3.4), the distribution of the overall accuracy (Eq. 3.1), recall (Eq. 3.2), precision (Eq. 3.3), F1-Score (Eq. 3.4) and ROC-AUC (Fawcett 2006) was derived. Score metrics were then used to select the best performing model to be further optimised with hyperparameters tuning. Since F1-score considers not only the number of prediction errors that the model makes, but also looks at the type of errors that are made, it works well on imbalanced data. For this reason, we chose the F1-score to select the most suitable model to predict susceptibility levels in the whole area. GridSearchCV function of the Scikit-learn library (Pedregosa et al. 2011) was exploited to determine the optimal set of hyperparameter values, thus minimizing both variance and bias. To

better understand and explain how the model works, we implemented the permutation feature importance method to assess feature relevance avoiding bias from categorical variables (Altmann et al. 2010).

Table 3.2: Caption.

ML model	Abbreviation	Type of algorithm
Extra Trees Classifier	ET	Ensemble of decision trees
Light Gradient Boosted Machine	LightGBM	Ensemble of decision trees
Random Forest	RF	Ensemble of decision trees
Gradient Boosting Classifier	GBC	Ensemble of decision trees
Ada Boost Classifier	ADA	Ensemble of decision trees
Decision Tree	DT	Single tree structure
Logistic Regression	LR	Sigmoid function
Naïve Bayes	NB	Bayes theorem
Linear Discriminant Analysis	LDA	Linear decision boundary
K Nearest Neighbours	KNN	k-nearest neighbours vote

Once tested, the trained model was applied to predict the probability of each raster cell being a landslide initiation point, thus resulting in the LS map with values ranging from 0 to 1, representing minimum and maximum susceptibility respectively.

All the machine learning model development and deployment were coded in Python 3 taking advantage of Scikit-learn and PyCaret libraries among others (Pedregosa et al. 2011; Ali 2020).

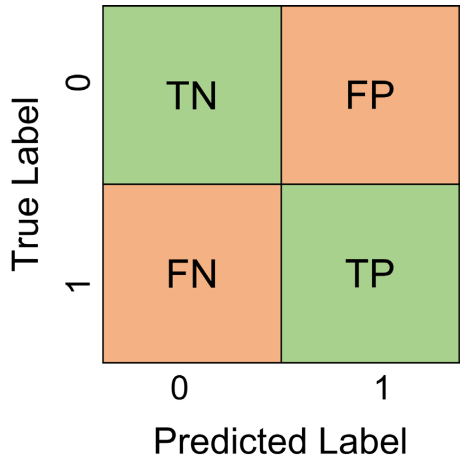


Figure 3.4: Confusion matrix for binary classification where TP and TN denote the number of positive and negative examples that are classified correctly, while FN and FP denote the number of misclassified positive and negative examples respectively.

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN}, \quad (3.1)$$

$$Recall = \frac{TP}{TP + FN}, \quad (3.2)$$

$$Precision = \frac{TP}{TP + FP}, \quad (3.3)$$

$$F1 - score = 2 \times \frac{Recall \times Precision}{Recall + Precision}, \quad (3.4)$$

3.3.2. Accuracy assessment and data fusion

This section focuses on evaluating and merging the available open-source LS maps originally produced at different scales of analysis to boost their performance. To this aim, the assessment of both similarity and accuracy of open-source LSZ with reference to the urban-scale map produced in this study was required. According to Cascini (2008), different scales of analysis denote different purposes. However, regional- to urban- scale susceptibility maps are often lacking. Therefore, there is a need to assess the reliability and potentials of small- to medium- scale LSZ in case more detailed studies have not yet been conducted. For that purpose, five LS maps at different scales and mapping units, including regional, national, European, and global scale products, were collected and compared with the one produced in this study. Figure 3.2b illustrates maps and workflow implemented to carry out the performance assessment and to achieve multiscale-based LSZ products.

FranaRisk LSZ (Argentieri et al. 2018) covers the entire province of Rome with a cell size of 20 m. ISPRA 2013 (Trigila et al. 2013) and CNR-IRPI 2022 shallow LSZ (Loche et al. 2022) both cover the whole Italian territory but adopting different mapping units. The first is grid-based (500 m resolution) while the latter relies on geomorphological slope units (hereafter named SU). ELSUSv2 represents the official European LSZ (Wilde et al. 2018); it has a pixel size of 200 m. Lastly, NASA LSZ covers the entire globe with a grid size of 1 km (Stanley and Kirschbaum 2017).

Out of the six LS maps, two are already categorized in five classes (i.e., NASA, ELSUSv2), while the others have continuous numerical susceptibility values. Regarding the ISPRA 2013 LS map, we rescaled it in the range of 0-1 to make it comparable. The vector-based map utilizing SU was rasterized to the resolution of URGENT Rome LSZ (5m grid cell). Among the methods reported by the literature to select class thresholds (Baeza, Lantada, and Amorim 2016), we decided to focus on the capacity of the model to detect true landslides (i.e., true positives). Empirical cumulative distribution functions (ECDF) were computed specifically from the predictions made on the event-based test samples as well as on the overall landslides polygons stored in the database. Class

thresholds were defined by detection rates of 10%, 15%, 25% and 50% to classify the probabilities as ‘very low’, ‘low’, ‘moderate’, ‘high’, and ‘very high’ landslide susceptibility. The detection rate percentiles were chosen following the same values adopted for ELSUSv2 Pan-European LSZ (Wilde et al. 2018). This step was necessary to standardize the LSZs and make them quantitatively comparable.

Following the LSZ reclassification, a similarity evaluation process was implemented by means of Wasserstein distance metric together with spatialized accuracy assessment. We exploited recorded landslides polygons to evaluate the ability of each LSZ to predict landslide hazardous areas by quantifying the difference between the ECDF. Furthermore, we also compared the distributions of the event-based test samples, i.e., the January 2014 rainfall-triggered landslides, related to non-ordinary conditions (Alessi et al. 2014). This was achieved by means of the Wasserstein distance metric (Piccoli and Rossi 2016). The Wasserstein Distance represents the area between the curves of the cumulative distribution of the two groups. It tells how much, on average, you should move each point of one group to get the other group, while maintaining the quantile of the points; for categorical features, it expresses the difference in classes between the tested LSZ map and a benchmark, here assumed in the URGENT Rome, which is the most detailed and updated one.

Even though these statistical approaches allow us to determine quantitatively which is the closest LS map, spatial data also needs spatial analysis to reveal trends and patterns, thus visualizing the spatial location of differences and similarities. To do so, we achieved a spatialized accuracy assessment for each collected LSZ with the benchmark map (i.e., URGENT Rome). For this purpose, the respective excerpts of the categorized LSZ were binarized by merging the two highest susceptibility classes (“high” and “very high”) as susceptible areas with the rest as non-susceptible terrain in each case (Günther et al. 2014). Overlying binarized maps allow incorrectly classified terrains to be derived in terms of false positives and false negatives. This results in a new raster map for each comparison showing true/false positives and true/false negatives in the space, i.e., an accurate or inaccurate prediction. We also report the overall confusion matrix together with metrics scores, thus allowing a performance analysis of each open-source LSZ compared to URGENT Rome. The analysis ended up with an F1-score value for each LSZ which has been considered an indicator of its reliability in Rome. F1-scores were then used as weighting parameters when fusing the LS maps. Two approaches were tested (Figure 3.2b) with the aim to evaluate the improvement of predictive performance by data fusion approaches: firstly, a cumulative fusion of all LSZ was implemented from smallest to largest scale, assessing the performance at each step; then, a pairwise fusion was performed between the two maps built at closer scale. Eq. 3.5 represents the adopted criteria to merge the maps, rounding decimals to the ceil value. Where LSI is the fused landslide susceptibility index; F1 and LS denote the F1-score and landslide susceptibility class relative to the i-th LSZ, respectively.

$$LSI = \frac{1}{\sum_i F1_i} \times \sum_i (F1_i \times LS_i), \quad (3.5)$$

3.4. Results

3.4.1. Urban-scale LSZ

F1 scores calculated from the validation folds of cross-validation phase for each separate model are provided in Figure 3.5. Table 3.3 represents performance scores based on the 10 independent validation sets of candidate models. The best four models (i.e., ET, LightGBM, RF, and GBC) are characterised by a distribution of F1 scores always greater than 0.8, with the ET model standing out better in the mean and maximum value, and throughout the overall distribution (i.e., lower variance). The other models have lower scores both in terms of mean values and whisker length, with KNN being the worst. By considering also other metrics the ET model remains the best among candidates.

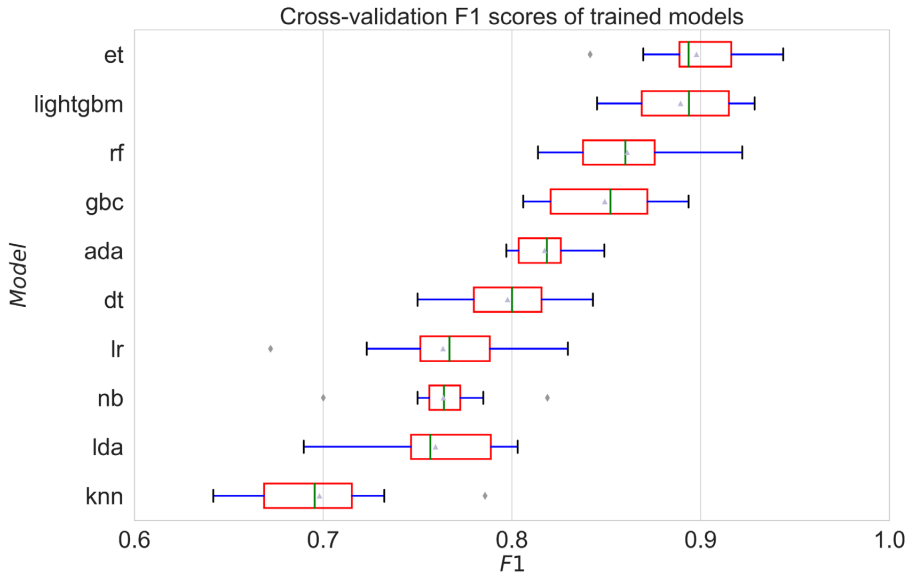


Figure 3.5: Distribution of cross-validation F1 scores obtained during model training.

Table 3.3: The mean and standard deviation (SD) of scoring metrics resulting from K-fold cross validation analysis for candidate models. Data are sorted by F1-score.

Model	Accuracy		AUC		Recall		Precision		F1	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
et	0.933	0.018	0.979	0.007	0.906	0.034	0.886	0.034	0.896	0.028
lightgbm	0.928	0.019	0.977	0.006	0.903	0.033	0.876	0.043	0.889	0.029
rf	0.915	0.019	0.970	0.009	0.875	0.032	0.861	0.034	0.868	0.029
gbc	0.904	0.017	0.964	0.007	0.856	0.046	0.843	0.027	0.849	0.042
ada	0.088	0.009	0.953	0.008	0.836	0.030	0.810	0.025	0.817	0.014
dt	0.869	0.019	0.854	0.019	0.814	0.034	0.783	0.041	0.797	0.026
lr	0.857	0.020	0.931	0.013	0.732	0.059	0.799	0.031	0.661	0.054
nb	0.828	0.022	0.907	0.021	0.873	0.030	0.679	0.036	0.763	0.028
lda	0.859	0.014	0.935	0.011	0.705	0.055	0.826	0.030	0.759	0.032
knn	0.808	0.025	0.879	0.021	0.699	0.052	0.698	0.046	0.558	0.057

Figure 3.6 shows an instance of ET decision regions pre- and post-optimisation by finding the best combination of hyperparameter values to minimize both variance and bias. The pre-optimization decision regions (Figure 3.6a) highlight some overfitting represented by small regions that split only a few isolated samples of one class embedded in the other. Those erroneous regions suggest overfitting and need to be removed to allow the model to better generalize to new data. Among other hyperparams (e.g., max depth, number of estimators, etc.), the tuned ET model was set to not split below a minimum number of samples, thus removing those isolated regions as shown in Figure 3.6b.

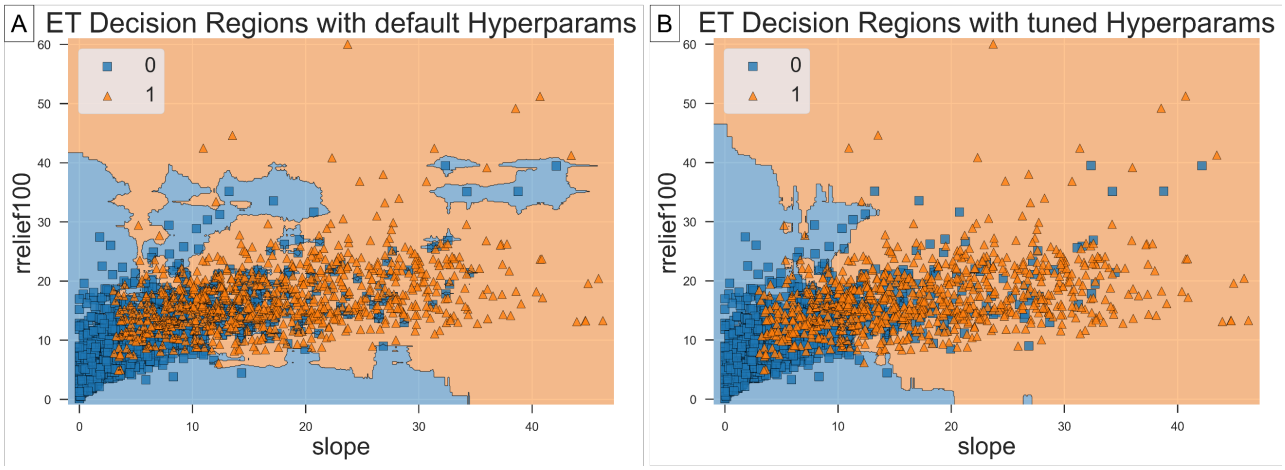


Figure 3.6: Decision regions between slope angle and relative relief computed by the ExtraTreesClassifier before (A) and after (B) the optimization process achieved with hyperparameters tuning, which returned optimal parameter values, including number of trees, minimum number of samples to split and maximum depth equal to 60, 2, and 25 respectively.

To understand how the model works, the importance of each variable was sorted from the most to least predictive (Figure 3.7). The higher the slope and relative relief the higher the landslide susceptibility, as also shown by the decision regions plot. Land use results as the third most important variable. Ordinary annual rainfall, distance to streams, litho-technical units, distance to permeability limits as well as TWI follow respectively. It should be noted the high consistency of variable importance given the slight standard deviation displayed. Permutation features importance results also denote the positive contribution of almost all the predictor variables to model performance, thus indicating that the selected variables should be kept. Flow accumulation is the only feature that appears irrelevant to the model.

To represent the ability of the model to generalise to never seen data, the confusion matrix together with kernel density of probabilities (Figure 3.8) have been computed on the 20% of the data we held before creating the model (i.e., out-of-sample test set). The tabular confusion matrix denotes the number of predictions made by the model where it classified the classes correctly or incorrectly. Using a probability threshold of 0.5, we have almost 90% of correctly classified data for both 0s (i.e., stable points) and 1s (i.e., LIP). The reliability of the model is also certified by the overall precision, recall and F1-score which have values of 0.88 or 0.89. The kernel density plot represents the

distribution of predicted probabilities for the withheld test data. This plot suggests a strong separation between LIP samples and stable points data. LIP samples are almost never getting a probability below 0.4. The median probabilities stand close to 0 and 0.9 for stable and LIP samples respectively (Figure 3.8).

The resulting prediction across the whole mapped extent and some example areas at a larger scale is provided in Figure 3.9. Red areas are those that are predicted to have a high likelihood of slope failure occurrence while green areas are predicted as having a low likelihood.

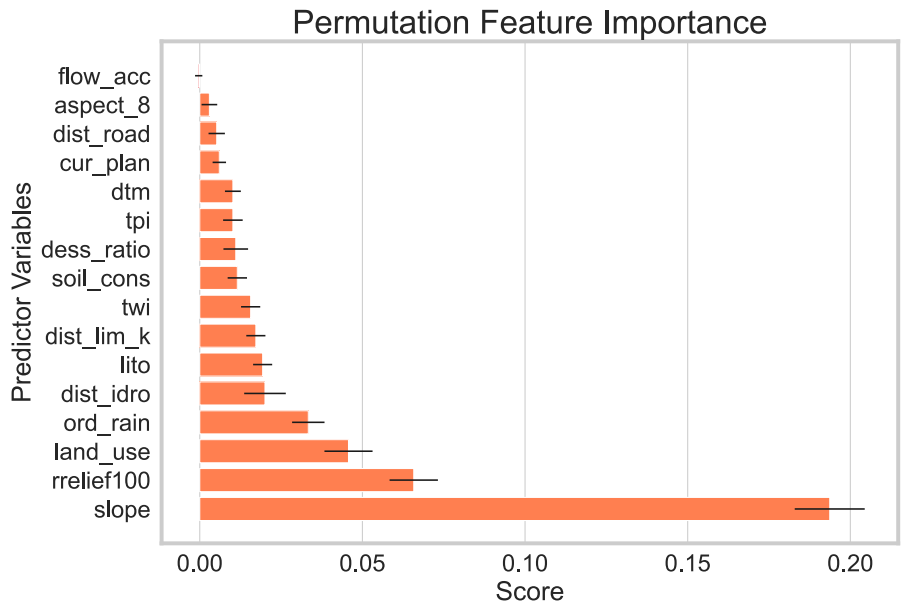


Figure 3.7: Average permutation feature importance with standard deviations derived by repeating the shuffling process 15 times.

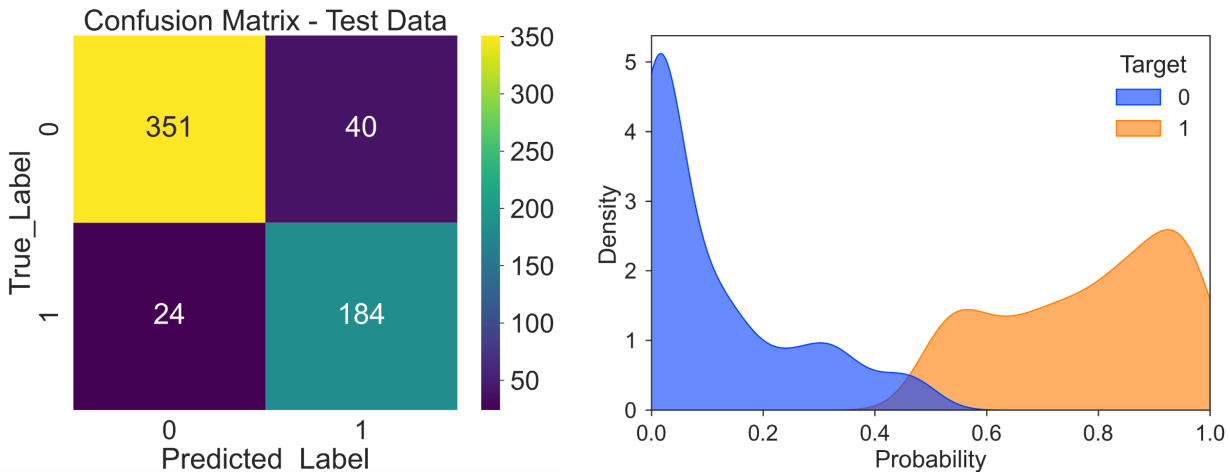


Figure 3.8: Confusion matrix plot (left) and classification report (right) of predicted labels for the test dataset. Kernel density plot of predicted probabilities for the test samples.

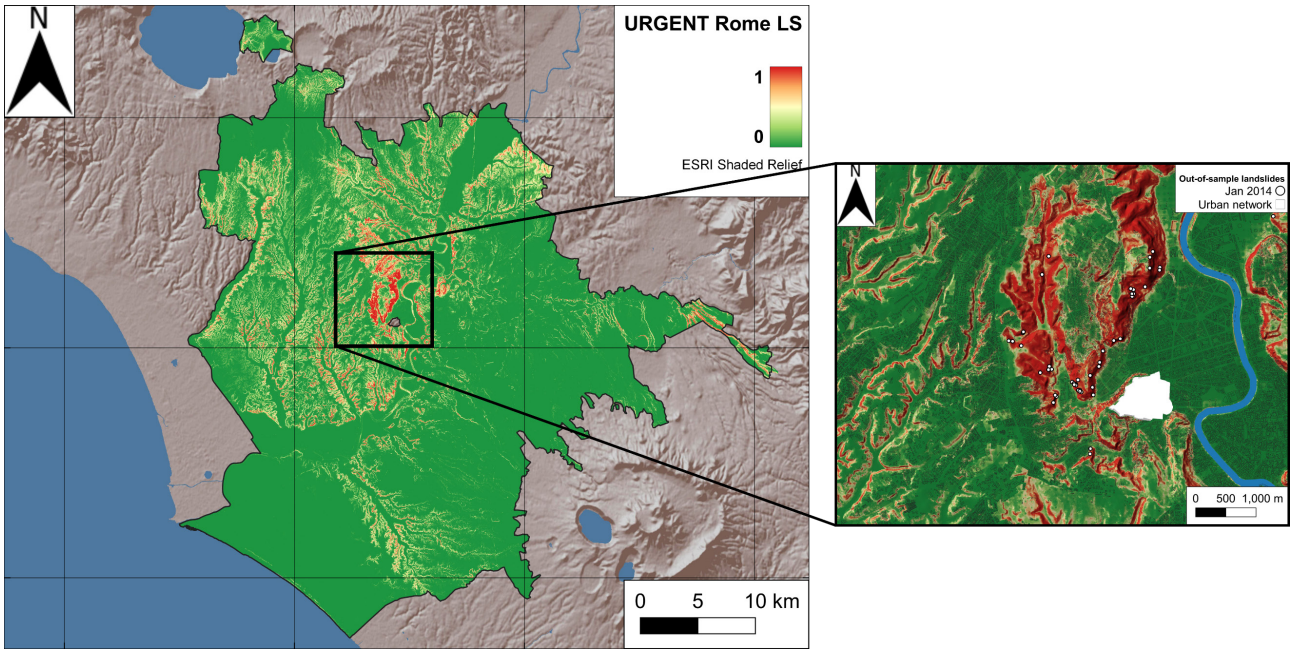


Figure 3.9: Urban-scale landslide susceptibility of Rome (namely URGENT Rome) derived with the tuned Extra Trees Classifier (left). A detailed example of Monte Mario hill with superimposed the event-based test samples (right).

3.4.2. Multiscale LSZ reliability assessment

Figure 3.10 shows the empirical cumulative distribution of susceptibility values in the overall landslide polygons for each quantitative LSZ. The distributions appear significantly different among all scales of analysis. The national scale maps (i.e., ISPRA 2013 and CNR-IRPI 2022) have a totality of values below 0.2. It should be noted that the distribution of URGENT Rome predictions is typically greater than the others, with FranaRisk close behind. The FranaRisk LSZ performs better than others until the 15th percentile and then decreases with respect to the URGENT Rome values.

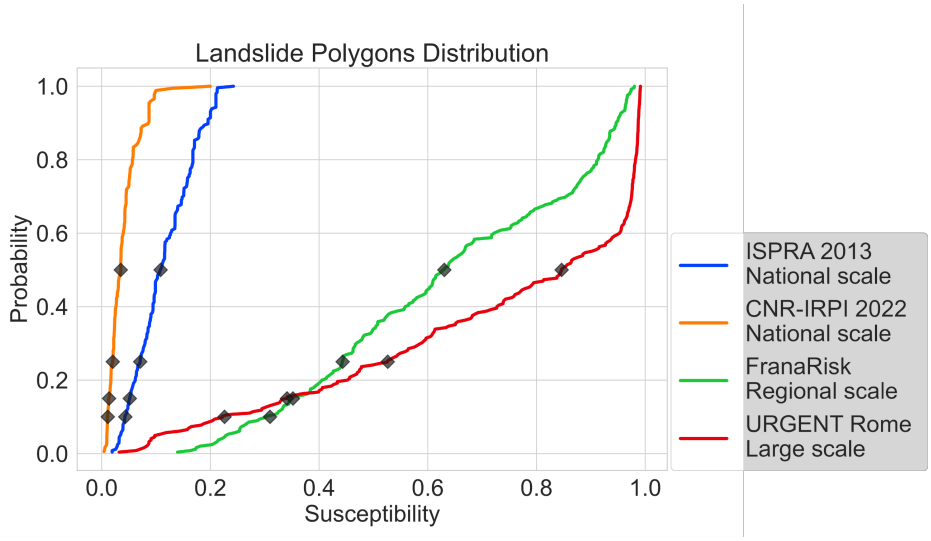


Figure 3.10: ECDF of mean susceptibilities predicted to landslides stored in the database as polygons. Black dots represent the percentile thresholds employed to derive the five classes.

Once reclassified, similarity analysis of the pair of LSZ for the event-based test samples LIP (Figure 3.11) and the overall landslide polygons (Figure 3.12) allowed us to compute a mean distance in susceptibility class distribution between the existing maps and URGENT Rome LSZ (i.e., the Wasserstein distance). Regarding the event-based test samples, the NASA Global LSZ as well as the ELSUSv2 Pan-European map are the furthest from URGENT Rome with an average distance of 1.84 and 1.82 classes respectively. CNR-IRPI 2022 and ISPRA 2013 maps gained much more predictive capacity as reported by the Wasserstein distance results of 0.33 and 1.12. FranaRisk confirms to be the closest to URGENT Rome with only 0.12 classes of distance.

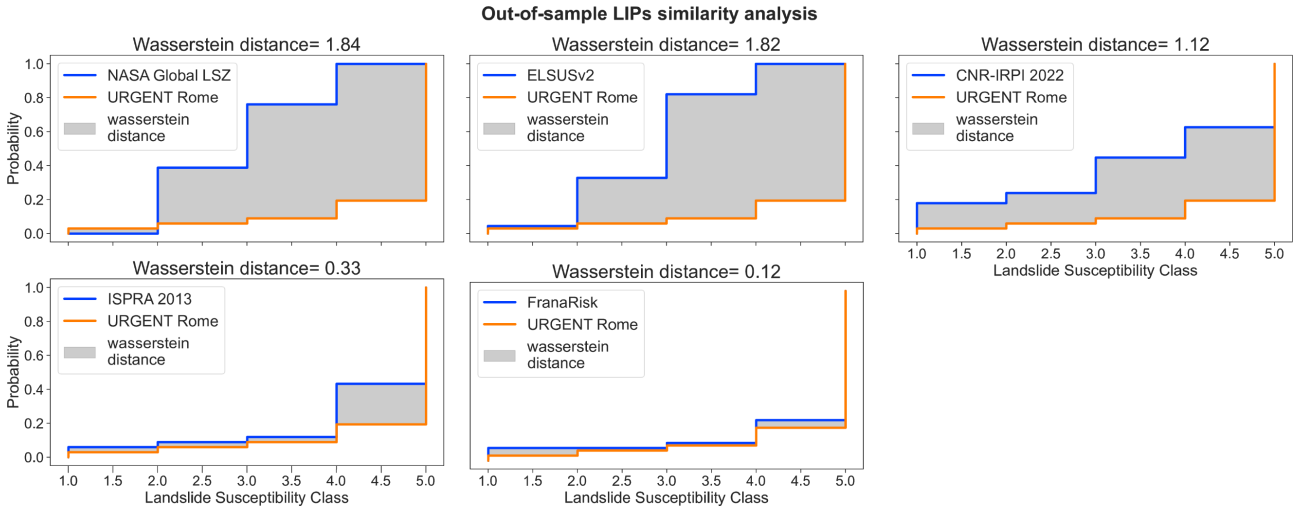


Figure 3.11: Wasserstein distance as the area between cumulative distributions of 2014 event-based LIPs as predicted by LSZ. The Wasserstein Distance of categorical features expresses the difference in landslide susceptibility classes between the tested map and the benchmark (i.e., URGENT Rome).

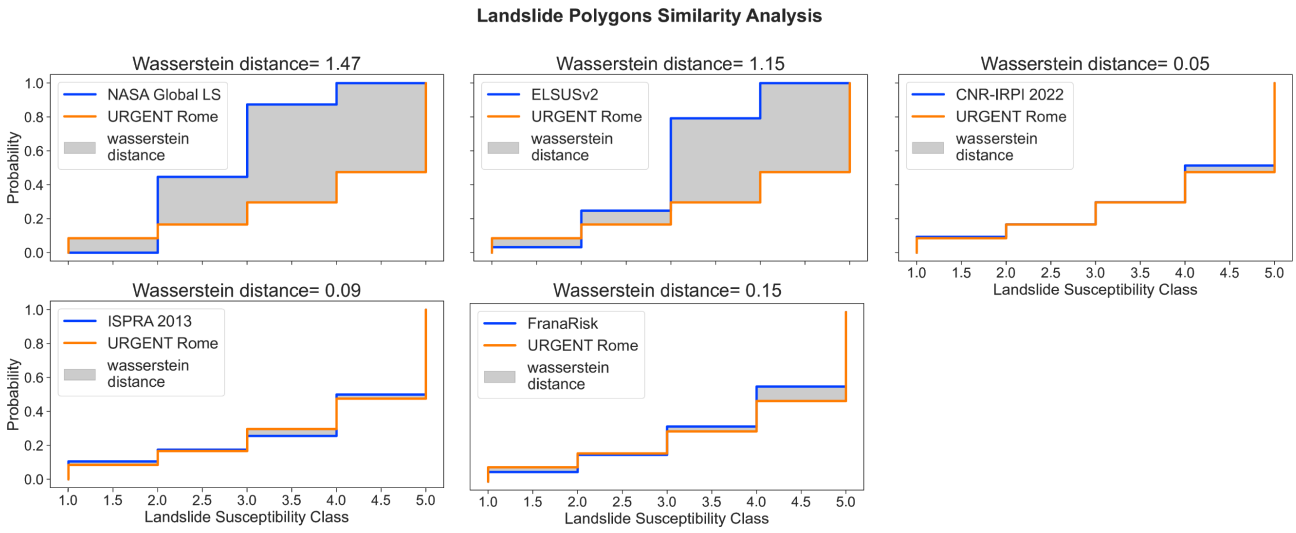
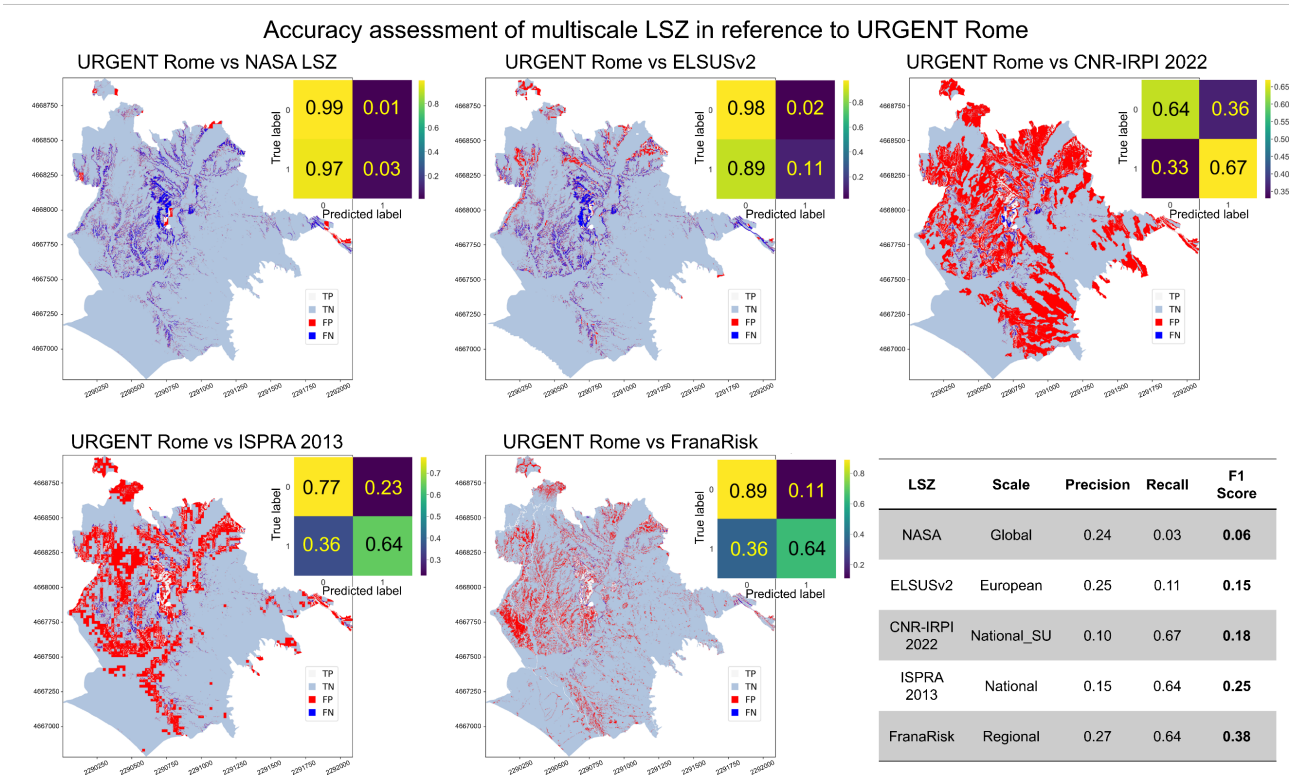


Figure 3.12: Wasserstein distance as the area between cumulative distributions of the overall landslides as predicted by LSZ. The Wasserstein Distance of categorical features expresses the difference in landslide susceptibility classes between the tested map and the benchmark (i.e., URGENT Rome).

As regards landslide polygons, the NASA Global LS map and ELSUSv2 still remain the furthest, while ISPRA 2013 and CNR-IRPI 2022 got even closer than FranaRisk, with the CNR-IRPI 2022 being almost identical (0.05 classes).

Spatialized accuracy assessment confirms the low capacity of LSZ at European and global scales to detect true landslides since 97% and 89% of false negatives were recorded by NASA LSZ and ELSUSv2 respectively (Figure 3.13). The national scale LSZ (i.e., ISPRA 2013 and CNR-IRPI 2022) improve the ability to detect true landslides. However, they report a high landslide susceptibility level even in areas not prone to landslides (i.e., false positives). Whereas FranaRisk LSZ keeps the predictive capabilities for true landslides but improves the detection of true negatives compared to the nationwide maps. As a result, the spatial accuracy assessment returned scoring metrics (e.g., F1-score) that increase proportionally with the scale of the LSZ (Figure 3.14).



LSZ	Scale	Precision	Recall	F1 Score
NASA	Global	0.24	0.03	**0.06**
ELSUSv2	European	0.25	0.11	**0.15**
CNR-IRPI 2022	National_SU	0.10	0.67	**0.18**
ISPRA 2013	National	0.15	0.64	**0.25**
FranaRisk	Regional	0.27	0.64	**0.38**

Figure 3.13: Spatialized performance assessment of each open-source LSZ with reference to URGENT Rome LSZ.

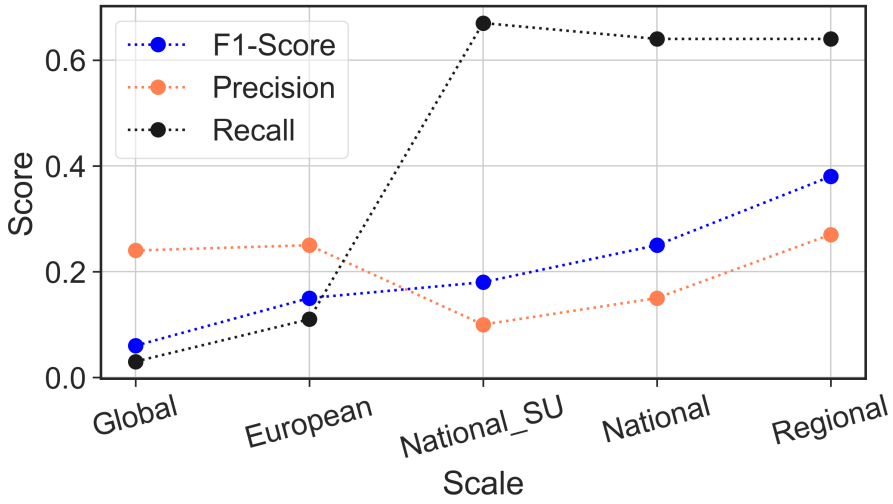


Figure 3.14: Line plot of evaluation metrics obtained from the spatialized accuracy assessment of open-source LSZ at different scales of investigation.

Performance metrics relative to original and merged LS maps (both cumulative and pairwise) are summarized in Figure 3.15. Precision scores (Figure 3.15c) highlight low performance of all the considered maps. Among the others, the original LSZ at regional scale (i.e., FranaRisk) stands as the closer to the reference, and it is closely followed by its integration with the national scale map (i.e., ISPRA 2013). Eventually, precision values of multiscale maps remain close to the original ones.

On the other hand, recall scores (Figure 3.15b) are always higher than original maps, with the fusion of the two nationwide maps (i.e., ISPRA 2013 and CNR-IRPI 2022) producing the highest performance (0.75). Anyway, the cumulative fusion of all LSZ is only 0.02 away.

By bringing together recall and precision, the resulting F1 scores (Figure 3.15a) provided an overall picture of the performance. As a result, the map that comes closest to URGENT Rome LSZ is provided by the fusion of FranaRisk and ISPRA 2013 maps (F1-score of 0.39), improving the performance of the best original map.

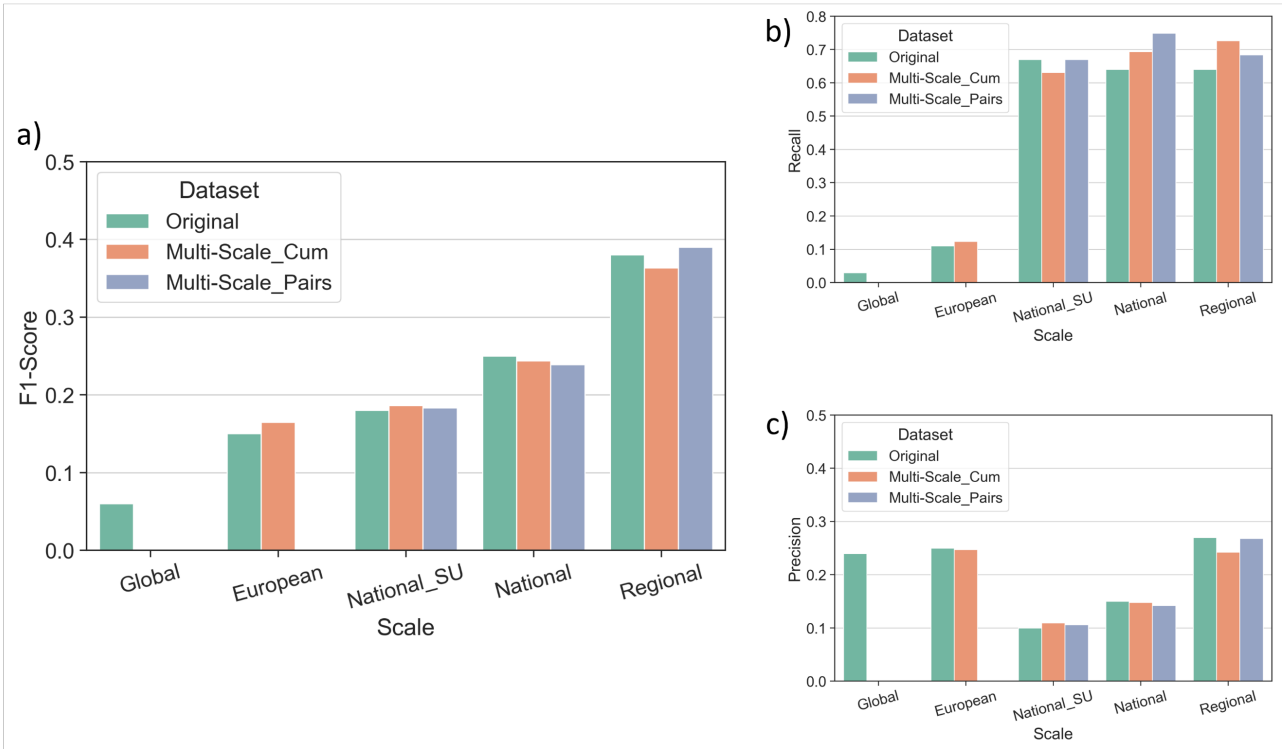


Figure 3.15: Comparison of F1-score (a), Recall (b), and Precision (c) scores for both original and multiscale landslide susceptibility maps with reference to URGENT Rome LSZ.

3.5. Discussion

The several supervised machine learning classifiers individually trained and validated by means of the K-fold cross validation technique allowed us to identify the ET model as the best performing one among others. Generally, all the trained models perform well, but the ET outperform the others in any calculated metric. We stress the importance of model optimization by tuning its hyperparameters, thus finding the optimal minimum between bias and variance. Moreover, tree-based models have a strong tendency to overestimate the importance of continuous numerical or high cardinality categorical features because they provide more opportunity for the models to split the data in half than the discrete features. Therefore, the default feature importance function of tree-based models returns biased results. Indeed, to overcome this problem we recommend using the permutation feature importance method, which also lists the features that worsen the model performance. In urban areas like the city of Rome, morphological variables such as slope angle and relative relief play the most important role in the prediction of the landslide susceptibility value, which is consistent with previous studies (Maxwell et al. 2020; Titti et al. 2022; Baeza, Lantada, and Amorim 2016). This is probably due to urbanization activities which generate man-made steep slopes close to the almost flat urban network. Nevertheless, cumulated annual rainfall, land use and litho-technical units play an important role as preparatory and predisposing factors for shallow landslides.

The final optimized model tested with the held 20% of the samples correctly classified almost 90% of both positive and negative samples, thus confirming its generalization capabilities and allowing its deployment to predict landslide susceptibility for the whole area of the Municipality of Rome. Indeed, incorrectly classified samples never get susceptibility further than 0.1 from the binary classification threshold (i.e., 0.5).

The urban-scale LS map (i.e., URGENT Rome) highlights the hotspots of Monte Mario and Monte Ciocci hills. This was indeed confirmed by considering the event-based test samples related to the landslides of January 2014 that occurred mostly along the south-eastern slope of the hill. We stress that the rainfall-triggered landslides in winter 2014 represent an out-of-sample set that the model has never seen. Other high susceptibility areas are represented by steep slopes near the streams, especially in the northern sector of the municipality, where volcanic deposits have been strongly eroded.

Once obtained the urban-scale URGENT Rome LSZ has been compared to other available open-source LSZ built at different scales of investigation. Firstly, we applied the empirical cumulative distribution function to reclassify quantitative maps in 5 hazard classes. This allowed us to make them comparable with the already qualitative maps (i.e., NASA and ELSUSv2 LSZ). Since the ECDF curve shows the probability of local landslides to be predicted, we stress that this reclassification method may strongly increase the predictive performance of hazard products made at small and medium scales. However, percentiles values to be used as classes thresholds are not standardised and they rely on expert judgment. This is proved by the similarity analysis, in which the national scale maps show reduced differences with the URGENT Rome LSZ. The similarity analysis carried out on both 2014 event-based LIPs and landslide polygons highlighted the effect of mapping unit as well as pixel size, and therefore the overall scale of analysis. Among the maps that were compared, grid-based ones perform better when predicting landslide initiation points, and the larger the scale the closest the map with respect to the reference. The SU-based LSZ turns out to be the closest to URGENT Rome when predicting the entire area potentially affected. This is because it assigns a uniform hazard level to the whole slope. Although the similarity analysis carried out with the Wasserstein distance metric concerns true positives only, the ability of hazard products to detect stable areas (i.e., true negatives) must be also evaluated. With this purpose, we carried out a spatialized accuracy assessment which returned the number of true/false positives and true/false negatives throughout the map. The results confirm the low capacity of global and European scale maps to detect true landslides, thus resulting in a significant underestimation of landslide hazard in Rome. As suggested by the similarity analysis, national scale maps may be adequate to detect areas susceptible to landsliding, but they overestimate them, thus resulting in high false positive rates (i.e., low precision). The regional scale map has the most high and balanced scores. It’s worth noting that landslide inventory is partially overlapped with the one considered in URGENT urban susceptibility

zoning. Although it has the biggest F1-score, it is still far from URGENT Rome. According to Cascini (2008) and Corominas et al. (2013), the larger the scale the most reliable the zoning is, thus changing its purposes. Hereby, we spotted the linear increment of F1-scores with the scale of analysis. We stress that this is probably due to the resolution and distribution of input data which inevitably are biased and incomplete when very large areas must be considered. Landslide hazard assessment at a large or detailed scale is often lacking. However, the proposed method for merging a variety of low accuracy products represents an attempt to improve the predictive performance of the single ones and gaining reliability. The several LSZ were integrated cumulatively and pairwise to evaluate the accuracy at each integration. The closest reliability and classification scores can be achieved by merging the FranaRisk and ISPRA 2013 maps (i.e., regional, and national scale). Despite a general limited improvement reached by merging the multiple scale maps, the implemented method can enhance the proportion of actual landslides correctly classified (i.e., the recall metric). With this perspective, the adopted data fusion criteria increased the ability to detect occurred landslides while maintaining precision. However, the number of false positives remains high as for original maps, thus affecting the overall F1 scores.

As an example, Figure 3.16 shows how the Monte Mario hill is classified by the different LS maps. In contrast to URGENT Rome (Figure 3.16a), the fusion of lower resolution original LSZ (Figure 3.16g, h, i), although they properly classify all LIPs, also misclassify some terrains surrounding the slopes that may not be prone to landslide. This reflects on the number of buildings that would be exposed to landslide risk within the municipality. In fact, it goes from a truthful 3% of the urban-scale URGENT Rome map up to about 25% recorded with the fusion of the two nationwide maps. With this perspective, site-specific urban scale landslide susceptibility analyses are always required for a balanced assessment of landslide hazard when detailed risk analysis and/or land-use plannings are carried out.

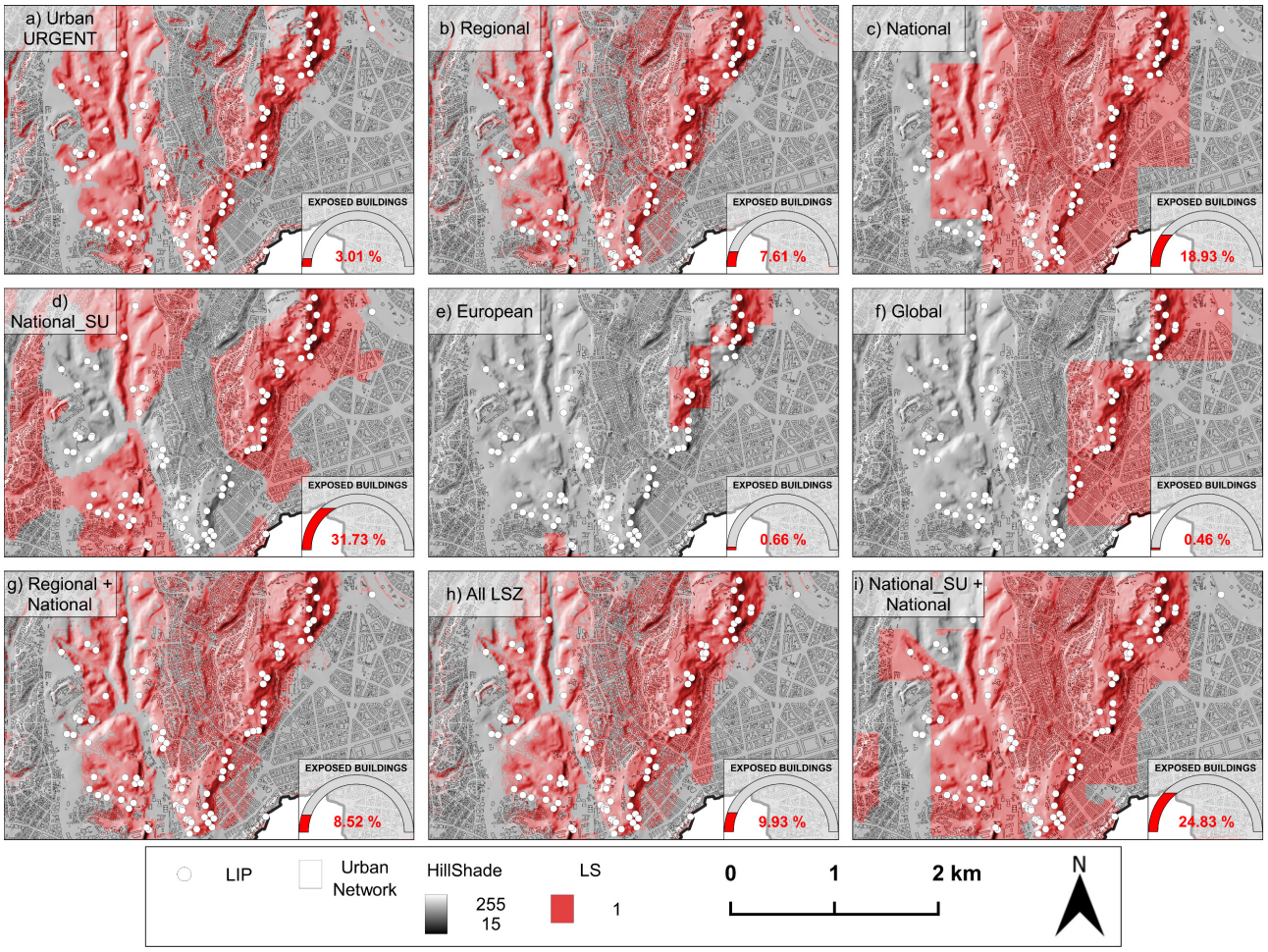


Figure 3.16: Details of predicted landslide susceptibility at Monte Mario hill and overall percentage of exposed buildings within the municipality of Rome according to each product. The URGENT Rome map (a) compared with the original open-source LS products built at different scales of analysis (b – f) and with the top performing products obtained with the fusion criteria: regional and national scale (b), all available maps (c), and the two nationwide maps based on different mapping unit (d). Landslide Initiation Points (LIP) and the Roman urban network are plotted on each map.

3.6. Conclusion

This study reports the complete and comprehensive workflow that should be implemented to obtain urban-scale LS maps when dealing with data-driven methods as supervised machine learning (ML) algorithms. Thorough pre-processing stages are fundamental to prepare data for machine learning. In this research, we tested various ML models to select the optimal and most performing one for landslide and stability classification in the urban area of Rome (Italy). Through empirical cumulative distribution functions applied to the out-of-samples landslides we selected susceptibility class thresholds based on model performance to correctly classify landslides. Once validated, the largest scale LSZ available in Rome was compared with regional, national, European and global scale LS products.

ECDF were used to standardize the reclassification of continuous LS maps once values of percentiles are determined. Comparisons between LSZ reveal differences in the ability to predict instability and stability with respect to the reference. The main outcomes revealed how the pixel size as well as the adopted mapping unit play a crucial role in the result. Specifically, wide grids fail to accurately localise potential landslides and stable areas. Whereas LSZ based on slope unit may give comprehensive information about the overall area potentially affected by the landslide evolution, they do not allow the direct definition of the source areas, which can be improved by merging grid and SU-based products. The implemented multiscale zoning method assigns a specific accuracy value to individual maps, which is appropriately accounted for when merging.

In a general perspective, the fusion of multiple low-resolution maps increased the ability to correctly classify occurred landslides; however, with the adopted method, the improvement in general performance is not enough to justify its deployment due to the high number of false positives. Nevertheless, when both urban and regional scale products are lacking, the fusion of national, European, and global maps might significantly help in better predict landslides occurrence in urban areas.

Disclosure Statement: The authors declare no conflict of interest.

Data availability: We share the necessary data to make our study transparent and reproducible. The landslide database and the urban-scale landslide susceptibility map of Rome reported in this work are available at the following repository: 10.5281/zenodo.7589881.

Funding: This research is framed within the PRIN 2017 project – URGENT “URban Geology and geohazards: Engineering geology for safer, resilieNt and smart ciTies” – UR SAPIENZA.

References

Alessi D, Bozzano F, Lisa A, Esposito C, Fantini A, Loffredo A, Martino S, Mele F, Moretto S, Noviello A, et al. 2014. Geological risks in large cities: The landslides triggered in the city of Rome (Italy) by the rainfall of 31 January-2 February 2014. Italian Journal of Engineering Geology and Environment. 1. https://doi.org/10.4408/IJEGE.2014-01.O-02

Ali, Moez. 2020. 'PyCaret: An Open Source, Low-Code Machine Learning Library in Python'. PyCaret Version 2.

Altmann, André, Laura Toloşi, Oliver Sander, and Thomas Lengauer. 2010. 'Permutation Importance: A Corrected Feature Importance Measure'. Bioinformatics 26 (10): 1340–1347. doi:10.1093/bioinformatics/btq134.

Alvioli, Massimiliano, Fausto Guzzetti, and Ivan Marchesini. 2020. 'Parameter-Free Delineation of Slope Units and Terrain Subdivision of Italy'. Geomorphology 358 (June): 107124. doi:10.1016/j.geomorph.2020.107124.

Amanti, M., V. Chiessi, and P. M. Guarino. 2012. 'The 13 November 2007 Rock-Fall at Viale Tiziano in Rome (Italy)'. Natural Hazards and Earth System Sciences 12 (5). Copernicus GmbH: 1621–1632. doi:10.5194/nhess-12-1621-2012.

Argentieri, A., C. Esposito, M. Fabiani, G. M. Marmoni, M. Piro, G. Rotella, G. Scarascia Mugnozza, and P. Vitali. 2018. 'The "Franarisk" Project in Rome Metropolitan Area: A Tool for Land Planning and Management and for Preliminary Risk Assessment of Infrastructures and Buildings.' In . Catania, SGI-SIMP 2018.

Baeza, Cristina, and Jordi Corominas. 2001. 'Assessment of Shallow Landslide Susceptibility by Means of Multivariate Statistical Techniques'. Earth Surface Processes and Landforms 26 (12): 1251–1263. doi:10.1002/esp.263.

Baeza, Cristina, Nieves Lantada, and Samuel Amorim. 2016. 'Statistical and Spatial Analysis of Landslide Susceptibility Maps with Different Classification Systems'. Environmental Earth Sciences 75 (19): 1318. doi:10.1007/s12665-016-6124-1.

Basu, Tirthankar, and Swades Pal. 2018. 'Identification of Landslide Susceptibility Zones in Gish River Basin, West Bengal, India'. Georisk: Assessment and Management of Risk for Engineered Systems and Geohazards 12 (1). Taylor & Francis: 14–28. doi:10.1080/17499518.2017.1343482.

Bhuyan, Kushanav, Hakan Tanyaş, Lorenzo Nava, Silvia Puliero, Sansar Raj Meena, Mario Floris, Cees van Westen, and Filippo Catani. 2023. 'Generating Multi-Temporal Landslide Inventories through a General Deep Transfer Learning Strategy Using HR EO Data'. Scientific Reports 13 (1). Nature Publishing Group: 162. doi:10.1038/s41598-022-27352-y.

Bordoni, M., V. Vivaldi, L. Lucchelli, L. Ciabatta, L. Brocca, J. P. Galve, and C. Meisina. 2021. 'Development of a Data-Driven Model for Spatial and Temporal Shallow Landslide Probability of Occurrence at Catchment Scale'. Landslides 18 (4): 1209–1229. doi:10.1007/s10346-020-01592-3.

Bozzano, F., A. Andreucci, M. Gaeta, and R. Salucci. 2000. 'A Geological Model of the Buried Tiber River Valley beneath the Historical Centre of Rome'. Bulletin of Engineering Geology and the Environment 59 (1): 1–21. doi:10.1007/s100640000051.

Bozzano, Francesca, Carlo Esposito, Stefania Franchi, Paolo Mazzanti, Daniele Perissin, Alfredo Rocca, and Emanuele Romano. 2015. 'Analysis of a Subsidence Process by Integrating Geological and Hydrogeological Modelling with Satellite InSAR Data'. In Engineering Geology for Society and Territory - Volume 5, edited by Giorgio Lollino, Andrea Manconi, Fausto Guzzetti, Martin Culshaw, Peter Bobrowsky, and Fabio Luino, 155–159. Cham: Springer International Publishing. doi:10.1007/978-3-319-09048-1_31.

Brabb, Earl E. 1984. Innovative Approaches to Landslide Hazard and Risk Mapping. Proceedings of the 4th international symposium on landslides, Toronto, Canada, 1, pp. 30-32.

Cascini, Leonardo. 2008. 'Applicability of Landslide Susceptibility and Hazard Zoning at Different Scales'. Engineering Geology 102 (3–4): 164–177. doi:10.1016/j.enggeo.2008.03.016.

Ciotoli, Giancarlo, E. Loreto, Lorenzo Liperi, Fabio Meloni, Stefania Nisio, and Adelaide Sericola. 2014. 'Carta Dei Sinkhole Naturali Del Lazio 2012 e Sviluppo Futuro Del Progetto Sinkholes Regione Lazio'.

Corominas, J., C. van Westen, P. Frattini, L. Cascini, J.-P. Malet, S. Fotopoulou, F. Catani, et al. 2013. 'Recommendations for the Quantitative Analysis of Landslide Risk'. Bulletin of Engineering Geology and the Environment, November. doi:10.1007/s10064-013-0538-8.

Crosta, G. B., S. Imposimato, D. Roddeman, S. Chiesa, and F. Moia. 2005. 'Small Fast-Moving Flow-like Landslides in Volcanic Deposits: The 2001 Las Colinas Landslide (El Salvador)'. Engineering Geology 79 (3): 185–214. doi:10.1016/j.enggeo.2005.01.014.

Cui, Yifei, Deqiang Cheng, Clarence E. Choi, Wen Jin, Yu Lei, and Jeffrey S. Kargel. 2019. 'The Cost of Rapid and Haphazard Urbanization: Lessons Learned from the Freetown Landslide Disaster'. Landslides 16 (6): 1167–1176. doi:10.1007/s10346-019-01167-x.

Del Monte, Maurizio, Maurizio D'Orefice, Gian Marco Luberti, Roberta Marini, Alessia Pica, and Francesca Vergari. 2016. 'Geomorphological Classification of Urban Landscapes: The Case Study of Rome (Italy)'. Journal of Maps 12 (sup1). Taylor & Francis: 178–189. doi:10.1080/17445647.2016.1187977.

Esposito, Carlo, Niccolò Belcecchi, Francesca Bozzano, Alessandro Brunetti, Gian Marco Marmoni, Paolo Mazzanti, Saverio Romeo, Flavio Cammillozzi, Giancarlo Cecchini, and Massimo Spizzirri. 2021. 'Integration of Satellite-Based A-DInSAR and Geological Modeling Supporting the Prevention from Anthropogenic Sinkholes: A Case Study in the Urban Area of Rome'. Geomatics, Natural Hazards and Risk 12 (1). Taylor & Francis: 2835–2864. doi:10.1080/19475705.2021.1978562.

Esposito, Carlo, Giandomenico Mastrantoni, Gian Marco Marmoni, Benedetta Antonelli, Patrizia Caprari, Alessia Pica, Luca Schilirò, and Francesca Bozzano. Submitted. 'From Theory to Practice: Optimisation of Available Information for Landslide Hazard Assessment in Rome Relying on Official, Fragmented Data Sources.' Landslides.

Fawcett, Tom. 2006. 'An Introduction to ROC Analysis'. Pattern Recognition Letters, ROC Analysis in Pattern Recognition, 27 (8): 861–874. doi:10.1016/j.patrec.2005.10.010.

Fell, Robin, Jordi Corominas, Christophe Bonnard, Leonardo Cascini, Eric Leroi, and William Z. Savage. 2008. 'Guidelines for Landslide Susceptibility, Hazard and Risk Zoning for Land-Use Planning'. Engineering Geology 102 (3–4): 99–111. doi:10.1016/j.enggeo.2008.03.014.

Flentje, Phillip N, Anthony Miner, Graham Whitt, and Robin Fell. 2007. 'Guidelines for Landslide Susceptibility, Hazard and Risk Zoning for Land Use Planning' 42 (1): 28.

Funiciello, R., and G. Giordano. 2008. 'The Geological Map of Rome: Lithostratigraphy and Stratigraphic Organization'. La Geologia Di Roma. Dal Centro Storico Alla Periferia II. Mem Desc Carta Geologica d'Italia 80: 39–85.

Günther, Andreas, Miet Van Den Eeckhaut, Jean-Philippe Malet, Paola Reichenbach, and Javier Hervás. 2014. 'Climate-Physiographically Differentiated Pan-European Landslide Susceptibility Assessment Using Spatial Multi-Criteria Evaluation and Transnational Landslide Information'. Geomorphology 224 (November): 69–85. doi:10.1016/j.geomorph.2014.07.011.

Guzzetti, Fausto, Alberto Carrara, Mauro Cardinali, and Paola Reichenbach. 1999. 'Landslide Hazard Evaluation: A Review of Current Techniques and Their Application in a Multi-Scale Study, Central Italy'. Geomorphology 31 (1): 181–216. doi:10.1016/S0169-555X(99)00078-1.

Hungr, Oldrich, Serge Leroueil, and Luciano Picarelli. 2014. 'The Varnes Classification of Landslide Types, an Update'. Landslides 11 (2): 167–194. doi:10.1007/s10346-013-0436-y.

Hutchinson, Michael F., Tingbao Xu, and John A. Stein. 2011. 'Recent Progress in the ANUDEM Elevation Gridding Procedure'. Geomorphometry 2011. International Society for Geomorphometry Redlands: 19–22.

Iadanza, Cacace, Del Conte, Spizzichino, Cespa, and Alessandro Trigila. 2013. 'Cultural Heritage, Landslide Risk and Remote Sensing in Italy'. In Landslide Science and Practice: Volume 6: Risk Assessment, Management and Mitigation, edited by Claudio Margottini, Paolo Canuti, and Kyoji Sassa, 491–499. Berlin, Heidelberg: Springer. doi:10.1007/978-3-642-31319-6_65.

Iadanza, Trigila, Starace, Dragoni, Biondo, and Roccisano. 2021. 'IdroGEO: A Collaborative Web Mapping Application Based on REST API Services and Open Data on Landslides and Floods in Italy'. ISPRS International Journal of Geo-Information 10 (2). Multidisciplinary Digital Publishing Institute: 89. doi:10.3390/ijgi10020089.

Iadanza, Trigila, Vittori, and Serva. 2009. 'Landslides in Coastal Areas of Italy'. Geological Society, London, Special Publications 322 (1). Geological Society of London: 121–141. doi:10.1144/SP322.5.

Jurgiel, Borys. 2013. 'Borysiasty/Pointsamplingtool'. Python. https://github.com/borysiasty/pointsamplingtool.

Kainthura, Poonam, and Neelam Sharma. 2022. 'Machine Learning Driven Landslide Susceptibility Prediction for the Uttarkashi Region of Uttarakhand in India'. Georisk: Assessment and Management of Risk for Engineered Systems and Geohazards 16 (3). Taylor & Francis: 570–583. doi:10.1080/17499518.2021.1957484.

Kiersch, G. 1964. 'Vaiont Reservoir Disaster'. Civil Engineering, ASCE 34: 32–47.

Kotsiantis, Sotiris B., Dimitris Kanellopoulos, and Panagiotis E. Pintelas. 2006. 'Data Preprocessing for Supervised Leaning'. International Journal of Computer Science 1 (2). Citeseer: 111–117.

Kuhn, Max, and Kjell Johnson. 2013. Applied Predictive Modeling. Vol. 26. Springer.

La Vigna, Francesco, Roberto Mazza, Marco Amanti, Cristina Di Salvo, Marco Petitta, and Luca Pizzino. 2015. 'The Synthesis of Decades of Groundwater Knowledge: The New Hydrogeological Map of Rome'. Acque Sotterranee - Italian Journal of Groundwater 4 (4). doi:10.7343/as-128-15-0155.

Loche, Marco, Massimiliano Alvioli, Ivan Marchesini, Haakon Bakka, and Luigi Lombardo. 2022. Landslide Susceptibility Maps of Italy: Lesson Learnt from Dealing with Multiple Landslide Classes and the Uneven Spatial Distribution of the National Inventory. Preprint. Physical Sciences and Mathematics. doi:10.31223/X5Q92S.

Luberti, Gian Marco, Francesca Vergari, Roberta Marini, Alessia Pica, and Maurizio Del Monte. 2018. 'ANTHROPOGENIC MODIFICATIONS TO THE DRAINAGE NETWORK OF ROME (ITALY): THE CASE STUDY OF THE AQUA MARIANA'. Alpine and Mediterranean Quaternary 31 (2): 119–132. doi:10.26382/AMQ.2018.08.

Luo, Junyao, Lulu Zhang, Haoqing Yang, Xin Wei, Dongsheng Liu, and Jiabao Xu. 2022. 'Probabilistic Model Calibration of Spatial Variability for a Physically-Based Landslide Susceptibility Model'. Georisk: Assessment and Management of Risk for Engineered Systems and Geohazards 16 (4). Taylor & Francis: 728–745. doi:10.1080/17499518.2021.1988986.

Martino, S., F. Bozzano, P. Caporossi, D. D'Angiò, M. Della Seta, C. Esposito, A. Fantini, et al. 2019. 'Impact of Landslides on Transportation Routes during the 2016–2017 Central Italy Seismic Sequence'. Landslides 16 (6): 1221–1241. doi:10.1007/s10346-019-01162-2.

Mastrantoni, Giandomenico, Patrizia Caprari, Carlo Esposito, Gian Marco Marmoni, Paolo Mazzanti, and Francesca Bozzano. 2022. Data Requirements and Scientific Efforts for Reliable

Large-Scale Assessment of Landslide Hazard in Urban Areas. EGU22-4669. Copernicus Meetings. doi:10.5194/egusphere-egu22-4669.

Mateos, Rosa María, Juan López-Vinielles, Eleftheria Poyiadji, Dimetrios Tsagkas, Michael Sheehy, Kleopas Hadjicharalambous, Pavel Liscák, et al. 2020. 'Integration of Landslide Hazard into Urban Planning across Europe'. Landscape and Urban Planning 196 (April): 103740. doi:10.1016/j.landurbplan.2019.103740.

Maxwell, Aaron E., Maneesh Sharma, James S. Kite, Kurt A. Donaldson, James A. Thompson, Matthew L. Bell, and Shannon M. Maynard. 2020. 'Slope Failure Prediction Using Random Forest Machine Learning and LiDAR in an Eroded Folded Mountain Belt'. Remote Sensing 12 (3). Multidisciplinary Digital Publishing Institute: 486. doi:10.3390/rs12030486.

Mazzanti, P., and F. Bozzano. 2011. 'Revisiting the February 6th 1783 Scilla (Calabria, Italy) Landslide and Tsunami by Numerical Simulation'. Marine Geophysical Research 32 (1): 273–286. doi:10.1007/s11001-011-9117-1.

Ngadisih, Ryuichi Yatabe, Netra P. Bhandary, and Ranjan K. Dahal. 2014. 'Integration of Statistical and Heuristic Approaches for Landslide Risk Analysis: A Case of Volcanic Mountains in West Java Province, Indonesia'. Georisk: Assessment and Management of Risk for Engineered Systems and Geohazards 8 (1). Taylor & Francis: 29–47. doi:10.1080/17499518.2013.826030.

Oguchi, Takashi. 1997. 'Drainage Density and Relative Relief in Humid Steep Mountains with Frequent Slope Failure'. Earth Surface Processes and Landforms 22 (2): 107–120. doi:10.1002/(SICI)1096-9837(199702)22:2<107::AID-ESP680>3.0.CO;2-U.

Parotto, M. 2008. 'Evoluzione Paleogeografica Dell'area Romana: Una Breve Sintesi.' Special volume "La geologia di Roma. Dal centro storico alla periferia", Memorie descrittive della Carta Geologica d'Italia, 80: 25–38.

Pedregosa, Fabian, Gael Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, et al. 2011. 'Scikit-Learn: Machine Learning in Python'. MACHINE LEARNING IN PYTHON, 6.

Persichillo, Maria Giuseppina, Massimiliano Bordoni, Claudia Meisina, Carlotta Bartelletti, Michele Barsanti, Roberto Giannecchini, Giacomo D'Amato Avanzi, et al. 2017. 'Shallow Landslides Susceptibility Assessment in Different Environments'. Geomatics, Natural Hazards and Risk 8 (2). Taylor & Francis: 748–771. doi:10.1080/19475705.2016.1265011.

Piccoli, Benedetto, and Francesco Rossi. 2016. 'On Properties of the Generalized Wasserstein Distance'. Archive for Rational Mechanics and Analysis 222 (3). Springer: 1339–1365.

Reichenbach, Paola, Mauro Rossi, Bruce D. Malamud, Monika Mihir, and Fausto Guzzetti. 2018. 'A Review of Statistically-Based Landslide Susceptibility Models'. Earth-Science Reviews 180 (May): 60–91. doi:10.1016/j.earscirev.2018.03.001.

Salvati, P., C. Bianchi, M. Rossi, and F. Guzzetti. 2010. 'Societal Landslide and Flood Risk in Italy'. Natural Hazards and Earth System Sciences 10 (3). Copernicus GmbH: 465–483. doi:10.5194/nhess-10-465-2010.

Saulnier, Georges-Marie, Keith Beven, and Charles Obled. 1997. 'Including Spatially Variable Effective Soil Depths in TOPMODEL'. Journal of Hydrology 202 (1): 158–172. doi:10.1016/S0022-1694(97)00059-0.

Schilirò, L., G. Poueme Djueyep, C. Esposito, and G. Scarascia Mugnozza. 2019. 'The Role of Initial Soil Conditions in Shallow Landslide Triggering: Insights from Physically Based Approaches'. Geofluids 2019 (June). Hindawi: e2453786. doi:10.1155/2019/2453786.

Soeters, Robert, and C. J. Van Westen. 1996. 'Slope Instability Recognition, Analysis and Zonation'. Landslides: Investigation and Mitigation 247. National Academy Press Washington, DC: 129–177.

Stanley, Thomas, and Dalia B. Kirschbaum. 2017. 'A Heuristic Approach to Global Landslide Susceptibility Mapping'. Natural Hazards 87 (1): 145–164. doi:10.1007/s11069-017-2757-y.

Su, Chenxu, Bijiao Wang, Yunhong Lv, Mingpeng Zhang, Dalei Peng, Bate Bate, and Shuai Zhang. 2022. 'Improved Landslide Susceptibility Mapping Using Unsupervised and Supervised Collaborative Machine Learning Models'. Georisk: Assessment and Management of Risk for Engineered Systems and Geohazards 0 (0). Taylor & Francis: 1–19. doi:10.1080/17499518.2022.2088802.

Titti, Giacomo, Alessandro Sarretta, Luigi Lombardo, Stefano Crema, Alessandro Pasuto, and Lisa Borgatti. 2022. 'Mapping Susceptibility With Open-Source Tools: A New Plugin for QGIS'. Frontiers in Earth Science 10 (March): 842425. doi:10.3389/feart.2022.842425.

Tonini, M., G. Pecoraro, K. Romailler, and M. Calvello. 2022. 'Spatio-Temporal Cluster Analysis of Recent Italian Landslides'. Georisk: Assessment and Management of Risk for Engineered Systems and Geohazards 16 (3). Taylor & Francis: 536–554. doi:10.1080/17499518.2020.1861634.

Trigila, Alessandro, Paolo Frattini, Nicola Casagli, Filippo Catani, Giovanni Crosta, Carlo Esposito, Carla Iadanza, et al. 2013. 'Landslide Susceptibility Mapping at National Scale: The Italian Case Study'. In Landslide Science and Practice: Volume 1: Landslide Inventory and Susceptibility and Hazard Zoning, edited by Claudio Margottini, Paolo Canuti, and Kyoji Sassa, 287–295. Berlin, Heidelberg: Springer. doi:10.1007/978-3-642-31325-7_38.

Trigila, Alessandro, Carla Iadanza, Carlo Esposito, and Gabriele Scarascia-Mugnozza. 2015. 'Comparison of Logistic Regression and Random Forests Techniques for Shallow Landslide Susceptibility Assessment in Giampilieri (NE Sicily, Italy)'. Geomorphology, Geohazard Databases: Concepts, Development, Applications, 249 (November): 119–136. doi:10.1016/j.geomorph.2015.06.001.

Trigila, Alessandro, Carla Iadanza, Michele Munafò, and Ines Marinosci. 2015. 'Population Exposed to Landslide and Flood Risk in Italy'. In Engineering Geology for Society and Territory - Volume 5, edited by Giorgio Lollino, Andrea Manconi, Fausto Guzzetti, Martin Culshaw, Peter Bobrowsky, and Fabio Luino, 843–848. Cham: Springer International Publishing. doi:10.1007/978-3-319-09048-1_163.

Trigila, Alessandro, Carla Iadanza, and Daniele Spizzichino. 2010. 'Quality Assessment of the Italian Landslide Inventory Using GIS Processing'. Landslides 7 (4): 455–470. doi:10.1007/s10346-010-0213-0.

van Westen, Cees J., Enrique Castellanos, and Sekhar L. Kuriakose. 2008. 'Spatial Data for Landslide Susceptibility, Hazard, and Vulnerability Assessment: An Overview'. Engineering Geology, Landslide Susceptibility, Hazard and Risk Zoning for Land Use Planning, 102 (3): 112–131. doi:10.1016/j.enggeo.2008.03.010.

Vijith, H., K.N. Krishnakumar, G.S. Pradeep, M.V. Ninu Krishnan, and G. Madhu. 2014. 'Shallow Landslide Initiation Susceptibility Mapping by GIS-Based Weights-of-Evidence Analysis of Multi-Class Spatial Data-Sets: A Case Study from the Natural Sloping Terrain of Western Ghats, India'. Georisk: Assessment and Management of Risk for Engineered Systems and Geohazards 8 (1). Taylor & Francis: 48–62. doi:10.1080/17499518.2013.843437.

Wilde, Martina, Andreas Günther, Paola Reichenbach, Jean-Philippe Malet, and Javier Hervás. 2018. 'Pan-European Landslide Susceptibility Mapping: ELSUS Version 2'. Journal of Maps 14 (2): 97–104. doi:10.1080/17445647.2018.143251

CHAPTER 4

4. A Novel Model for Multi-Risk Ranking of Buildings at City Level based on Open Data: The Test Site of Rome, Italy

Authors

Giandomenico Mastrantoni[1], Claudia Masciulli[1], Roberta Marini[2], Carlo Esposito[1], Gabriele Scarascia Mugnozza[1, 2], Paolo Mazzanti[1, 2]

[1] Dipartimento di Scienze della Terra e Centro di ricerca per i rischi geologici CERI, Sapienza Università di Roma, P.le Aldo Moro 5, 00185, Roma, Italy.

[2] NHAZCA S.r.l. start-up di Sapienza Università di Roma, P.le Aldo Moro 5, 00185, Roma, Italy.

Published in Geomatics, Natural Hazards and Risk (2023)

Abstract

In the context of population concentration in large cities, assessing the risks posed by geological hazards to enhance urban resilience is becoming increasingly important. This study introduces a robust and replicable procedure for assessing ground instability hazards and associated physical risks. Specifically, our comprehensive model integrates spatial hazard assessments, multi-satellite InSAR data, and physical features of the built environment to rank and prioritise assets facing multiple risks, with a focus on ground instabilities. The model generates risk scores based on hazard probability, potential damage, and displacement rates, aiding decision-makers in identifying high-risk buildings and implementing appropriate mitigation measures to reduce economic losses. The procedure was tested in Rome, Italy, where the analysis revealed that 60% of the examined buildings (90×10^3) are at risk of ground instability. Specifically, 33%, 22%, and 5% exhibit the highest multi-risk score for sinkholes, landslides, and subsidence, respectively. Landslide risk prevails among residential structures, while retail and office buildings face a higher risk of subsidence and sinkholes. Notably, our study identified a positive correlation between mitigation expenses and the multi-risk scores of nearby buildings, highlighting the practical implications of our findings for urban planning and risk management strategies.

Keywords: Multi-risk; Ground Instability Hazards; Urban Resilience; InSAR; Open Data.

4.1. Introduction

Large cities often face natural and geological hazards, engineering geological, and technological threats. The importance of the analysis and evaluation of geo-hazards for effective risk mitigation measures is evident in view of implementing effective strategies to make our cities and communities more resilient to natural disasters (Frankenberg et al. 2013). Due to the increased incidence of extreme weather events and human-made hazards related to accelerated urban growth (UNISDR 2012), built-up areas are specifically threatened and should therefore be the priority target of the action. Identifying the spatial distribution and concentration of risks in urban areas helps determine where and how preventive and corrective actions can reduce levels of vulnerability and exposure of urban populations (Johnson et al. 2016). Therefore, simplified vulnerability and risk assessment methodologies have been gaining space and importance (Liu et al. 2015; Romão et al. 2016; Liu and Chen 2019; Wei et al. 2022; Kappes et al. 2012a). Geo-hazards risk assessment provides strategic input for enhancing city resilience through mitigation design (McGlade et al. 2019). Nevertheless, losses from natural disasters continue to grow (White et al. 2001; Annual report 2022 | Munich Re). According to Kappes et al. (2012b), this may be due to a lack of interaction between science and practice in terms of knowledge transfer and applicability of results. To fill this gap, multi-risk assessment must support decision-makers by providing them with valuable information regarding the types and probabilities of hazards and their physical impacts, which will guide mitigation measures (Komendantova et al. 2014).

The statement that natural hazards almost never occur individually is of great importance: this leads to the notion of multi-risk (Gill et al. 2022). Multi-risk is generated from the presence of multiple hazards that may also be correlated, affecting the same elements exposed (Terzi et al. 2019). The interrelationships between these hazards and vulnerability level should be considered. A necessary condition for risk prevention, mitigation and reduction is its analysis, quantification, and assessment (Ward et al. 2020; Hochrainer-Stigler et al. 2023). However, assessing the risks and vulnerabilities of an entire city is a very demanding task, requiring considerable amount of data, technical knowledge and financial resources that are usually not available (van Westen et al. 2014; Julià and Ferreira 2021; Kappes et al. 2012b). Different temporal and spatial scales of hazardous events, and the potential interactions between hazards and socio-economic fragilities make multi-risk assessment problematic (Bell and Glade 2004; Kappes et al. 2012a). Existing risk assessment methods integrate large volumes of data and sophisticated analyses, as well as different approaches to risk quantification. In multi-risk assessment, three different approaches are commonly used to calculate compound risk: the first approach combines hazard, vulnerability, and exposure indexes through qualitative analysis or computation indices (Dilley 2005; Greiving 2006). The second approach combines quantitative single-hazard risk analysis methods to assess multiple hazards

(Bell and Glade 2004). These approaches often disregard hazard interactions and amplified risks. To address this, the third approach incorporates advanced modelling techniques that consider interdependencies and cascading effects (Terzi et al. 2019). Nowadays, a universally accepted procedure for multi-risk assessment has yet to be established.

This paper seeks to contribute to the field of urban resilience by providing a robust and replicable method for evaluating ground instability hazards (i.e., landslides, sinkholes, and subsidence) and associated building risks, ultimately leading to an effective framework for decision-makers in order to address the most appropriate mitigation measures, thus enhancing the resilience of urban environments. Risk assessment for landslides, subsidence, and sinkholes have been exploited by several researchers both qualitative and quantitative (Brabb 1984; Guzzetti 2000; Fell et al. 2008; Zisman 2012; Huang et al. 2012; Corominas et al. 2014; Chang and Hanssen 2014; Giampaolo et al. 2016; Shi et al. 2019; Mohebbi Tafreshi et al. 2021). However, in the existing literature, these ground instability processes have not yet been integrated with each other into a unified multi-risk assessment. In this study, we present a novel score-based approach which ranks assets exposed at multiple risks through the integration of multiple spatial hazard maps with multi-satellite and multi-frequency InSAR data and physical features of the urban built environment, thus guiding mitigation measures aimed at preventing future losses. The risk scores rely on three components, namely hazard, activity, and potential damage score engineered using data retrieved from susceptibility analyses, building displacement rates, census tracts and real estate market observatory. Consequently, each individual building is ranked according to its single- and multi-risk score, and the elements with the maximum values are the most relevant ones, as these are the geographic areas which are most affected by ground instabilities and potential economic losses.

4.2. Area Of Application

The area of interest (AOI) includes the Metropolitan City of Rome, extending from the historical centre to about the A90 highway (Figure 4.1). The city of Rome is situated in a hilly region characterized by a diverse geological and geomorphological setting. The geological formations consist of marine deposits (late Pliocene to lower Pleistocene), and continental sediments (middle Pleistocene – Holocene). These deposits, including marine clays, silt, silty sands, and transitional sediments, are prominently exposed on the hills along the right bank of the Tiber River. Additionally, volcanic deposits from the Colli Albani and Sabatini Volcanic Districts alternate with continental sediments, such as alluvial and palustrine deposits (middle to upper Pleistocene) (Figure 4.1).

The present-day landscape is shaped by the Tiber River and its tributaries, which have carved valleys and slopes. The river valleys are partially filled with alluvial deposits, reaching considerable thicknesses (Bozzano et al. 2000). The groundwater circulation is controlled by the superimposition

of permeable volcanic deposits over less permeable clayey and silty horizons, resulting in the development of ephemeral and perched water tables, such as those forming within the weathered soil covers and contributing to surface seepage in unsaturated soils. The different geological units on the left and right embankments of the Tiber River valley play a significant role in the response to natural and anthropogenic hazards affecting the urban area. Hazards such as subsidence, sinkholes, landslides, and floods have been previously experienced and assessed by several studies (Alessi et al. 2014; Bozzano et al. 2015a; Delgado Blasco et al. 2019; Esposito et al. 2021; Esposito et al. 2023).

Landslides are often localized within unsaturated soil covers originated from Plio-Pleistocene sedimentary units. Intense and prolonged rainfall events have been identified as triggering factors for many slope failures, causing significant damage to infrastructure such as pipelines, aqueducts, and roads (Amanti et al. 2008; Alessi et al. 2014; Esposito et al. 2023).

Landslides and natural or anthropogenic sinkholes are the most common and impactful processes in the urban area. Sinkholes concentrate primarily in sectors characterized by the presence of volcanic deposits that were extensively exploited as a mineral resource since ancient Roman times. Consequently, a dense network of underground cavities exists, the majority of which remain undiscovered (Ciotoli et al. 2014; Ciotoli et al. 2016; Esposito et al. 2021). According to Bianchi Fasani et al. (2011), these cavities experience an upward migration of the tunnel crown within the volcanic subsoil, eventually leading to the formation of sinkholes. Due to the predominantly brittle behaviour of cap rocks, particularly tuffs, abrupt failures following minimal deformations may occur, thus affecting effective monitoring for early-warning purposes (Esposito et al. 2021).

Numerous instances of land subsidence in Rome have been identified and quantified using multi-temporal InSAR techniques (Campolunghi et al. 2007; Cigna et al. 2014; Bozzano et al. 2015b; Delgado Blasco et al. 2019). The subsidence processes primarily occur in the alluvial sediments along the Tiber River and are mainly attributed to the load imposed by relatively recent urban development on the unconsolidated alluvial deposits. The subsidence rate in the region varies depending on the age of the constructed structures and the geological characteristics of the underlying formations (Bozzano et al. 2018; Delgado Blasco et al. 2019).

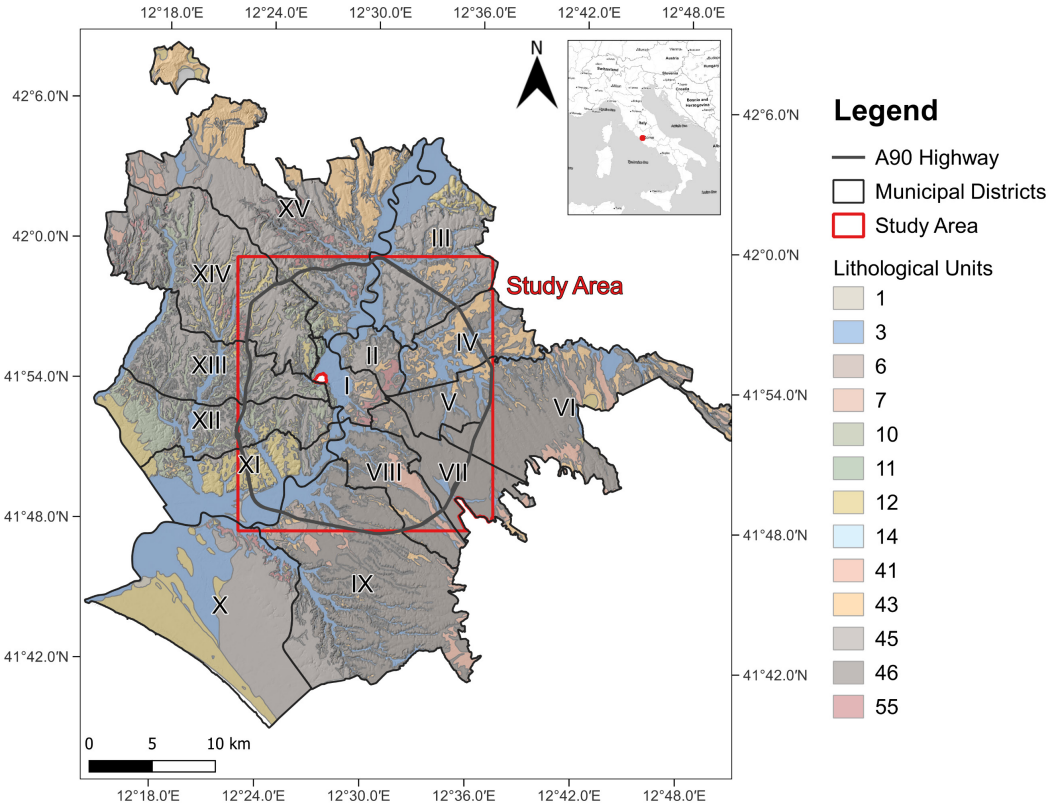


Figure 4.1: Lithological units in the Municipality of Rome with the municipal districts, the study area, and the A90 highway outlined. Key to legend: 1- Anthropic deposits; 3- recent and terraced sandy-gravelly alluvial deposits; 6- silty-sandy alluvial deposits, fluvio-lacustrine deposits; 7- travertines; 10- Plio-Pleistocene clayey and silty deposits; 11- marine Pliocene clays; 12- debris and talus slope deposits, conglomerates and cemented breccias; 14- marls, marly limestones and calcarenites; 41- leucitic/trachytic lavas; 43- lithoid tuffs, pomiceous ignimbritic and phreatomagmatic facies; 45- welded tuffs, tufites; 46- pozzolanic sequence; 55- alternance of loose and welded ignimbrites. EPSG:4326.

4.3. Material And Methods

In this work, multi-risk assessment is conceived as the process of assessing different independent hazards threatening individual buildings. We assume landslides, sinkholes, and subsidence as compound threats with no interrelationships between them (Tilloy et al. 2019; Hochrainer-Stigler et al. 2023; Kappes et al. 2012b).

In this framework, a new workflow to identify and rank urban assets at risk (e.g., buildings) integrating hazard maps, multi-frequency PS-InSAR data and building features was developed. We underline that cultural heritage and religious structures were neglected in the present study due to the complexity of estimating their economic and vulnerability value. The workflow of the proposed methodology is reported in Figure 4.2. It consists of three main sections where geo-hazards probability for ground instabilities (a), building features (b), and multi-frequency InSAR displacement rates (c) are analysed and integrated to achieve both single- and multi-risk values. Each section

results in a score that makes up the specific risk and subsequently the multi-risk ranking of analysed buildings.

A detailed description of the implemented workflow and methods of achieving risk components is discussed in the following chapters. The entire data analysis process was automated via an implemented Python code freely available on the author's GitHub page (https://github.com/gmastrantoni/mhr) together with the input data (Table 4.1).

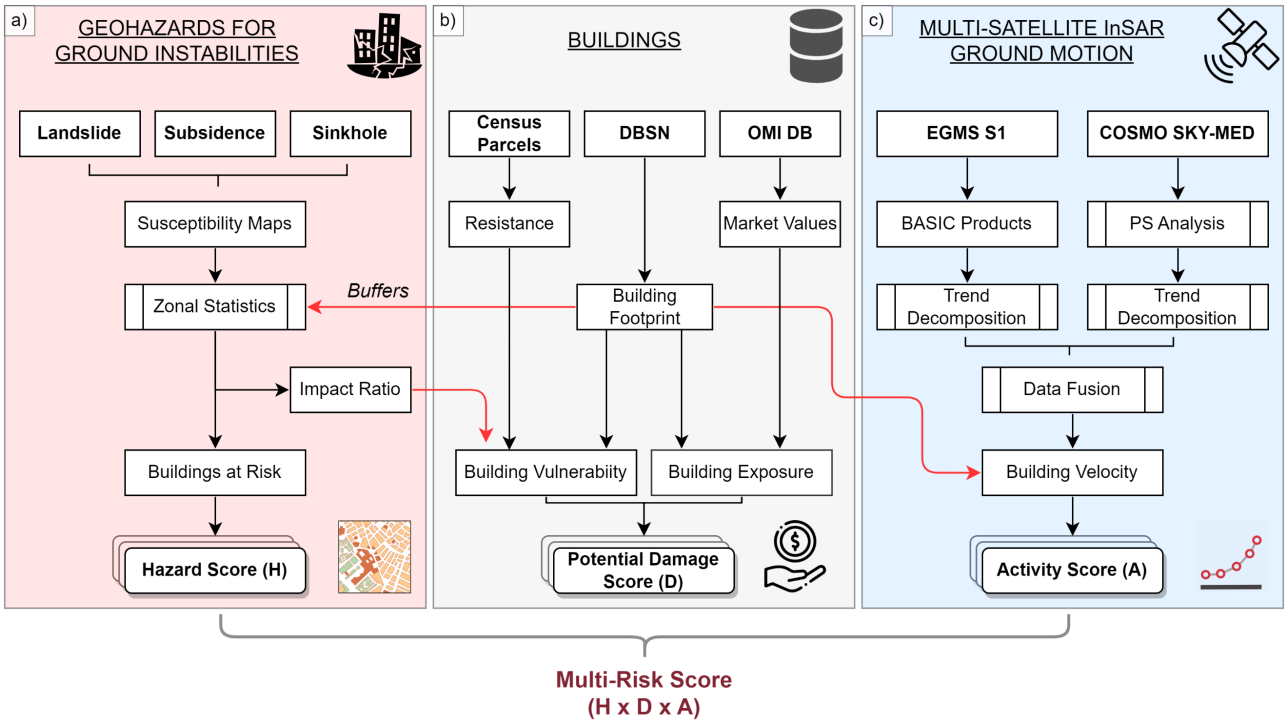


Figure 4.2: Flowchart of the adopted methodology to compute buildings' single- and multi-risk scores. Coloured subplots report the workflow to derive the hazard (a), potential damage (b) and activity (c) components of the multi-risk equation. In bold the main inputs and outputs data.

Table 4.1: Input data with features and role in the multi-risk ranking. ISTAT (Istituto nazionale di STATistica) stands for "National Institute of Statistics". DBSN (DataBase di Sintesi Nazionale) stands for "national synthesis database". OMI (Osservatorio Mercato Immobiliare) stands for "real estate market observatory". ReNDiS (Repertorio Nazionale degli interventi per la Difesa del Suolo) stands for "national inventory of soil protection interventions.

Input data	Description	Risk component	Format	Website/Reference
Landslide susceptibility map	Spatial probability of landslides occurrence	Hazard	Raster (5x5m)	Esposito et al., 2023, Mastrantoni et al., 2023.
Subsidence susceptibility map	Spatial probability of subsidence occurrence	Hazard	Raster (5x5m)	Research contract CERI – ACEA "Development of a monitoring and prevention plan of the geo-hydrological instability of the city of Rome"

Sinkhole susceptibility map	Spatial probability of sinkhole occurrence	Hazard	Raster (5x5m)	Esposito et al., 2021, and Research contract CERI – ACEA.
ISTAT census tracts	Spatial distribution of building features	Vulnerability	Vector (Polygons)	https://www.istat.it/
DBSN database	Building footprints and features	Building type and footprint	Vector (Polygons)	https://www.igmi.org/
OMI database	Asset market value	Exposure	Vector (Polygons)	https://www.agenziaentrate.gov.it/portale/
Sentinel-1 PS time series	Spatial distribution of ground motion rate	Activity	Vector (Points)	https://land.copernicus.eu/pan-european/european-ground-motion-service
Cosmo-SkyMed PS time series	Spatial distribution of ground motion rate	Activity	Vector (Points)	https://www.asi.it/en/earth-science/cosmo-skymed/
DTM	Elevation	Topography	Raster (5x5m)	Esposito et al., 2023
ReNDiS	Mitigation Measures	Validation	Vector (Points)	http://www.rendis.isprambiente.it/rendisweb/

4.3.1. Spatial multi-hazard assessment

This section describes the materials and methods applied for the assessment of multiple single hazard scores related to elements exposed to ground instabilities, as depicted in Figure 4.2a.

4.3.1.1. Hazard Score

In this work, we considered the spatial component of the hazard, which is the susceptibility, since the analysed ground instabilities are based on completely different processes with peculiar characteristics. Hence, the need to standardise the data to common information arises. Among the three geohazards considered in the present study, the temporal component of the hazard factor is conceptually definable for landslides only. Given the nature of subsidence, it occurs as a continuous (and non-linear) process that may be triggered by overloads and/or vary its intensity with time (e.g., velocity of settlements) up to the end of the process, thus preventing the assignment of return time values, as for landslides. A similar reasoning applies to sinkholes. A sinkhole can be interpreted as a unique shock event occurring in a confined place, which will no longer experience it. For this reason, sinkholes do not have a return time, nor have significant relationship with recurrent triggers pointed out so far. Given that, we decide to use high-resolution susceptibility maps as the hazard component of risk. Susceptibility maps report the spatial probability of occurrence of a certain event without considering the temporal dimension. They represent one of the products that lead to a

comprehensive hazard assessment (Cascini 2008; Fell et al. 2008; Corominas et al. 2014). According to Corominas et al. (2014), susceptibility assessment can be considered an end product that can be used in land-use planning and environmental impact assessment. As demonstrated by (Mastrantoni et al. 2023), a measurable loss in accuracy, and thus in reliability for specific purposes (Cascini 2008), is proportional to the decrease in zoning scale.

Susceptibility maps are represented in raster format with a resolution of 5 metres and cover the entire centre of Rome (Figure 4.3). Each raster cell gives the level of probability of being affected by a landslide (Figure 4.3a), a sinkhole (Figure 4.3b), and a subsidence (Figure 4.3c). The value of susceptibility ranges from 1 to 5, where 1 stands for negligible and 5 for very high probability. The employed landslide and sinkhole susceptibility maps derive from already published studies (Esposito et al. 2021; Mastrantoni et al. 2023; Esposito et al. 2023). Susceptibility to subsidence was commissioned to Sapienza University of Rome by ACEA ATO2 Spa that funded a project for a better knowledge of hazards aimed at safeguarding the underground service network in Rome.

The landslide susceptibility map was obtained by applying the extra trees classifier (Pedregosa et al. 2011), a supervised machine learning model that has proved to outperform several other algorithms (Mastrantoni et al. 2023). The model was trained and tested with a database of about 3000 landslide initiation and stable points distributed throughout the Municipality of Rome.

The sinkhole susceptibility map refers to anthropogenic sinkholes of Rome mainly due to the presence of underground cavity networks in areas featured by relevant thickness of volcanic deposits (Bianchi Fasani et al. 2011) and to “piping-like” processes due to shallow groundwater circulation and related erosion of silty fraction of layers in area dominated by sedimentary sequences. The assessment was carried out through the logistic regression multivariate statistical technique taking advantage of a database containing about 1000 sinkholes that occurred in the period 1875-2018.

Lastly, the subsidence susceptibility map was realised by reconstructing the extent and thickness of recent alluvial deposits. This process was based on the interpolation of available stratigraphic logs throughout the valley of the Tiber River. The derived thickness map of recent alluvial deposits was then weighted based on their type, thickness, and depth within the stratigraphic sequence, considered as proxies of compressibility, and thus proneness to develop consolidation settlements.

To derive the multiple single-hazard scores, the built environment was analysed with the zonal statistics method, thus retrieving the maximum value of hazard class within the perimeter of individual buildings. Elements with a hazard score above 2 were considered exposed to a specific single risk. We stress that to account for landslide run-outs, a buffer of 20m was drawn around the buildings. The buffer size was calibrated on the mapped landslides within the territory of Rome, as reported by Alessi et al. 2014; and Del Monte et al. 2016.

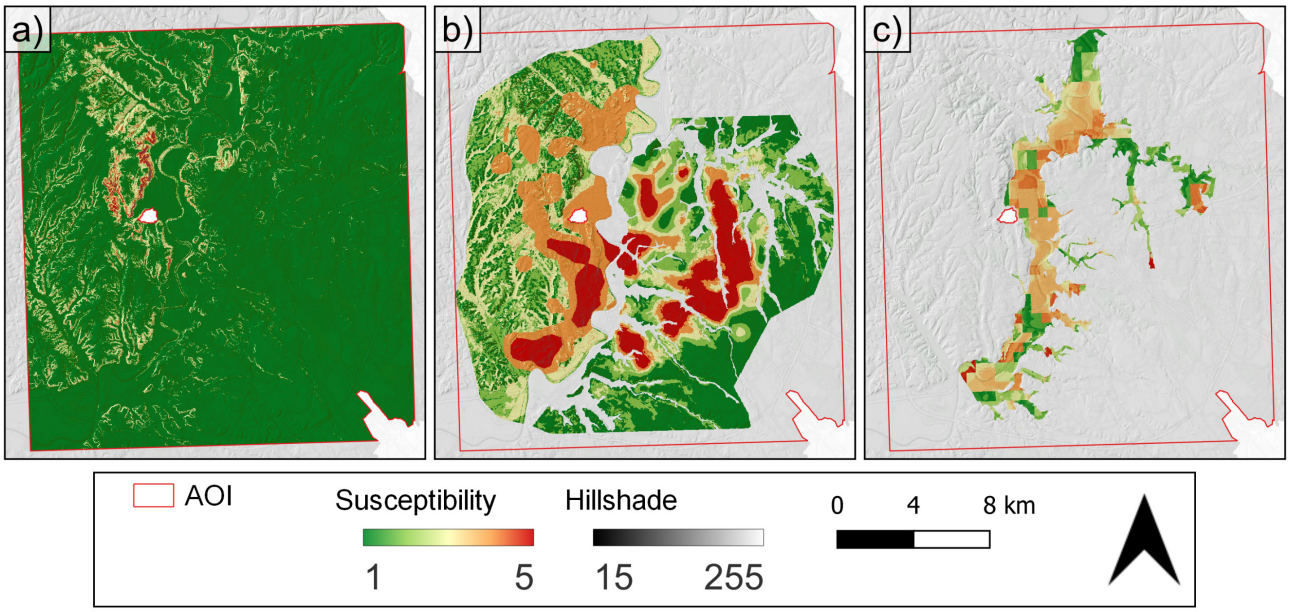


Figure 4.3: Susceptibility map for landslides (a), sinkholes (b), and subsidence (c) within the area of interest (AOI).

4.3.1.2. Impact Ratio

Since susceptibility maps do not directly provide magnitude information in terms of dimension and volume of the process, a strategy to overcome this limitation and define a proxy for hazard intensities was needed. Consequently, it was assumed that the potential magnitude of the hazards impacting the built environment is directly related to the percentage of the building's area exposed to the threat. To achieve this information, building footprints were overlaid on susceptibility maps and a categorical zonal statistic was computed, thus counting, and weighting the hazard classes within the polygons. By setting the threshold for the hazard class equal to 2 (i.e., low susceptibility) the weighted area of the element exposed to risk was derived, and consequently the ratio of the total area. Such information, hereafter called impact ratio (I_b), is then coupled in the following phases with structural resistance for the assignment of physical vulnerability values to the specific ground instability hazard. The values of the impact ratio were classified into four classes using equal intervals. I0 is defined by an impact ratio equal to 0, I1 for $0<Ib\leq0.25$; I2 for $0.25<Ib\leq0.5$; I3 for $0.5<Ib\leq0.75$ and I4 for $I_b>0.75$. Therefore, each element within the area of interest has one or more impact ratio classes following the threats to which it is exposed.

4.3.2. Multi-hazard consequence analysis

This section describes the materials and methods applied for quantifying the potential multi-hazard damage scores of the built environment, as depicted in Figure 4.2b.

4.3.2.1. Vulnerability

Building vulnerability was defined according to the impact ratio of the i-th hazard and the structural resistance of buildings. The definition of structural resistance relies on the approach proposed by Li et al. 2010 for landslide hazards and modified by Caleca et al. 2022. In this study, we assume that structural resistance to landslides can also be considered valid for sinkhole and subsidence hazards.

The applied approach exploits census tract data produced by ISTAT in 2011 (Table 4.1) and impact ratio data. It consists of three phases. First, the estimation of the structural resistance for each census tract within the area of interest was made by using a modified equation proposed by Li et al. 2010; Caleca et al. 2022:

$$R_{str} = \left(\varepsilon_{sty} \times \varepsilon_{smn} \times \varepsilon_{sht}\right)^{1/3} \quad (4.1)$$

where ε_{sty}, ε_{smn}, and ε_{sht} are resistance factors of structure type, maintenance state and number of floors, respectively. These parameters are derived from census tract data (Table 4.2). The Italian census tracts are delineated as geographically contiguous areas within the country's territory that demonstrate a notable degree of homogeneity. Each census tract represents either an entire municipality, a specific portion thereof, or a cluster of municipalities characterized by comparable environmental and socioeconomic features. They provide comprehensive information pertaining to the characteristics of buildings within the area, encompassing factors such as typology, quantity, materials, and other relevant attributes. Furthermore, census tracts offer valuable insights into population distribution. Values of resistance factors are set according to that suggested by previous studies (Li et al. 2010; Caleca et al. 2022) (Table 4.2), and combined with impact ratio values to retrieve the vulnerability level.

Table 4.2: Values of resistance factors employed in the structural resistance assessment (from Caleca et al. 2022).

Resistance Factor	Parameter	Typology	Value
Structure type	ε_{sty}	Productive and commercial	0.1
		Residential with light structure	0.2
		Residential with brick walls	0.8
		Residential reinforced concrete	1.5
Maintenance state	ε_{smn}	Productive and commercial	0.1

		Residential in very poor condition	0.1
		Residential in a medium condition	0.6
		Residential in a good condition	1.2
		Residential in a very good condition	1.5
Number of floors	ε_{sht}	Productive and commercial	0.1
		Single-storey residential	0.1
		Two-storey residential	0.4
		Three-storey residential	0.9
		Number of floors ≥4	1.5

The second step of the procedure consists of the attribution of the structural resistance values (R_{str}) computed from the census tracts to the corresponding building depending on its position and type (i.e., residential/not residential). Following that, a classification was conducted, categorizing the built environment into six distinct classes based on their respective level of structural resistance (Table 4.3). Footprints and categories of use of the built environment were extracted from the DBSN database (*Database di Sintesi Nazionale*) (Table 4.1). It is freely available under the Open Data Commons Open Database License (ODbL). The DBSN database is a geographic database containing territorial information primarily derived from regional geo-topographic data. These data have been harmonized and standardized within the structure to ensure national homogeneity while preserving the original level of detail. Administrative boundaries are derived from cadastral data, ensuring congruence with municipal and state administrative borders. The built environment is classified into detailed categories of use.

The third phase of the applied approach involved the quantification of the multi-hazard vulnerability values for each element at risk. These values were obtained using the contingency matrix proposed by Caleca et al. 2022 and tailored to this study, to established a link between the classes of the impact ratio for the i-th hazard type and the classes of structural resistance (Table 4.3). The contingency matrix defines five distinct vulnerability classes: V_0 (null vulnerability), V_1 (low vulnerability), V_2 (medium vulnerability), V_3 (high vulnerability), and V_4 (very high vulnerability). To calculate quantitative values for vulnerability, which were necessary for assessing the economic potential damage, numerical values ranging from 0 to 1 were assigned to each vulnerability class as

follows: $V_0 = 0$, $V_1 = 0.25$, $V_2 = 0.5$, $V_3 = 0.75$, $V_4 = 1$. Hence, each building has been assigned one or more vulnerability values depending on the number and type of hazards to which it is exposed.

Table 4.3: Contingency matrix for the assessment of building vulnerability by means of impact ratio and structural resistance classes (modified from Caleca et al. 2022).

Resistance	Impact Ratio				
	I0	**I1**	**I2**	**I3**	**I4**
R4	0	0	0.25	0.25	0.50
R3	0	0.25	0.25	0.50	0.50
R2	0	0.25	0.50	0.50	0.75
R1	0	0.25	0.50	0.75	0.75
R0	0	0.25	0.75	0.75	1
N.A.	0	0	0	0	0

4.3.2.2. Exposure

In this study, exposure of elements at risk (E_b) is deemed as the economic value per square meter, taking the real estate's market value as a reference. To obtain this information, we exploited the OMI national-scale open access database (which stands for *Osservatorio Mercato Immobiliare* – real estate market observatory) (Table 4.1). It is managed and updated every 6 months by the National Revenue Agency under the Italian Ministry of Economy and Finance. The OMI dataset encompasses comprehensive information regarding the minimum and maximum market values for various building typologies, expressed in euros per square meter (€/m^2). This data is aggregated at the municipality level and further stratified into smaller subdivisions, known as "OMI zones", within the respective municipality. The study area includes 186 zones. Market data related to the first semester of 2022 were downloaded and pre-processed to define for each OMI zone the average market value of six peculiar building typologies: economic, retail, residential, office, and industrial (Figure C1 in Appendix C). Thereafter, categories of use from the DBSN database were linked to typologies from OMI zones, and the respective economic values were associated according to location and type of buildings.

4.3.2.3. Potential Damage Score

To quantify the potential damage, multi-hazard vulnerabilities were combined with the previously derived exposure values (Marzocchi et al. 2012), thus obtaining the expected economic loss per square meter ($LOSS_b$):

$$LOSS_b(h) = V_b(h) \times E_b \quad (4.2)$$

where $V_b(h)$ is the vulnerability value of building exposed to the hazard *h*, and E_b represents its economic value expressed in euros per square meter.

Once the potential multi-hazard loss data were derived, the empirical cumulative distribution function (eCDF) was applied to them. With this approach, we sorted and ranked the damage data without making assumptions about the underlying probability distribution. The eCDF quantifies the proportion of data points that are less or equal to a given value. By exploiting the cumulative distribution of potential economic losses, class thresholds were set following Equation 3 and potential damage scores (D_b) ranging from 1 to 5 were defined.

$$D_b(h) = eCDF^{-1}\big(p,\ LOSS(b|h)\big),\ p = (20,40,60,80) \quad (4.3)$$

where $eCDF^{-1}$ denotes the inverse of the empirical cumulative distribution function, p is the specific value of the proportion or percentile for which the damage score D_b was calculated. $LOSS(b|h)$ represents the distribution of economic loss values of buildings (b) considering only those exposed to the hazard h. These class thresholds assist in delineating the severity levels of the potential economic loss associated with multiple hazards. We acknowledge that the ranking will be sensitive to the threshold values chosen to delimit the classes. However, the strength of it lies in being able to customize and optimize class thresholds according to the proportion of items to be contained in each class.

4.3.3. Multi-satellite SAR interferometry

This section describes the materials and methods applied for assessing the multi-hazard activity scores of the built environment, as depicted in Figure 4.2c.

4.3.3.1. Interferometric analysis

In this study, the Advanced-DInSAR (A-DInSAR) technique (Hanssen 2001) was employed to derive displacement rates throughout the area of interest. The A-DInSAR method is increasingly applied to study the temporal evolution (by means of time series) of ground displacements for objects with long-term stability in terms of reflectivity, known as Persistent Scatterers (PS) (Ferretti et al. 2001). The PS-InSAR technique extracts this information from a huge collection of SAR images, forming an interferometric stack that enables the derivation of displacement patterns over the desired time span. The dataset accessed in this study was acquired from two currently active satellite SAR missions that provide freely available historical archive data for scientific research: Sentinel-1 (by ESA) and Cosmo-SkyMed (by ASI). The main differences between these images lie in the resolution, both temporal and geometrical, and in the acquisition bandwidth (C- and X-band respectively). Sentinel-

1 (S1) has a pixel resolution of 5x20 m on the ground, while Cosmo-SkyMed (CSK) has a resolution of 3×3 m (for the Interferometric Wide images used and available from the archive). The revisit time of the S1 constellation is higher than CSK. The former being approximately 6 days (until December 23, 2021, end date of the Sentinel-1B mission), whereas the latter being 16 days on average. This is due to its dual-purpose mission for both civilian and military applications. Moreover, it may have sparse missing data.

The S1 PSs derive from the European Ground Motion Service products (EGMS). The EGMS is part of the Copernicus Land Monitoring Service and applies the PS-InSAR technology to monitor ground deformations over most of Europe (Costantini et al. 2021). The data employed in this study refers to Level 2a, which is the first of three product levels (Crosetto et al. 2021; Costantini et al. 2021). This level is also called Basic Product, it includes displacement rates and time series measured along the Line of Sight (LOS) for both the ascending and descending orbital geometries. According to Crosetto et al. 2021, these data are suitable to study local deformation phenomena such as subsidence, landslides, tectonic effects, and earthquakes, which impact the stability of slopes, buildings and infrastructures. The resultant PS time series data covers a period of 7 years (from 10/02/2015 to 23/12/2021), with a temporal sampling of one image every 6 days.

The CSK dataset, not yet available from the EGMS project, was elaborated using the PS approach starting from a total of 358 SAR images retrieved from the archives of the Italian Space Agency (ASI), which provides free of charge data for academic purposes. The selected time span ranges from 2010 to 2022 for both orbital geometries:

- Ascending orbital geometry: 196 images in Single Look Complex (SLC) format acquired by the CSK satellites from 04 July 2010 to 19 July 2022.
- Descending orbital geometry: 162 images in Single Look Complex (SLC) format acquired by the CSK satellites from 27 July 2010 to 19 June 2022.

For each geometry in each dataset, a master image was selected as the reference for the calculation of phase differences in other images within the dataset. The master image for the descending dataset was captured on December 17, 2017, and for the ascending dataset on May 28, 2014. The displacement rate values (here expressed in mm/year) estimated by the A-DInSAR analysis for all the PS are relative to the selected reference point. The reference point for the ascending dataset was located at LAT=41.8868, LON=12.4914 and for the descending dataset at LAT=41.8879, LON=12.4405 (EPSG:4326). Finally, the analyses were calibrated and validated with displacement data acquired from GNSS stations (Ferretti et al. 2022). The greater ground resolution of the CSK SAR images provided a denser PS coverage than the S1 products. Indeed, the total number of PS within the area of interest is about 4.65×10^6 and 1.46×10^6 for CSK and S1 descending geometry respectively.

Displacement time series can be divided into three parts that include seasonal, trend, and remainder components. The latter containing random fluctuations and noise (Hyndman and Athanasopoulos 2018; Costantini et al. 2018). Hence, the need to decompose PS time series to extract the long-term trend arises. To this purpose, we applied a Seasonal-Trend Decomposition (STL) technique (Cleveland et al. 1990) based on a LOcally wEighted regreSsion Smoother (LOESS) (Figure 4.4), which is implemented within the PS-ToolBox plugin for QGIS developed by Nhazca (https://www.sarinterferometry.com/ps-toolbox/).

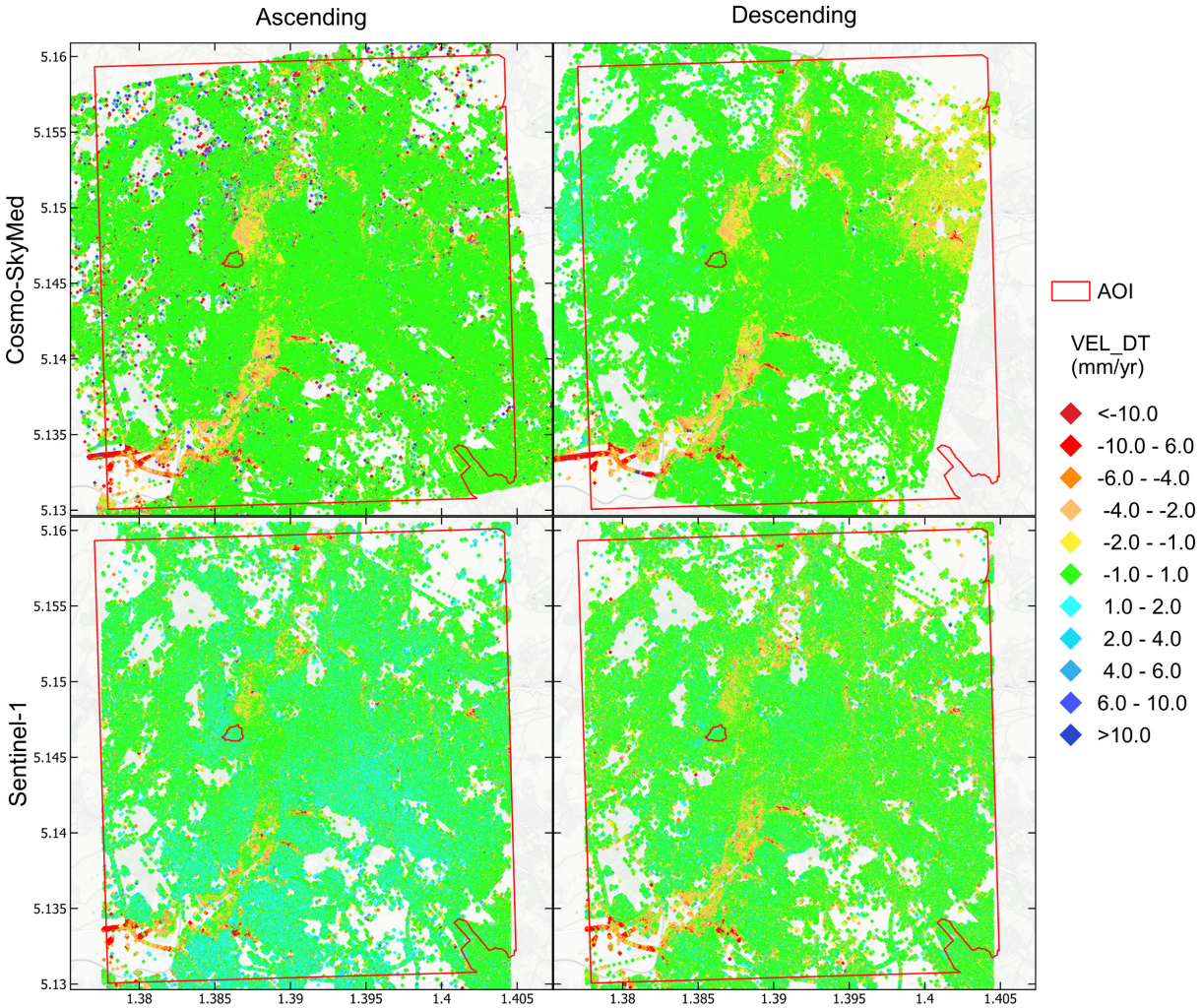


Figure 4.4*:* PSs map of average velocity along LOS computed from long-term trends extracted from Sentinel-1 and Cosmo-SkyMed for ascending and descending orbits. EPSG:3857.

4.3.3.2. Data fusion

S1 and CSK PS time series were then merged for cross-validation and to gain reliability, as well as to increase spatial coverage of displacement information. The implemented data fusion method within the PS-ToolBox allows the integration of multi-mission PS data. The combination of multiple sources measurements, with different orbital geometries, exploits the strain tensor (Guglielmino et al. 2011) through the ESISTEM (Extended Simultaneous and Integrated Strain Tensor Estimation

from Geodetic and Satellite Deformation Measurements) (Luo and Chen 2016) method to generate synthetic datasets that incorporate multi-band information. The estimated displacement ($D_{LOS}^{k}(\mathrm{X})$, $k = 1, \dots, K$) of each synthetic data point ($\mathrm{X} = (x, y)$) along the *k* LOS is determined by combining the ground deformation components (i.e., the vertical Up-Down and horizontal East-West displacement, $f_i(X), i = (1,2)$) and the direction cosine ($S_i^k, \mathrm{i} = (1,2)$) of the *k* LOS:

$$D_{LOS}^{k}(X) = S_1^k f_1(X) + S_2^k f_2(X), \; k = 1, \dots, K \tag{4.4}$$

The deformation components are estimated using the known *N* neighbour PS points ($X_n = (x_n, y_n)$, n=1, …, N) along each *k* LOS through the formulation of a first order Taylor polynomial:

$$D_{LOS}^{k}(X_n) = D_{LOS}^{k}(X) + \nabla D_{LOS}^{k}(X) \Delta X_n \tag{4.5}$$

where ΔX_n contains the relative distances between the X_n point components and the synthetic point X. The displacement map along the multi-satellite LOS is found by solving the Taylor system for the *K* LOS. The data fusion formulation considers two parameters: the maximum distance within which nearby PS measurements are selected to estimate deformations, and the locality factor that weights the contribution of each measurement based on its distance from the point estimation. The synthetic datasets of Up-Down and East-West components are derived into a regular grid of squared cells from the vector decomposition of the displacement map along the multi-satellite LOS.

4.3.3.3. Activity score

The vertical and horizontal components of ground displacement achieved by the fusion of S1 and CSK PS data were further processed with building footprints. Points contained within each polygon were interpolated to produce a raster with mean velocity values over the time span at 10 meters grid resolution. This procedure results in a series of velocity values for each element within the area of interest. With the aim of quantifying the state of activity of each element in reference to the hazard to which it is exposed, three specific values of velocity were retrieved from individual building statistics and assigned to the whole element at risk. For the landslide hazard, the maximum absolute velocity between the vertical and horizontal components was selected. Regarding the subsidence hazard, we computed the range between minimum and maximum vertical velocity, thus focusing on buildings experiencing distortion; and concerning the sinkhole hazard, the maximum negative velocity of the vertical component was preferred.

Once the specific hazard velocities were assigned to the entire built environment, we applied the same approach employed to rank potential damage. Therefore, Equation 4.6 was applied to derive class thresholds of the building velocity. According to the eCDF in Eq. 4.6, class thresholds were set at the respective percentile (p) value, and activity scores (A_b) ranging from 1 to 5 were defined.

$$A_b(h) = eCDF^{-1}\big(p,\ VEL(b|h)\big),\ p = (50, 82, 95, 97.5) \quad (4.6)$$

4.3.4. Multi-risk ranking

To derive multi-risk scores, we firstly ranked the specific risk of single hazards by combining the three risk components described above. Equation 4.7 defines the convolution of the hazard ($H_b(h)$), activity ($A_b(h)$), and potential damage ($D_b(h)$) that compose the single-risk score (R) of each exposed building (b).

$$R_b = H_b(h) \times A_b(h) \times D_b(h) \quad (4.7)$$

Secondly, starting from the single-risk evaluations (R_b), the multi-risk score at building level (MR_b) is defined as their summation corrected by a multiplication factor (F_h) (Equation 4.8). This coefficient weights the risk score by the corresponding hazard type, depending on whether it is a shock or a stress process. This is because stress phenomena usually cause fewer losses than shock ones and are also harder to mitigate. Regarding shock events such as landslides and sinkholes, the F_h coefficient is set equal to 1, thus keeping the risk score unchanged. Whereas, in the case of stress processes (e.g., subsidence) we decided to fix the F_h coefficient equal to 0.5.

$$MR_b = \sum_{All\ risks}(R_b \times F_h) \quad (4.8)$$

Finally, a spatial distribution of dominant risk among the built environment was retrieved by selecting the maximum value across single risk scores and assigning the corresponding hazard to it.

Once obtained, both the single- and multi-risk scores were rescaled to five classes using the eCDF based method and percentiles described above, thus clustering assets with risk level in the top 2.5%, 5%, 18%, 50%, and lower than median into scores ranging from 5 to 1, respectively.

Finally, to validate the results, financial data related to expenses incurred for remediation and mitigation measures within the AOI were retrieved from the ReNDiS database (Table 4.1) and compared with the multi-risk scores of nearby buildings.

4.4. Results

4.4.1. Hazard

The exposure analysis to ground instabilities of urban assets within the area of interest assigned a binary classification (i.e., exposed/not exposed) to each individual building. As a result, the threat to which most buildings are exposed proves to be the sinkhole hazard (Figure 4.5), as more than half of the elements were found to be at risk. Additionally, about 10% of the built-up area might be

affected by subsidence processes, while landslides pose a potential risk to about one-third of the buildings.

Further investigation into the specific elements exposed to each threat revealed their hazard score, which was subsequently mapped (Figure 4.6). Figure 4.6d reports the statistical distribution of the elements at risk across assigned hazard scores. Landslide-prone buildings are mainly ranked with a hazard score of 3, with just a few having the maximum hazard. Spatially, they spread throughout the northwestern and southern slopes (Figure 4.6a). The sinkhole hazard score seems to distribute almost equally, and the same applies to subsidence hazard. However, subsidence hazard spread along the valleys and paleo-valleys of the Tiber and Aniene Rivers (Figure 4.6b), whereas the highest sinkhole hazard scores concentrate among the S-E part of the city (Figure 4.6c).

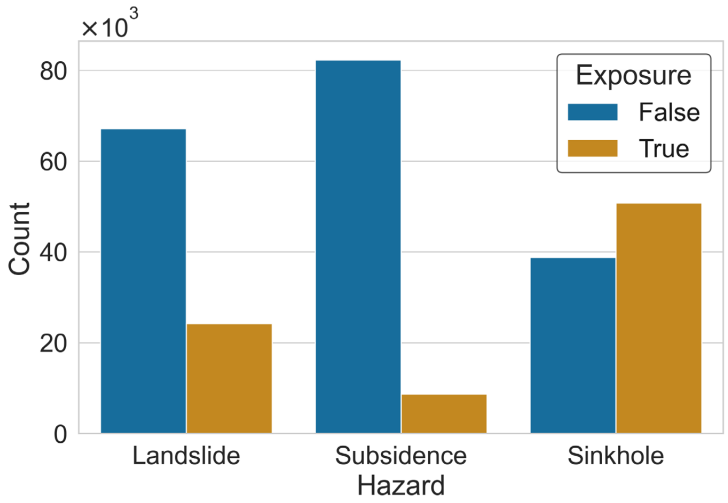


Figure 4.5: Overall number of buildings exposed (true) and unexposed (false) for each type of hazard investigated.

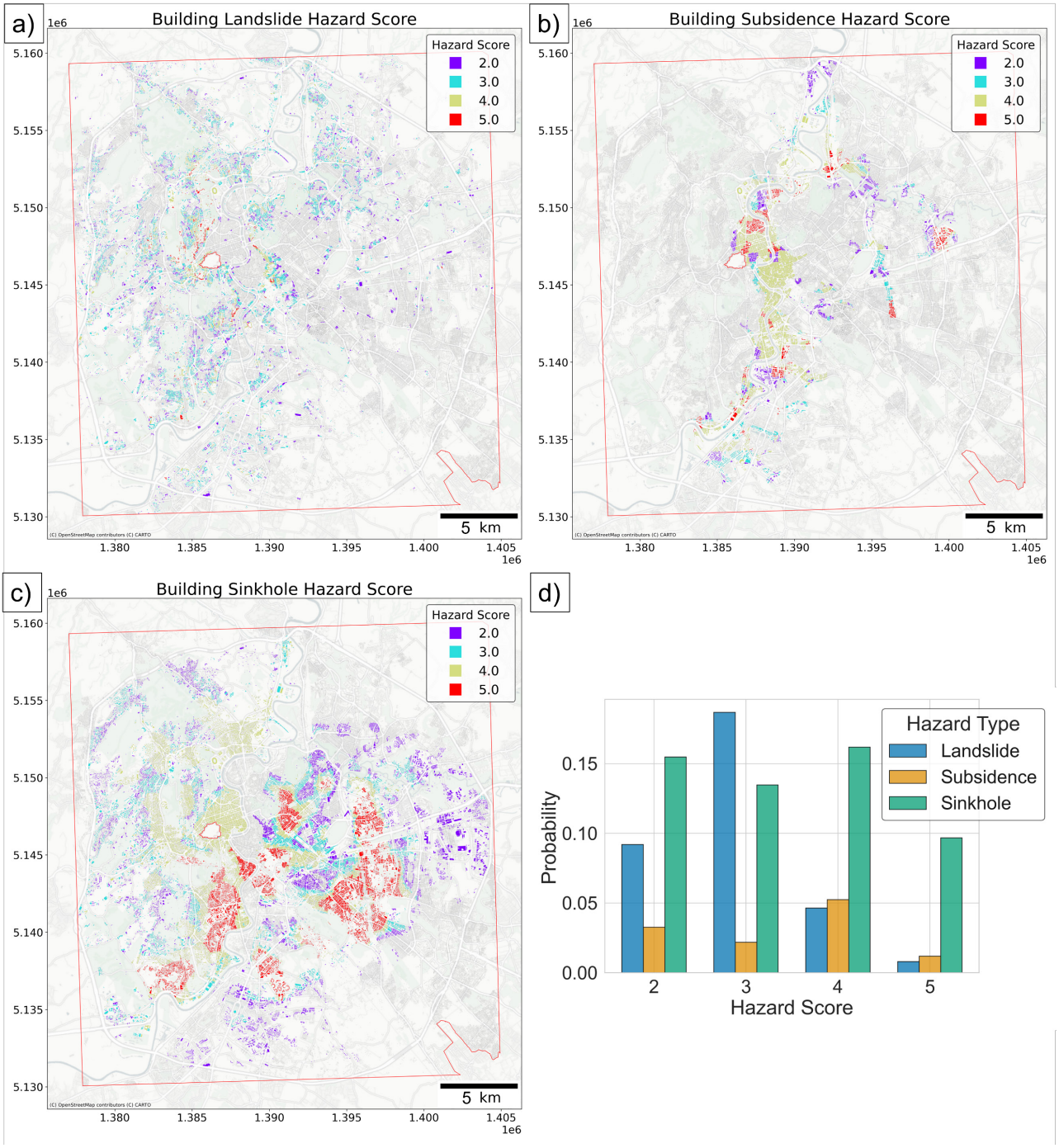


Figure 4.6: Hazard score of buildings for landslides (a), subsidence (b), and sinkholes (c). Statistical distribution of hazard score for elements at risk by hazard type (d). EPSG:3857.

4.4.2. Potential damage

Figure 4.7a illustrates the probability density function (PDF) of impact ratio values by different ground instability types. The distributions of impact ratios exhibit similarities between landslide and subsidence, with values spread over the range. A slight deviation towards higher values is observed for subsidence. Conversely, the impact ratio values for sinkholes tend to concentrate around 0.8.

Structural resistance values were obtained for both residential and non-residential buildings (i.e., productive, and commercial). In Table 4.2, the reported parameters reveal that resistance values for non-residential constructions consistently equate to 0.1, as no structural characteristics are included in the census tract data. Contrarily, for residential buildings, the resistance values can vary between 0 and 1.5. Figure 4.7b displays the histograms along with the PDF of computed structural resistance values specifically for residential buildings, which have a mean value of 0.9 and a standard deviation of 0.28.

Once reclassified, impact ratio and structural resistance were merged to form vulnerability classes using the contingency matrix presented in Table 4.3. The pairs of impact ratio and resistance values are illustrated in Figure 4.8a, with distinct colours indicating the assigned vulnerability value. This graphical representation effectively shows the boundaries of the vulnerability classes. The resultant multi-hazard physical vulnerability values (Figure 4.8b) exhibit a tendency to concentrate at 0 and 0.5 for landslides, with about 75% of the structures within the two classes. In the case of sinkhole events, about 55% of the buildings demonstrate a vulnerability value of 0.5 and 20% of 0.75. However, vulnerability to subsidence appears to be evenly distributed among values of 0, 0.5, and 0.75 for 80% of the buildings. Furthermore, it should be noted that among the others, subsidence hazard has the highest percentage of elements with vulnerability of 0.75 and 1.0, about 25% and 10% respectively.

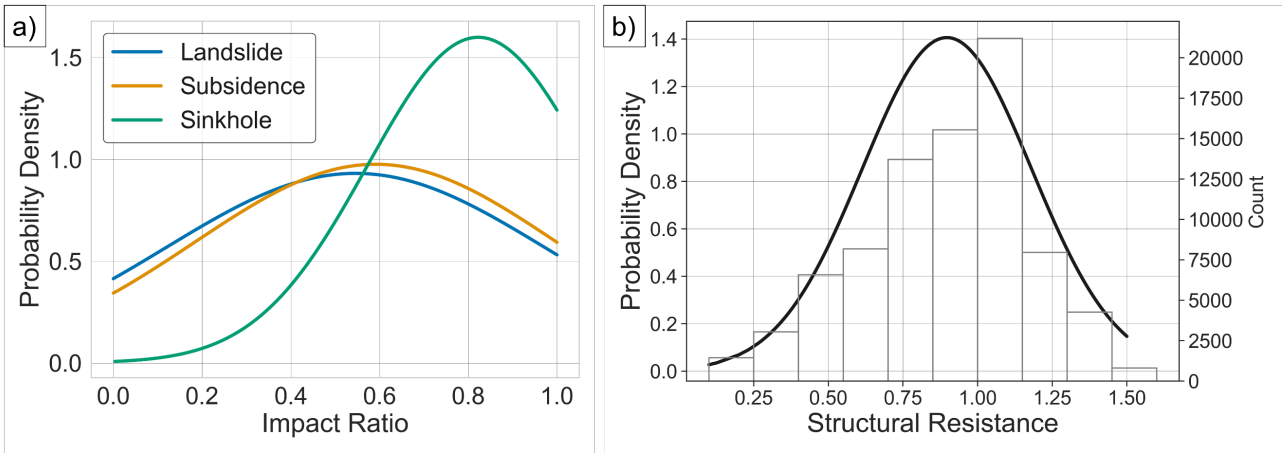


Figure 4.7: Probability density functions of impact ratio values by hazard type (a), and of structural resistance values for residential buildings.

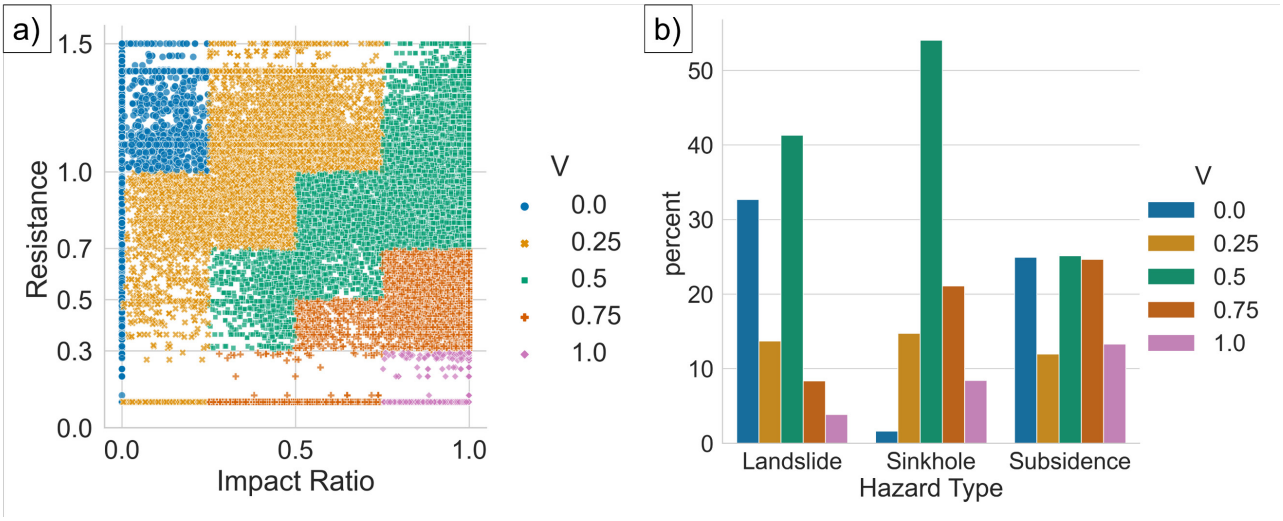


Figure 4.8: Graphical representation of the vulnerability value (V) assignment based on the contingency matrix reported in Table 3 (a). Distribution of vulnerability value of elements at risk by hazard type (b).

The analysis of real estate market data enabled the determination of economic value per square metre for each urban asset. Figure 4.9 presents the exposure of the built environment in the designated area of interest. Notably, there is a distinctive radial distribution pattern observed, with the highest values concentrated within the historic centre. These peak values exceed 7000 euro/m^2 and gradually decrease as one moves towards the suburbs. In suburban areas, the exposed value can be as low as 1000 euro/m^2. This applies particularly to the eastern sector, where a significant number of industrial activities are located.

The integration of multi-hazard vulnerability and exposure values enabled the computation of potential economic losses per square meter. Subsequently, these losses were reclassified into potential damage scores, as depicted in Figure 4.10. Each element at risk was assigned a damage score based on the specific threats it is exposed to. The mapped damage scores for landslide, subsidence, and sinkhole can be observed in Figure 4.10a, b, and c, respectively. Furthermore, Figure 4.10d presents the eCDF of resultant potential economic losses. These eCDF were utilized to identify the score class thresholds, which are indicated by markers along the curves.

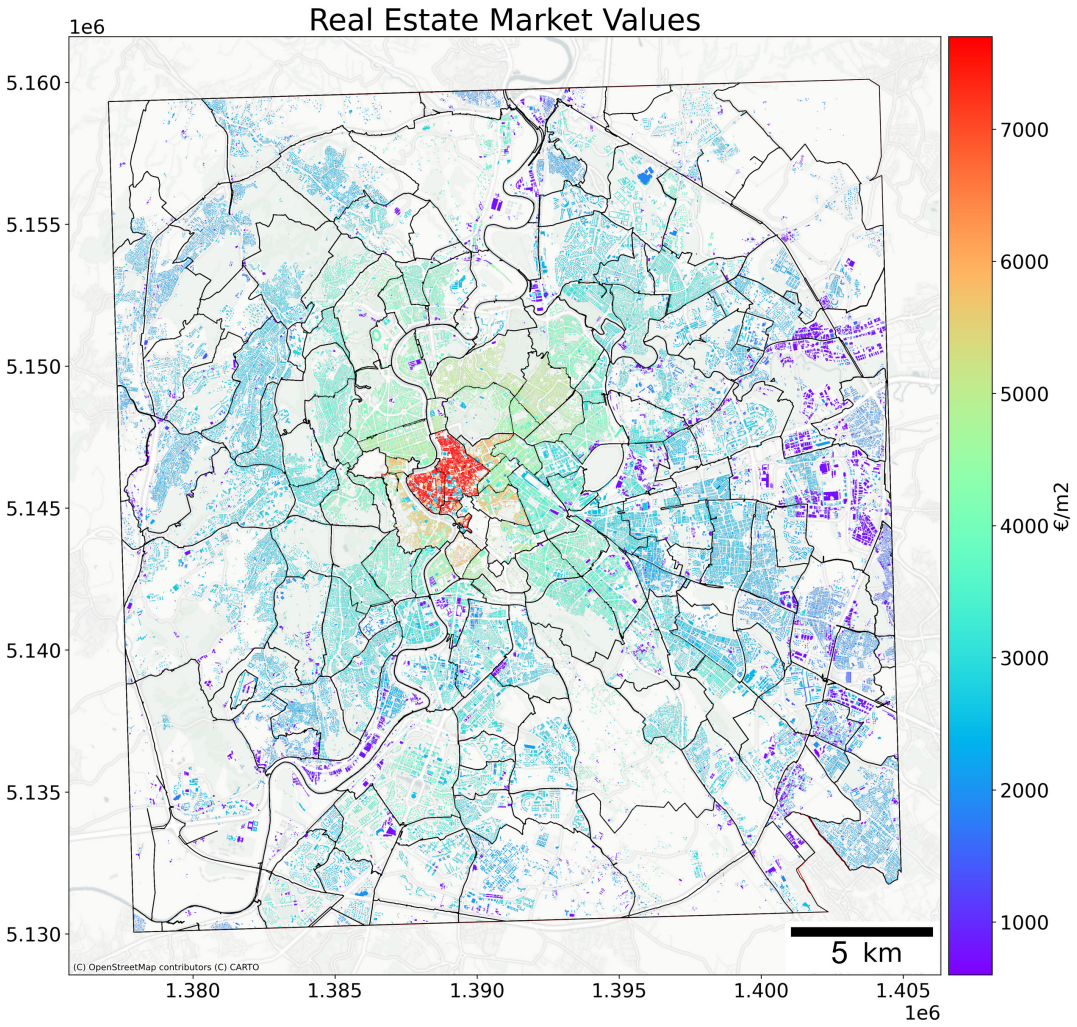


Figure 4.9: Real estate market values of buildings per square meter (€/m^2) within the study area according to their category of use and position. EPSG:3857.

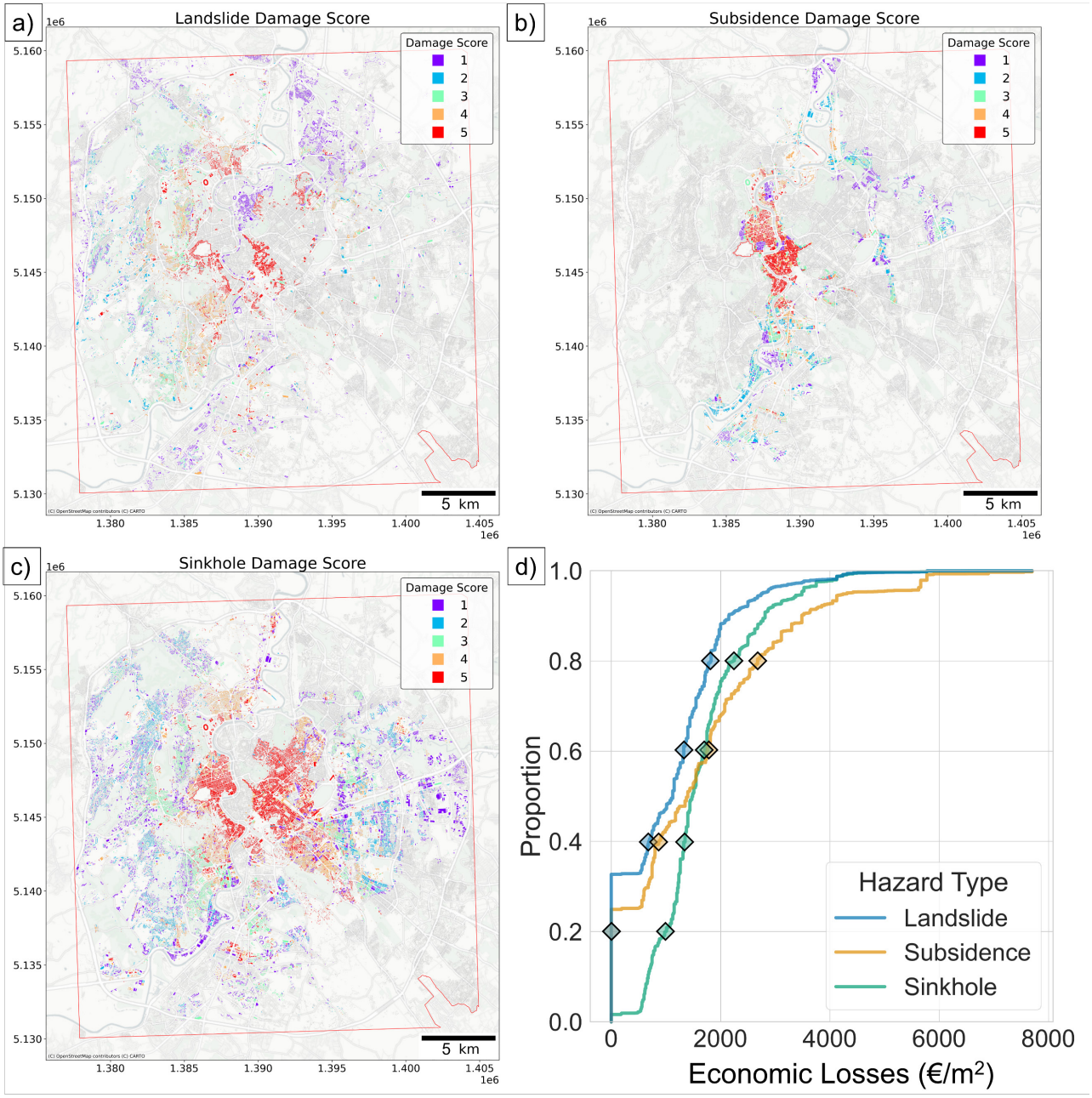


Figure 4.10: Potential damage score of buildings for landslides (a), subsidence (b), and sinkholes (c). Empirical cumulative distribution of economic losses per square metre selected according to hazard type (d). The coloured diamonds define percentile thresholds used to derive the five score classes. EPSG:3857.

4.4.3. State of activity

The synthetic datasets of vertical and horizontal components derived from the data fusion process of S1, and CSK PS time series are reported in Figure C2 (see Appendix C). The resulting displacement rates of the built environment are reported in Figure 4.11. A comparison between the two components reveals that displacements predominantly occur along the vertical axis (Figure 4.11a). The horizontal displacements exhibit minimal rates, peaking at a few mm per year (Figure 4.11b). Notably, the largest rates can be observed for assets situated along the Tiber River valley.

Based on the hazard-specific value distribution, activity scores were assigned to the building displacement rates. Figure 4.12a, b, and c depict the resulting activity scores for landslide, subsidence, and sinkhole, respectively. Figure 4.12d presents the eCDF of the quantitative absolute velocity values, which were assigned to individual assets based on the hazard under consideration.

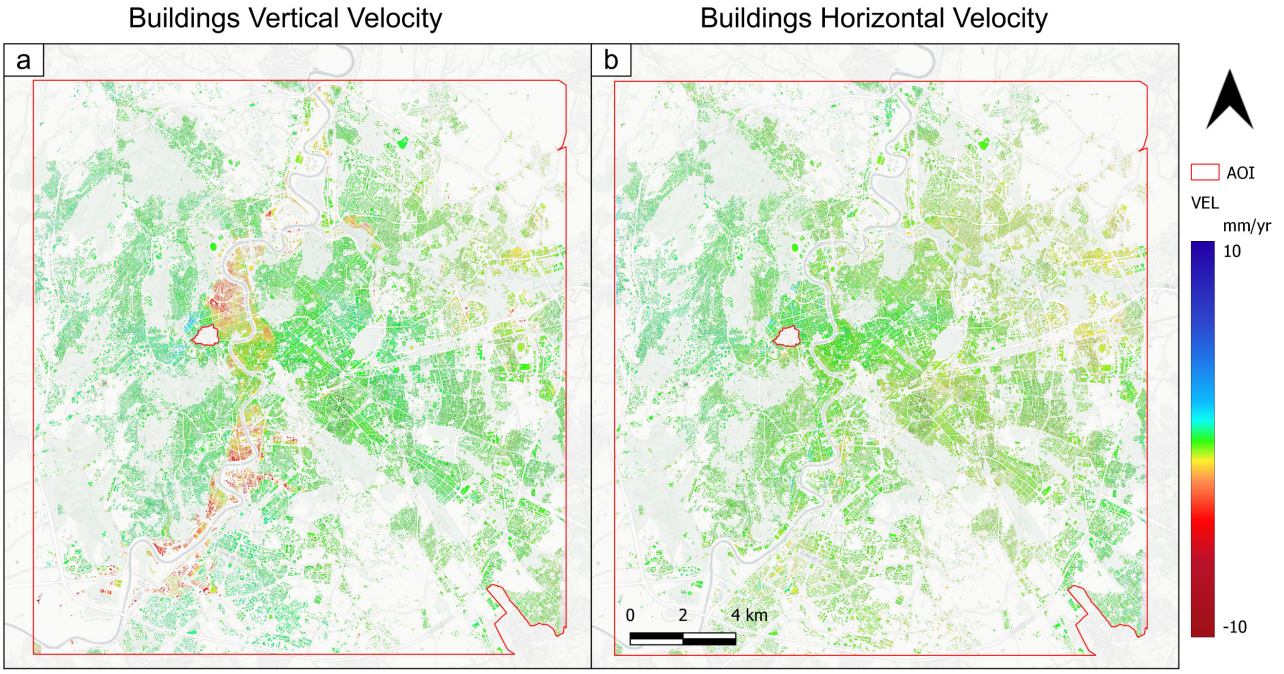


Figure 4.11: Buildings' vertical (a) and horizontal (b) displacement rates (i.e., velocity) derived from the grid-based synthetic datasets (Figure A2) and assigned to the entire built-environment under investigation. EPSG:4326.

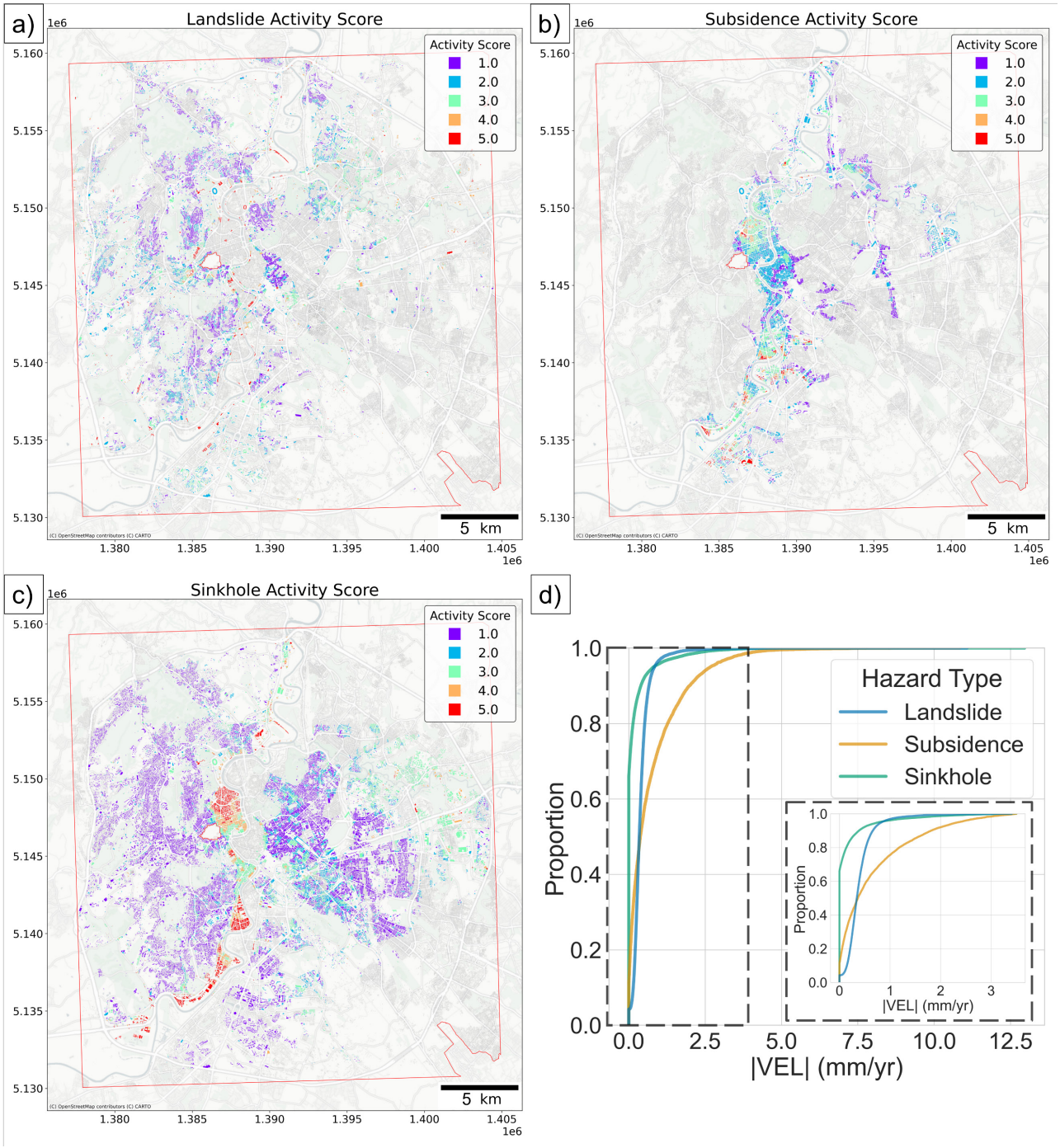


Figure 4.12: Activity score of buildings for landslides (a), subsidence (b), and sinkholes (c). Empirical cumulative distribution of absolute velocities values selected according to hazard type (d) employed to set up class thresholds. EPSG:3857.

4.4.4. Multi-risk

The rankings of landslide, subsidence, and sinkhole risks for individual buildings, derived from the combination of the corresponding hazard, damage, and activity scores, are illustrated in Figure 4.13a, b, and c, respectively. Moreover, the aggregation of single-risk scores using the appropriate weighting factor produced a multi-risk map, as shown in Figure 4.13d. Examining the building use categories alongside their associated risk scores reveals that retail and economic structures exhibit

the highest proportion of buildings with elevated risk scores, closely followed by offices (Figure C3 in Appendix C). In the case of residential buildings, they are distributed relatively evenly across the risk classes, although there is a slightly higher concentration within risk score 1. The dominant risk type of each element at risk is depicted in Figure 4.14a. Out of the approximately 90×10^3 urban assets analysed, it was determined that 60% of them are at risk. Within this group, 33% exhibit the highest risk score attributed to sinkhole phenomena, while 22% and 5% possess the highest score related to landslides and subsidence processes, respectively (Figure 4.14b). Delving into buildings' use categories, it results that the dominant risk type is not evenly distributed among them, thus landslide risk prevails for residential buildings, while subsidence and sinkhole risks predominate for retail and offices, respectively (Figure 4.14c).

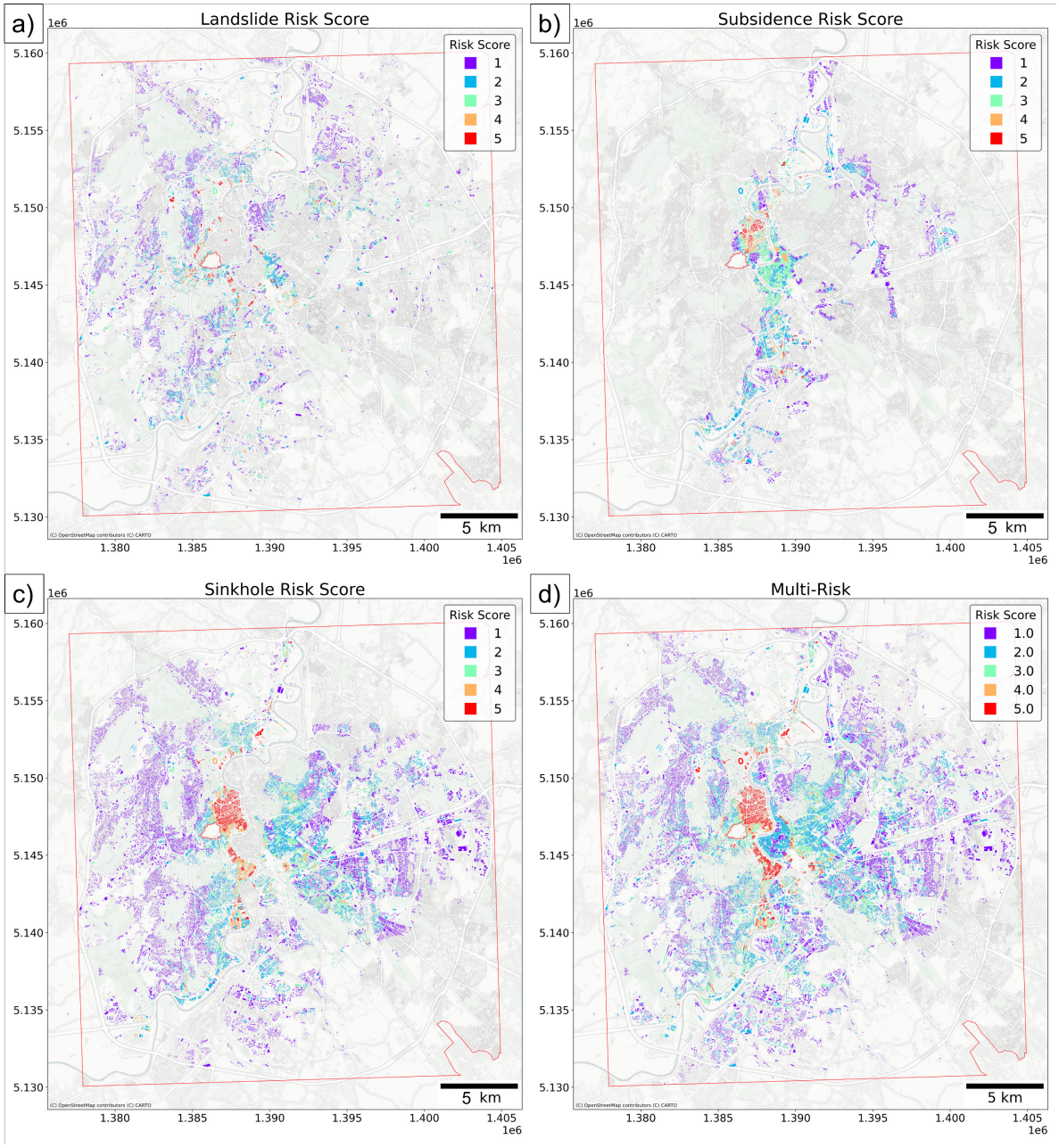


Figure 4.13: Risk score of buildings for landslides (a), subsidence (b), sinkholes (c) and multi-hazard (d).

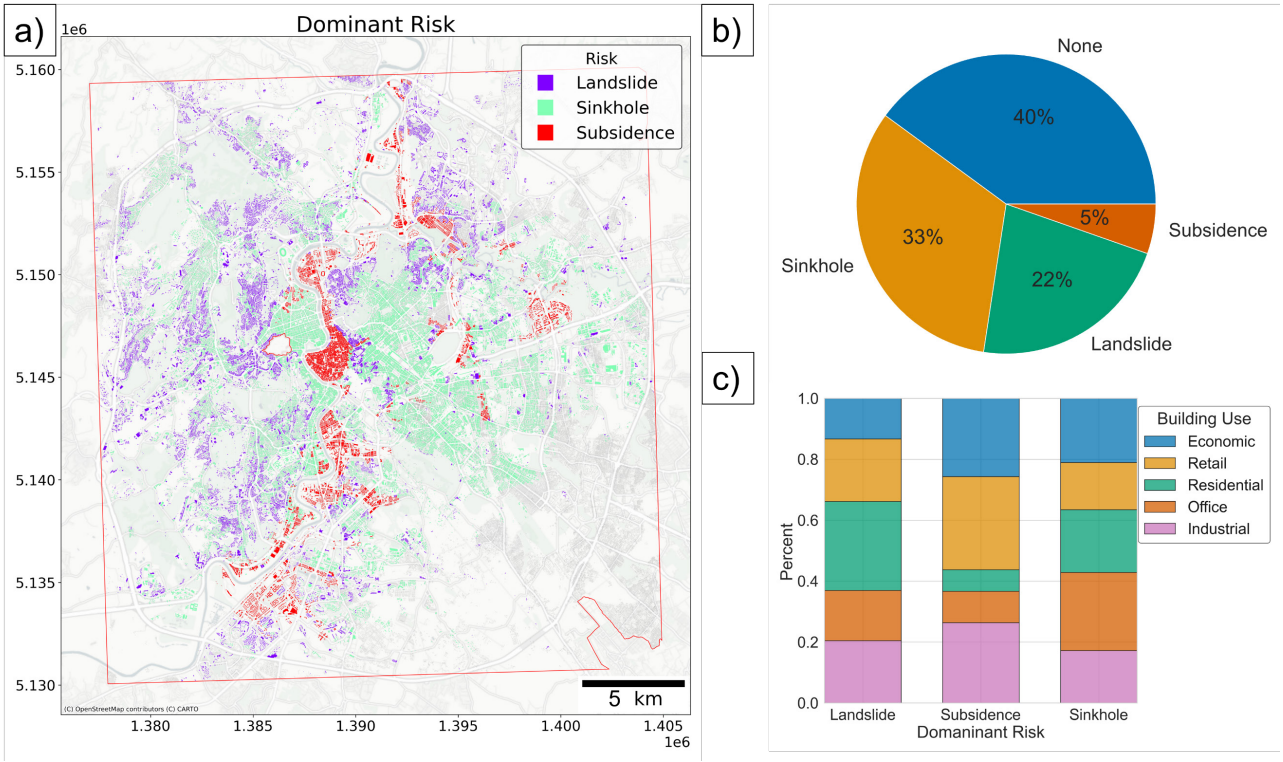


Figure 4.14: Spatial distribution of dominant risk threatening individual buildings (a) with a focus on the hazard type ratio among the overall elements (b), and among the buildings use categories (c).

4.5. Discussion

Risk assessment of natural hazards can guide preventive and corrective actions to reduce levels of vulnerability and hazard exposure of urban assets and populations. In this study, a semiquantitative multi-risk assessment was developed to provide a replicable and customisable framework for ranking and prioritising elements at risk based on official open-source data. The approach relies on three factors (i.e., hazard, potential damage, and state of activity) that are combined to compute the specific- and multi-risk score of urban buildings. Albeit both temporal and spatial information should be considered to perform a proper hazard assessment (Fell et al. 2008), data on the recurrence and magnitude of ground instabilities in Rome are lacking, except for landslides (Esposito et al. 2023). The investigated ground instabilities rely on completely different geological processes that affect the possibility of estimating return periods. Moreover, they don't seem to be triggered by any environmental variables. Hence, the hazard related score factors are only based on the spatial component (i.e., susceptibility) which was engineered to derive i) the probability of individual elements being involved in an event in the near future (hazard score), and ii) the building's area potentially affected and weighted by susceptibility levels, as a proxy for the expected magnitude, and thus for the destructiveness of the event (impact ratio). The potential damage score is built upon building's physical vulnerability and its economic value, which provide valuable results to quantify the potential economic losses. With respect to others (Papathoma-Köhle et al. 2007; Kappes et al.

2012a; Romão et al. 2016; Bianchini et al. 2017; Huang et al. 2023) deriving the physical vulnerability from structural resistance values and hazard intensity facilitated the computation at large scale, since it exploits features of the built environment based on open-source official census tracts of Italy. Moreover, structural resistance parameters can be easily modified to suit specific hazards. The last risk component, i.e., activity score, provides a dynamic layer that relies on displacement rates of the elements within hazardous area. It employs the distribution of building velocity values retrieved from multi-satellite PS time series. The activity score is based on the fusion of S1 and CSK PSs, thus exploiting the strengths and overcoming the weaknesses of each dataset. According to Costantini et al. 2017, PS density from CSK images (X-band) is higher by one order of magnitude respect to C-band SAR (e.g., S1), especially for surfaces with reduced radar backscattering coefficient and without strong point-like scatterers, thus helping to better detect ground deformation phenomena affecting small areas (Esposito et al. 2021; Festa et al. 2022). Moreover, CSK sensor has higher sensitivity in detecting small movements. Conversely, S1 has significantly higher frequency and regularity, thus increasing the applicability for early detection of failure precursors. Hence, the fusion of PSs retrieved from both satellites ensured higher resolution and spatial coverage to catch ground instability processes in detail. Both potential damage and activity scores derive from the exploitation of the empirical cumulative distribution function, which allows the selection of threshold values equal to the specified fraction of observations (e.g., percentile). According to Mastrantoni et al. (2023), percentiles must be chosen based on values distribution and expert judgment. Therefore, class thresholds can be easily adapted to other cities. For the city of Rome, the top 2.5% of buildings at multi-risk were selected as the highest class. However, the number of elements in the highest priority class may be set depending on the availability of funds for mitigation measures. The ReNDiS database states a total amount of expenses financed for mitigation measures within the area of interest of approximately 97 million euros, encompassing 71 projects. This data enables us to compare and validate the multi-risk outcomes using official financial records (Figure 4.15). The average cost of mitigation rises progressively as the multi-risk class of the nearby buildings increases, thus revealing a significant correlation between mitigation expenses and calculated multi-risk score that ensure the reliability of the proposed method.

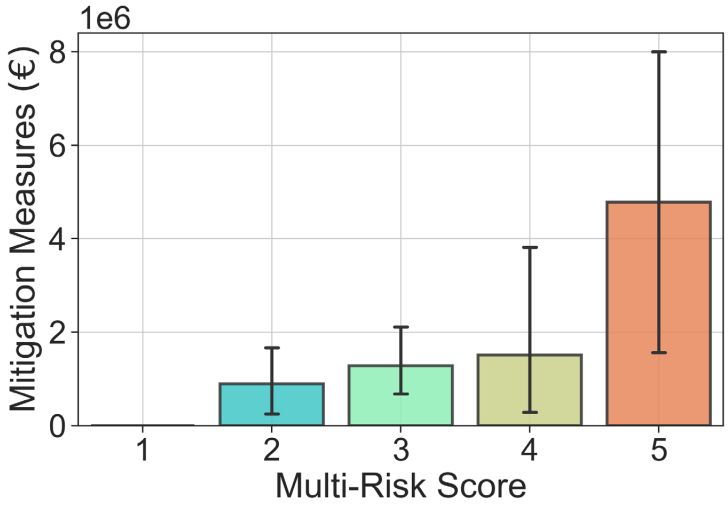


Figure 4.15: Average expenses (M€) and standard deviations of incurred mitigation measures within the study area compared with the multi-risk score of nearby buildings as computed in this work.

Multi-risk scores can also be analysed prior to reclassification, thus retrieving the buildings with the absolute highest indices (Figure 4.16). As expected, the most at-risk buildings may face all three ground instabilities investigated, although landslide and sinkhole risk generally overwhelm subsidence risk. This is mainly related to the F_h coefficient halving the risk of subsidence, since stress phenomena, such as subsidence, are hereby considered less urgent and impactful in terms of mitigation measures and expected economic losses respectively.

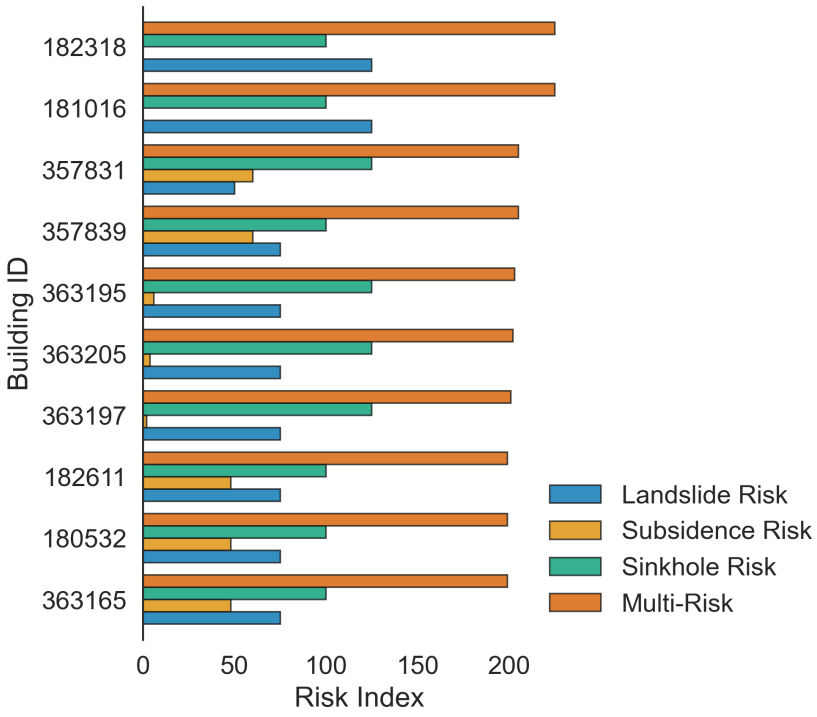


Figure 4.16: Ranking of the top 10 buildings with the highest absolute risk within the study area.

The developed model is intended as an alternative to other risk assessment approaches requiring very demanding tasks, such as quantitative probabilistic hazard scenarios and fragility curves. The workflow is completely automated and customisable. More hazards related to ground deformations can be easily added, albeit not considering potential interrelationships. Structural resistance parameters can be adapted to suit the hazard characteristics and percentile thresholds defining the scores can be modified to fit stakeholders needs. The data availability at national scale unlocks its replicability throughout Italy, thus enhancing the scalability of the approach. Moreover, by exploiting individual urban assets and their features allowed us to perform the analyses at the real scale (i.e., 1:1). Further advantages of the multi-risk ranking approach presented in this study are i) its complete objectivity in conducting the analyses without any assumption on the input data, ii) its site-specific risk prioritisation of structures for natural hazards mitigation, and iii) the possibility of updating the results using up-to-date satellite InSAR monitoring data. However, it has some limitations. Hazards interactions, and therefore amplified vulnerability and risk, are not evaluated due to the challenges of performing multi-hazard scenarios modelling for ground instabilities. The hazard component of the risk equation is derived from spatial probability without considering the temporal factor, thus static risk and multi-risk scores are defined. Although the application in the selected case study did not account for structural resistance parameters variations depending on hazard type, the optimization of such parameters based on historical data and expert judgement may improve the quality of results. Thus, future research and development will focus on the integration of additional geohazards, such as earthquakes and floods, and the optimization of building data, with the aim of also assessing cultural heritage, currently omitted due to the lack of official data on economic value and vulnerability. A huge effort will also be devoted to multi-hazard scenario modelling aimed at correlating the three components of the hazard (i.e., space, time, and intensity) and evaluating potential hazard interactions, which will unlock a quantitative multi-risk assessment. A key point is represented by the expected costs of mitigation actions. Their estimation will have a significant impact on stakeholders involved in risk management for large cities. Nowadays, ever more accurate data is being collected and released through databases often with open access. The new census tracts of Italy from ISTAT will contain data about hospitals, schools and other non-residential structures that may allow to better constrain their vulnerability to natural hazards. Eventually, knowing the heights of buildings will also be crucial to better model the physical vulnerability of buildings and therefore quantify the potential losses.

4.6. Conclusion

A novel semiquantitative model for multi-risk ranking of buildings exposed to ground instability hazards is presented and tested in this study. The developed approach, which is replicable on a national scale, was tested and validated in the centre of Rome, where it assessed approximately

90×10^3 buildings at a 1:1 scale. The integration of hazard, vulnerability, and activity score maps has yielded the first comprehensive multi-risk assessment for ground instability in Rome. Notably, 60% of the city's assets were found to be at risk, with residential buildings, offices, and retail primarily associated with landslide, sinkhole, and subsidence risk, respectively. The proposed procedure offers valuable support to risk managers and decision-makers by providing objective prioritization of elements at risk and identifying the dominant threats. This guidance can inform mitigation strategies, ultimately enhancing urban resilience and minimizing future economic losses. We achieved this by integrating and fusing various official databases and monitoring data accessible at the national scale, allowing for scalability and reproducibility of the analysis in Italian cities and potentially beyond. Our incorporation of activity rate data into the risk equation, as proposed by Varnes (1984), introduces a dynamic dimension to the multi-risk assessment, enabling regular updates over time. Empirical cumulative distribution functions and score classes based on percentile thresholds were applied to each risk component, facilitating the ranking of urban assets by single- and multi-risk levels. Financial data of incurred mitigation measures were exploited to validate the results. The findings of our study provide crucial insights into the potential hazards facing urban communities, enabling the identification of vulnerable assets and prioritization of mitigation measures. As urbanization continues to drive population concentration in large cities, our research carries significant implications for infrastructure management, disaster preparedness, and urban planning. By applying the developed multi-risk ranking model, city planners and policymakers can better adapt to the challenges posed by ground instabilities, enhancing urban resilience and safeguarding lives and critical infrastructure. Moreover, the methodology presented here can serve as a model for conducting similar assessments globally, fostering a collaborative effort to address geological multi-hazard risks and promote sustainable urban development. Ultimately, the integration of scientific research and geospatial data science methodologies, as demonstrated in this work, plays a vital role in building resilient societies prepared to face natural hazards.

Acknowledgements: The authors are grateful to the Italian Space Agency (ASI) for providing the Cosmo-SkyMed SAR images free of charge to the CERI research centre for academic research purposes (project card ID: 615).

Disclosure Statement: The authors report there are no competing interests to declare.

Data availability: All input data and the code are accessible online at the author's GitHub page (https://github.com/gmastrantoni/mhr).

Funding: This research is framed within the project "Geo-multi-hazard analysis through geospatial data driven methods for resilient cities" funded by Sapienza University of Rome. Project protocol number: RM12117A86DC0211.

References

Alessi D, Bozzano F, Lisa A, Esposito C, Fantini A, Loffredo A, Martino S, Mele F, Moretto S, Noviello A, et al. 2014. Geological risks in large cities: The landslides triggered in the city of Rome (Italy) by the rainfall of 31 January-2 February 2014. Italian Journal of Engineering Geology and Environment. 1. https://doi.org/10.4408/IJEGE.2014-01.O-02

Amanti M, Cesi C, Vitale V. 2008. Le frane nel territorio di Roma - Landslides distribution in the Roma municipality.

Annual report 2022 | Munich Re. [accessed 2023 Jun 20]. https://www.munichre.com/en/company/investors/reports-and-presentations/annual-report.html

Bell R, Glade T. 2004. Quantitative risk analysis for landslides ‒ Examples from Bíldudalur, NW-Iceland. Natural Hazards and Earth System Sciences. 4(1):117–131. https://doi.org/10.5194/nhess-4-117-2004

Bianchi Fasani G, Bozzano F, Cercato M. 2011. The underground cavity network of south-eastern Rome (Italy): an evolutionary geological model oriented to hazard assessment. Bull Eng Geol Environ. 70(4):533–542. https://doi.org/10.1007/s10064-011-0360-0

Bianchini S, Solari L, Casagli N. 2017. A GIS-Based Procedure for Landslide Intensity Evaluation and Specific risk Analysis Supported by Persistent Scatterers Interferometry (PSI). Remote Sensing. 9(11):1093. https://doi.org/10.3390/rs9111093

Bozzano F, Andreucci A, Gaeta M, Salucci R. 2000. A geological model of the buried Tiber River valley beneath the historical centre of Rome. Bull Eng Geol Env. 59(1):1–21. https://doi.org/10.1007/s100640000051

Bozzano F, Esposito C, Franchi S, Mazzanti P, Perissin D, Rocca A, Romano E. 2015a. Analysis of a Subsidence Process by Integrating Geological and Hydrogeological Modelling with Satellite InSAR Data. In: Lollino G, Manconi A, Guzzetti F, Culshaw M, Bobrowsky P, Luino F, editors. Engineering Geology for Society and Territory - Volume 5. Cham: Springer International Publishing; p. 155–159. https://doi.org/10.1007/978-3-319-09048-1_31

Bozzano F, Esposito C, Franchi S, Mazzanti P, Perissin D, Rocca A, Romano E. 2015b. Understanding the subsidence process of a quaternary plain by combining geological and hydrogeological modelling with satellite InSAR data: The Acque Albule Plain case study. Remote Sensing of Environment. 168:219–238. https://doi.org/10.1016/j.rse.2015.07.010

Bozzano F, Esposito C, Mazzanti P, Patti M, Scancella S. 2018. Imaging Multi-Age Construction Settlement Behaviour by Advanced SAR Interferometry. Remote Sensing. 10(7):1137. https://doi.org/10.3390/rs10071137

Brabb EE. 1984. Innovative approaches to landslide hazard and risk mapping. [place unknown]: Proceedings of the 4th international symposium on landslides, Toronto, Canada, 1, pp. 30-32.

Caleca F, Tofani V, Segoni S, Raspini F, Rosi A, Natali M, Catani F, Casagli N. 2022. A methodological approach of QRA for slow-moving landslides at a regional scale. Landslides. 19(7):1539–1561. https://doi.org/10.1007/s10346-022-01875-x

Campolunghi MP, Capelli G, Funiciello R, Lanzini M. 2007. Geotechnical studies for foundation settlement in Holocenic alluvial deposits in the City of Rome (Italy). Engineering Geology. 89(1):9–35. https://doi.org/10.1016/j.enggeo.2006.08.003

Cascini L. 2008. Applicability of landslide susceptibility and hazard zoning at different scales. Engineering Geology. 102(3–4):164–177. https://doi.org/10.1016/j.enggeo.2008.03.016

Chang L, Hanssen RF. 2014. Detection of cavity migration and sinkhole risk using radar interferometric time series. Remote Sensing of Environment. 147:56–64. https://doi.org/10.1016/j.rse.2014.03.002

Cigna F, Lasaponara R, Masini N, Milillo P, Tapete D. 2014. Persistent Scatterer Interferometry Processing of COSMO-SkyMed StripMap HIMAGE Time Series to Depict Deformation of the Historic Centre of Rome, Italy. Remote Sensing. 6(12):12593–12618. https://doi.org/10.3390/rs61212593

Ciotoli G, Di Loreto E, Finoia MG, Liperi L, Meloni F, Nisio S, Sericola A. 2016. Sinkhole susceptibility, Lazio Region, central Italy. Journal of Maps. 12(2):287–294. https://doi.org/10.1080/17445647.2015.1014939

Ciotoli G, Loreto E, Liperi L, Meloni F, Nisio S, Sericola A. 2014. Carta dei Sinkhole Naturali del Lazio 2012 e sviluppo futuro del Progetto Sinkholes Regione Lazio.

Cleveland RB, Cleveland WS, McRae JE, Terpenning I. 1990. STL: A seasonal-trend decomposition. J Off Stat. 6(1):3–73.

Corominas J, van Westen C, Frattini P, Cascini L, Malet J-P, Fotopoulou S, Catani F, Van Den Eeckhaut M, Mavrouli O, Agliardi F, et al. 2014. Recommendations for the quantitative analysis of landslide risk. Bull Eng Geol Environ. 73(2):209–263. https://doi.org/10.1007/s10064-013-0538-8

Costantini M, Ferretti A, Minati F, Falco S, Trillo F, Colombo D, Novali F, Malvarosa F, Mammone C, Vecchioli F, et al. 2017. Analysis of surface deformations over the whole Italian territory by interferometric processing of ERS, Envisat and COSMO-SkyMed radar data. Remote Sensing of Environment. 202:250–275. https://doi.org/10.1016/j.rse.2017.07.017

Costantini M, Minati F, Trillo F, Ferretti A, Novali F, Passera E, Dehls J, Larsen Y, Marinkovic P, Eineder M, et al. 2021. European Ground Motion Service (EGMS). In: 2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS. [place unknown]; p. 3293–3296. https://doi.org/10.1109/IGARSS47720.2021.9553562

Costantini M, Zhu M, Huang S, Bai S, Cui J, Minati F, Vecchioli F, Jin D, Hu Q. 2018. Automatic Detection of Building and Infrastructure Instabilities by Spatial and Temporal Analysis of Insar Measurements. In: IGARSS 2018 - 2018 IEEE International Geoscience and Remote Sensing Symposium. [place unknown]; p. 2224–2227. https://doi.org/10.1109/IGARSS.2018.8518270

Crosetto M, Solari L, Balasis-Levinsen J, Bateson L, Casagli N, Frei M, Oyen A, Moldestad DA, Mróz M. 2021. DEFORMATION MONITORING AT EUROPEAN SCALE: THE COPERNICUS GROUND MOTION SERVICE. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences. XLIII-B3-2021:141–146. https://doi.org/10.5194/isprs-archives-XLIII-B3-2021-141-2021

Del Monte M, D'Orefice M, Luberti GM, Marini R, Pica A, Vergari F. 2016. Geomorphological classification of urban landscapes: the case study of Rome (Italy). Journal of Maps. 12(sup1):178–189. https://doi.org/10.1080/17445647.2016.1187977

Delgado Blasco JM, Foumelis M, Stewart C, Hooper A. 2019. Measuring Urban Subsidence in the Rome Metropolitan Area (Italy) with Sentinel-1 SNAP-StaMPS Persistent Scatterer Interferometry. Remote Sensing. 11(2):129. https://doi.org/10.3390/rs11020129

Dilley M. 2005. Natural Disaster Hotspots: A Global Risk Analysis. [place unknown]: World Bank Publications.

Esposito C, Belcecchi N, Bozzano F, Brunetti A, Marmoni GM, Mazzanti P, Romeo S, Cammillozzi F, Cecchini G, Spizzirri M. 2021. Integration of satellite-based A-DInSAR and geological modeling supporting the prevention from anthropogenic sinkholes: a case study in the urban area of Rome. Geomatics, Natural Hazards and Risk. 12(1):2835–2864. https://doi.org/10.1080/19475705.2021.1978562

Esposito C, Mastrantoni G, Marmoni GM, Antonielli B, Caprari P, Pica A, Schilirò L, Mazzanti P, Bozzano F. 2023. From theory to practice:

optimisation of available information for landslide hazard assessment in Rome relying on official, fragmented data sources. Landslides [Internet]. [accessed 2023 Jun 26]. https://doi.org/10.1007/s10346-023-02095-7

Fell R, Corominas J, Bonnard C, Cascini L, Leroi E, Savage WZ. 2008. Guidelines for landslide susceptibility, hazard and risk zoning for land-use planning. Engineering Geology. 102(3–4):99–111. https://doi.org/10.1016/j.enggeo.2008.03.014

Ferretti A, Fumagalli A, Passera E, Rucci A. 2022. Insar Data Calibration in Wide Area Processing. In: IGARSS 2022 - 2022 IEEE International Geoscience and Remote Sensing Symposium. [place unknown]; p. 5101–5104. https://doi.org/10.1109/IGARSS46834.2022.9884822

Ferretti A, Prati C, Rocca F. 2001. Permanent scatterers in SAR interferometry. IEEE Transactions on Geoscience and Remote Sensing. 39(1):8–20. https://doi.org/10.1109/36.898661

Festa D, Bonano M, Casagli N, Confuorto P, De Luca C, Del Soldato M, Lanari R, Lu P, Manunta M, Manzo M, et al. 2022. Nation-wide mapping and classification of ground deformation phenomena through the spatial clustering of P-SBAS InSAR measurements: Italy case study. ISPRS Journal of Photogrammetry and Remote Sensing. 189:1–22. https://doi.org/10.1016/j.isprsjprs.2022.04.022

Frankenberg E, Sikoki B, Sumantri C, Suriastini W, Thomas D. 2013. Education, Vulnerability, and Resilience after a Natural Disaster. Ecology and Society [Internet]. [accessed 2023 Jun 20] 18(2). https://doi.org/10.5751/ES-05377-180216

Giampaolo V, Capozzoli L, Grimaldi S, Rizzo E. 2016. Sinkhole risk assessment by ERT: The case study of Sirino Lake (Basilicata, Italy). Geomorphology. 253:1–9. https://doi.org/10.1016/j.geomorph.2015.09.028

Gill JC, Duncan M, Ciurean R, Smale L, Stuparu D, Schlumberger J, de Ruiter M, Tiggeloven T, Torresan S, Gottardo S, et al. 2022. D1.2 Handbook of multi-hazard, multi-risk definitions and concepts [Internet]. [accessed 2023 Jun 18]. https://nora.nerc.ac.uk/id/eprint/533237/

Greiving S. 2006. Integrated risk assessment of multi-hazards: a new methodology. Special Paper-Geological Survey of Finland. 42:75.

Guglielmino F, Nunnari G, Puglisi G, Spata A. 2011. Simultaneous and Integrated Strain Tensor Estimation From Geodetic and Satellite Deformation Measurements to Obtain Three-Dimensional Displacement Maps. IEEE Transactions on Geoscience and Remote Sensing. 49(6):1815–1826. https://doi.org/10.1109/TGRS.2010.2103078

Guzzetti F. 2000. Landslide fatalities and the evaluation of landslide risk in Italy. Engineering Geology. 58(2):89–107. https://doi.org/10.1016/S0013-7952(00)00047-8

Hanssen RF. 2001. Radar Interferometry: Data Interpretation and Error Analysis. [place unknown]: Springer Science & Business Media.

Hochrainer-Stigler S, Šakić Trogrlić R, Reiter K, Ward PJ, de Ruiter MC, Duncan MJ, Torresan S, Ciurean R, Mysiak J, Stuparu D, Gottardo S. 2023. Toward a framework for systemic multi-hazard and multi-risk assessment and management. iScience. 26(5):106736. https://doi.org/10.1016/j.isci.2023.106736

https://www.sarinterferometry.com/ps-toolbox/. SAR PS ToolBox [Internet]. [accessed 2023 Jul 26]. https://www.sarinterferometry.com/ps-toolbox/

Huang B, Shu L, Yang YS. 2012. Groundwater Overexploitation Causing Land Subsidence: Hazard Risk Assessment Using Field Observation and Spatial Modelling. Water Resour Manage. 26(14):4225–4239. https://doi.org/10.1007/s11269-012-0141-y

Huang S, Dou H, Jian W, Guo C, Sun Y. 2023. Spatial prediction of the geological hazard vulnerability of mountain road network using machine learning algorithms. Geomatics, Natural Hazards and Risk. 14(1):2170832. https://doi.org/10.1080/19475705.2023.2170832

Hyndman RJ, Athanasopoulos G. 2018. Forecasting: principles and practice. [place unknown]: OTexts.

Johnson K, Depietri Y, Breil M. 2016. Multi-hazard risk assessment of two Hong Kong districts. International Journal of Disaster Risk Reduction. 19:311–323. https://doi.org/10.1016/j.ijdrr.2016.08.023

Julià PB, Ferreira TM. 2021. From single- to multi-hazard vulnerability and risk in Historic Urban Areas: a literature review. Nat Hazards. 108(1):93–128. https://doi.org/10.1007/s11069-021-04734-5

Kappes MS, Keiler M, von Elverfeldt K, Glade T. 2012b. Challenges of analyzing multi-hazard risk: a review. Nat Hazards. 64(2):1925–1958. https://doi.org/10.1007/s11069-012-0294-2

Kappes MS, Papathoma-Köhle M, Keiler M. 2012a. Assessing physical vulnerability for multi-hazards using an indicator-based methodology. Applied Geography. 32(2):577–590. https://doi.org/10.1016/j.apgeog.2011.07.002

Komendantova N, Mrzyglocki R, Mignan A, Khazai B, Wenzel F, Patt A, Fleming K. 2014. Multi-hazard and multi-risk decision-support tools as a part of participatory risk governance: Feedback from civil protection stakeholders. International Journal of Disaster Risk Reduction. 8:50–67. https://doi.org/10.1016/j.ijdrr.2013.12.006

Li Z, Nadim F, Huang H, Uzielli M, Lacasse S. 2010. Quantitative vulnerability estimation for scenario-based landslide hazards. Landslides. 7(2):125–134. https://doi.org/10.1007/s10346-009-0190-3

Liu X, Chen H. 2019. Integrated assessment of ecological risk for multi-hazards in Guangdong province in southeastern China. Geomatics, Natural Hazards and Risk. 10(1):2069–2093. https://doi.org/10.1080/19475705.2019.1680450

Liu Z, Nadim F, Garcia-Aristizabal A, Mignan A, Fleming K, Luna BQ. 2015. A three-level framework for multi-risk assessment. Georisk: Assessment and Management of Risk for Engineered Systems and Geohazards. 9(2):59–74. https://doi.org/10.1080/17499518.2015.1041989

Luo H, Chen T. 2016. Three-Dimensional Surface Displacement Field Associated with the 25 April 2015 Gorkha, Nepal, Earthquake: Solution from Integrated InSAR and GPS Measurements with an Extended SISTEM Approach. Remote Sensing. 8(7):559. https://doi.org/10.3390/rs8070559

Marzocchi W, Garcia-Aristizabal A, Gasparini P, Mastellone ML, Di Ruocco A. 2012. Basic principles of multi-risk assessment: a case study in Italy. Nat Hazards. 62(2):551–573. https://doi.org/10.1007/s11069-012-0092-x

Mastrantoni G, Marmoni GM, Esposito C, Bozzano F, Scarascia Mugnozza G, Mazzanti P. 2023. Reliability assessment of open-source multiscale landslide susceptibility maps and effects of their fusion. Georisk: Assessment and Management of Risk for Engineered Systems and Geohazards. 0(0):1–18. https://doi.org/10.1080/17499518.2023.2251139

McGlade J, Bankoff G, Abrahams J, Cooper-Knock SJ, Cotecchia F, Desanker P, Erian W, Gencer E, Gibson L, Girgin S. 2019. Global assessment report on disaster risk reduction 2019.

Mohebbi Tafreshi G, Nakhaei M, Lak R. 2021. Land subsidence risk assessment using GIS fuzzy logic spatial modeling in Varamin aquifer, Iran. GeoJournal. 86(3):1203–1223. https://doi.org/10.1007/s10708-019-10129-8

Papathoma-Köhle M, Neuhäuser B, Ratzinger K, Wenzel H, Dominey-Howes D. 2007. Elements at risk as a framework for assessing the vulnerability of communities to landslides. Natural Hazards and Earth System Sciences. 7(6):765–779. https://doi.org/10.5194/nhess-7-765-2007

Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V, et al. 2011. Scikit-learn: Machine Learning in Python. MACHINE LEARNING IN PYTHON.:6.

Romão X, Paupério E, Pereira N. 2016. A framework for the simplified risk analysis of cultural heritage assets. Journal of Cultural Heritage. 20:696–

708. https://doi.org/10.1016/j.culher.2016.05.007

Shi Y, Tang Y, Lu Z, Kim J-W, Peng J. 2019. Subsidence of sinkholes in Wink, Texas from 2007 to 2011 detected by time-series InSAR analysis. Geomatics, Natural Hazards and Risk. 10(1):1125–1138. https://doi.org/10.1080/19475705.2019.1566786

Terzi S, Torresan S, Schneiderbauer S, Critto A, Zebisch M, Marcomini A. 2019. Multi-risk assessment in mountain regions: A review of modelling approaches for climate change adaptation. Journal of Environmental Management. 232:759–771. https://doi.org/10.1016/j.jenvman.2018.11.100

Tilloy A, Malamud BD, Winter H, Joly-Laugel A. 2019. A review of quantification methodologies for multi-hazard interrelationships. Earth-Science Reviews. 196:102881. https://doi.org/10.1016/j.earscirev.2019.102881

UNISDR W. 2012. Disaster risk and resilience. Thematic think piece, UN system task force on the post-2015 UN development agenda.

Varnes DJ. 1984. LANDSLIDE HAZARD ZONATION: A REVIEW OF PRINCIPLES AND PRACTICE. Natural Hazards.(3):63.

Ward PJ, Blauhut V, Bloemendaal N, Daniell JE, de Ruiter MC, Duncan MJ, Emberson R, Jenkins SF, Kirschbaum D, Kunz M, et al. 2020. Review article: Natural hazard risk assessments at the global scale. Natural Hazards and Earth System Sciences. 20(4):1069–1096. https://doi.org/10.5194/nhess-20-1069-2020

Wei L, Hu K, Hu X, Wu C, Zhang X. 2022. Quantitative multi-hazard risk assessment to buildings in the Jiuzhaigou valley, a world natural heritage site in Western China. Geomatics, Natural Hazards and Risk. 13(1):193–221. https://doi.org/10.1080/19475705.2021.2004244

van Westen C, Kappes MS, Luna BQ, Frigerio S, Glade T, Malet J-P. 2014. Medium-Scale Multi-hazard Risk Assessment of Gravitational Processes. In: Van Asch T, Corominas J, Greiving S, Malet J-P, Sterlacchini S, editors. Mountain Risks: From Prediction to Management and Governance [Internet]. Dordrecht: Springer Netherlands; [accessed 2023 Jun 28]; p. 201–231. https://doi.org/10.1007/978-94-007-6769-0_7

White GF, Kates RW, Burton I. 2001. Knowing better and losing even more: the use of knowledge in hazards management. Global Environmental Change Part B: Environmental Hazards. 3(3):81–92. https://doi.org/10.3763/ehaz.2001.0308

Zisman ED. 2012. A Method of Quantifying Sinkhole Risk. :278–287. https://doi.org/10.1061/41003(327)27.

CHAPTER 5

5. Conclusions and Perspectives

Throughout this doctoral thesis, we embarked on a comprehensive exploration of ground instability hazards in urban areas, with a particular focus on the Municipality of Rome, Italy. Ground instability hazards pose a significant threat to urban environments, and our research provided a replicable and customizable framework to address this challenge. We aimed to offer practical guidance for decision-makers and risk managers by providing a thorough understanding of shallow landslide hazard and building multi-risk. We have investigated the synergistic power of machine learning, historical-statistical pluviometric analysis, and official open data for the purpose of shallow landslide hazard assessment. Recognizing the importance of the temporal component of the hazard, we introduced temporal attributes and ranked hazard zones. Additionally, we explored the impact of mapping units and analysis scale on prediction accuracy, and the potential of merging multiscale maps. Lastly, we developed a novel semi-quantitative multi-risk model that harnesses the power of open data integration to rank and prioritise elements at risk from landslide, sinkhole, and subsidence processes.

Below I summarized the main findings related to the research questions posed in Chapter 1.3.

(1) HOW CAN OFFICIAL OPEN DATA BE EXPLOITED TO DEVELOP RELIABLE LANDSLIDE HAZARD MODELS VALUABLE FOR STATUTORY PURPOSES?

Open data can serve as a foundation for developing landslide hazard models at the urban scale. In Chapter 2 we acknowledged the intrinsic challenges of data quality, including incompleteness, heterogeneity, and errors in open-source landslide inventories. Thus, we initiated a comprehensive approach to address these issues. We recognised that improving reliability necessitated the integration of low-accuracy datasets in terms of both quality and quantity. However, this may still not be sufficient to obtain an appropriate geodatabase. Hence, we developed innovative methodologies to derive synthetic landslide initiation points (LIPs), thereby enhancing the overall consistency of the database and facilitating more accurate modelling (Figure 5.1). Machine learning algorithms represent a solution to resolve spatial hazard, enabling the creation of accurate landslide susceptibility zoning. Multi-annual landslide inventories and historical-statistical analysis of pluviometric records complements these approaches by providing insights into hazard recurrence and temporal patterns (Figure 5.2). By integrating these components, it becomes possible to create valuable landslide hazard maps that capture landslide scenarios of specific intensities and pluviometric events, thus informing statutory decision-making and risk management.

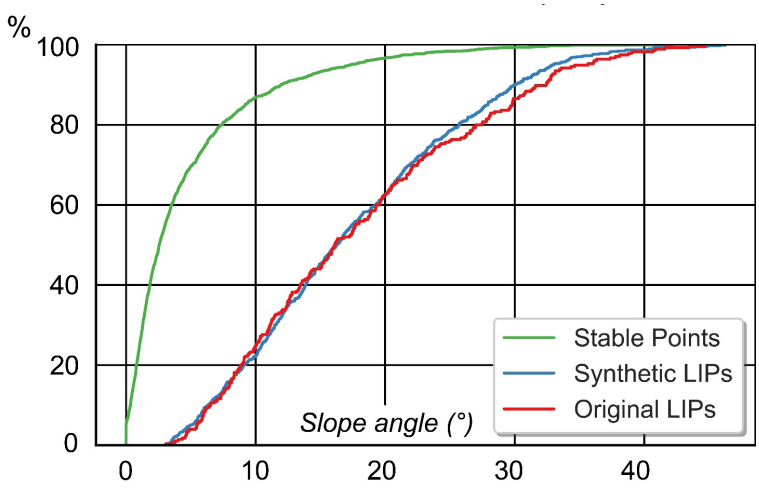


Figure 5.1: Distribution of Landslide Initiation Points (LIPs) and stable points vs slope angle adopted in the susceptibility model.

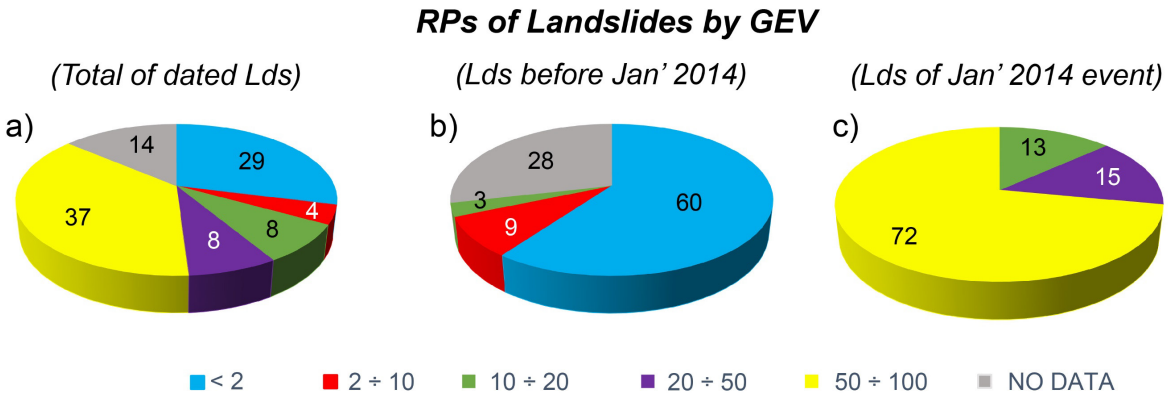


Figure 5.2: Pie chart showing the percentage of landslides with different Return Periods (RPs) considering the entire database (a), before the 2014 event (b), and during January 2014 meteorological event (c).

(2) WHAT IS THE IMPACT OF MAPPING UNIT AND ANALYSIS SCALE ON THE ACCURACY OF LANDSLIDE SUSCEPTIBILITY ZONING (LSZ)?

In Chapter 3 we extended beyond model development and incorporated a comparative analysis of our urban-scale landslide susceptibility zoning (LSZ) with other available open-source maps at varying spatial scales. By employing empirical cumulative distribution functions, statistical and spatial similarity analyses, we assessed the predictive performance of these maps. Figure 5.3 shows the proportional increasement of the F1-score with the scale of analysis ranging from global to regional. Therefore, the employed scale plays a crucial role in accurately localizing potential landslide sources and stable areas. Among the mapping units investigated, grid based LSZ perform better when predicting landslide initiation points, and the larger the scale the higher the accuracy of predictions. The slope unit LSZ turns out to be the closest to the benchmark when predicting the entire area potentially affected, represented by landslide polygons, and this is due to the uniform hazard level assigned to the whole slope. However, all the investigated maps struggled when it comes to detect stable areas, thus resulting in high false positive rates (i.e., low precision) and a significant overestimation of landslide hazard. This is probably due to the resolution and distribution of input data which inevitably are biased and incomplete when very large areas must be considered.

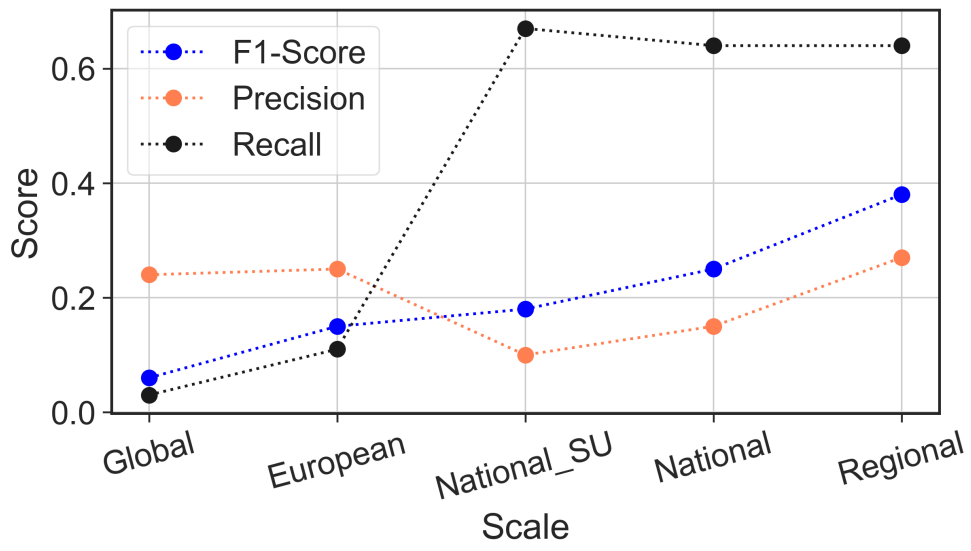


Figure 5.3: Evaluation metrics obtained from the spatialized accuracy assessment of open-source LSZ at different scales of analysis.

(3) CAN THE FUSION OF MULTIPLE LOW-RESOLUTION LSZ OVERCOME THE LACK OF SITE-SPECIFIC STUDIES?

In Chapter 3, we have observed the poor performance of low-resolution landslide susceptibility maps when applied to urban environments. Hence, we introduced a data fusion approach for merging a variety of low accuracy products with the purpose of improving their predictive performance and thus gaining reliability. The retrieved F1-scores provided an overall picture of the performance (Figure 5.4). This approach has the potential to improve the classification of landslide occurrences (i.e., recall) while maintaining precision. As a result, the map that comes closest to the benchmark is provided by the fusion of regional and national grid based maps, improving the performance of the best original map. However, false positive rate remains as high as for original maps, thus affecting the overall performance. Hence, the effectiveness of LSZ data fusion may vary depending on the specific maps being fused and the criteria used for fusion. Therefore, the achieved performance gains, although promising, are not sufficient to justify the deployment of the developed criteria, due to the high number of false positives. Nevertheless, in the absence of urban and regional scale products, the fusion of national, European, and global maps can contribute to better predicting landslide occurrence in urban areas.

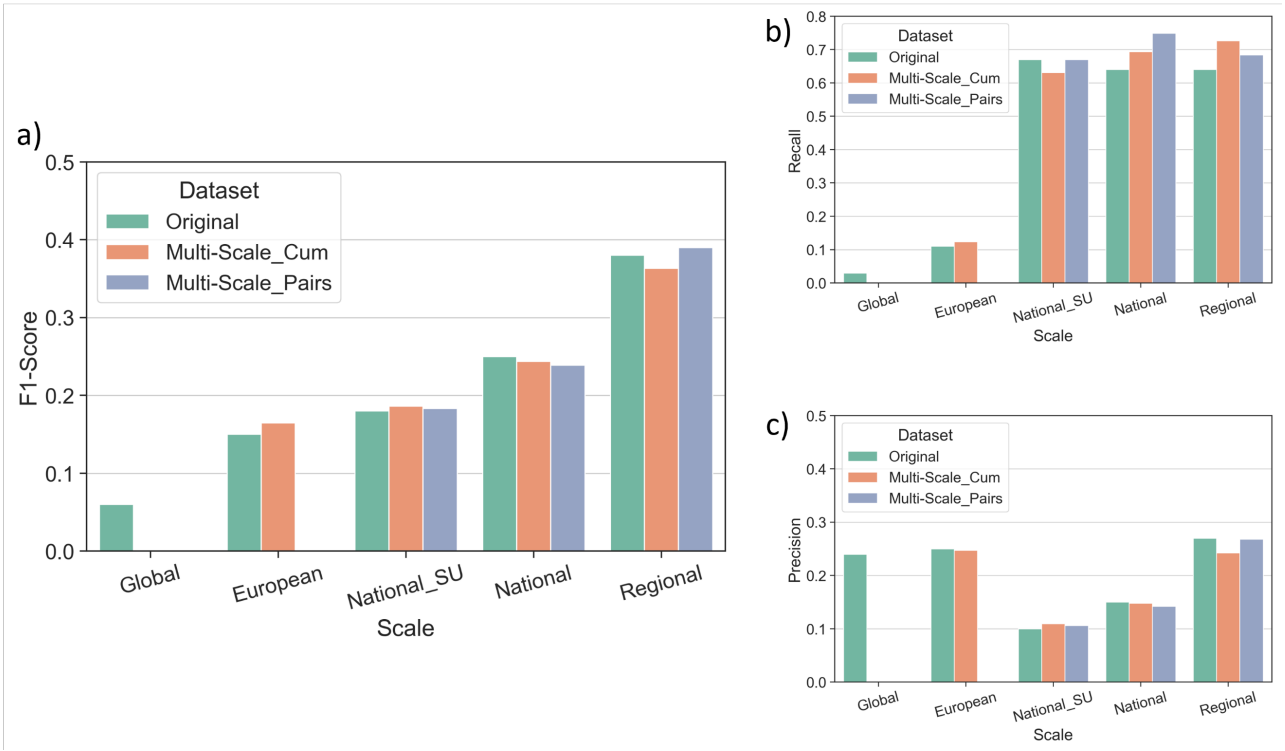


Figure 5.4: Performance achieved by original and multiscale landslide susceptibility maps expressed as F1-score (a), Recall (b), and Precision (c), computed with reference to the urban scale zonation carried out in this research.

(4) HOW CAN SUSCEPTIBILITY MAPS, CENSUS DATA, AND SATELLITE INTERFEROMETRY CONTRIBUTE TO MULTI-RISK ASSESSMENT IN THE URBAN ENVIRONMENT?

In chapter 4, we have developed a novel semi-quantitative model for multi-risk assessment of individual buildings which takes as input hazard maps of ground instabilities, building characteristics, assets market values, and satellite InSAR PS time series of displacement. Each dataset was processed to compute the respective factor within the risk equation. The factors are spatial hazard, potential damage, and the state of activity, respectively, that are combined to compute the specific- and multi-risk level of urban buildings (Figure 5.5). With these data at hand, an automated and scalable score-based approach was designed to obtain a multi-risk ranking of urban assets. By taking advantage of empirical cumulative distribution functions, percentile thresholds were set to specific fraction of observation to retrieve the respective priority score. Recognizing the limitations in data availability related to ground instability hazards, we focused on the spatial component of the hazard. Hence, the hazard score is based on susceptibility, allowing us to estimate the probability of individual buildings being involved in an event in the near future, and its potential magnitude. Potential economic losses are based on vulnerability and exposure values, which are derived from structural resistance parameters, hazard intensity, and real estate market values, respectively. This process was allowed from data contained within Italian census tracts, which facilitated the computation at large scale. Displacement time series were processed to retrieve the mean velocities of buildings along the horizontal and vertical axis, which generated hazard-specific activity scores. Incorporating the activity score contributes a dynamic dimension, enabling regular updates over time, thus capturing real-time changes in hazard activity. Therefore, by computing and integrating semi-quantitative scores, various types of open data, such as the ones used in this research, can be harmonized to create a multi-risk priority index valuable to guide mitigation strategies.

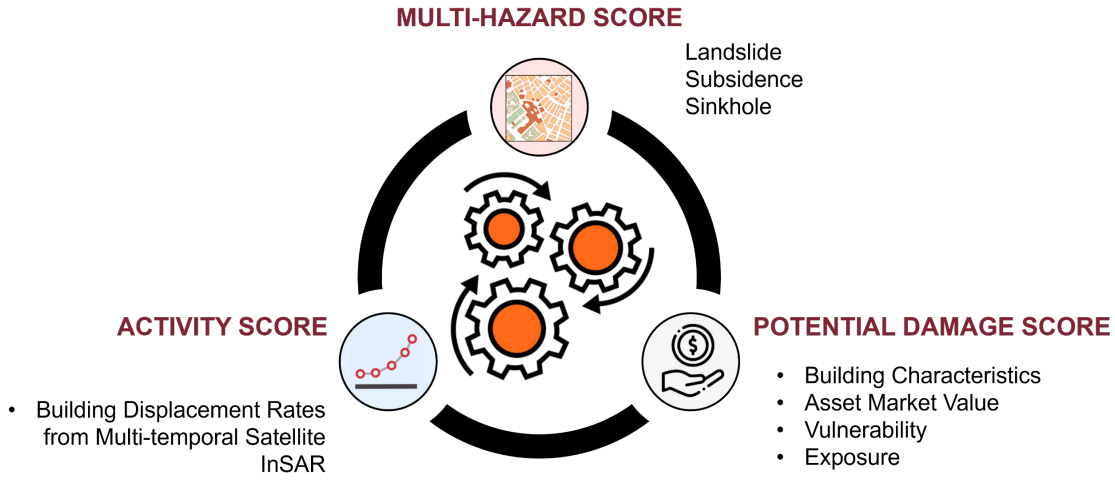


Figure 5.5: Framework for the semi-quantitative multi-risk modelling of buildings at city level based on the integration of three risk factors namely hazard, potential damage, and activity score.

(5) HOW CAN SINGLE- AND MULTI-RISK ANALYSIS INFORM STRATEGIES FOR ENHANCING URBAN RESILIENCE AND MINIMIZING ECONOMIC LOSSES?

In order to be implemented on the operational level by policy making institutions, multi-risk assessment must provide easily interpretable outcomes informing new solutions and prioritization of investments for prevention and mitigation measures aimed at reducing future economic losses. To this purpose, semi-quantitative approaches may be the best choice for combining rigorous data-driven analysis and comprehensible results. The presented model for ranking urban assets exposed to multiple risks offers a valuable support to risk managers and decision-makers by identifying the dominant threats affecting vulnerable assets and providing objective prioritization to multi-risk mitigation measures. This means that risk managers can allocate resources effectively, develop targeted resilience strategies, and safeguard both human lives and infrastructures from the impact of ground instability hazards. A further advantage of the presented approach includes the scalability and reproducibility throughout Italy thanks to the availability of input data and to the automated processing. Moreover, the introduction of the activity rate into the risk equation introduces a dynamic dimension given by satellite remote sensing monitoring data, thus enabling regular updates of the ranking over time. Hence, the model's scalability, customizability, and objectivity make it a practical tool for urban risk assessment, representing a practical alternative to other risk assessment approaches that require extensive data and complex probabilistic hazard scenarios. Eventually, by adopting this approach, similar assessments can be carried out globally, fostering collaboration in addressing natural hazards and promoting a safer and more sustainable urban future.

The insights garnered from this doctoral research have far-reaching implications for the field of geological risk analysis and assessment. Chapter 2 underscores the importance of integrating fragmented data sources for hazard assessment. This approach can be extended to various geological phenomena, enhancing data quality and reliability for geoscientific research. Chapter 3 has demonstrated the efficacy of machine learning techniques in landslide susceptibility mapping. The optimization and fusion of predictive maps can enhance overall accuracy, providing valuable tools for landslide hazard assessment in urban areas and beyond. Chapter 4 introduces a pioneering model for multi-risk assessment of urban buildings. By prioritizing elements at risk, this model empowers decision-makers and urban planners to proactively address ground instability hazards, ultimately enhancing urban resilience. In conclusion, this thesis represents a comprehensive exploration of geological multi-hazard risk at urban scale, highlighting the potential of data-driven approaches, open data integration, and advanced risk assessment models to address geological hazards in urban environments. The insights generated from these studies contribute to our understanding of the field and offer practical solutions for mitigating risks associated with ground instability, promoting a closer collaboration between scientific research, operative civil protection,

and political decision-making. As we continue to grapple with urbanization and the challenges it poses, the knowledge and methodologies presented here pave the way for more informed decision-making and safer, more resilient cities worldwide.

5.1. Future Research

While this study has made significant advancements in addressing multi-risk assessment for ground instability hazards in urban areas, several limitations and avenues for future research should be acknowledged. These include the need to consider interactions between different hazards, incorporate temporal factors into hazard assessment, and optimise structural vulnerability through expert judgement and fragility curves. Hence, future research and development efforts will centre on integrating additional geohazards within the data-driven model, such as earthquakes and floods, and refining building data, with the goal of including assessments for cultural heritage. Currently, this aspect is omitted due to the unavailability of official data on economic value and vulnerability. We will also dedicate considerable effort to multi-hazard scenario modelling, aiming to establish correlations between the three components of the hazard (space, time, and intensity) and assess potential interactions, ultimately enabling a quantitative multi-risk assessment. Of outmost importance is the estimation of mitigation costs in various scenarios, as this will significantly impact stakeholders involved in risk management for large cities. In light of the growing availability of accurate data through open-access databases, we anticipate that the upcoming census tracts from ISTAT in Italy will provide valuable data about hospitals, schools, and other non-residential assets, allowing for a more comprehensive assessment of their vulnerability to natural hazards. Additionally, having access to building height data will be crucial for enhancing the modelling of physical vulnerability, ultimately leading to better quantification of potential economic losses.

A further addition would be the integration of population factors. Key aspects to consider include: i) social vulnerability, and ii) population density dynamics. The former can be assessed by examining demographic factors like age, socioeconomic status, and ethnicity. This involves understanding how different social groups are differently affected by hazards, considering factors such as access to resources, health status, and resilience to disaster impacts. The latter is crucial as it affects exposure levels to various hazards at different times. For instance, a commercial area may have a high population density during the day but low at night, altering the risk profile significantly. This could involve developing models that predict population movement and density at different times of the day or in different scenarios, thereby providing a more accurate assessment of exposure to hazards. By understanding where and when the population is most vulnerable, resources can be allocated more effectively, and response plans can be more precisely targeted.

Appendix A

Supplementary material from Chapter 1.

Table A1: Repository of geospatial resources collected on the web during the research work.

DB	*Reference*	*Category*	*Covered Area*	*Type*	*Access Type*	*Key Words*
Agenzia Entrate. Quotazioni immobiliari	https://www.agenziaentrate.gov.it	Buildings	Italy	Inventory	Open	Real estate market
DBSN	https://www.igmi.org/it/dbsn-database-di-sintesi-nazionale/	Cartography	Italy	Inventory	No	geodatabase, cartography, topography, aerial photos
Global Administrative Units	https://gadm.org/	Cartography	Global	Inventory	Open	Global Administrative Units
ISTAT- basi territoriali	https://www.istat.it	Cartography	Italy	Inventory	Open	Administrative limits, census, Italian localities, toponymy
OpenStreet Map	https://www.openstreetmap.org	Cartography	Global	Inventory	Open	buildings, street map
C3S	https://climate.copernicus.eu/	Climate	Global	Inventory + hazard + risk	Open	atmosphere, land, ocean, climate indices, forecast
MISTRAL	https://www.cineca.it/news/mistral-portale	Climate	Italy	Inventory	Open (Account needed)	meteorological dataset,
Catalogo parametrico dei terremoti italiani	https://emidius.mi.ingv.it/CPTI15-DBMI15/	Earthquakes	Italy	Inventory	Open	Italian Historical Earthquakes
Classificazione Sismica aggiornata al 2022	https://rischi.protezionecivile.gov.it/	Earthquakes	Italy	Hazard	Open	Seismic classification, Seismic zones

Database Macrosismico Italiano	https://emidius.mi.ingv.it/CPTI15-DBMI15/	Earthquakes	Italy	Inventory	Open	Italian Earthquakes, macroseismic intensity
DISS	https://diss.ingv.it/	Earthquakes	Italy	Inventory	Open	seismogenic sources, active faults, subduction
ESM database	https://esm-db.eu/	Earthquakes	Europe	Inventory	Open	waveforms, ground-motion data
ESM database	https://esm-db.eu/	Earthquakes	Middle East	Inventory	Open	waveforms, ground-motion data
Europe Earthquake Hazard	https://datacatalog.worldbank.org/	Earthquakes	Europe	Hazard	Open	earthquake hazard, Europe, seismic hazard model
European Database of Seismogenic Faults (EDSF)	https://data.ingv.it/	Earthquakes	Europe	Inventory	Open	seismogenic faults, Europe,
European Fault-Source Model 2020	https://seismofaults.eu/	Earthquakes	Europe	Inventory	Open	seismogenic sources, active faulting, fault geometry
FRA.SI	https://frasi-project.irpi.cnr.it/	Earthquakes	Italy	Inventory	Open	seismic-triggered landslides, Italy
GEM - Global Active Faults	https://github.com/GEMScienceTools/gem-global-active-faults	Earthquakes	Global	Inventory	Open	active faults dataset
GEM - OpenQuake Engine	https://github.com/gem/oq-engine	Earthquakes	Global	Inventory + Hazard	Open	seismic hazard, seismic risk, global earthquakes
GEM - OpenQuake Explore	https://platform.openquake.org/	Earthquakes	Global	Inventory + Hazard	Open (Account needed)	global, earthquake

GEM- Global Earthquake Maps	https://www.globalquakemodel.org/	Earthquakes	Global	Inventory + Hazard	Partially	global earthquake
GEM- Global Seismic Hazard Map	https://maps.opennquake.org/	Earthquakes	Global	Hazard	Open	earthquake hazard, global earthquake map
Global Earthquake Hazard Distribution	https://sedac.ciesin.columbia.edu/	Earthquakes	Global	Hazard	Open	global earthquake hazard, peak ground acceleration
Hazard. FEHR	http://hazard.efehr.org/	Earthquakes	Europe	Hazard	Open	seismic hazard, seismic risk, hazard spectra, curves, maps
HIKE European Fault Database	https://geoera.eu/	Earthquakes	Europe	Inventory	Open	tectonics, faults system, fault data database
ITACA	https://itaca.mi.ingv.it/ItacaNet_32/	Earthquakes	Italy	Inventory	Open (Account needed)	accelerometric waveforms, seismic motion
Italy Vs30	https://data.mendeley.com/	Earthquakes	Italy	Inventory	Open	Geotechnical Engineering, Seismic Hazard Assessment, Building Safety Risk Assessment
MPS04-S1	http://esse1-gis.mi.ingv.it/	Earthquakes	Italy	Hazard	Open	Seismic hazard, Exceedance probability
Global Significant Earthquake Database	https://www.ngdc.noaa.gov/hazel/	Earthquakes	Global	Inventory	Open	global significant earthquakes
NOAA Natural Hazards	https://www.ncei.noaa.gov/	Earthquakes	Global	Inventory	Open	Global database, Natural Hazard, Historical Earthquakes, Tsunamis, Volcanoes
Risk.EFEHR	http://risk.efehr.org/	Earthquakes	Europe	Hazard + risk	Open	seismic risk, earthquake scenarios, exposure,

						vulnerability, fragility
USGS-Latest Earthquake	https://earthquake.usgs.gov/	Earthquakes	Global	Inventory	Open	Earthquakes Map, World, Seismicity, Seismotectonic
USGS-Earthquake Catalog	https://earthquake.usgs.gov/	Earthquakes	Global	Inventory	Open	earthquake catalogue, global earthquake map
Zone Sismiche INGV	http://zonesismiche.mi.ingv.it/	Earthquakes	Italy	Hazard	Open	Seismic zones, hazard, maximum ground acceleration
European Ground Motion Service	https://land.copernicus.eu/	EO data	Europe	Inventory	Open	interferometric data, ground motion
Geoportale Nazionale	http://www.pcn.minambiente.it/	EO data	Italy	Inventory	Open (Account needed)	lidar, interferometry, satellite data
Regione Toscana - PS	https://geoportale.lamma.rete.toscana.it/difesa_suolo/	EO data	Italy	Inventory	Open	Satellite monitoring, ground motion
Regione Valle D'Aosta - PSInSAR dati	http://geologiavda.partout.it/	EO data	Italy	Inventory	Open (Account needed)	Persistent scatter, deformation map
Regione Veneto - PS	https://idt2.regione.veneto.it/	EO data	Italy	Inventory	Open	Satellite interferometry, deformation maps, Persistent Scatter
Global Rainfall Erosivity	https://esdac.jrc.ec.europa.eu/	Erosion	Global	Inventory	Open	rainfall erosivity, R-factor map, land use
Global Soil Erosion Map	https://data.jrc.ec.europa.eu/	Erosion	Global	Inventory	Open	land use, soil erosion
Soil erosion by water	https://esdac.jrc.ec.europa.eu/	Erosion	Europe	Inventory	Open	soil erodibility, rain erosivity

Fire Danger Forecast	https://effis.jrc.ec.europa.eu/	Fires	Europe	Inventory	Open	fire hazard, fire danger
GEOPORTALE INCENDI BOSCHIVI	https://geoportale.incendiboschivi.it/	Fires	Italy	Inventory	Open (Account needed)	fire hazard, Italian fire catalogue
ISTAT-STATISTICHE GEOGRAFICHE SUI COMUNI	https://www.istat.it/	Geography	Italy	Inventory	Open	Census, population, municipality, territory
Global Lithological Map Database v1.0	https://doi.pangaea.de/	Geology	Global	Inventory	Open	lithological map,
One Geology	http://www.onegeology.org	Geology	Global	Inventory	Open	geological maps, geoscience data
Slope Unit	https://geomorphology.irpi.cnr.it/	Geomorphology	Italy	Inventory	Open	slope unit, geomorphology, terrain classification
Catasto frane alpi	https://geoclimalp.irpi.cnr.it	Hydrogeological instabilities	Italy	Inventory	Open	Landslide
European Flood Awareness System	https://www.efas.eu/	Hydrogeological instabilities	Europe	Inventory + Hazard	Partially	flood hazard, rainfalls
European Landslide Susceptibility Map version 2 ELSUS	https://esdac.jrc.ec.europa.eu/	Hydrogeological instabilities	Europe	Inventory + Hazard	Open	ELSUS, landslide susceptibility mapping, Europe
Global Active Archive of Large Flood Events	https://data.amerigeoss.org/	Hydrogeological instabilities	Global	Inventory	Open	flood events, flood hazard
Global Flood Database v1	https://global-flood-database.cloudtostreet.ai	Hydrogeological instabilities	Global	Inventory	Open	global, flood database, flood events
Global Flood Map	https://www.ambientalrisk.com/	Hydrogeological instabilities	Global	Hazard	No	flood hazard, flood map, potential flood risk

HydroSHEDS	https://www.hydrosheds.org/	Hydrogeological instabilities	Global	Inventory	Open	hydro-environmental model,
IdroGEO	https://idrogeo.isprambiente.it/	Hydrogeological instabilities	Italy	Inventory + Hazard	Open	Landslides, Floods
ITALICA	https://zenodo.org/	Hydrogeological instabilities	Italy	Inventory	Open	Landslide, Rainfall, Italy
Landslide Hazard Program	https://www.usgs.gov/	Hydrogeological instabilities	USA	Inventory	Open	Landslide hazard, ground failure, landslide inventory
NASA - Global Landslide Catalog	https://maps.nccs.nasa.gov/	Hydrogeological instabilities	Global	Inventory	Open	Landslide hazard, global landslide map
NASA- LHASA	https://github.com/nasa/lhasa	Hydrogeological instabilities	Global	Hazard	Open	Landslide hazard, real time, probability of occurrence, potential exposure
PAI ABDAC	https://www.autoritadistrettoac.it/	Hydrogeological instabilities	Italy	Inventory + Hazard	Open	Hydrogeological plan, landslides, hydraulic risk
River Flood Hazard	https://data.jrc.ec.europa.eu/	Hydrogeological instabilities	Europe	Inventory	Open	flood hazard, flood map
SaferPlaces Global Platform	https://saferplaces.co/	Hydrogeological instabilities	Global	Inventory + Hazard	No	flood events, flood hazard, climate resilience, flood risk intelligence
Sistema Informativo sulle Catastrofi Idrogeologiche	http://sici.irpi.cnr.it	Hydrogeological instabilities	Italy	Inventory	Open	Hydrogeological instability, Landslides, floods, archives
Agri Nimbo Earth	https://agri.nimbo.earth	Land cover and use	Europe	Inventory	Open	land cover, land use, crop map
Copernicus Land Monitoring Service	https://land.copernicus.eu/	Land cover and use	Global	Inventory	Open	land use, land cover, hot spot monitoring, EGMS,

						biophysical parameters
Copernicus Land Monitoring Service (local)	https://land.copernicus.eu/	Land cover and use	Europe	Inventory	Open	hotspot areas, urban zones, coastal areas, nature conservation
Copernicus Land Monitoring Service (pan)	https://land.copernicus.eu/	Land cover and use	Europe	Inventory	Open	Corine land cover, biophysical parameters
Corine Land Cover	https://land.copernicus.eu/	Land cover and use	Europe	Inventory	Open	Copernicus, Land, Satellite Image, Corine Land Cover
DRYAD	https://datadryad.org/	Land cover and use	USA	Inventory	Open	land value, market value
I-COLT	https://sites.google.com/arpae.it/	Land cover and use	Italy	Inventory	Open	Crops classification, Italy
INFC	https://www.inventarioforestale.org/it/	Land cover and use	Italy	Inventory	Open (Account needed)	Forest cover, Italy
ISPRA- Dati consumo suolo	https://www.isprambiente.gov.it/	Land cover and use	Italy	Inventory	Open	land use, Italy, indicators
Land Cover Products	https://www.esa-landcover-cci.org/	Land cover and use	Global	Inventory	Open	land cover, land use, climate change initiative
Sentinel 2 LULC	https://www.arcgis.com/	Land cover and use	Global	Inventory	Open	land cover, land use, sentinel data
SINA	https://groupware.sinanet.isprambiente.it/	Land cover and use	Italy	Inventory	Open	land use, Italy
Urban Atlas	https://land.copernicus.eu/	Land cover and use	Europe	Inventory	Open	Copernicus, Land Cover. Land Use, Functional Urban Areas
Geoportale regionale Piemonte	https://www.geoportale.piemonte.it/	Multiple	Italy	Inventory	Open	

CEDIT	https://gdb.ceri.uniroma1.it/	Multiple	Italy	Inventory	Open	earthquake induced ground failures, earthquake hazard
DisasterAWARE	https://disasteraware.com/	Multiple	Global	Hazard Early Warning	No	early warning system, global hazards
EU - Risk Data Hub	https://drmkc.jrc.ec.europa.eu/	Multiple	Europe	Inventory + hazard + risk	Open	disaster risk assessment, qualitative assessment, decision support system
European Geological Data Infrastracture	http://www.europe-geology.eu/	Multiple	Europe	Inventory	Open	Pan-European and national geological datasets
GAPMINDER	https://www.gapminder.org/data/	Multiple	Global	Exposure Indicators	Open	
Geoportale Liguria	https://geoportal.regione.liguria.it/	Multiple	Italy	Inventory	Open	
Geoportale Lombardia	https://www.geoportale.regione.lombardia.it	Multiple	Italy	Inventory	Open	
Geoportale Nazionale	http://www.pcn.minambiente.it/	Multiple	Italy	Inventory	Open	Environment, Territory
Geoportale Regionale Valle D'Aosta	https://mappe.regione.vda.it/	Multiple	Italy	Inventory	Open	
Global Risk Data Platform	https://wesr.unepgrid.ch/	Multiple	Global	Inventory + hazard + risk	Open	global risk, multiple hazards
ISTAT- mappa rischi comuni	https://www.istat.it/it/mappa-rischi	Multiple	Italy	Inventory + hazard + risk	Open	Italy, geological risks, elements at risk
Munich Re's NatCatSERVICE	https://www.munichre.com/	Multiple	Global	Hazard	No	digital catalogue, natural disasters

NASA Disaster Mapping Portal	https://maps.disasters.nasa.gov/	Multiple	Global	Inventory	Open	disaster portal, multiple hazards
Portale Servizio Geologico Italia	http://portalesgi.isprambiente.it/	Multiple	Italy	Inventory	Open	geological service portal, natural hazard, geology, land use and cover, hydrogeology, infrastructure
SEDAC	https://sedac.ciesin.columbia.edu/	Multiple	Global	Inventory	Open	socioeconomic and earth science data
Think Hazard!	https://www.thinkhazard.org/	Multiple	Global	Hazard	Open	global hazard, hazard levels
Sinkhole	http://sgi.isprambiente.it/sinkholeweb/	Sinkhole	Italy	Inventory	Open	sinkhole dataset
ALOS Global Digital Surface Model	https://www.eorc.jaxa.jp/	Topography	Global	Inventory	Open	DSM, digital surface model
EarthEnv-DEM90	https://www.earthenv.org/	Topography	Global	Inventory	Open	DEM, digital elevation model
ETOPO 2022	https://data.noaa.gov/	Topography	Global	Inventory	Open	elevation dataset, surface elevation
European Digital Elevation Model	https://land.copernicus.eu/	Topography	Europe	Inventory	Open	DEM, digital elevation model
FABDEM	https://data.bris.ac.uk/	Topography	Global	Inventory	Open	elevation map, Forests and buildings removed Copernicus DEM
GMRT	https://www.gmrt.org/	Topography	Global	Inventory	Open	DEM, digital elevation mode, grid
NASADEM	https://lpdaac.usgs.gov/	Topography	Global	Inventory	Open	DEM, digital elevation model
NEXTMap®	https://www.intermap.com/	Topography	Global	Inventory	No	3D terrain dataset, DTM, DSM

SRTM	https://dwtkns.com/	Topography	Global	Inventory	Open	DEM, digital elevation data
Tinitaly	https://tinitaly.pi.ingv.it	Topography	Italy	Inventory	Open	DEM, digital elevation model
Tsunami Map Viewer	http://sgi2.isprambiente.it/tsunamimap/	Tsunami	Italy	Inventory	Open	tsunami, early warning
Africa Groundwater Atlas	https://www2.bgs.ac.uk/	Water resources	Africa	Inventory	Open	groundwater, African hydrogeology
BIGBANG	https://www.isprambiente.gov.it/	Water resources	Italy	Inventory	Open	hydrology, basin
Bollettino siccità ISPRA	https://www.isprambiente.gov.it/	Water resources	Europe	Bulletin	Open	drought, precipitation index
HIS Central platform	http://www.hiscentral.isprambiente.gov.it/	Water resources	Italy	Inventory	Open (Account needed)	
Severità idrica ISPRA	https://www.isprambiente.gov.it/	Water resources	Italy	Bulletin	Open	drought

Appendix B

Supplementary material from Chapter 2.

Table B1: Rainfall stations used for the definition of mean annual rainfall and the analysis of severe rainfall events. Time sampling of first and last data in the studied period (*) and coverage period (years) of the pluviometric data are also reported.

Rain Gage	***First Record (dd/mm/yyyy)***	***Last Record* (dd/mm/yyyy)***	***Daily rainfall Time Coverage(yrs)***	***Hourly rainfall Time Coverage(yrs)***
Acqua Acetosa	25/09/1992	31/03/2021	25.6	14
Aniene a Lunghezza	01/03/1992	31/03/2021	27.5	19
Aniene a Ponte Salario	01/01/1992	31/03/2021	29.0	21
Aurelio	25/09/1992	31/03/2021	25.5	14
Capannacce	25/09/1992	31/03/2021	16.7	5
Casilino	26/11/1991	31/03/2021	26.8	12
Cassiodoro	25/09/1992	31/03/2021	25.7	15
Collegio Romano	01/01/1951	31/03/2021	69.3	31
Eleniano	25/09/1992	31/03/2021	24.6	12
EUR	25/09/1992	31/03/2021	23.5	12
Flaminio	01/01/2005	31/03/2021	14.4	4
Fosso di Pratolungo	01/01/1981	31/03/2021	36.8	17
Monte Mario	25/09/1992	31/03/2021	25.2	9
Ostiense	25/09/1992	31/03/2021	24.4	11

Ottavia	25/09/1992	31/03/2021	25.7	13
Regillo	25/09/1992	31/03/2021	17.7	10
Roma Bufalotta	01/01/1987	31/03/2021	24.0	13
Roma est	25/09/1992	31/03/2021	25.6	0
Roma EUR	01/01/1951	31/03/2021	58.6	30
Roma Flaminio	01/01/1956	31/03/2021	49.2	29
Roma Macao	01/01/1952	31/03/2021	67.1	45
Roma Monte Mario	01/01/2008	31/03/2021	13.2	1
Roma Monte Mario (Villa Millerose)	01/01/1965	31/12/1999	35.0	11
Roma Nord	25/09/1992	31/03/2021	24.3	11
Roma sud	25/09/1992	31/03/2021	25.7	10
Rosolino Pilo	01/01/1987	31/03/2021	19.9	5
Salone	01/01/1951	31/03/2021	28.7	3
Tevere a Castel Giubileo	01/01/1951	31/03/2021	52.8	24
Tevere a Fidene	01/01/2007	31/03/2021	14.0	8
Tevere a Porta Portese	01/01/1994	31/03/2021	23.6	10
Tor Marancia	01/01/1991	31/03/2021	15.3	5
Tor Vergata	21/04/1994	31/03/2021	25.9	9
Via Marchi	01/01/1995	31/03/2021	21.4	3

Appendix C

Supplementary material from Chapter 4.

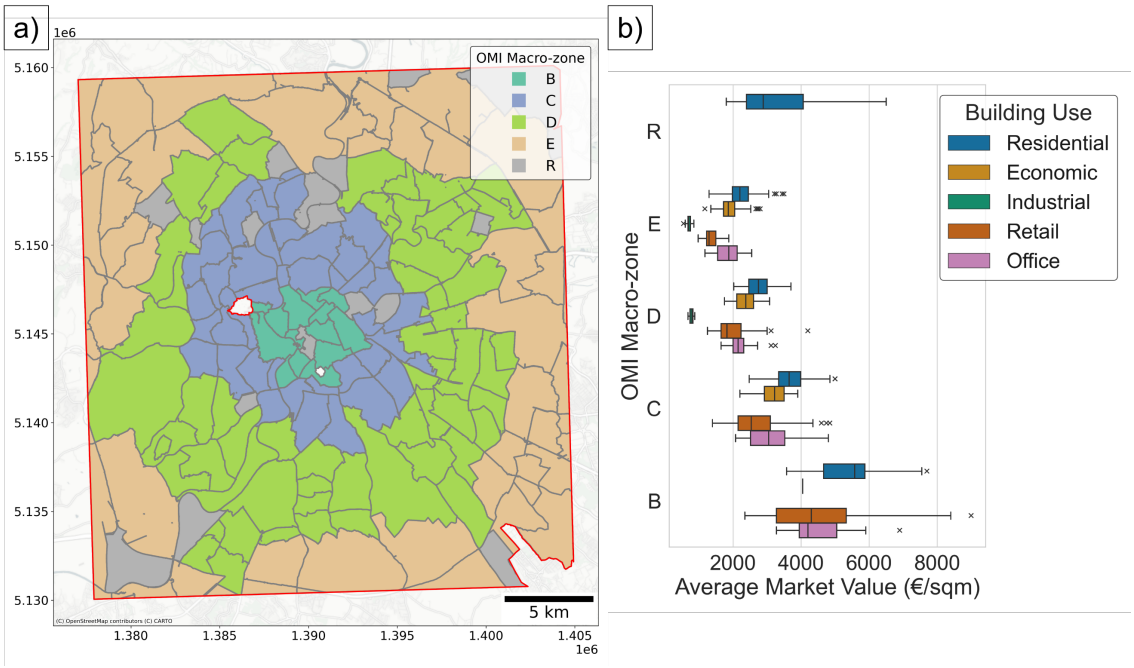


Figure C1: OMI macro-zones (a) and distribution of real-estate market values of buildings in Rome according to category of use and zone.

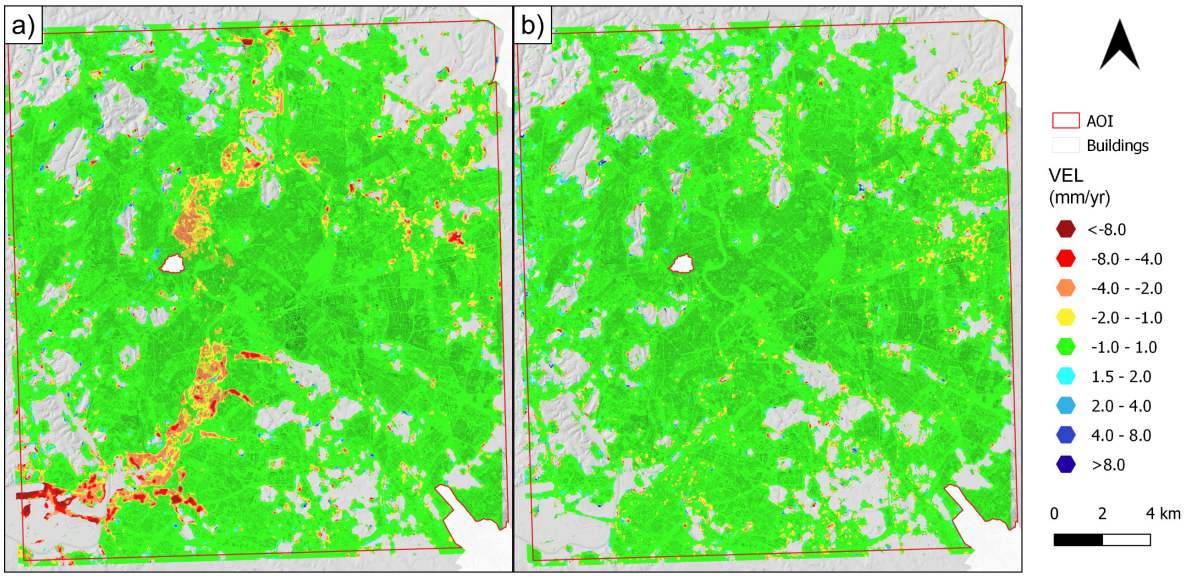


Figure C2: Synthetic datasets of Up-Down (a) and East-West (b) ground displacement rates derived from the data fusion of PSs resulting from A-DInSAR analyses of S1 and CSK SAR images.

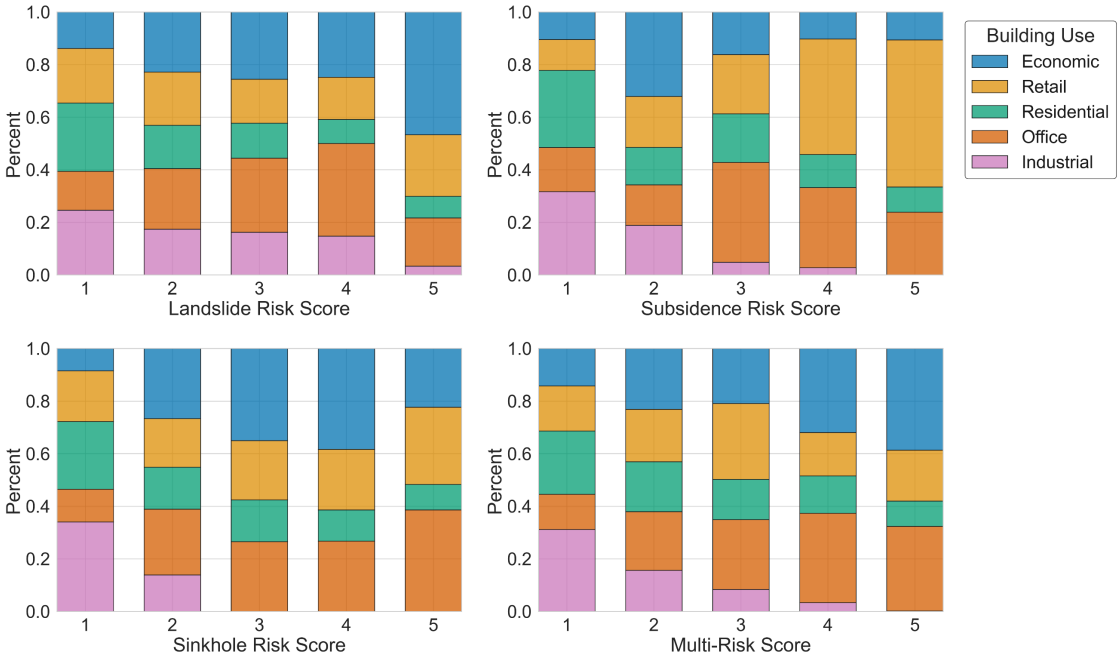


Figure C3: Distribution of building use categories according to single- and multi-risk scores.

Acknowledgements

I would like to express my heartfelt gratitude to the following people and organisations who played a crucial role in my doctoral journey:

First and foremost, I extend my sincere appreciation to Prof. Paolo Mazzanti and Prof. Carlo Esposito for their invaluable guidance and mentorship, which have been the bedrock of my academic development.

I am deeply thankful to the CERI research center for their funding and unwavering support throughout this research endeavor. Their belief in the importance of this work has been pivotal to its successful completion.

I would also like to acknowledge my exceptional colleagues and friends with whom I had the privilege of collaborating over the past three years. Their passion, intellect, and teamwork have enriched my academic experience, and I feel fortunate to have worked with such outstanding people.

My family deserves a special mention for their constant support, love, and encouragement, which sustained me through the challenges of this journey. I deeply appreciate their sacrifices and belief in me.

I want to express my profound gratitude to Valeria for her patience, understanding, and constant faith in me. Her presence in my life has made this journey even more meaningful and fulfilling.

To each of these people, I extend my heartfelt thanks for their integral role in my doctoral experience and for helping me achieve this significant milestone in my academic career.