multi-Risk sciEnce for resilienT commUnities undeR a changiNgclimate

Codice progetto MUR: **PE00000005** – CUP J33C22002840002

Deliverable title: Rationale for selecting and scale-dependent weighing of predisposing factors

Deliverable ID: DV 2.2.5

Due date: 30/11/2023

Submission date: 30/11/2023

AUTHORS

Matteo Berti (UNIBO); Giuseppe Ciccarese (UNIBO); Riccardo Fanti (UNIFI); Isabella Lapietra (UNIBA); Mario Parise (UNIBA); Isabella Serena Liso (UNIBA); Carlo Tacconi Stefanelli (UNIFI)

1. Technical references

Project Acronym	RETURN
Project Title	multi-Risk sciEnce for resilienT commUnities undeR a changiNg climate
Project Coordinator	Domenico Calcaterra UNIVERSITA DEGLI STUDI DI NAPOLI FEDERICO II domcalca@unina.it
Project Duration	December 2022 – November 2025 (36 months)

Deliverable No.	DV2.2.5
Dissemination level*	PU
Work Package	WP2 - State of the art and knowledge base to define impact-oriented hazard indicators
Task	T2.2.3 - Spatial analysis of proneness to ground instabilities: statistical and deterministic approaches
Lead beneficiary	UNIBO
Contributing beneficiary/ies	UNIFI, UNIBA.

* PU = Public

PP = Restricted to other programme participants (including the Commission Services)

RE = Restricted to a group specified by the consortium (including the Commission Services)

CO = Confidential, only for members of the consortium (including the Commission Services)

Document history

Version	Date	Lead contributor	Description
0.1	22.11.2023	Matteo Berti (UNIBO) Riccardo Fanti (UNIFI) Mario Parise (UNIBA)	First draft (General structure)
0.2	23.11.2023	Giuseppe Ciccarese (UNIBO)	Draft of Chapter 5
0.3	24.11.2023	Riccardo Fanti (UNIFI)	Draft of Chapter 4
0.4	24.11.2023	Matteo Berti (UNIBO)	Chapters 1-3, Review and Proofreading of Chapters 4 and 5, Chapter 6
0.5	27.11.2023	Salvatore Martino, Francesca Bozzano (UniRoma1); Domenico Calcaterra, Diego Di Martire (UniNapoli), Riccardo Fanti (UniFI)	Edits for approval
1.0	28.11.2023	Participants to TK2.2.3	Final version

2. ABSTRACT

This Deliverable, part of Milestone 2.2 of Spoke 2 in the Extended Partnership RETURN Project, deals with the theme "Identification of impact-oriented indicators" as outlined in the Executive Work Plan – Milestone 2.1. It summarizes the scientific research activities conducted from January to November 2023 by Task 2.2.3 ("Spatial analysis of proneness to ground instabilities: statistical and deterministic approaches") of Work Package 2.2 ("State of the art and knowledge base to define impact-oriented hazard indicators"). This task is a component of the vertical spoke VS2, "Ground Instabilities", and involves 57 researchers from various institutions.

The focus of WP2 is on detecting and analysing predisposing factors to ground instabilities, while WP3 and WP4 concentrate on preparatory factors, and triggering and multiple geohazards cascading scenarios (MULTI-HAZARD), respectively. These work packages collectively aim to quantify ground instabilities' effects on territories, buildings, and communities, and to develop an IT platform for the spatial and temporal analysis of these instabilities.

A significant phase within Task 2.2.3 involved defining Ground Instability categories, which were categorized initially into landslides, subsidence, liquefaction, and sinkholes. A more detailed differentiation was later made, particularly distinguishing between slow and fast types of ground instability in subaerial phenomena. These categories are detailed in Table 4.1 and have been fundamental in guiding the project's direction.

The methodological approach for weighting predisposing factors, crucial for assessing ground instability susceptibility, is outlined in Figure 5.1. This process, led by the Task T2.2.3 leader and the RTDa PNRR team, started with insights from prior work package outputs and learning examples from partners. It evolved into the creation of an initial list of weighting methods, which was refined and finalized through expert panel discussions. This list is categorized into expert-based, data-driven, and physically-based methods.

Expert-based methods, while transparent and reproducible, were excluded from the Proof Of Concept due to their high subjectivity. Physically-based methods, although detailed, were also set aside because of their extensive data requirements and limited scalability. The focus, therefore, shifted to data-driven methods, known for their ability to handle large datasets and provide insights where physical parameters are less understood. The selection of a specific data-driven method was recognized as dependent on various factors, such as the type of ground instability, data availability, and study objectives.

To support the Proof of Concept phase, a "popularity index" was developed to assess the prevalence of different methods in addressing ground instability processes. This index was derived from an analysis of scholarly databases. The results, presented in Tables 5.1 and 5.2, show a notable preference for methods like Logistic Regression and Artificial Neural Networks.

The Deliverable also addresses the connections to WP3 and WP4, emphasizing the critical role of this work in the broader project framework. It lays the foundation for subsequent analyses and risk mitigation strategies by identifying susceptible areas, which then lead to the examination of preparatory processes in WP3 and triggering mechanisms in WP4. This comprehensive and systematic approach ensures effective management and mitigation strategies for ground instability risks.

3. Table of contents

List of Tables

List of Figures

4. Introduction

4.1 Project framework

This Deliverable is drawn up as part of Milestone 2.2 of Spoke 2 having as its topic (from the Executive Work Pan – Milestone 2.1) "Identification of impact-oriented indicators".

The Deliverables of Spoke 2 for this Milestone have therefore set themselves as an overall objective the identification of rationales, starting from specific learning examples of literature, for identifying both the ground instabilities through macrocategories of factors (predisposing, preparatory, triggers) and the construction of analytical tools which, arranged in a specific logical-executive order (tool-chain), should lead to the design of an IT platform for the restitution in the PoC of the spatial overlap (multiple-hazard) or the temporal succession (multi-hazard, i.e. chain effects) of Ground Instability processes (Figure 4.1). This will allow quantifying the ground instabilities effects on the territory with a view to their impact on buildings and communities also evaluating their suitability and reliability.

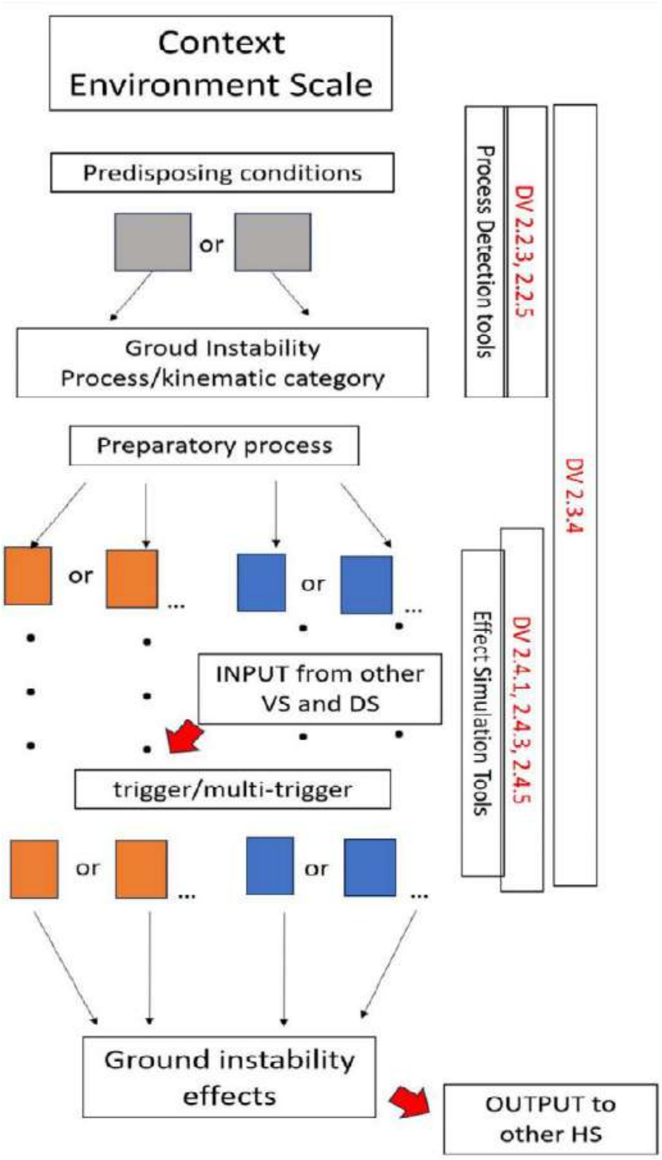

Figure 4.1: Diagram illustrating the methodology implemented in Spoke 2 for quantifying ground instabilities and delivering scenario impacts to other spokes

This report summarizes the scientific research activities carried out in the period January 2023 - November 2023 by the **Task 2.2.3** *"Spatial analysis of proneness to ground instabilities: statistical and deterministic approaches"* (hereinafter referred to as **TK3**) of the **Work Package 2.2** *"State of the art and knowledge base to define impact-oriented hazard indicators"* (hereinafter referred to as **WP2**) within the vertical spoke **VS2** "Ground Instabilities" of the Extended Partnership RETURN.

It should be noted that VS2 structured the work packages WP2, WP3 and WP4 by identifying the following areas of interest for each of them:

- WP2 focuses on the detection and analysis of **PREDISPOSING** factors to ground instabilities;
- WP3 targets **PREPARATORY** factors to ground instabilities;
- WP4 is centered on **TRIGGERING** and multiple geohazards cascading scenarios (**MULTI-HAZARD**).

In accordance with the definitions given within the VS2, the distinction between predisposing, preparatory and triggering factors/processes is made on a temporal basis: the predisposing factors are considered invariable on the observation scale, while the preparatory factors show changes or cyclical trends during the same period. As a consequence, a trigger is considered as a process that acts in a very short and well-defined time.

The partners involved in the **WP2** are ENEA, OGS, POLITO, UNIBA, UNIBO, UNIFI, UNIGE, UNINA, UNIPA, UNIPD and UNIROMA1. **WP2** leaders are Riccardo Fanti (UNIFI) and Mario Parise (UNIBA), **TK1** leader is Francesco Maria Chiocci (UNIROMA1), **TK2** leader is Mario Parise (UNIBA), **TK3** leader is Matteo Berti (UNIBO). 57 researchers participate in the activities of **WP2/TK3** (i.e. TK 2.2.3): 4 from ENEA, 1 from OGS, 3 from POLITO, 6 from UNIBA, 5 from UNIBO, 7 from UNIFI, 3 from UNIGE, 7 from UNINA, 10 from UNIPA, 7 from UNIPD and 4 from UNIROMA1.

The goal of **TK2** (*Spatial analysis of proneness to ground instabilities: statistical and deterministic approaches*) and the issue of **DV 2.2.5** (*Rationale for selecting and scale-dependent weighing of predisposing factors*) have been interpreted within the framework of the entire Spoke work process.

According with the main idea of the Project and of **VS2**, the learning phase had the objective of building a rationale for preparatory processes to be used as input to the Proof of Concept (PoC). This phase has been articulated in three stages:

i) Inventory of Learning Examples (LE).

ii) Individuation of the preparatory processes analyzed in each LE.

iii) Definition of a Rationale for each process based on the available LEs.

This **DV 2.2.5** represents (together with the **DV 2.2.3**) the description of the transition from phase ii) above to phase iii).

In this sense, the work is a gradual transition from the exemplary level represented by the synthesis of the LE (i.e. recent experience of each partner, comprising leading-edge analyzes on the topic of characterization of predisposing factors and spatial and temporal quantification of susceptibility) to an exhaustive level consisting of a synthesis useful for the purposes of drafting a real Rationale.

To achieve this objective, we progressed through an "internal recall" phase, aimed at identifying, among the experiences of **TK3** participants, an additional set of LEs capable of integrating case studies of phenomena and approaches intended to better complete the emerging panorama during the initial phase (refer to DV 2.2.1 and DV 2.2.2). Subsequently, an additional phase of analyzing global case studies was added, too, involving a bibliographic review on the topic of **TK3**, to arrive at a final product (the present Rationale) that can serve as a valid and comprehensive support for the subsequent phases of the project.

In the following steps, **TK3** an in-depth examination into the synergistic application of different methods (geographically weighted multivariate analyses, statistical methodologies, and cutting-edge Artificial Intelligence (AI) and Machine Learning (ML) algorithms) aimed to highlight the connections between predisposing and triggering factors and the resulting ground instabilities. Simultaneously, an overview on

Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA

the optimization of methods tailored to the diverse geomorphological settings and scale has been carried out.

In this way, **TK3** wanted to examine the question of the transition from a qualitative enumeration of the predisposing factors selected for each phenomenon of ground instability to a quantitative evaluation of the weight to be attributed to each of them within the framework of a quantitative or semi-quantitative procedure, aimed at defining a complete territorial evaluation also adapted to the scale of observation.

In order to achieve this fundamental goal, it is considered important to preface this introduction with a significant operational phase carried out within **TK3**, fully shared by **TK1** and **TK2**. This phase involves the definition of Ground Instability categories, a topic to which the following paragraph is dedicated.

4.2 Ground Instability Typologies

The topic of defining the typologies of Ground Instability has been addressed since the beginning of the Project (see the Executive Working Plan). As a preliminary step, a distinction was made between landslides, subsidence, liquefaction, and sinkholes. Based on this categorization, we proceeded with the identification of the LEs and their typological analysis, and all documents from the initial phase of the Project contain this distinction. Subsequently, **VS2** deemed it necessary to make a more detailed differentiation, considering the kinematics of the phenomena as the primary discriminating element, especially distinguishing between slow and fast typologies of ground instability in subaerial phenomena.

Ground Instabilities	Subaerial Landslides	Subaerial Slow Landslides Typologies	Slow Flows (Earthflows)
			Slow Slides (Rotational and Planar Slides, Soil slips)
			Slow Spread & Slow Slope Deformations (Spread (except Liquefaction), Rock/Soil Slope Deformations, Creep, DsGSD)
		Subaerial Rapid Landslides Typologies	Rapid Flows (Debris flows, Mudflows)
			Rapid Slides (Rock Slides, Rock Avalanches)
			Falls & Topples (Rock Falls, Rock Topples)
	Submarine Landslides	*Submarine Landslides Typologies*	*Canyon head Landslides*
			Inner shelf/Insular shelf Landslides
	Sinkholes	Slow Sinkholes Typologies	Slow Sinkholes (All Types)
		Rapid Sinkholes Typologies	Rapid Sinkholes (All Types)
	Subsidence	Subsidence Typologies	Subsidence (All Types)
	Liquefaction	Liquefaction Typologies	Liquefaction (All Types)

Table 4.1: Typologies of Ground Instabilities.

Following a series of collegial discussions at the level of the entire **WP2** (thus collaboratively involving **TK1**, **TK2**, and **TK3**), a subdivision of Ground Instability functional to this Project was established. It is crucial to note that this subdivision should not be interpreted as a proposal for a new classification of landslides, subsidence, sinkholes, and liquefaction.

The considered types of Ground Instabilities are presented in Table 4.1 and serve as a reference for this document. Different technical meetings have been then organized, dedicated to each ground instabilities (respectively, Slow landslides, Fast landslides, Sinkholes/Subsidence/Liquefaction, and Marine Instabilities), aimed at sharing ideas and experiences before writing the Rationales. The meetings were attended by at least a reference person per each institution, with highly productive discussions among the attendees.

5. Methodological approaches for weighting Predisposing Factors

5.1 Methods

The Task T2.2.3 leader, in collaboration with the RTDa PNRR team, adopted an expert-based methodology to identify suitable methods for weighting Predisposing Factors affecting ground stability, as depicted in Figure 5.1. This process began with insights from previous outputs of Work Package 2 (WP2), specifically DV2.2.1 and DV2.2.2, and incorporated learning examples from partners. This collaborative effort led to the creation of an initial list of Weighting Methods, which was refined through discussions with expert panels, resulting in a consensus on the final list showcased in Figure 5.1.

The literature review revealed numerous potential methods for weighting these predisposing factors, broadly categorized into expert-based, data-driven, and physically-based methods.

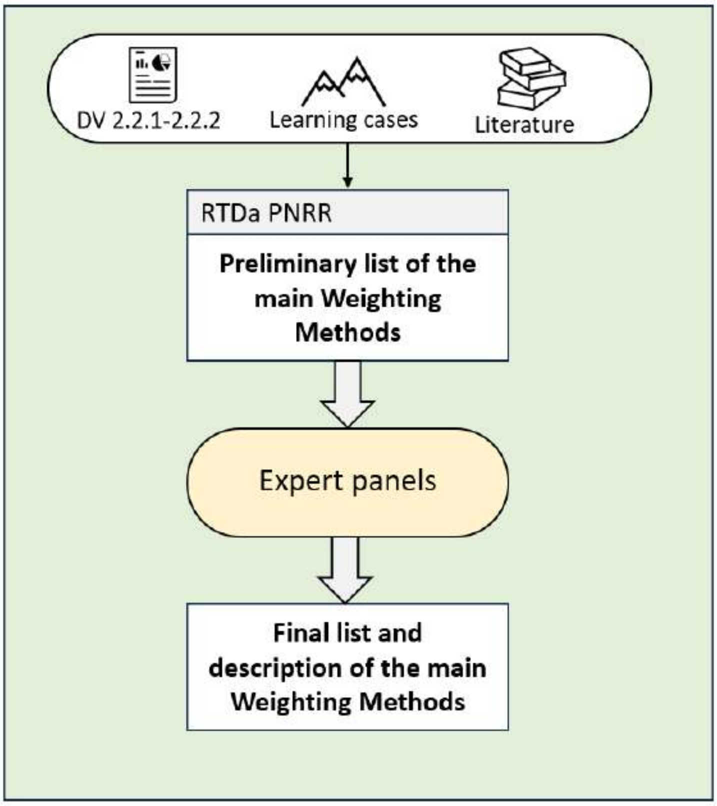

Figure 5.1: flowchart of the approach for identifying the weighting methods for Predisposing Factors

Expert-based methods rely on the knowledge and judgment of experts to assess susceptibility in a specific area. These methods often involve the qualitative assessment of various factors and parameters that contribute to Ground Instability occurrence. Some common expert-based methods include geomorphological mapping, Multi-Criteria Decision Analysis, and Analytical Hierarchy Process. Expert-based methods are often used in the absence of comprehensive data or as a supplement to quantitative modeling approaches. Although these methods are transparent and reproducible, their degree of subjectivity remains high. For this reason, they have been excluded from those suitable for the Proof Of Concept.

Physically-based methods are approaches that use mathematical and physical models to evaluate the likelihood of ground instability occurrence. These methods rely on understanding the physical processes that lead to instability. Key aspects of physically-based methods are the use of mechanical properties of soils and rocks, the quantitative measurements of slope geometry, and the explicit definitions of groundwater flow. Despite their detailed, site-specific risk assessments, these methods require extensive data and their scalability is limited. Furthermore, they intrinsically incorporate predisposing factors, eliminating the need for a separate weighting procedure.

Consequently, the focus shifted solely to data-driven methods. Data-driven methods focus on analyzing historical data and observed patterns to predict future ground instability occurrences. These approaches typically involve statistical analysis, machine learning, or artificial intelligence techniques to identify correlations between various factors and instability events. Data-driven methods are powerful for their ability to handle large and diverse datasets and for providing insights where physical parameters are not well understood or difficult to measure. Nonetheless, their effectiveness hinges on data quality and quantity and might not fully capture the physical mechanisms involved.

Selecting a specific data-driven method depends on various factors, including the type of ground instability, data availability and quality, study objectives, and available resources. Given the diversity in geological conditions, computational resources, and study aims, it's impractical to prescribe a universally applicable guideline for choosing the most appropriate data-driven method. In essence, there's no universally "best method" for a particular ground instability type.

To address this issue and support the Proof of Concept phase, we explored a variety of methods, assessing their prevalence in addressing different ground instability processes by means of a “popularity index”. This index, based on the frequency of method usage for specific processes, suggests that more commonly used methods are likely more tailored and reliable for certain processes. They are also likely to require standard data types, computational resources, and technical skills, simplifying their application in large-scale susceptibility evaluations.

The "popularity index" was computed by analyzing two scholarly databases, Scopus and Google Scholar. We counted the papers that mentioned both a methodological approach and a type of ground instability (for example, 'logistic regression' and 'rock falls'). It's crucial to note that this index doesn't reflect an exact count of papers applying a specific method to a process, as keywords might be mentioned in contexts like introductions, not central to the methodology. Nevertheless, this index serves as an informative measure of the approaches frequently used in literature for weighting predisposing factors across various ground instability types.

5.2 Results

The results of the work are comprehensively presented in Tables 5.1 and 5.2, and elaborated upon in subsequent pages. These tables categorize 15 data-driven methods across columns and various ground instability types along rows. Each row displays the frequency of scientific papers associating a particular ground instability with a specific data-driven method, based on Google Scholar data. Similar findings are obtained from other academic databases like Scopus or WOS.

The key distinction between the two tables lies in their color-coding schemes, denoting the popularity of methods. Table 5.1 highlights the three most frequently used methods in red, orange, and yellow, while the rest are in green. Table 5.2 employs a percentile-based color scheme, with shades from red to green representing decreasing percentile classes from 100%-75% to 25%-0%.

As can be seen, the analysis reveals a notable preference for two methods, Logistic Regression and Artificial Neural Networks. These methods are predominantly used across nearly all types of ground instabilities, suggesting their adaptability and effectiveness. Other methods like Fuzzy Logic, Random Forest, and Support Vector Machine also show considerable popularity. The high popularity index of these methods suggests extensive testing and proven effectiveness in diverse studies, leading to their recurrent use in the scientific community. Their popularity has probably fostered a rich body of supporting literature, case studies, tutorials, and forums, which often guide new researchers in the field, perpetuating the preference for these methods.

However, it is crucial to recognize that no single method is universally optimal. The suitability of a method depends on the study's specific goals, the characteristics of the area under study, and the nature of the available data. Less commonly used methods should not be dismissed as inferior; they might be more recent, complex, or data-intensive, but they could be the most appropriate choice for certain cases.

Following the tables, the report provides detailed descriptions of each method. These descriptions are applicable to any ground instability type under study, offering a reference for selecting the most appropriate methodological approach for ground instability susceptibility assessment.

				QUANTITATIVE														
				data-driven methods														
				Weights of evidence (WOE)	Logistic regression (LR)	Frequency ratio (FR)	Certainty factors (CF)	Evidential belief function (EBF)	Analytic Hierarchy Process (AHP)	Fuzzy logic (FL)	Random forest (RF)	Artificial neural networks (ANN)	K-Nearest Neighbor (KNN)	Support Vector Machine (SVM)	Decision trees (DT)	Naïve Bayes (NB)	Linear Discriminant Analysis (LDA)	Quadratic Discriminant Analysis (QDA)
Ground Instabilities	Subaerial Landslides	Subaerial Rapid Landslides Typologies	Rapid Flows (Debris flows, Mudflows)	1540	5270	2450	158	592	2510	2140	2450	3100	362	2550	1320	776	285	75
			Rapid Slides (Rock Slides, Rock Avalanches)	187	453	253	17	58	139	190	139	232	16	163	77	43	9	3
			Falls & Topples (Rock Falls, Rock Topples)	454	1270	693	43	176	514	644	435	679	79	497	292	175	64	8
		Subaerial Slow Landslides Typologies	Slow Flows (Earthflows)	272	543	133	41	115	191	242	212	338	38	241	152	108	44	12
			Slow Slides (Rotational and Planar Slides, Soil slips)	122	307	181	23	50	92	138	87	163	9	112	50	38	12	4
			Slow Spread/DsGSD	90	765	156	7	27	103	219	261	460	49	269	207	45	60	6
	Submarine Landslides	Submarine Landslides Typologies	Rapid Landslides	41	150	69	3	23	58	95	85	114	11	79	42	23	7	0
			Slow Landslides															
	Sinkholes	Slow Sinkholes Typologies	Slow Sinkholes (All Types)	246	1250	320	25	121	487	1440	1250	1020	480	1480	791	814	181	27
		Rapid Sinkholes Typologies	Rapid Sinkholes (All Types)															
	Subsidence	Subsidence Typologies	Subsidence (All Types)	1480	6780	2310	127	825	4070	4210	4600	6090	739	4610	2480	928	561	89
	Liquefaction	Liquefaction Typologies	Liquefaction (All Types)	249	4400	861	51	101	1700	2870	2090	6070	459	3140	1390	507	377	61

Table 6.1: "popularity index" of different data-drive methods to weight Predisposing Factors (columns) for different type of Ground Instabilities (rows). Colors indicate the three most frequently used methods in red, orange, and yellow respectively.

				QUANTITATIVE														
				data-driven methods														
				Weights of evidence (WOE)	Logistic regression (LR)	Frequency ratio (FR)	Certainty factors (CF)	Evidential belief function (EBF)	Analytic Hierarchy Process (AHP)	Fuzzy logic (FL)	Random forest (RF)	Artificial neural networks (ANN)	K-Nearest Neighbor (KNN)	Support Vector Machine (SVM)	Decision trees (DT)	Naïve Bayes (NB)	Linear Discriminant Analysis (LDA)	Quadratic Discriminant Analysis (QDA)
Ground Instabilities	Subaerial Landslides	Subaerial Rapid Landslides Typologies	Rapid Flows (Debris flows, Mudflows)	1540	5270	2450	158	592	2510	2140	2450	3100	362	2550	1320	776	285	75
			Rapid Slides (Rock Slides, Rock Avalanches)	187	453	253	17	58	139	190	139	232	16	163	77	43	9	3
			Falls & Topples (Rock Falls, Rock Topples)	454	1270	693	43	176	514	644	435	679	79	497	292	175	64	8
		Subaerial Slow Landslides Typologies	Slow Flows (Earthflows)	272	543	133	41	115	191	242	212	338	38	241	152	108	44	12
			Slow Slides (Rotational and Planar Slides, Soil slips)	122	307	181	23	50	92	138	87	163	9	112	50	38	12	4
			Slow Spread/DsGSD	90	765	156	7	27	103	219	261	460	49	269	207	45	60	6
	Submarine Landslides	Submarine Landslides Typologies	Rapid Landslides / Slow Landslides	41	150	69	3	23	58	95	85	114	11	79	42	23	7	0
	Sinkholes	Slow Sinkholes Typologies / Rapid Sinkholes Typologies	Slow Sinkholes (All Types) / Rapid Sinkholes (All Types)	246	1250	320	25	121	487	1440	1250	1020	480	1480	791	814	181	27
	Subsidence	Subsidence Typologies	Subsidence (All Types)	1480	6780	2310	127	825	4070	4210	4600	6090	739	4610	2480	928	561	89
	Liquefaction	Liquefaction Typologies	Liquefaction (All Types)	249	4400	861	51	101	1700	2870	2090	6070	459	3140	1390	507	377	61

Table 6.2: "popularity index" of different data-drive methods to weight Predisposing Factors (columns) for different type of Ground Instabilities (rows). Colors represent decreasing percentile classes: red= 75%-100%; orange=50%-75%; yellow=25%-50%; green=0%-25%.

Description of the main weighting methods	
Weights of evidence (WOE)	Weights of Evidence (WOE) is a statistical technique used for analyzing the relationship between a binary target variable and one or more predictor variables. In the context of ground instabilities, WOE serves as a valuable method for assessing susceptibility and predicting potential hazards in specific geographic areas. WOE operates by examining the evidence provided by predictor variables to determine their strength and influence on the likelihood of the binary event, such as ground instabilities, occurring or not occurring. In the case of ground instabilities, WOE can be applied to evaluate susceptibility by analyzing the contribution of geological, topographical, and environmental factors to the likelihood of instabilities. WOE is well-suited for handling geospatial and environmental data, making it a valuable tool for assessing ground instabilities susceptibility. It allows for the quantification of the weight or impact of various factors on susceptibility by analyzing their evidence. One of the strengths of WOE is its ability to provide clear insights into the factors that significantly affect susceptibility, aiding in decision-making related to land use planning, hazard assessment, and risk management in regions prone to geohazards.
Logistic regression (LR)	Logistic Regression (LR) is a statistical method used for modeling the probability of a binary outcome or event based on one or more predictor variables. In the context of ground instabilities, Logistic Regression provides a valuable approach for assessing susceptibility and predicting potential hazards in specific geographic areas. LR works by creating a model that estimates the log-odds of the binary outcome, such as the occurrence of ground instabilities (1) or non-occurrence (0), as a linear combination of predictor variables. This model allows for the estimation of the probability of the event happening, offering insights into the factors influencing ground instabilities. LR is well-suited for handling geospatial and environmental data, making it a valuable tool for assessing ground instabilities susceptibility. It can accommodate various factors, including geological, topographical, and environmental variables, to create predictive models that highlight the relationships between these factors and the probability of ground instabilities. One of the strengths of Logistic Regression is its interpretability, providing a clear understanding of the factors influencing susceptibility and supporting decision-making related to land use planning, hazard assessment, and risk management in regions susceptible to geohazards.
Frequency ratio (FR)	Frequency Ratio (FR) is a statistical technique used to analyze and quantify spatial patterns and relationships within datasets. In the context of ground instabilities, FR provides a valuable method for assessing susceptibility and predicting potential hazards in specific geographic areas. FR operates by comparing the frequency of an event, such as ground instabilities, in relation to specific factors or variables. When applied to ground instabilities, FR assesses susceptibility by examining the relationship between geological, topographical, and environmental variables and the frequency of ground instabilities in different areas. FR is well-suited for handling geospatial and environmental data, making it a valuable tool for assessing ground instabilities susceptibility. It allows for the quantification of the impact of various factors on susceptibility by analyzing their frequencies.

	One of the strengths of FR is its simplicity and interpretability, providing clear insights into the relationships between factors and the frequency of ground instabilities in different locations. In the field of ground instabilities, Frequency Ratio (FR) serves as a practical and effective tool for decision-making related to land use planning, hazard assessment, and risk management. It offers a straightforward approach to assess susceptibility, supporting informed choices and resource allocation in regions susceptible to geohazards.
Analytic Hierarchy Process (AHP)	The Analytic Hierarchy Process (AHP) is a mathematical approach used in various fields, including civil engineering and environmental studies, to evaluate complex and hierarchical decision-making. In the context of assessing ground instabilities susceptibility, AHP provides a structured decision-making framework for considering and weighing various factors that may contribute to such susceptibility. The AHP approach begins with the creation of a hierarchy of criteria and sub-criteria that define ground instabilities susceptibility. These criteria may include geological, topographical, hydrological, and environmental factors, among others. Decision-makers then assign relative importance ratings to each criterion and sub-criterion through pairwise comparisons. After obtaining the relative ratings, AHP uses mathematical calculations to assign weights to each criterion and sub-criterion, allowing for the quantification of their relative importance. These weights are then applied to assess and rank the importance of factors in ground instabilities susceptibility. Once the weights are obtained, the AHP model can be used to evaluate susceptibility in different locations, taking into account the values of relevant factors in each area. The result is often a susceptibility map that indicates areas with higher or lower risk of ground instabilities. The AHP approach is particularly useful when dealing with complex and hierarchical decisions that involve numerous factors and when it's important to weigh the relative importance of these factors. Its application in the field of ground instabilities can support land planning, risk mitigation, and decision-making based on a structured assessment of susceptibility.
Certainty factors (CF)	Certainty Factors (CF) is a knowledge representation and reasoning technique used in the field of artificial intelligence and expert systems. In the context of ground instabilities, CF serves as a valuable method for assessing susceptibility and predicting potential hazards in specific geographic areas. CF operates by quantifying and managing the uncertainty associated with information and evidence. It allows for the expression of degrees of belief or certainty in the truth of propositions or hypotheses. In the case of ground instabilities, CF can be applied to evaluate susceptibility by assigning and updating certainty values to various geological, topographical, and environmental factors influencing ground instabilities. CF is well-suited for handling geospatial and environmental data, making it a valuable tool for assessing ground instabilities susceptibility while considering the inherent uncertainty in the data. It enables the representation of the certainty or confidence in the factors contributing to susceptibility.

	One of the strengths of Certainty Factors is its ability to capture and manage uncertainty in a structured manner, supporting decision-making related to land use planning, hazard assessment, and risk management in regions susceptible to geohazards.
Evidential belief function (EBF)	The Evidential Belief Function (EBF) is a mathematical framework used to handle uncertainty and incomplete information in decision-making processes. It is particularly valuable when dealing with problems where information is uncertain or not fully precise. In the context of assessing ground instabilities susceptibility, EBF provides a structured approach for modeling and managing uncertainty. For ground instabilities susceptibility assessment, EBF serves as a tool to represent information with basic belief assignments, which are essentially mass functions applied to a frame of discernment encompassing all possible hypotheses or states related to ground instabilities. This framework allows for the treatment of complex and uncertain data. EBF offers mechanisms to manage conflicts and inconsistencies in the available evidence. When different pieces of evidence provide conflicting information, EBF provides methods to reconcile and update belief assignments accordingly. It allows for the fusion of multiple sources of evidence, combining their information using Dempster's rule of combination. This fusion process addresses uncertainty and conflicts in the evidence, resulting in a more comprehensive assessment of ground instabilities susceptibility. EBF also offers tools to quantify and represent uncertainty associated with the assessment, aiding in the understanding of the confidence level in the results. In the context of ground instabilities susceptibility, EBF can effectively model complex interactions between various factors contributing to susceptibility. It can account for dependencies and correlations among different pieces of evidence. The results obtained using EBF can be used to support decision-making in areas prone to ground instabilities. The assessment of susceptibility, with uncertainty and conflict management, can inform decisions related to land use planning, risk mitigation, and resource allocation.
Random forest (RF)	Random Forest (RF) is a machine learning technique used for data analysis and prediction. It belongs to the ensemble learning category, combining multiple decision trees to make more accurate and robust predictions. In the context of ground instabilities, RF can be employed to assess susceptibility and evaluate potential hazards. It leverages a collection of decision trees, each trained on a different subset of the data, to provide a more comprehensive and reliable prediction of ground instabilities. RF excels in handling complex and large datasets, where multiple factors can contribute to ground instabilities susceptibility. It's particularly useful for modeling the intricate relationships between geological, topographical, and environmental variables, making it an effective tool for assessing and mapping ground instabilities. The key advantage of RF is its ability to capture the influence of various factors on ground instabilities susceptibility, offering insights into which factors are the most significant. It can support decision-making processes related to land use planning, risk mitigation, and hazard assessment in regions vulnerable to ground instabilities.

Fuzzy logic (FL)	Fuzzy Logic (FL) is a mathematical approach used in data analysis and decision-making. It departs from traditional binary logic by allowing for degrees of truth and uncertainty. In the context of ground instabilities, FL offers a flexible framework for modeling and reasoning with imprecise or vague information. For ground instabilities susceptibility assessment, Fuzzy Logic can be applied to accommodate and represent the uncertainty and fuzziness inherent in geohazard data. It provides a way to handle and process data that may not fit neatly into traditional binary categories. Instead, FL assigns degrees of membership to different categories, allowing for a more nuanced understanding of the factors influencing susceptibility. In FL, linguistic variables and fuzzy sets are used to capture the imprecise nature of ground instabilities data. This approach is particularly valuable when dealing with complex, multidimensional factors contributing to susceptibility. FL enables the creation of fuzzy inference systems that can model the relationships between various parameters and their impact on ground instabilities. These systems offer a more comprehensive and flexible way of understanding the complex interactions involved in hazard assessment. The benefit of Fuzzy Logic in ground instabilities assessment lies in its ability to handle and represent uncertainty, providing a more realistic and adaptable approach to modeling susceptibility. It assists decision-makers in understanding and managing the risks associated with ground instabilities, supporting informed choices in land use planning, hazard mitigation, and resource allocation in vulnerable regions.
Artificial neural networks (ANN)	Artificial Neural Networks (ANN) are computational models inspired by the structure and function of the human brain. In the context of ground instabilities, ANNs are a powerful tool for data analysis and prediction, particularly when dealing with complex geospatial and environmental data. For ground instabilities susceptibility assessment, ANNs can be utilized to capture intricate relationships between various factors contributing to susceptibility. These networks consist of interconnected nodes (neurons) that process and transform data, allowing for the modeling of nonlinear and complex patterns in the data. ANNs excel in learning from large datasets and can uncover hidden associations between geological, topographical, and environmental variables, thus enhancing the understanding of susceptibility factors. This makes them valuable for assessing and mapping ground instabilities. The strength of ANNs lies in their ability to adapt and self-optimize through training. They can identify patterns and trends in the data, making them valuable for predictive modeling. ANNs offer insights into the influence of different parameters on ground instabilities, aiding in decision-making related to land use planning, risk assessment, and hazard mitigation. In the realm of ground instabilities, ANNs are a sophisticated computational tool that can provide valuable support for informed decisions, thanks to their capacity to model complex relationships and extract meaningful information from large and diverse datasets.

K-Nearest Neighbor (KNN)	K-Nearest Neighbor (KNN) is a machine learning algorithm used for classification and regression tasks. In the context of ground instabilities, KNN provides a simple yet effective approach to assess susceptibility and evaluate potential hazards in specific geographic areas. KNN operates by determining the class or value of a data point based on the majority class or average value of its K-nearest neighbors within a given dataset. In the case of ground instabilities, this algorithm can be applied to evaluate susceptibility by comparing the characteristics of a location with its neighboring data points. KNN is particularly useful when dealing with geospatial and environmental data. It can consider multiple factors such as geological, topographical, and historical data to make predictions about the ground instabilities susceptibility of a specific location. One of KNN's strengths is its ease of use and interpretability. It doesn't require complex model training and offers a straightforward way to estimate susceptibility based on the characteristics of nearby data points. In the field of ground instabilities, KNN serves as a practical tool for decision-making related to land use planning, hazard assessment, and risk management in regions prone to geohazards. It allows for the quick assessment of susceptibility, supporting informed choices and resource allocation.
Support Vector Machine (SVM)	Support Vector Machine (SVM) is a powerful machine learning algorithm used for classification and regression tasks. In the context of ground instabilities, SVM serves as an effective tool for assessing susceptibility and predicting potential hazards in specific geographical areas. SVM operates by finding a hyperplane that best separates data points into different classes or predicts a continuous value. In the case of ground instabilities, SVM can be applied to evaluate susceptibility by identifying the factors that best distinguish susceptible areas from non-susceptible ones. SVM is well-suited for handling geospatial and environmental data, making it valuable for assessing ground instabilities susceptibility. It considers various factors, such as geological, topographical, and environmental variables, to create a model that can predict susceptibility for different locations. One of SVM's strengths lies in its ability to handle complex data and find the most informative features for making predictions. It is particularly useful in cases where the relationship between factors and susceptibility is nonlinear or when there are multiple factors at play. In the field of ground instabilities, SVM is a robust tool for decision-making related to land use planning, hazard assessment, and risk management. It offers the capability to create accurate susceptibility models, assisting in informed choices and resource allocation in regions susceptible to geohazards.

Decision trees (DT)	Decision Trees (DT) are a popular and intuitive machine learning approach used for both classification and regression tasks. In the context of ground instabilities, Decision Trees provide a versatile tool for assessing susceptibility and predicting potential hazards in specific geographic areas. DT operates by recursively splitting the data into subsets based on the most significant attributes, leading to a tree-like structure. In the case of ground instabilities, Decision Trees can be applied to evaluate susceptibility by identifying the key factors that distinguish susceptible areas from non-susceptible ones. Decision Trees are well-suited for handling geospatial and environmental data, making them valuable for assessing ground instabilities susceptibility. They consider various factors, such as geological, topographical, and environmental variables, to create a model that can predict susceptibility for different locations. One of the strengths of Decision Trees is their ease of interpretability. They provide clear and easily understandable rules for making predictions, which can be valuable for decision-makers and stakeholders. In the field of ground instabilities, Decision Trees serve as a practical and insightful tool for decision-making related to land use planning, hazard assessment, and risk management. They allow for the quick and interpretable assessment of susceptibility, supporting informed choices and resource allocation in regions prone to geohazards.
Naïve Bayes (NB)	Naïve Bayes (NB) is a probabilistic machine learning algorithm used primarily for classification tasks. In the context of ground instabilities, Naïve Bayes provides a reliable approach for assessing susceptibility and predicting potential hazards in specific geographical areas. NB operates by applying Bayes' theorem to estimate the probability of a data point belonging to a particular class based on its features. In the case of ground instabilities, Naïve Bayes can be applied to evaluate susceptibility by considering various geological, topographical, and environmental factors. Naïve Bayes is well-suited for handling geospatial and environmental data, making it valuable for assessing ground instabilities susceptibility. It can model complex relationships between factors and predict susceptibility by estimating the likelihood of an event occurring given the available evidence. One of the strengths of Naïve Bayes is its simplicity and efficiency. It is often used when dealing with large datasets and is known for its quick training and prediction times. In the field of ground instabilities, Naïve Bayes serves as a practical and effective tool for decision-making related to land use planning, hazard assessment, and risk management. It provides a probabilistic framework for assessing susceptibility, supporting informed choices and resource allocation in regions susceptible to geohazards.
Linear Discriminant Analysis (LDA)	Linear Discriminant Analysis (LDA) is a dimensionality reduction and classification technique widely used in the field of pattern recognition and machine learning. In the context of ground instabilities, LDA offers a valuable method for assessing susceptibility and predicting potential hazards in specific geographical areas. LDA works by finding a linear combination of features that best discriminates between different classes or categories. In the case of ground instabilities, LDA can be applied to

	assess susceptibility by identifying the factors that most effectively separate areas prone to instabilities from those that are not. LDA is well-suited for handling geospatial and environmental data, making it a valuable tool for assessing ground instabilities susceptibility. It can reduce the dimensionality of the data while preserving the most relevant information, thus aiding in the modeling of complex relationships between geological, topographical, and environmental variables. One of the strengths of LDA is its ability to uncover the key factors that contribute to the differentiation between different susceptibility classes. It provides clear linear combinations of features that can be used to make informed predictions. In the field of ground instabilities, Linear Discriminant Analysis serves as a powerful tool for decision-making related to land use planning, hazard assessment, and risk management. It helps in understanding the factors that influence susceptibility and supports informed choices and resource allocation in regions susceptible to geohazards.
Quadratic Discriminant Analysis (QDA)	Quadratic Discriminant Analysis (QDA) is a statistical technique utilized for classification and discrimination tasks. In the context of ground instabilities, QDA presents a valuable method for assessing susceptibility and predicting potential hazards in specific geographical areas. QDA operates by estimating the discriminant functions based on the probability distributions of different classes. In the case of ground instabilities, QDA can be applied to assess susceptibility by considering various geological, topographical, and environmental factors and modeling their probabilistic relationships. QDA is well-suited for handling geospatial and environmental data, making it a valuable tool for assessing ground instabilities susceptibility. It can capture complex relationships between factors and offers a flexible framework for understanding the probabilistic aspects of susceptibility. One of the strengths of QDA is its capacity to model the underlying probability distributions of different susceptibility classes. It provides a more nuanced view of the data compared to linear methods. In the field of ground instabilities, Quadratic Discriminant Analysis serves as an effective tool for decision-making related to land use planning, hazard assessment, and risk management. It allows for a probabilistic assessment of susceptibility, supporting informed choices and resource allocation in regions susceptible to geohazards.

6. Connections to WP3 and WP4

The efforts undertaken in Work Package 2 (WP2) have been primarily dedicated to identifying, quantifying, and weighting the Predisposing Factors that contribute to ground instabilities.

The outcomes of these efforts are concisely encapsulated in the deliverables submitted in July 2023 (DV2.2.1 and DV2.2.2), and the more recent submissions (DV2.2.3 and DV2.2.5). The present deliverable focuses specifically on the methods prevalently employed to weight these Predisposing Factors across different types of ground instabilities, utilizing data-driven approaches.

This work forms the initial segment of the comprehensive framework depicted in Figure 6.1, which outlines the Process Detection Tools. The primary objective of these tools is to discern potential sites of Ground Instability in a designated area. This phase of the project is critical, as it lays the groundwork for subsequent analyses and risk mitigation strategies.

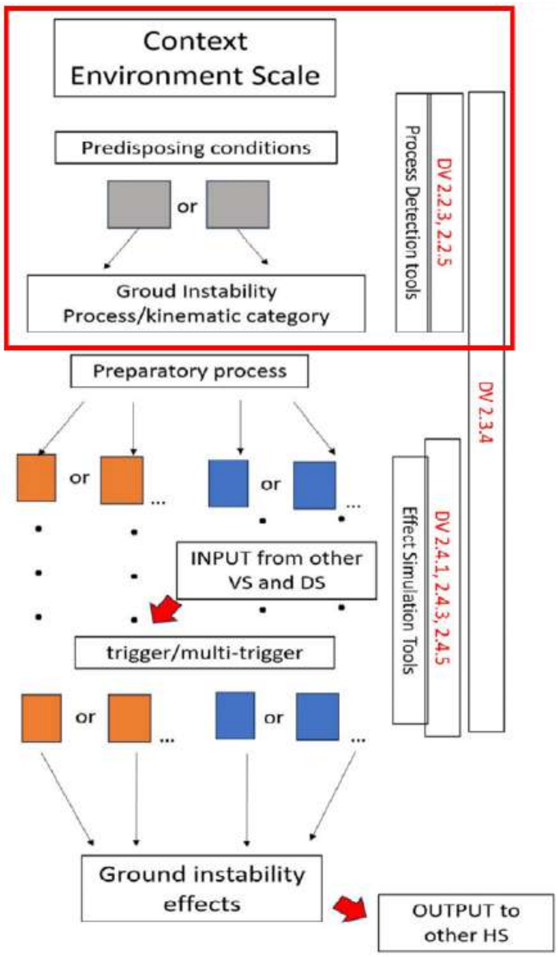

Figure 6.1: Diagram illustrating the contribution of the deliverables of WP2 to the methodology implemented in Spoke 2

Delving deeper into this initial segment, the analysis encompasses several key steps:

1. Identification of Ground Instability Processes: This crucial first step involves the identification of potential types of ground instability in the target area. It relies heavily on expert judgment, combining historical records of past instabilities with a detailed examination of the area's geological, geomorphological, and hydrogeological characteristics.

2. Selection and Quantification of Predisposing Factors: For each identified type of ground instability, it's essential to determine the most significant Predisposing Factors, again drawing upon expert insight. The Process-Factors list in DV2.2.3 serves as a guiding framework for this selection. These factors must then be measured in line with their nature — whether qualitative, semi-quantitative, or quantitative — as recommended in the deliverable.

3. Assessment of Susceptibility: This phase involves appraising the susceptibility of various instability processes using an appropriate data-driven method. The insights from this deliverable offer guidance in selecting a method, taking into account its characteristics, data availability, and its prevalence in academic literature.

Once the areas prone to susceptibility are detected, the analysis can progress to the next stages, which include examining the preparatory processes (in WP3) and the triggering mechanisms (in WP4). This sequential approach ensures a comprehensive and systematic assessment of ground instability risks, paving the way for effective management and mitigation strategies.

7. References

Ballabio C. & Sterlacchini S. (2012). Support vector machines for landslide susceptibility mapping: the Staffora River Basin case study, Italy. Mathematical geosciences, 44, 47-70.

Bruno E., Calcaterra D. & Parise M. (2008). Development and morphometry of sinkholes in coastal plains of Apulia, southern Italy. Preliminary sinkhole susceptibility assessment. Engineering Geology, 99(3-4), 198-209.

Carrara A., Crosta G. & Frattini P. (2008). Comparing models of debris-flow susceptibility in the alpine environment. Geomorphology, 94(3-4), 353-378.

Ciccarese G., Mulas M. & Corsini A. (2021). Combining spatial modelling and regionalization of rainfall thresholds for debris flows hazard mapping in the Emilia-Romagna Apennines (Italy). Landslides, 18(11), 3513-3529.

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.

Conforti M., Pascale S., Robustelli G. & Sdao, F. (2014). Evaluation of prediction capability of the artificial neural networks for mapping landslide susceptibility in the Turbolo River catchment (northern Calabria, Italy). Catena, 113, 236-250.

Goetz J. N., Brenning A., Petschko H. & Leopold, P. (2015). Evaluating machine learning and statistical prediction techniques for landslide susceptibility modeling. Computers & geosciences, 81, 1-11.

Lanfredi Sofia C., Oliveira S. C., Pereira S., Zêzere J. L. & Corsini A. (2018). A comparison between bivariate and multivariate methods to assess susceptibility to liquefaction-related coseismic surface effects in the Po Plain (Northern Italy). Geomatics, Natural Hazards and Risk, 9(1), 108-126.

Pradhan B. (2013). A comparative study on the predictive ability of the decision tree, support vector machine and neuro-fuzzy models in landslide susceptibility mapping using GIS. Computers & Geosciences, 51, 350-365.

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.

Taheri K., Gutiérrez F., Mohseni H., Raeisi E. & Taheri M. (2015). Sinkhole susceptibility mapping using the analytical hierarchy process (AHP) and magnitude–frequency relationships: A case study in Hamadan province, Iran. Geomorphology, 234, 64-79.

Tien Bui D., Tuan T. A., Klempe H., Pradhan B. & Revhaug I. (2016). Spatial prediction models for shallow landslide hazards: a comparative assessment of the efficacy of support vector machines, artificial neural networks, kernel logistic regression, and logistic model tree. Landslides, 13, 361-378.

Trigila A., Iadanza C., Esposito C. & Scarascia-Mugnozza G. (2015). Comparison of Logistic Regression and Random Forests techniques for shallow landslide susceptibility assessment in Giampilieri (NE Sicily, Italy). Geomorphology, 249, 119-136.

Tufano R., Guerriero L., Annibali Corona M., Bausilio G., Di Martire D., Nisio S. & Calcaterra D. (2022). Anthropogenic sinkholes of the city of Naples, Italy: An update. Natural Hazards, 112(3), 2577-2608.

Vakhshoori V. & Zare M. (2016). Landslide susceptibility mapping by comparing weight of evidence, fuzzy logic, and frequency ratio methods. Geomatics, Natural Hazards and Risk, 7(5), 1731-1752.

Yilmaz I. (2009). Landslide susceptibility mapping using frequency ratio, logistic regression, artificial neural networks and their comparison: a case study from Kat landslides (Tokat—Turkey). Computers & Geosciences, 35(6), 1125-1138.

Zhang K., Wu X., Niu R., Yang K. & Zhao L. (2017). The assessment of landslide susceptibility mapping using random forest and decision tree methods in the Three Gorges Reservoir area, China. Environmental Earth Sciences, 76, 1-20.

Zêzere, J. L., Pereira, S., Melo, R., Oliveira, S. C. & Garcia R. A. (2017). Mapping landslide susceptibility using data-driven methods. Science of the total environment, 589, 250-267.

Zumpano V., Pisano L., Filice F., Ugenti A., de Lucia D., Wasowski J., Santaloia F & Lollino P. (2022). Regional-scale seismic liquefaction susceptibility mapping via an empirical approach validated by site-specific analyses. Geosciences, 12(5), 215.